THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF ECONOMICS

DEVELOPING A MODEL: MAXIMIZING DOCTOR UTILITY GIVEN PATIENT SEVERITY AND INVASIVENESS OF TREATMENT

KELLY TAMBURO
SPRING 2019

A thesis
submitted in partial fulfillment
of the requirements
for baccalaureate degrees
in Economics and Risk Management (Actuarial Science option)
with honors in Economics

Reviewed and approved* by the following:

Edward Green
Professor Emeritus, Department of Economics
Thesis Supervisor

Russell Chuderewicz
Teaching Professor, Department of Economics
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

One of the challenges in a doctor-patient-insurer relationship is finding the right balance between costly, invasive treatment, and cheap, noninvasive treatment for a given patient's severity level. This leads to two key questions: (1) From the perspective of the doctor and patient, what is the right severity threshold at which to prescribe the most invasive treatment; and (2) From the insurer perspective, can we estimate from the decisions made by a doctor over time, what is the threshold used by that doctor to prescribe the most invasive treatment? This paper will utilize the Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality National Inpatient Sample from 2014. A two-part model that optimizes patient life outcomes is used in this paper and consists of an ordered logit regression model and conditional logit discrete choice model. This model determines a doctor's utility from medical versus surgical treatment based on a patient's severity level and life outcome.

In the model findings, medical treatment stochastically dominates outcomes. Additionally, in the model, doctor utility preferences result in a ranking of outcomes as semi-cured (lowest), death, and cured (highest). This discrepancy may be driven by two explanations. The assumption that doctors maximize a utility function that depends only on, and that is increasing in, the patient's outcome is violated, or the assumption that severity, as measured in the data, is a sufficient statistic for the doctor's best judgement about the appropriate treatment, given all the evidence that the doctor sees does not hold. There is evidence that suggests that the first model assumption could be violated through regional disparities and further research would be needed to determine if the second model assumption is violated.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Dr. Edward Green: Thank you for your support and guidance throughout the thesis process.

Dr. James Tybout: Thank you for your feedback and advice throughout the writing process.

Family and Friends: Thank you for your encouragement throughout my college career.

Schreyer Honors College: Thank you for your generous funding to purchase the data necessary to move forward with my research.

Chapter 1

Paper Goals

In the healthcare industry, doctors and insurance companies have to match patient severity[1] for an illness with the correct treatment. The challenge for patients, doctors, and insurers is finding the balance between costly, invasive treatment, and cheap, noninvasive treatment for a patient's given severity. In this paper, we want to understand: (1) from the perspective of the doctor and patient, what is the right severity threshold at which to prescribe the most invasive treatment; and (2) from the insurer perspective, can we estimate from the decisions made by a doctor over time, what is the threshold used by that doctor to prescribe the most invasive treatment?

The intuition behind the model begins with the assumption that a patient comes in with a severity of disease that the doctor must treat. The doctor does two things: (1) observes the severity with error, and (2) prescribes a treatment of either high or low invasiveness. In this scenario, severity refers to the ease of cure, not to painfulness of symptoms. Invasiveness refers to some weighted combination of the medical invasiveness and the financial cost of the treatment. It is the cost-side of the cost-benefit analysis. However, costs do not directly need to be incorporated into the invasiveness metric, since it is assumed that costs increase as the invasiveness of treatment increases.

Both severity and chosen treatment will be observed in the data. Invasiveness of treatment may be estimated from some combination of things observed in a data record such as

[1] For this model, severity will be defined as the degree of difficulty to cure with simple treatment.

average recovery time for a given procedure and dollar cost, and severity may be estimated from some combination of things observed in a medical record, such as stage in illness and the patient's actual recovery time. In this study, we will use All Patient Refined Diagnosis Related Groups (APRDRG) severity codes from the Healthcare Cost and Utilization Project (HCUP), Agency for Healthcare Research and Quality (AHRQ) National Inpatient Sample (NIS) from 2014 to determine the patient's given severity level. Severity is adjusted for age and other comorbidities. Invasiveness of treatment will be determined based on APRDRG codes for traumatic brain injuries to distinguish between medical and surgical treatment.

Traumatic brain injuries (TBI) are the chosen illness being studied in this paper. TBI was chosen due to its high percentage of trauma admissions and due to its higher mortality rates. These factors allow us to use discharge status of patients as a proxy for patient life outcomes. Additionally, TBI leads to high indirect costs as more symptoms have delayed onset. Therefore, addressing ways to optimize patient care for this illness can help reduce indirect costs and lead to better life outcomes. More about TBI and variable definitions will be discussed during Chapter 3.

Next, the patient's post-treatment health is a function of the treatment given and patient severity. This health level will most likely be unobserved. However, there will be an observable, decision-relevant outcome based on the health variable. For simplicity, assume it is whether the patient is cured, semi-cured, and dead. This outcome will be measured through the proxy variable DISPUNIFORM, which accounts for patient discharge status in the HCUP-NIS data set.

Finally, the functional relationship among severity, treatment, and post-treatment health will be found by latent-variable estimation through an ordered logit regression model. Doctor

utility from a given treatment will be found using a discrete choice model (DCM). Specifically, the chosen DCM will be a conditional logit, where the expected utility is derived from the conditional probabilities of a given severity level and treatment on life outcomes from the prior ordered logit regression model. The model description can be found in Chapter 3 and the model outcomes can be found in Chapter 4. Implications of these results will be discussed in Chapter 5.

Chapter 2

Prior Relevant Work

This paper focuses on developing a structural econometric model of how a doctor makes decisions about a patient's treatment given his or her severity level. Despite a thorough search, very few, if any, prior studies have sought to answer a similar question as the one posed in this paper. However, studies can be found on developing severity indexes without the treatment component. Studies focused on drawing patterns from realized decisions by doctors, such as ones on regional disparity, can be found as well. Very limited research can also be found on doctor moral hazard, which aids in understanding how doctors may be influenced in their decision-making process. Therefore, to better understand how this paper's model was developed, we will explore prior studies on severity and treatment thresholds for acute otitis media and double moral hazard between doctors and patients. It is important to keep in mind that this is a pilot study that is answering a relatively new topic of research, so these studies are provided to compare and contrast model development.

Severity and Treatment Thresholds for Acute Otitis Media

A study by Froom (2001) looks at the risk factors, severity, and treatment of acute otitis media[2] (AOM) at the first hospital visit. This study evaluates patient severity of children and

[2] Acute otitis media is the inflammation of the middle ear, where there is fluid in the middle ear accompanied by signs or symptoms of an ear infection.

chosen treatment from three countries: the United States (and some cases from Canada), the United Kingdom, and the Netherlands. Both the United Kingdom and the Netherlands have national treatment guidelines that physicians are expected to follow based on corresponding severity levels, whereas the United States physicians use their formal training and personal judgement to aid in diagnosis. This study uses a logistic regression model to adjust for the separate networks and variations in sample sizes across the networks. It also accounts for individual risk factors, severity, and selected treatment across the networks (Froom, 2001). Froom found no difference in treatment for increased severity of patient age groups across the countries, except for ages 25 to 180 months; for patients ages 25 to 180 months, a North American physician was more likely to prescribe the more expensive second-line antimicrobial treatment than their British and Dutch counterparts (Froom, 2001). Froom contributed the difference in severity levels to cultural differences in the prevalence in risk factors across the networks.

When treating an increased severity level patient, both Dutch and North American doctors alter their treatment decisions, but Dutch doctors often choose a first-line antimicrobial, whereas North American doctors tend to choose a second-line antimicrobial. British physicians did not alter their treatment methods, and continued to follow their national guidelines. Duration of treatment remained consistent with national standards regardless of severity level across the networks (Froom, 2001). Interestingly, despite the differences in treatment methods for increased severity, both the Dutch and North American approach led to similar outcomes; therefore, if North American physicians were to change their prescribing habits "at an average cost of $115 for an episode of AOM, a similar reduction of the 20,009,000 office visits for AOM in the United States would result in an annual saving of about $185 million" (Froom, 2001).

The study on AOM uses microeconomic data by studying a single disease similar to the strategy used in this paper. Froom's study also utilized a logistic regression similar to the approach taken in the paper; however, this paper uses ordering, and takes the model a step further by analyzing doctor utility from a given treatment using a discrete choice model. The study on AOM highlights how developing a model to test for optimal treatment decisions can aid in the diagnosis process through patient welfare maximization. In certain cases, this can lead to reduction in costs as the less invasive treatment may be the more optimal treatment decision.

Doctor Behavior and Moral Hazard

This study questions whether a doctor is wastefully prescribing the expensive, more invasive treatments when the cheaper, less invasive treatment would lead to similar life outcomes. A potential reason for overprescribing more invasive treatment could stem from moral hazard[3] in the doctor-patient relationship. For example, Lundin (2000) explored moral hazard in physician prescription behavior. The pharmaceutical market shields the true cost of the prescription drug, since patients do not observe the extra cost associated with having a brand-name drug compared to a generic one; and if physicians act more in the interest of the patient than in the interest of the payer, physicians also do not worry about costs when prescribing. Therefore, without considering costs, doctors may make their prescribing decisions on perceived quality differences or brand loyalty.

[3] Moral hazard is the instance where one party increases their involvement in risky events knowing that he or she is protected against the risk and the other party will incur the cost. The decision to behave in risky behavior is derived from greater perceived utility from the one party, ignoring morality factors of right and wrong. Double moral hazard consists of asymmetric information on both sides, leading to both parties maximizing utility functions with the negative cost borne on the opposing party.

Brand loyalty or perceived quality differences can stem from the transaction cost of researching a generic drug when it becomes available. If the doctor is compensated for his or her time to research the generic drug and the drug is a commonly prescribed treatment, there would be moral hazard present if the doctor still refuses to do the research. This is because a large number of patients would receive a benefit from the generic prescription, but fail to receive this benefit due to the lack of an action on the doctor's part. On the other hand, if it is a rare illness that has a generic drug alternative, a doctor may not find it worth his or her time to research the product due to its limited scalability. In this case, if the cost of research time is greater than the cost savings of the generic drug, this would not be a profitable transaction and moral hazard would not be present in this scenario. This is because this transaction does not maximize a joint doctor-patient welfare function, which should account for both the doctor's and patient's welfare. However, if the doctor could not bill his or her time and the cost savings of the generic drug outweighs the cost of research time, then the doctor selfishly maximizes their personal utility function instead of a joint doctor-patient utility function if they choose not to research the generic drug, in turn leading to moral hazard.

The study also found that if more of the costs are shifted to patients through higher out-of-pocket cost-sharing, then patients are less likely to have the brand-name drugs compared to patients with lower out-of-pocket cost-sharing (Lundin, 2000). This suggests physician moral hazard through both information asymmetry and third-party financing in patient care, where the doctor has to act in the interest of both the patient and the insurer to achieve efficiency in allocations. This also suggests that doctors are trying to maximize a joint utility function by considering patient costs in addition to their own costs such as time.

Chapter 3

Research Methodology

Background on Traumatic Brain Injuries

Traumatic brain injuries (TBI) can range in severity from mild, moderate, and severe, and can be classified as an open injury, a closed injury, a skull fracture, and/or a penetrating injury. Mild TBI (mTBI) has a much higher likelihood of recovery compared to moderate or severe TBI, but many patients with mTBI are left with disabilities that negatively impact their quality of life. Given its higher prevalence, there is a large social burden caused from mTBI at least equivalent to that resulting from severe TBI (Vos, 2015). As a result, mTBI is being thought of as a lifelong chronic health condition.

Additionally, many survivors of TBI have a higher risk of developing other neurological disorders, such as Alzheimer's disease[4] and Parkinson's disease[5], and/or disorders of the hypothalamic-pituitary axis[6] leading to long-term consequences, including sleep disorders. As a consequence, survivors of moderate-to-severe TBI have an approximately 4-7 years reduction in life expectancy (Vos, 2015). Therefore, it is important to consider the effectiveness of treatment by the degree of recovery achieved by the patient from the chosen treatment of TBI to help decrease this social burden. However, the unique combination of patient-related factors, such as age and medical history, and surgeon-related factors, such as experience and expertise,

[4] Alzheimer's is a type of dementia that causes problems with memory, thinking and behavior (Alzheimer's Association, 2019).

[5] Parkinson's disease is a progressive nervous system disorder that affects movement (Mayo Clinic, 2018).

[6] The hypothalamus is part of the brain that monitors many aspects of the state of the body systems, integrating a large amount of information from many sensory pathways. The anterior pituitary contains a number of secretory cells that release hormones. These hormones are released in response to stimulation by the appropriate releasing hormones. These primarily impact stress and sleep responses (Endocrine Surgeon, 2019).

contribute to the complexity of choosing the optimal treatment, and narrow the treatment options based on what is technically feasible and which risks are acceptable.

The aim of this study is to develop a broadly applicable methodology to understand a doctor's preferences regarding less invasive to more invasive treatment given a patient's severity level and likeliness of certain life outcomes. This model was built using a specific injury, in this case TBI, in order to better understand the doctor's tradeoffs in the decision-making process. Once developed, this model should be able to be broadly used across different APRDRG codes that have tradeoffs between medical or surgical treatment options. This study used an ordered logit regression model and a discrete choice model in order to estimate the utility of treatment options from the doctor's perspective.

Model Design

The discrete choice model was analyzed by recording each choice among the two treatment options (medical or surgical). The observations were analyzed by an ordered logit regression model. This model was implemented in SAS software (Version 9.4, SAS Institute Inc., Cary, NC, USA). The first phase with the ordered logit regression model was the following:

$$o_i^* = \beta_0 + \beta_1 * severity + \beta_2 * treatment + \varepsilon_i$$

and

$$\begin{aligned} o_i = 0 \quad if \quad -\infty \quad < o_i^* \leq \quad \mu_0 \\ o_i = 1 \quad if \quad \mu_0 \quad < o_i^* \leq \quad \mu_1 \\ o_i = 2 \quad if \quad \mu_1 \quad < o_i^* \leq \quad +\infty \end{aligned}$$

where,

- o_i^* is the latent variable for patient outcome. Patient outcome is one of three outcomes: death (0), semi-cured (1), or cured (2).
- The error term (ε_i) represents any unobserved variables impacting patient outcome.
- β_0 is the intercept that represents the constant numerical differences between the outcome levels and represents the cut-offs between outcomes.
- β_1 is a coefficient that represents the relative importance of severity on the outcome level.
- β_2 is a dummy variable of 'treatment'. For treatment, the base level is when the patient is treated medically not surgically[7].
- Cured is when DISPUNIFORM = {1, 2}, or when the patient is transferred to their home as a routine discharge, or transferred to a short-term hospital, respectively. Semi-cured is considered when DISPUNIFORM = {5, 6}, or when the patient is discharged to another facility, such as a skilled nursing facility or intermediate care, and home health care, respectively. Death is when DISPUNIFORM = 20. As a result of symptoms showing delayed onset, TBI is starting to be viewed as a chronic condition. Therefore, the discharge status is used as a proxy for the patient's outcome, since there is not knowledge on the future development of other health problems for patients within this data set.
- Patients aged 70 or older were excluded from this study. At an older age, a skilled nursing facility or home health care could be an older patient's home before their hospital admission, and a discharge to one of those facilities would not signify a better or worse recovery outcome. Using discharge status as a proxy for recovery for patients older than

[7] Surgical treatments are considered ones performed in the operating room. Non-surgical medical procedures are used to diagnose, measure, monitor, or treat problems. They are generally not highly invasive and do not involve cutting.

70 years old is ambiguous and therefore, it is excluded from this study. However, there can be instances of ambiguity for younger ages, but would be significantly less common.

- Severity is the sum of the APRDRG_Risk_Mortality score and the APRDRG_Severity score. The combined severity score ranges from the lowest value of 2 to the highest value of 8. No-class-specified discharges are coded as APRDRG_Risk_Mortality = 0 and/or APRDRG_Severity = 0 and therefore, these instances will be excluded.
- Medical treatment for TBI is when APRDRG = {55, 56} and surgical treatment for TBI is when APRDRG = {20, 910}. APRDRG 55 is coded for head traumas with a coma greater than one hour or a hemorrhage. APRDRG 56 is coded for brain contusion or brain laceration and complicated skull fractures, a coma less than one hour, or no coma. APRDRG 20 is coded for craniotomy for trauma. APRDRG 910 is coded for craniotomy for multiple significant traumas. APRDRG 57 (concussion, closed skull fracture, uncomplicated intracranial injury) was excluded from the study due the less severe treatment required for this level of brain injury. These categories describe the observed treatment decision.
- See Table 1 for more detailed variable definitions.

Assuming that all choices have independent influence on a doctor's preferences, the following discrete choice model was estimated:

$$Doctor\ Utility\ and\ Treatment\ Decision = Max(U_{i,0}, U_{i,1})$$

$$U_{i,1} = \sum_{k=1}^{2} \beta_k * P[outcome = k|severity, treatment] + \varepsilon_{i,1}$$

$$U_{i,0} = \sum_{k=1}^{2} \beta_k * P[outcome = k|severity, treatment] + \varepsilon_{i,0}$$

where:

- U_{ij} represents the total hypothetical utility or relative positive consequences resulting from a preference for a certain type of treatment for a given patient status. It is a combination of observed factors, and unobserved, random factors in the error term for decision-maker i, i = 1, 2, …, n. $U_{i,1}$ is the utility for a surgical treatment, and $U_{i,0}$ is the utility for medical treatment.
- $\sum_{k=1}^{2} \beta_k P[outcome = k|severity, treatment]$ is the expected utility of the outcome of that treatment for the patient in question. The conditional probabilities are the likeliness of that outcome occurring given the patient's severity level and chosen treatment for a given patient. β_k is the weighting factor. β_0 is normalized to 0 for the outcome of death.

Using these coefficients, the doctor's hypothetical utility scores will be generated for the different patient status and treatment combinations and subsequently ranked. The higher the hypothetical utility score, the stronger the preference of the doctor for that chosen treatment. The utility score allows comparisons across different patient status by converting the preferences to one scale. The resulting values can be compared with each other to deduce preferred treatment options for specific types of conditions and patient severity.

The sign of the coefficient reflects if the attribute has a positive or negative impact on utility, and the value of the attribute indicates the relative importance to overall utility. Therefore, a statistically significant coefficient implies that it is an important factor to the doctor's decision-making. In this study, the sample of TBI patient discharges can be seen in Table 2 and the sample of TBI patient discharges of people aged less than 70 can be seen in Table 3.

Table 1: HCUP Variable Definitions

Variable Names	Variable	Variable Definition
APRDRG	All patient defined DRG[8]	APR-DRGs and APS-DRGs are DRG-based severity measurement systems. APR-DRGs (version 20) include 316 base disease categories. The following measures are included for APR-DRGs: base APR-DRG, severity of illness subclass, and risk of mortality subclass within each base APR-DRG.
APRDRG_Risk_Mortality	All patient refined DRG: risk of mortality subclass	The APRDRG risk of mortality, uniform coding: (0) No class specified, (1) Minor likelihood of dying, (2) Moderate likelihood of dying, (3) Major likelihood of dying, (4) Extreme likelihood of dying.
APRDRG_Severity	All patient refined DRG: severity of illness subclass	The APRDRG severity, uniform coding: (0) No class specified, (1) Minor loss of function (includes cases with no comorbidity or complications), (2) Moderate loss of function, (3) Major loss of function, (4) Extreme loss of function.
DISCWT	Discharge weights	This weight is used to create national estimates for all analyses.
DISPUNIFORM	Disposition of patient (discharge status)	Disposition of patient, uniform coding: (1) routine, (2) transfer to short-term hospital, (5) other transfers, including skilled nursing facility, intermediate care, and another type of facility, (6) home health care, (7) against medical advice, (20) died in hospital, (99) discharged alive, destination unknown.

[8] Diagnosis Related Groups (DRGs) are payment categories that are used to classify for the purpose of reimbursing hospitals for each case in a given category with a fixed fee regardless of the actual costs incurred.

Table 2: Distribution of TBI Patient Discharges in 2014 (n=29,387)

		Effectiveness of Treatment		
Treatment	**Severity**	**Death**	**Semi-cured**	**Fully cured**
Medical where APRDRG = 55 or APRDRG = 56	2	13	498	4,183
	3	32	1,348	3,587
	4	91	3,181	1,979
	5	69	1,241	762
	6	222	2,021	1,039
	7	416	622	387
	8	996	884	332
Total Medical		**1,839**	**9,795**	**12,269**
Surgical where APRDRG = 20 or APRDRG = 910	2	xxx[9]	201	645
	3	xxx	378	438
	4	11	202	197
	5	13	243	139
	6	61	720	233
	7	133	362	126
	8	645	649	147
Total Surgical		**874**	**2,755**	**1,925**
Total (Medical, Surgical)		**2,713**	**12,550**	**14,194**

Statistical Modeling Techniques

Discrete choice models depend on the following assumptions: (1) the existence of well-defined utility indexes or the completeness of rankings, (2) rationality or utility maximization, and (3) axioms of revealed preferences or the idea that revealed choices do not reveal utility, only rankings which are scale invariant. Under utility maximization assumptions, it is assumed that the doctor's decision was the optimal choice. Additionally, given independence from irrelevant alternatives (IIA) holds, then overall choice preference is unaffected by the set of choices offered.

[9] HCUP prohibits publication of cell counts less than or equal to 10. These values have been scrubbed in the table, but are still used in the statistical model.

Table 3: Distribution of TBI Patients in 2014 (Age < 70) (n=15,008)

		Effectiveness of Treatment		
Treatment	**Severity**	**Death**	**Semi-cured**	**Fully cured**
Medical where APRDRG = 55 or APRDRG = 56	2	xxx[10]	220	3,572
	3	xxx	467	2,696
	4	xxx	340	678
	5	14	278	478
	6	29	384	649
	7	134	239	335
	8	525	389	259
Total Medical		**712**	**2,317**	**8,667**
Surgical where APRDRG = 20 or APRDRG = 910	2	xxx	104	548
	3	xxx	93	317
	4	xxx	71	153
	5	xxx	124	118
	6	28	244	167
	7	72	236	110
	8	423	361	126
Total Surgical		**540**	**1,233**	**1,539**
Total (Medical, Surgical)		**1,252**	**3,550**	**10,206**

First, the DCM needs a choice-set, or the set of all choices the subject might have chosen. The choice set is typically mutually exclusive, exhaustive, and/or a finite set of all alternatives. In this case, the choice set is the two treatment options (medical or surgical) that the doctor can choose from. Next, the DCM fits a model with utility on the left-hand side and characteristics of the subject, characteristics of the choices, and an error term on the right-hand side. The utility is the value being estimated in this study. The characteristics of the subject are patient information such as severity and patient outcome. The characteristics of the choices are the treatment options. There are other factors observed by the doctor but are not observed in the data set that are accounted for in the model through the error term.

[10] HCUP prohibits publication of cell counts less than or equal to 10. These values have been scrubbed in the table, but are still used in the statistical model.

By using categorical data for patient outcome, this model aims to estimate conditional probabilities instead of conditional means. As a result, an ordered logit regression model is used, which also utilizes an underlying latent regression model. The underlying latent variable (o_i^*) is observed in the data that helps explain the observed ordinal response (o_i). The logit model assumes an underlying logit regression model for the latent variable (o_i^*) with a constant standard deviation. It also implies stochastic orderings at different explanatory variable levels and is designed to detect location rather than dispersion effects.

Some advantages of utilizing ordinality of response include: (1) no interval-scale assumption about distances between response categories, (2) greater variety of models, and (3) greater statistical power for testing effects, because the model can focus the effect on small degrees of freedom as ranking reduces the total number of variables in the model. The ordered model creates cutoffs that are used to estimate the conditional probability of the estimated value to fall under a certain ordinal response (life outcome) given the parameters (severity and treatment) in the latent variable.

However, there are some limitations that come with these models. The ordered logit model falls susceptible to a few inconsistent estimations of the parameters if any of the following exist: (1) omitted variables even if they are orthogonal to the regressors, (2) heteroscedasticity[11], (3) incorrect distribution assumption, (4) endogeneity[12], and/or (5) omission of latent heterogeneity. These potential problems will be further discussed in more detail later with the results of the model section.

[11] Heteroscedasticity when the standard deviations of a variable are non-constant over a specific amount of time.

[12] A parameter or variable is said to be endogenous when there is a correlation between the parameter or variable and the error term. This can be caused from, for example, measurement error or omitted variables.

2014 HCUP National Inpatient Sample

This study examines the doctor's decision-making when choosing a treatment, considering the severity of the patient and the outcome for TBI using data from the 2014 AHRQ-HCUP National Inpatient Sample (NIS). According to the HCUP (2016), the NIS is a "database of hospital inpatient stays derived from billing data submitted by hospitals to statewide data organizations across the U.S. These inpatient data include clinical and resource use information typically available from discharge abstracts." The 2014 NIS sampling frame consists of 44 States and the District of Columbia, covers more than 96 percent of the U.S. population, and includes more than 94 percent of discharges from U.S. community hospitals. The 2014 NIS sample is divided across four files: inpatient core file[13], hospital weights file[14], disease severity measures file[15], and diagnosis and procedure groups file[16]. For this paper, the inpatient core file, the disease severity measure file, and the diagnosis and procedure groups file were used.

HCUP-NIS Potential Measurement Problems

There are several potential measurement problems with NIS such as double-counting patient records, handling missing values, and handling variance. Since NIS records are discharges, patients could contribute multiple entries to the study if hospitalized more than once

[13] The inpatient core file contains a sample of hospital discharge records from participating States, recorded as an inpatient stay record.

[14] The hospital weights file contains one observation for each hospital included in the NIS, used for weighting and variance estimation purposes.

[15] The disease severity measures file contains information about sets of disease severity measures in terms of inpatient stay record to be used with the inpatient core file.

[16] The diagnosis and procedure groups file contains data elements based on the ICD-9-CM diagnostic and procedure information in units of an inpatient stay record. The ICD-9-CM (International Classification of Diseases, Ninth Revision, Clinical Modification) is the U.S. health system's adaptation of international ICD-9 standard list of six-character alphanumeric codes to describe diagnoses.

within the year. This measurement problem can also be considered from the transfer in and out of participating hospitals. If the patient is transferred into a participating hospital, they could be recorded twice within the study, which could lead to overestimation. However, if the patient is transferred into a non-participating hospital and dies there, the study could be missing a record of a key outcome. Additionally, if the outcome of discharges with missing values is different from the outcome for discharges with valid values, then sample estimates for that outcome will be biased and inaccurately represent the discharge population.

Finally, variance estimates must consider both the sampling design and the form of the statistic. For example, discharges are clustered by hospital and should be treated as clusters when calculating statistics. In most cases, standard formulas for a stratified, single-stage cluster sample without replacement should still be used. Due to the sample design, any estimates that attempt to accurately describe characteristics and interrelationships among hospitals and discharges during a specific year should be governed by finite-sample theory[17].

HCUP-NIS Limitations

The NIS is the largest all-payer inpatient care database in the U.S., but still faces some limitations in regards to clinical detail, patient tracking, and hospital billing. First, the records lack clinical detail, such as stage of the disease and patient's vitals data, and laboratory and pharmacy data. The clinical detail depends on the how the physician interprets and records the ICD-9-CM codes, and it is subject to changing coding practices.

[17] Finite-sample theory suggests that the intent of the estimation process is to obtain estimates that are precise representations of the national population at a specific point in time.

Additionally, the NIS contains discharge-level records, not patient-level records, so patients cannot be tracked throughout the year if admitted multiple times. This can impact studies on conditions that require hospitalization multiple times in a single year. Patients cannot be tracked across years and studies cannot be broken down by state for analysis.

HCUP also faces some limitations in terms of hospital billing. Some limitations of hospital billing data are the limited clinical details and lack of reimbursed claims information. HCUP also does not include all hospital types, does not show a complete episode of care, does not have data on individuals outside of the hospital system, cannot link national databases to external sources, and faces differences in coding across hospitals.

Chapter 4

Results

The final model chosen as the optimal fit was a cumulative ordered logit model for the first phase of the model and a conditional logit discrete choice model for the second phase of the model. Three models were fitted for phase one of the model and two models were fitted for phase two of the model. The three models tested for phase one were a cumulative ordered probit, cumulative ordered logit, and a multinomial logit model. The two models tested for phase two were a conditional logit discrete choice model and a multinomial probit discrete choice model. The other model variations results can be found in Appendix C.

Preliminary Statistics

Due to missing values under patient discharge status (DISPUNIFORM), 15,008 of the 15,313 TBI patient records were evaluated in this study using 2014 HCUP-NIS data. The severity variable has a mean value of 4.211. This suggests a left skew, since a uniform distribution would have a mean value of 3 {(8-2)/2 = 3}. The treatment variable has a mean value of 0.217. This suggests a right skew, since a uniform distribution would have a mean value of 0.5 {(1-0)/2 = 0.5}. Finally, the outcome variable has a mean value of 1.596. This suggests a left skew, since a uniform distribution would have a mean value of 1 {(2-0)/2 = 1}. This means the data set reflects on average, less severe cases to more severe cases of TBI, more medical treatment decisions than surgical treatment decisions, and more favorable outcomes (cured and semi-cured) than less favorable death outcomes. These results can be seen in Table 4.

Table 4: Preliminary Variable Statistics as Observed in the 2014 NIS Sample

Variable	N	Mean	Std Dev	Minimum	Maximum
Severity	15313	4.2111931	2.1747548	2.0	8.0
Treatment	15313	0.2177888	0.4127565	0.0	1.0
Outcome	15008	1.5966151	0.6383863	0.0	2.0

Phase 1: Cumulative Ordered Logit Model

The cumulative ordered logit provided the best fit for the first phase of the model due to intuitive reasoning and generally similar results in output. However, all models did not produce a strong fit overall. In the cumulative logit, the variable Outcome was used as the response variable with three levels. The optimization technique used was Fisher's scoring. In the model, Outcome = 0 was used as the base level and the probabilities modeled are cumulated over the lower ordered values. Additionally, the model did satisfy convergence criterion[18] and thus, can be interpreted for statistical significance.

The score test for the proportional odds assumption leads us to reject the null hypothesis in favor of the alternative (Chi-square = 440.9609, DF = 2, Pr > Chi-Square: < 0.0001), suggesting that the ordered logit coefficients are not equal across the levels of the outcome response variable and a less restrictive model may result in a better fit. A less restrictive model such as a multinomial logit would provide a less restrictive fit. This model was tested and did not provide stronger results. Additionally, a less restrictive model does not make intuitive sense, since life outcomes has a clear ordering in practice. A less restrictive model provides an opportunity for a rank such as Cured > Death > Semi-cured or Semi-cured > Cured > Death.

[18] When the convergence criterion converges, this means that the maximum likelihood algorithm converged. In this case, the maximum likelihood algorithm was Fisher's scoring.

Since a less restrictive multinomial logit model violates a clear assumption in a doctor-patient relationship, the cumulative logit model was used.

Next, the large values across the model fit statistics (Intercept Only: AIC[19] = 24,330.175, SC[20] = 24,345.407, -2 Log L = 24,326.175; Intercept and Covariates: AIC = 18,514.392, SC = 18,544.858, -2 Log L = 18,506.392) suggest this model has a weak fit. Typically, the criterion produces large values due to penalties from a large quantity of predictor variables and/or number of levels of the response variable. However, due to the low quantity of both the predictor variables and levels of the response variable, the large values are being generated from a weak fit. However, the lower criterion values with the intercepts and covariates suggest that the model is fitted better with those included.

Additionally, testing global null hypothesis, or the assumption that all of the predictors' regression coefficients are equal to zero in the model, leads us to reject the null hypothesis in favor of the alternative. This suggests that at least one of the predictors' regression coefficients is not equal to zero. This can be seen in Table 5.

Next, looking at the maximum likelihood estimates, note that Intercept 2 and Intercept 1 are the estimated ordered logits for the adjacent levels of the dependent variable cured v. semi-cured and death, and cured and semi-cured v. death, respectively, when the independent variables are evaluated at zero. The first intercept β_0 was set to zero as the baseline, or first threshold. Intercept 2 is the estimated log odds for cured versus semi-cured and death outcomes

[19] Akaike's information criterion (AIC) compares the quality of a set of statistical models to each other. The AIC will take each model and rank them from best to worst. The "best" model will be the one that neither under-fits nor over-fits.

[20] The Bayesian information criterion (BIC) or Schwarz criterion (SC) is a criterion for model selection among a finite set of models. When fitting models, it is possible to increase the likelihood by adding parameters, but doing so may result in overfitting. The BIC resolves this problem by introducing a penalty term for the number of parameters in the model. The penalty term is larger in BIC than in AIC.

Table 5: Cumulative Logit Global Hypothesis Test

Testing Global Null Hypothesis: BETA = 0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	5819.7824	2	< 0.0001
Score	5264.8123	2	< 0.0001
Wald	4544.3337	2	< 0.0001

when the predictor variables are evaluated at zero. The log-odds of cured v. semi-cured and death outcomes for medical treatment and severity score 0 is 3.8324. Intercept 1 is the estimated log odds for cured and semi-cured v. death outcome when the predictor variables are evaluated at 0. The log-odds of cured and semi-cured v. death outcomes for medical treatment and severity score 0 is 6.1284. Additionally, if a patient were to increase his or her severity score by one point, his or her outcome score would be expected to result in a 0.6498 unit decrease in the order log-odds scale while the other variables in the model are held constant. If a doctor were to use surgical treatment, or increase a patient's treatment score by one point, the patient outcome would be expected to result in a 0.5391 unit decrease in the ordered log-odds scale while the other variables in the model are held constant. This means both an increase in severity and treatment would decrease the likelihood of a cured patient outcome. Finally, using the Wald Chi-Square test statistic[21], we would reject the null hypothesis in favor of the alternative and conclude that the regression coefficient for Severity and Treatment have been found to be statistically different from zero in estimating Outcome. These estimates can be found in Table 6.

Evaluating the odds ratio estimates, for a one unit increase in severity, the odds of cured versus the combined semi-cured and death outcomes are 0.522 times greater, given that all of the other variables in the model are held constant. Likewise, the odds of the combined cured and

[21] The Wald Chi-square Test Statistic is the squared ratio of the estimate to the standard error of the respective predictor.

Table 6: Cumulative Logit Analysis of Maximum Likelihood Estimates

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Std Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept	2	1	3.8324	0.0543	4988.1271	< 0.0001
Intercept	1	1	6.1284	0.0713	7389.3963	< 0.0001
Severity		1	-0.6498	0.0102	4051.0494	< 0.0001
Treatment	1	1	-0.5391	0.0448	145.0822	< 0.0001

semi-cured versus death is 0.522 times greater, given that all of the other variables in the model are held constant. For a one unit increase in treatment, or the movement from medical to surgical treatment, the odds of cured versus the combined semi-cured and death outcomes are 0.583 times greater, given that all of the other variables in the model are held constant. Likewise, the odds of the combined cured and semi-cured outcomes versus death is 0.583 times greater, given that all of the other variables in the model are held constant. These estimates can be seen in Table 7.

Next, the association of predicted probabilities and observed responses all show a somewhat strong association. The Somer's D[22] test (= 0.654) and the C[23] method (=0.827) suggest a relatively strong association between ordinal variables. These association results are problematic as the conditional probabilities will be used during phase two to determine doctor utility for a given treatment based on severity.

Finally, the confusion matrix (Table 8) showed that more observations are being predicted as cured outcome classifications than the original outcome classification distribution. Overall, 71.14872% of observations were correctly classified, which is a relatively accurate classification rate for the model.

[22] Somer's D is used to determine the strength and direction of relation between pairs of variables with values ranging from -1.0 (all pairs disagree) to 1.0 (all pairs agree).

[23] C is another measure of rank correlation of ordinal variables. It ranges from 0 (no association) to 1 (perfect association).

Table 7: Cumulative Logit Odds Ratio Estimates

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
Severity	0.522	0.512	0.533
Treatment 1 vs 0	0.583	0.534	0.637

Table 8: Cumulative Logit Confusion Matrix

Output Order Frequency Percent Row Percent Column Percent	Table of Outcome by _INTO_			
	Outcome	**_INTO_ (Formatted Value of Predicted Responses)**		
		1	**2**	**Total**
	0	2,364	140	2,504
		7.88	0.47	8.34
		94.41	5.59	
		32.40	0.62	
	1	2,938	4,162	7,100
		9.79	13.87	23.65
		41.38	58.62	
		40.27	18.32	
	2	1,994	18,418	20,412
		6.64	61.36	68.00
		9.77	90.23	
		27.33	81.07	
	Total	7,296	22,720	30,016
		24.31	75.69	100.00

Phase 2: Conditional Logit Discrete Choice Model

Next, the conditional probabilities (Table 9) needed in phase two were calculated using the cumulative ordered logit model from phase one. The discrete choice model also converged and thus, can be interpreted for statistical significance. The optimization technique used was the Dual Quasi-Newton method.

This model resulted in medical treatment being chosen the majority of the time (response frequency = 77.93%). However, connecting this with the conditional probabilities, medical treatment stochastically dominates over surgical treatment. This can be seen when comparing the conditional probability of fully cured for medical treatment to the conditional probability of fully cured for surgical treatment at the same severity level. Surgical treatment is only chosen as the optimal decision when comparing death versus semi-cured patient outcomes, but surgical treatment is never chosen as the optimal treatment decision leading to a cured patient outcome. This leads to a stochastic dominance of medical treatment in doctor decision-making and drives some of the results seen in the discrete choice model. A model summary can be seen in Table 10.

Additionally, McFadden's likelihood ratio index (LRI)[24] (= 0.2281) suggests a weak fit of the model. In general, a weak fit is also suggested from the several different goodness-of-fit

Table 9: Conditional Probabilities of Outcome Given Severity and Treatment

		Conditional Probabilities of Outcome		
Treatment	**Severity**	**Death 0**	**Semi-cured 1**	**Fully cured 2**
Medical 0	2	0.0079	0.0657	0.9264
	3	0.0151	0.1169	0.8680
	4	0.0285	0.1971	0.7744
	5	0.0532	0.3049	0.6419
	6	0.0971	0.4194	0.4835
	7	0.1708	0.5009	0.3283
	8	0.2829	0.5138	0.2033
Surgical 1	2	0.0135	0.1064	0.8801
	3	0.0256	0.1813	0.7931
	4	0.0479	0.2852	0.6669
	5	0.0878	0.4011	0.5111
	6	0.1557	0.4912	0.3531
	7	0.2610	0.5172	0.2218
	8	0.4035	0.4669	0.1296

[24] McFadden's likelihood ratio index ranges from 0 (no fit) to 1 (perfect fit).

measures seen in Table 11. Looking at parameter estimates, both parameters are significant at an alpha level of 0.001. This would lead us to reject the null hypothesis in favor the alternative, suggesting that both parameters have nonzero coefficients. This can be seen in Table 12.

Table 10: Discrete Choice Model Summary

Model Fit Summary	
Dependent Variable	Decision
Number of Observations	15008
Number of Cases	30016
Log Likelihood	-8030
Log Likelihood Null (LogL(0))	-10403
Maximum Absolute Gradient	0.0000123
AIC	16064
Schwarz Criterion	16079

Table 11: Discrete Choice Model Goodness-of-Fit Measures

Goodness-of-Fit Measures		
Measures	**Value**	**Formula**
Likelihood Ratio (R)	4745.4	2*(LogL – LogL0)
Upper Bound of R (U)	20806	-2*LogL0
Aldrich-Nelson	0.2402	R / (R+N)
Cragg-Uhler 1	0.2711	1 – exp(-R/N)
Cragg-Uhler 2	0.3614	(1-exp(-R/N))/(1-exp(-U/N))
Estrella	0.3015	1 – (1-R/U)^(U/N)
Adjusted Estrella	0.3013	1 – ((LogL-K)/LogL0)^(-2/N*LogL0)
McFadden's LRI	0.2281	R/U
Veall-Zimmermann	0.4135	(R*(U+N))/(U*(R+N))
N = # of observations, K = # of regressors		

Table 12: Discrete Choice Model Parameter Estimates

Conditional Logit Parameter Estimates					
Parameter	**DF**	**Estimate**	**Std Error**	**t Value**	**Approx Pr > |t|**
PSemi_cured	1	-13.2020	0.4544	-29.05	< 0.0001
PCured	1	6.9436	0.3133	22.16	< 0.0001

Limitations of the Model

As with any model, there are some limitations with its current use and design. First, APRDRGs are used to classify the chosen illness and identify a patient's severity score. This is a design of the AHRQ-HCUP and is limited to data from their database. Therefore, with a different data set as the input, a new proxy for severity and/or doctor's best judgement would be needed to replace the current predictor variable Severity. However, the assumption that severity was an appropriate proxy for doctor's best judgement on choosing an optimal treatment could be violated in this paper's model (see Chapter 5), and thus will need further research regardless.

Next, the model does not account for the weighting in the HCUP-NIS data set. This is needed to make accurate assumptions about variance in the model as well as to account for the 5% of U.S. records that are not accounted for in this sample. This is the necessary adjustment needed for the stratified design. However, since this adjustment was not made in this paper's model due to time constraints, this would be an area of improvement and development in the model moving forward. Details about the sampling methodology and stratification can be found in Appendix A and Appendix B.

Finally, there are some limitations with the model design itself that come naturally. These were initially mentioned in Chapter 3. The ordered logit model falls susceptible to a few inconsistent estimations of the parameters if any of the following exist: (1) omitted variables even if they are orthogonal to the regressors, (2) heteroscedasticity, (3) incorrect distribution assumption, (4) endogeneity, and/or (5) omission of latent heterogeneity. In regards to omitted variables, there appears to be support for missing variables with a relatively weak overall model fit despite statistically significant parameters in the model. Heteroscedasticity would be more evident in this model if it were tracked over several years as opposed to only one year, but in this

case, it is relatively mitigated. In regards to an incorrect distribution assumption, that could be a potential error in the model, since only logit and probit models were tested. A multinomial model was also tested, which provides a less restrictive fit, but that led to some problematic logic as the order of life outcomes has a very clear ranking that the model needed to account for when calculating estimators. Endogeneity should be relatively mitigated, since there is a limited number of predictors in the model that show very weak correlation. For example, a cost variable was excluded from the model, since positive correlation between invasiveness of treatment and cost was suspected.

Chapter 5

Implications on Patient Care

Overall, these results suggest that a doctor would only receive a higher expected utility from surgical treatment at a severity level of 8. Using our current data, this would mean that surgical treatment is chosen 13.88%[25] of the time compared to medical treatment. This is about 8.19% less than the observed 22.07%[26] surgical treatments in the data.

Additionally, the relationship between the conditional probabilities of outcomes is concerning. Since the discrete choice model is normalized for death, the beta for the conditional probability for death is 0. The beta for the conditional probability for semi-cured is -13.2020, and the beta for the conditional probability for cured is 6.9436. The magnitude between the betas for semi-cured and death compared to the magnitude between the betas for cured and death, and the low standard errors of the parameters are problematic. These results show statistical significance, and they also suggest that a doctor has a stronger preference towards a patient death to a semi-cured patient than the feeling between a cured patient to a patient death. Clearly, there is a strong driver in the model through either inaccurate fit from insufficient predictor variables in phase one or a violation of utility assumptions leading to an unfavorable estimated outcome in doctor decision-making.

Expanding on prior sentiments, the two key assumptions of the model are the following: (1) doctors maximize a Von Neumann-Morgenstern utility function[27] that depends only on, and

[25] See Table 3 for counts. (Total Medical and Surgical Treatment at Severity 8)/N = 2,083/15,008 = 13.88%

[26] See Table 3 for counts. (Total Surgical Treatment)/N = 3,312/15,008 = 22.07%

[27] A Von Neumann-Morgenstern utility function shows that when a consumer is faced with a choice of outcomes subject to various levels of chance, the optimal decision will be the one that maximizes the expected value of the utility derived from the choice made. Expected value is the sum of the products of the various utilities and their associated probabilities. The consumer is expected to be able to rank the outcomes in terms of preference, but the expected value will be conditioned by their probability of occurrence (Britannica, 2018).

that is increasing in, the patient's life outcome, and (2) severity, as measured in the data, is a sufficient statistic for the doctor's best judgement about the appropriate treatment, given all of the evidence that the doctor sees that is not recorded in the data. Either or both of these assumptions must have been violated for a such a result to appear in the model. Both assumptions will be further inspected for possible explanations driving their invalidity in the model.

Assumption 1: Doctors Maximize a Von Neumann-Morgenstern Utility Function

This section will explore potential reasons that the assumption that doctors maximize a Von Neumann-Morgenstern utility function that depends only on, and that is increasing in, the patient's outcome was violated. To violate this assumption, either doctors are considering something, such as profitability, other than patients' welfare in mind, or doctors were following what they had been taught in medical school, during residency, and/or early in their career. If doctors were following earlier practices in their career, they may not be accounting for available evidence on the relative values of the various treatment options today. Additionally, this assumption could be violated through the inconsistency in patient care across regions. This was mentioned briefly in Chapter 3. Additionally, this can be seen through company involvement in selecting provider networks[28] for their employees in order to choose high quality doctors that are not influenced by misaligned financial incentives.

[28] For example, Walmart requires employees to travel to specific hospitals for certain treatments, such as spinal surgery, in an effort to lower health care costs. It will pay for surgeries at well-known facilities, such as Mayo Clinic and Geisinger, and will cut its budget for procedures at clinics closer to employees' homes. Walmart is trying to offset health care cost and encourage employees to choose quality treatment by offering to pay for procedures and travel costs. Walmart hopes that this will cut down on the costs of some unnecessary procedures. Walmart is one example of companies making a switch like this (Wall Street Journal, 2018).

Therefore, a new question for this paper, but one that other researchers continue to study, persists: is there a regional disparity in quality of care in the United States? Researchers have been interested in this question as it highlights a clear violation of the assumption that doctors make treatment decisions based on maximizing patient welfare. This sentiment has been questioned in a study showing that there is a large variation across regions in pediatric Ears, Nose, and Throat (ENT) surgeons' likelihood to perform tonsillectomies. During a speech to Congress in September 2009, President Obama reflected on these concerns through his following statement:

> Part of what we want to do is to make sure that those decisions are being made by doctors and medical experts based on evidence, based on what works…. Right now, doctors a lot of times are forced to make decisions based on the fee payment schedule that's out there. ... the doctor may look at the reimbursement system and say to himself, 'You know what? I make a lot more money if I take this kid's tonsils out … I'd rather have that doctor making those decisions based on whether you really need your kid's tonsils out, or whether …something else would make a difference…So part of what we want to do is to free doctors, patients, hospitals to make decisions based on what's best for patient care. (AAO-HNS, 2009).

President Obama's statement brings up concerns that surgeons in certain regions prescribe their treatments based on revenue maximization rather than maximization of patient welfare through evidence-based medicine. In response to President Obama's statement, the American Academy of Otolaryngology - Head and Neck Surgery (AAO-HNS) refutes the suggestion that surgeons act based on revenue maximization by highlighting the inadequate financial incentive of surgical intervention. AAO-HNS states that tonsillectomy reimbursement

ranges from approximately $180-$300 across all payers, and this reimbursement rate includes the follow-up care after the procedure. Using this reimbursement, doctors also must pay malpractice insurance costs, overhead costs, such as staff salaries and benefits, and utilities (AAO-HNS, 2009). However, AAO-HNS does not refute regional disparity in treatment, but instead, attributes this variation to differences in the medical practice and training of medical staff across the country in the management of recurrent tonsillitis and other conditions affecting the upper airway. This would suggest that surgeons are not following evidence-based medicine and instead are following their earlier training. This is contributing to regional disparity across the U.S. and the violation of assumption one.

Additionally, there was another study on regional disparity (Seabury, 2017) in the quality of care for stroke patients. Regional variations in care is especially pronounced in stroke care. Rural and underserved areas have the highest prevalence of strokes, but also have the worse care and patient life outcomes. In these areas, there is often a lack of clinical expertise that can prevent stroke-related disability or mortality. Clinical expertise is required for swift action in treatment, but sometimes, the distance to that type of care can be over an hour away. In order to receive quick treatment, a neurological consult is typically required. However, there is often a lack of neurological services in these areas, too. Therefore, improving access to neurological services could improve patient outcomes and care in rural and underserved areas.

Seabury's study (2017) confirmed their suspicions about lack of stroke-related services. Seabury found a positive association between quality of stroke care and Joint Commission stroke certification status regardless of geographic area; despite the positive association regardless of geographic area, there were treatment gaps evident between rural and urban settings, specifically

highlighted through the differences in treatment between certified and non-certified hospitals when comparing large metropolitan area hospitals and non-metropolitan area hospitals.

To help reduce this treatment gap, implementing telemedicine in low-access hospitals could be a more realistic alternative to hiring a full-time neurologist during a tight labor market with a shortage of neurologists. This would improve access to specialist services and provide higher quality care to patients. Additionally, distance continues to be a problem in addressing disparities of care with underserved areas. Mobile stroke units, ambulances equipped with CT scanners, and on-board or remote neurological services could reduce distance-related obstacles (Seabury, 2017). Finally, since strokes require similar neurological expertise and swift treatment as traumatic brain injuries, this study could serve as a proxy in highlighting similar reasons for assumption violations shown in this paper's model.

Overall, the regional disparity hypothesis could be tested using the same HCUP-NIS data set, and could be the next direction of research following this paper. For example, if regional disparity does not exist, the frequency of tonsillectomies to medicinal treatment should be approximately uniform across regions given a uniform mix of inflamed-tonsil cases across regions. If the distribution of severity is consistent across regions, then surgery is the appropriate treatment in some of those cases and should be reflected consistently in the frequency of tonsillectomies across regions. However, if there is a wide disparity across regions for this treatment, then in certain regions, doctors are either prescribing too many or too few tonsillectomies for patient welfare maximization. This could be tested across different illnesses to identify the largest areas for improvements and to continue building support for this case.

Assumption 2: Severity is a Sufficient Statistic for a Doctor's Best Judgement

This section will explore potential reasons that the assumption that severity, as measured in the data, is a sufficient statistic for the doctor's best judgement about the appropriate treatment, given all of the evidence that the doctor sees that is not recorded in the data, is violated. Despite evidence for a violation of assumption one, this paper considered these assumptions jointly, and therefore, a violation in assumption two could lead to the inaccurate fit of the model even if assumption one was not violated. To better understand if severity in the data set is an appropriate statistic to represent the doctor's best judgement about the appropriate treatment, we will explore how APRDRG severity scores are calculated. If these calculations accurately account for comorbidities and secondary diagnoses[29] that influence decision-making, then it would be appropriate to assume this variable is a sufficient proxy for doctor's judgement. If not, then the severity score is not a good proxy given other critical factors presented to a doctor, and those critical factors should be incorporated into the model as a dummy variable.

The information on APRDRGs derivation was received from AHRQ, which created this methodology. APRDRG have four severity levels with patient age used in severity leveling. APRDRGs are divided based on primary diagnosis. These are then subdivided further for pediatric and mortality distinctions. This leads to the final APRDRG classification. Next, these classifications are assigned one of four severity of illnesses subclasses and one of four risk of mortality subclasses. APRDRGs are frequently updated. Since its original release in December 1990, APRDRGs have had regular major clinical updates (typically every three to five years),

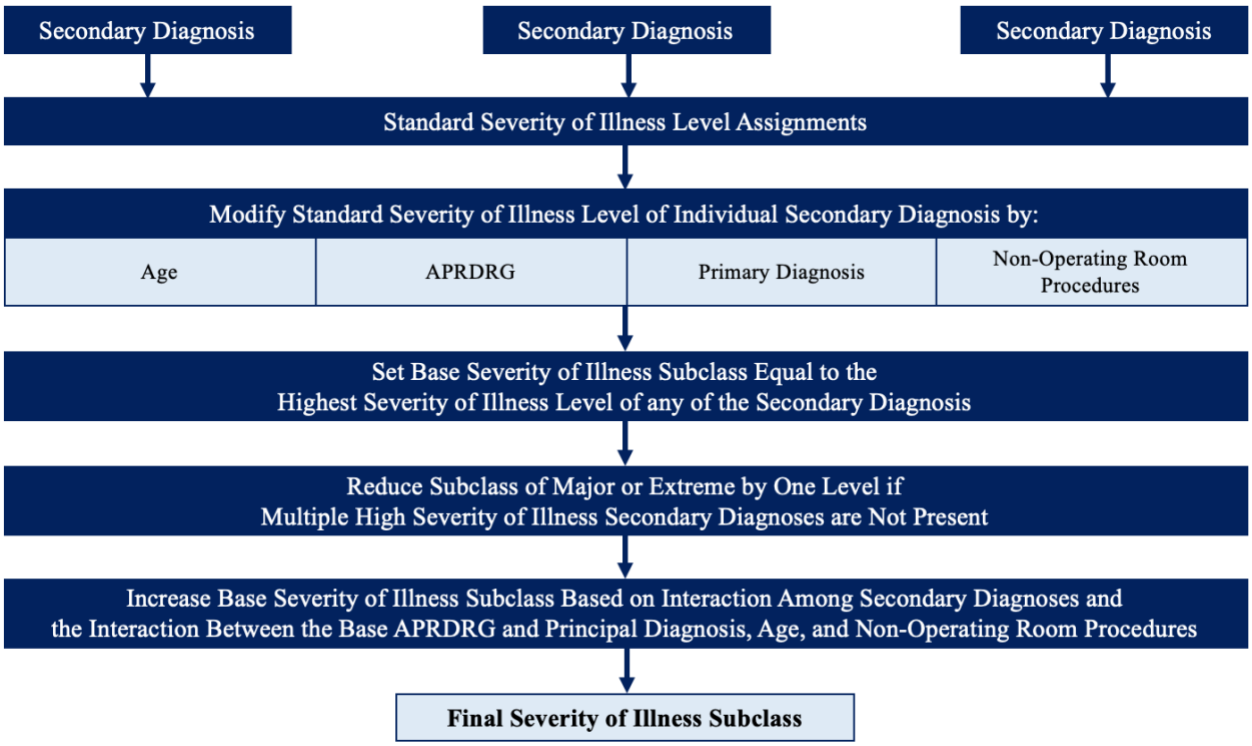
[29] The primary diagnosis is defined as "the condition, after study, which caused the admission to the hospital." (ICD-10-CM Official Guidelines for Coding and Reporting, 2016). Secondary diagnoses are "conditions that coexist at the time of admission, that develop subsequently, or that affect the treatment received and/or length of stay." (Uniform Hospital Discharge Data Set, 2016).

and version numbers correspond to the ICD-9-CM in which the APRDRGs are written to maintain consistency with Medicare DRGs. These codes are also updated every October to account for all ICD-9-CM code modifications. Therefore, year-to-year accuracy in APRDRGs classifications are not a concern. Additionally, APRDRGs have undergone some of the most intensive scrutiny of any severity system on the market with open, publicly available logic behind the development. The system was designed to be comprehensive and account for all payers, patients, and ages. As a result, APRDRGs are better at aligning payment and resource use. APRDRGs also allow payment and quality to be integrated through the use of tools that monitor complications and readmissions that occur as a result of the services and planning that are part of a patient's hospital stay and impending discharge.

Severity of illness is defined as the extent of physiologic decomposition or organ system loss of function, and the risk of mortality is defined as the likelihood of dying. These scores are disease specific and reflect the patient's underlying characteristics. The process flow of APRDRG severity of illness subclass assignment can be seen in Figure 1. A similar process flow is used to calculate the APRDRG risk of mortality subclass assignment. The design shows how the presence of multiple comorbid conditions in combination increases the severity of illness for a patient and the increase in severity accurately reflected the increased difficulty and costs involved in treating the patient.

Overall, the APRDRG severity of illness classification does account for the other diagnoses and complications that could arise in choosing appropriate treatment for traumatic brain injuries. However, with any classification model, the score assigned to each primary and secondary diagnoses face human error in reporting as the classification system could have slight variations in how doctors report them. Additionally, missing secondary diagnoses that could

Figure 1: Derivation of APRDRG Severity of Illness Subclass



increase the severity scores could lead to an inaccurate assignment. Finally, the severity scores are a clinical classification model and not a statistical classification model. Since the severity score is adjusted for the given primary diagnosis, the model developed in this paper may need to rank the primary diagnosis by severity as an additional predictor variable to account for the severity of the primary diagnosis on its own (For example, APRDRG 56 < 55 < 20 < 910 in terms of severity level of traumatic brain injuries). Therefore, in terms of a statistical model, this use of the severity level could lead to an inaccurate fit for doctor's best judgement about optimal treatment. Therefore, assumption two could also be violated, but would require more exploration in other proxy variables to build a stronger case for a violation in this assumption. This opens another opportunity for further research in more appropriate measures of doctor's judgement in treatment decision-making.

Next Steps

The goal of this paper was to better understand the following: (1) from the perspective of the doctor and patient, what is the right severity threshold at which to prescribe the most invasive treatment; and (2) from the insurer perspective, can we estimate from the decisions made by a doctor over time, what is the threshold used by that doctor to prescribe the most invasive treatment? The aim was to better understand these inquiries through the development a general model that combined an ordered logit regression and a discrete choice model to understand a doctor's utility from medical versus surgical treatment based on a patient's severity level and life outcome. This paper developed such a model by utilizing the Healthcare Cost and Utilization Project (HCUP), Agency for Healthcare Research and Quality National Inpatient Sample from 2014, and by testing the model on traumatic brain injuries as the chosen illness.

This study was successful in developing a methodology that could be applied to other illnesses. However, the ordered logit model in phase one has a relatively weak fit overall, but shows statistically significant parameters. This suggests a relationship between the parameters, but suggests that the model could also be missing a parameter that would better encompass the doctor's best judgement when choosing optimal treatment for a given patient.

Additionally, the conditional logit discrete choice model in phase two also shows a relatively weak fit overall, but shows statistically significant parameters. This could be driven by the conditional probabilities from phase one that show stochastic dominance for medical treatment. This stochastic dominance does not hold under the raw data set (violated at severity level 4), which could suggest a weakness in the developed model. This model also shows a decrease in surgical treatment chosen and suggests that surgical treatment will only be chosen at the highest patient severity level.

This model still needs further development by examining model assumption violations. The two model assumptions that were violated are the following: doctors maximize a Von Neumann-Morgenstern utility function that depends only on, and that is increasing in, the patient's outcome, and severity, as measured in the data, is a sufficient statistic for the doctor's best judgement about the appropriate treatment, given all of the evidence that the doctor sees that is not recorded in the data. Therefore, the next steps for research would be to focus on regional disparity in quality of care, and critical factors used in doctor decision-making.

For regional disparity, it would be important to focus on the drivers causing differences in care across regions. Is it access to care and/or clinical expertise that are increasing the likelihood of favorable outcomes in one region over the other? Where are the largest gaps in treatment across regions and how can we work to close some of those gaps?

Next, for doctor decision-making, there should be clearly laid out decision-trees similar to those utilized in other countries such as the National Institute for Health and Care Excellence (NICE) in the United Kingdom. This would provide a structured guide that follows evidence-based medicine for doctors to follow. It could also help to identify areas where doctors deviate and choose to follow their early training compared to current evidence-based clinical recommendations.

Appendix A

HCUP-NIS Sampling Methodology

The HCUP NIS uses five hospital characteristics (ownership/control, bed size, teaching status, urban/rural location, and the nine U.S. census divisions) to divide U.S. community hospitals into strata. The U.S. community hospitals are considered all non-Federal, short-term, general, and other specialty hospitals, excluding hospital units of institutions. Control is divided into government non-Federal (public), private not-for-profit (voluntary), and private investor-owned (proprietary). Bed size is broken into small, medium, or large categories so that approximately one-third of the hospitals in a given region, location, and teaching status combination would fall within each bed size category. Teaching status is considered as it can influence payments for services. Finally, the census division is broken into New England, Middle Atlantic, East North Central, West North Central, South Atlantic, South Central, Mountain, and Pacific to account for practice patterns that vary across regions.

This sampling design collects a sample of discharges drawn from all hospitals in the hospital universe. The sampling design resembles simple random sampling in a more efficient way by using a self-weighted sample design to ensure that the sample is representative of the population on the following factors: hospital, census division of hospital, hospital ownership, urban-rural location of hospital, hospital teaching status, number of beds in the hospital, diagnosis-related group (DRG) for the hospital stay, and admission month of the hospital stay (2014 Introduction to the NIS, 2016).

Within each stratum all discharges are sorted in the following order on patient-level control variables: encrypted hospital ID, DRG, admission month, and a random number. Within each stratum, discharges are sorted by hospital number, and then within each hospital, discharges are sorted by their DRG and their admission month. Additionally, within each stratum, a number of discharges proportionate to the number of discharges in the universe are selected from the sorted list. To ensure a self-weighted sample that has twenty percent of the universe within each stratum represented, sampling rates will vary within each stratum, depending on the proportion of the population of discharges covered by the discharges in the sampling frame. Therefore, the sampling rate is not always twenty percent within each stratum, but instead, the number of sampled discharges will equal twenty percent of the universe within each stratum (2014 Introduction to the NIS, 2016).

Appendix B

HCUP-NIS Stratum Calculations

The universe count of discharges was estimated within each stratum using the actual count of discharges contained in HCUP data, except for strata with HCUP hospitals that were open for the entire year but contributed less than a full year of data to HCUP. For these instances, the number of observed discharges was adjusted by a factor of 12 divided by the number of months for which the hospital contributed discharges to HCUP. Each discharge weight is equal to the number of target universe discharges that each sampled discharge represents in its stratum. Within stratum s, each NIS sample discharge's universe weight was calculated as:

$$DW_s(\text{Universe}) = DN_s(\text{Universe}) \div DN_s(\text{Sample}),$$

where DW_s(Universe) was the discharge weight, DN_s(Universe) was the number of discharges from community hospitals in the universe within stratum s, and DN_s(Sample) was the number of discharges selected for the NIS. Therefore, each discharge weight is equal to the number of universe discharges it represents in stratum *s* during that year. Note that these weights were not added in this study and would need added in future development of this model.

Appendix C

Tested Model Variations

Phase 1: Cumulative Ordered Probit Model

Model Information	
Data Set	THESIS.NIS_2014_ADDITIONS_SUBGROUP
Response Variable	Outcome
Number of Response Levels	3
Model	cumulative probit
Optimization Technique	Fisher's scoring

Number of Observations Read	15313
Number of Observations Used	15008

Response Profile		
Ordered Value	Outcome	Total Frequency
1	0	1252
2	1	3550
3	2	10206

Class Level Information		
Class	Value	Design Variables
Treatment	0	0
	1	1

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Score Test for the Equal Slopes Assumption		
Chi-Square	DF	Pr > ChiSq
289.9507	2	<.0001

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	24330.175	18301.220
SC	24345.407	18331.685
-2 Log L	24326.175	18293.220

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	6032.9551	2	<.0001
Score	5540.3048	2	<.0001
Wald	5315.4880	2	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Severity	1	4668.2447	<.0001
Treatment	1	145.5756	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	0	1	-3.5924	0.0372	9311.4678	<.0001
Intercept	1	1	-2.2547	0.0285	6247.6256	<.0001
Severity		1	0.3800	0.00556	4668.2447	<.0001
Treatment	1	1	0.3113	0.0258	145.5756	<.0001

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	79.2	Somers' D	0.654
Percent Discordant	13.7	Gamma	0.704
Percent Tied	7.1	Tau-a	0.311
Pairs	53453812	c	0.827

Phase 1: Multinomial Logit Model

Model Information	
Data Set	THESIS.NIS_2014_ADDITIONS_SUBGROUP
Response Variable	Outcome
Number of Response Levels	3
Model	generalized logit
Optimization Technique	Newton-Raphson

Number of Observations Read	15008
Number of Observations Used	15008

Response Profile		
Ordered Value	Outcome	Total Frequency
1	0	1252
2	1	3550
3	2	10206

Class Level Information		
Class	Value	Design Variables
Treatment	0	0
	1	1

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	24330.175	17950.359
SC	24345.407	17996.057
-2 Log L	24326.175	17938.359

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	6387.8159	4	<.0001
Score	5668.5804	4	<.0001
Wald	3118.0352	4	<.0001

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Severity	2	2734.1974	<.0001
Treatment	2	194.6959	<.0001

Analysis of Maximum Likelihood Estimates							
Parameter		Outcome	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	1	8.2842	0.3054	735.6285	<.0001
Intercept		2	1	11.5186	0.3069	1408.8402	<.0001
Severity		1	1	-1.0614	0.0408	677.0926	<.0001
Severity		2	1	-1.5290	0.0414	1361.0254	<.0001
Treatment	1	1	1	0.1117	0.0744	2.2554	0.1331
Treatment	1	2	1	-0.5857	0.0799	53.6796	<.0001

Odds Ratio Estimates				
Effect	Outcome	Point Estimate	95% Wald Confidence Limits	
Severity	1	0.346	0.319	0.375
Severity	2	0.217	0.200	0.235
Treatment 1 vs 0	1	1.118	0.966	1.294
Treatment 1 vs 0	2	0.557	0.476	0.651

Phase 2: Multinomial Probit Discrete Choice Model

Model Fit Summary	
Dependent Variable	Decision
Number of Observations	15008
Number of Cases	30016
Log Likelihood	-8066
Log Likelihood Null (LogL(0))	-10403
Maximum Absolute Gradient	1.10426E-6
Number of Iterations	8
Optimization Method	Dual Quasi-Newton
AIC	16136
Schwarz Criterion	16151
Number of Simulations	100
Starting Point of Halton Sequence	11

Discrete Response Profile			
Index	CHOICE	Frequency	Percent
0	1	11696	77.93
1	2	3312	22.07

Goodness-of-Fit Measures		
Measure	Value	Formula
Likelihood Ratio (R)	4673.4	2 * (LogL - LogL0)
Upper Bound of R (U)	20806	- 2 * LogL0
Aldrich-Nelson	0.2375	R / (R+N)
Cragg-Uhler 1	0.2676	1 - exp(-R/N)
Cragg-Uhler 2	0.3568	(1-exp(-R/N)) / (1-exp(-U/N))
Estrella	0.2972	1 - (1-R/U)^(U/N)
Adjusted Estrella	0.297	1 - ((LogL-K)/LogL0)^(-2/N*LogL0)
McFadden's LRI	0.2246	R / U
Veall-Zimmermann	0.4087	(R * (U+N)) / (U * (R+N))
N = # of observations, K = # of regressors		

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
PSemi_cured	1	-11.1668	0.3888	-28.72	<.0001
PCured	1	5.5984	0.2664	21.02	<.0001

Appendix D

SAS Model Code

```
/*************************************************************************
1. Joins the core with the severity
2. Joins the core + severity with the dx pr groups
3. Makes Treatment a dummy variable where Medical APRDRG={55,56} is the
        base level 0 and Surgical APRDRG={20,910} is 1, else '.'
4. Makes Outcome a ordered categorical variable where Death DISPUNIFORM={20}
        is the base level 0, Semi-cured DISPUNIFORM={5,6} is 1, and Cured
        DISPUNIFORM={1,2} is 2.
5. Makes Severity a new variable as APRDRG_Severity + APRDRG_Risk_Mortality
        excluding 0 and missing values.
6. Makes tables with APRDRG restrictions (and age restrictions) with new
columns
7. Understanding the Data Before Regressions (PROC MEANS & PROC FREQ)
8. Runs logistic function to get cut-offs
9. Prints Actual-by-Predicted Table (Confusion Matrix)
10. Creates the conditional probabilities (from the cumulative logit desc)
      New Variables:    pdeath1     pdeath2     psemi1      psemi2
                        pcured1     pcured2
      Treatment changes: Medical = 1 and Surgical = 2;
11. Puts data in format needed for Proc MDC ...doubles entries
12. Creates the Discrete Choice Model (Binary Conditonal Logit)
13. Forecasts Choice Probabilities
*************************************************************************/

LIBNAME thesis 'X:\THESIS\';

/*1. Joins Core with Severity*/
PROC SQL;
      title 'NIS 2014 Data with Severity Measures';
      create table thesis.nis_2014_core_severity as
      select a.*, b.APRDRG, b.APRDRG_Risk_Mortality, b.APRDRG_Severity
      from thesis.nis_2014_core as a, thesis.nis_2014_severity as b
      where a.KEY_NIS = b.KEY_NIS;
QUIT;

/*2. Joins Core + Severity with Diagnosis and Procedure Groups*/
PROC SQL;
      title 'NIS 2014 Data with Diagnosis & Severity';
      create table thesis.nis_2014_core_severity_dx_pr as
      select a.*,       b.DXMCCS1, b.PRMCCS1, b.PCLASS1,  b.PCLASS2,
b.PCLASS3,  b.PCLASS4, b.PCLASS5, b.PCLASS6,
                  b.PCLASS7, b.PCLASS8, b.PCLASS9, b.PCLASS10, b.PCLASS11,
b.PCLASS12, b.PCLASS13, b.PCLASS14, b.PCLASS15
      from thesis.nis_2014_core_severity as a, thesis.nis_2014_dx_pr_grps as
b
      where a.KEY_NIS = b.KEY_NIS
      order by HOSP_NIS;
QUIT;

/*3. Makes Treatment into a dummy variable*/
```

```
DATA thesis.nis_2014_additions_prelim;
      set thesis.nis_2014_core_severity_dx_pr;
           if APRDRG = 20  then Treatment = 1;
      else if APRDRG = 910 then Treatment = 1;
      else if APRDRG = 55  then Treatment = 0;
      else if APRDRG = 56  then Treatment = 0;
      else                      Treatment = .;
RUN;

/*4. Makes Outcome into an ordered categorical variable*/
DATA thesis.nis_2014_additions_prelim;
      set thesis.nis_2014_additions_prelim;
            if (DISPUNIFORM = 20)AND (APRDRG=55 OR APRDRG=56 OR APRDRG=20 OR
APRDRG=910) then Outcome = 0;
      else if (DISPUNIFORM = 6)  AND (APRDRG=55 OR APRDRG=56 OR APRDRG=20 OR
APRDRG=910) then Outcome = 1;
      else if (DISPUNIFORM = 5)  AND (APRDRG=55 OR APRDRG=56 OR APRDRG=20 OR
APRDRG=910) then Outcome = 1;
      else if (DISPUNIFORM = 2)  AND (APRDRG=55 OR APRDRG=56 OR APRDRG=20 OR
APRDRG=910) then Outcome = 2;
      else if (DISPUNIFORM = 1)  AND (APRDRG=55 OR APRDRG=56 OR APRDRG=20 OR
APRDRG=910) then Outcome = 2;
      else    Outcome = .;
RUN;

/*5. Makes Severity into a new column*/
DATA thesis.nis_2014_additions_prelim;
      set thesis.nis_2014_additions_prelim;
      if    (APRDRG_Severity > 0) AND (APRDRG_Risk_Mortality > 0)
      AND   (APRDRG = 55 OR APRDRG = 56 OR APRDRG = 20 OR APRDRG = 910)
      then Severity = (APRDRG_Severity + APRDRG_Risk_Mortality);
      else Severity = .;
RUN;

/*6. Makes tables with APRDRG restrictions (and age restrictions)*/
/*15,313 records AGE < 70*/
PROC SQL;
      create table thesis.nis_2014_additions_subgroup as
      select *
      from thesis.nis_2014_additions_prelim
      where (APRDRG = 55 OR APRDRG = 56 OR APRDRG = 20 OR APRDRG = 910) AND
(AGE < 70) AND (Outcome NE .);
QUIT;

/*29,820 records AGE = ALL*/
PROC SQL;
      create table thesis.nis_2014_additions_full as
      select *
      from thesis.nis_2014_additions_prelim
      where (APRDRG = 55 OR APRDRG = 56 OR APRDRG = 20 OR APRDRG = 910) AND
(Outcome NE .);
QUIT;

/*7. Understanding the Data Before Regressions*/
PROC MEANS data=thesis.nis_2014_additions_subgroup;
      var Severity Treatment Outcome;
```

```
RUN;

PROC FREQ data=thesis.nis_2014_additions_subgroup;
      tables Severity Treatment Outcome Treatment*Severity
Outcome*Treatment*Severity;
RUN;


/*8. Runs logistic function to get cut-offs*/
/*Run 1: Ordered Probit*/
PROC LOGISTIC data=thesis.nis_2014_additions_subgroup outest=mlogit_param
plots(only)=oddsratio(range=clip);
      class Outcome (ref= '0') Treatment (ref = '0') / param = ref;
      model Outcome = Severity Treatment / link= probit covb;
      effectplot interaction(sliceby=Outcome) / individual polybar;
            output out= thesis.pred pred=predicted predprobs=i;
RUN;

/*Run 2: Multinomial Logit*/
PROC LOGISTIC data=thesis.nis_2014_additions_subgroup outest=mlogit_param
plots(only)=oddsratio(range=clip);
      class Outcome (ref= '0') Treatment (ref = '0') / param = ref;
      model Outcome = Severity Treatment / link= glogit covb;
      effectplot interaction(sliceby=Outcome) / individual polybar;
      output out= thesis.pred pred=predicted predprobs=i;
RUN;

/*Run 3: Ordered Logit*/
PROC LOGISTIC data=thesis.nis_2014_additions_subgroup outest=mlogit_param
plots(only)=oddsratio(range=clip);
      class Outcome (ref= '0') Treatment (ref = '0') / param = ref;
      model Outcome = Severity Treatment / covb;
      effectplot interaction(sliceby=Outcome) / individual polybar;
      output out= thesis.pred pred=predicted predprobs=i;
RUN;

/*Run 4: Ordered Logit, desc includes Probabilities*/
PROC LOGISTIC data=thesis.nis_2014_additions_subgroup desc;
      class Outcome (ref= '0') Treatment (ref = '0') / param = ref;
      model Outcome = Severity Treatment / covb;
      output out= thesis.pred pred=predicted predprobs=i;
      estimate "Pr prob apply=2 at Treatment=0, Severity=2" intercept 1
               Severity 2 / ilink category='2';
      estimate "Pr prob apply=2 at Treatment=1, Severity=2" intercept 1
Treatment 1 Severity 2 / ilink category='2';
      estimate "Pr prob apply=2 at Treatment=0, Severity=3" intercept 1
Severity 3 / ilink category='2';
      estimate "Pr prob apply=2 at Treatment=1, Severity=3" intercept 1
Treatment 1 Severity 3 / ilink category='2';
      estimate "Pr prob apply=2 at Treatment=0, Severity=4" intercept 1
Severity 4 / ilink category='2';
      estimate "Pr prob apply=2 at Treatment=1, Severity=4" intercept 1
Treatment 1 Severity 4 / ilink category='2';
      estimate "Pr prob apply=2 at Treatment=0, Severity=5" intercept 1
               Severity 5 / ilink category='2';
```

```
      estimate "Pr prob apply=2 at Treatment=1, Severity=5" intercept 1
Treatment 1 Severity 5 / ilink category='2';
      estimate "Pr prob apply=2 at Treatment=0, Severity=6" intercept 1
              Severity 6 / ilink category='2';
      estimate "Pr prob apply=2 at Treatment=1, Severity=6" intercept 1
Treatment 1 Severity 6 / ilink category='2';
      estimate "Pr prob apply=2 at Treatment=0, Severity=7" intercept 1
              Severity 7 / ilink category='2';
      estimate "Pr prob apply=2 at Treatment=1, Severity=7" intercept 1
Treatment 1 Severity 7 / ilink category='2';
      estimate "Pr prob apply=2 at Treatment=0, Severity=8" intercept 1
              Severity 8 / ilink category='2';
      estimate "Pr prob apply=2 at Treatment=1, Severity=8" intercept 1
Treatment 1 Severity 8 / ilink category='2';
      estimate "Pr prob apply=1 or 2 at Treatment=0, Severity=2" intercept 1
               Severity 2 / ilink category='1';
      estimate "Pr prob apply=1 or 2 at Treatment=1, Severity=2" intercept 1
Treatment 1 Severity 2 / ilink category='1';
      estimate "Pr prob apply=1 or 2 at Treatment=0, Severity=3" intercept 1
Severity 3 / ilink category='1';
      estimate "Pr prob apply=1 or 2 at Treatment=1, Severity=3" intercept 1
Treatment 1 Severity 3 / ilink category='1';
      estimate "Pr prob apply=1 or 2 at Treatment=0, Severity=4" intercept 1
Severity 4 / ilink category='1';
      estimate "Pr prob apply=1 or 2 at Treatment=1, Severity=4" intercept 1
Treatment 1 Severity 4 / ilink category='1';
      estimate "Pr prob apply=1 or 2 at Treatment=0, Severity=5" intercept 1
               Severity 5 / ilink category='1';
      estimate "Pr prob apply=1 or 2 at Treatment=1, Severity=5" intercept 1
Treatment 1 Severity 5 / ilink category='1';
      estimate "Pr prob apply=1 or 2 at Treatment=0, Severity=6" intercept 1
               Severity 6 / ilink category='1';
      estimate "Pr prob apply=1 or 2 at Treatment=1, Severity=6" intercept 1
Treatment 1 Severity 6 / ilink category='1';
      estimate "Pr prob apply=1 or 2 at Treatment=0, Severity=7" intercept 1
               Severity 7 / ilink category='1';
      estimate "Pr prob apply=1 or 2 at Treatment=1, Severity=7" intercept 1
Treatment 1 Severity 7 / ilink category='1';
      estimate "Pr prob apply=1 or 2 at Treatment=0, Severity=8" intercept 1
               Severity 8 / ilink category='1';
      estimate "Pr prob apply=1 or 2 at Treatment=1, Severity=8" intercept 1
Treatment 1 Severity 8 / ilink category='1';
RUN;

/*9. Prints Actual-by-Predicted Table (Confusion Matrix)*/
PROC FREQ data=thesis.pred;
      table Outcome*_INTO_ / out=CellCounts;
RUN;

DATA thesis.CellCounts;
      set CellCounts;
      Match = 0;
      if Outcome = _INTO_ then Match = 1;
RUN;

PROC MEANS data=thesis.CellCounts mean;
```

```
      freq count;
      var Match;
RUN;

/*10. Creates the conditional probabilities*/
/*Example of probabilities: pdeath1 is the probability of death given medical
treatment*/
DATA thesis.nis_2014_additions_subgroup_t (keep= KEY_NIS Treatment Outcome
pdeath1 pdeath2 psemi1 psemi2 pcured1 pcured2 Severity);
      set thesis.nis_2014_additions_subgroup;
      if          Treatment = 1 then Treatment = 2;
      else if Treatment = 0 then Treatment = 1;
      if          (Severity = 2) then do;
                                          pdeath1 = 0.0079; pdeath2 = 0.0134;
                                          psemi1  = 0.0657; psemi2  = 0.1064;
                                          pcured1 = 0.9264; pcured2 = 0.8801;
                                    end;
      else if     (Severity = 3) then do;
                                          pdeath1 = 0.0151; pdeath2 = 0.0256;
                                          psemi1  = 0.1169; psemi2  = 0.1813;
                                          pcured1 = 0.8680; pcured2 = 0.7931;
                                    end;
      else if     (Severity = 4) then do;
                                          pdeath1 = 0.0285; pdeath2 = 0.0479;
                                          psemi1  = 0.1971; psemi2  = 0.2852;
                                          pcured1 = 0.7744; pcured2 = 0.6669;
                                    end;
      else if     (Severity = 5) then do;
                                          pdeath1 = 0.0532; pdeath2 = 0.0878;
                                          psemi1  = 0.3049; psemi2  = 0.4011;
                                          pcured1 = 0.6419; pcured2 = 0.5111;
                                    end;
      else if     (Severity = 6) then do;
                                          pdeath1 = 0.0971; pdeath2 = 0.1557;
                                          psemi1  = 0.4194; psemi2  = 0.4912;
                                          pcured1 = 0.4835; pcured2 = 0.3531;
                                    end;
      else if     (Severity = 7) then do;
                                          pdeath1 = 0.1708; pdeath2 = 0.2610;
                                          psemi1  = 0.5009; psemi2 = 0.5172;
                                          pcured1 = 0.3283; pcured2 = 0.2218;
                                    end;
      else if     (Severity = 8) then do;
                                          pdeath1 = 0.2829; pdeath2 = 0.4035;
                                          psemi1  = 0.5138; psemi2  = 0.4669;
                                          pcured1 = 0.2033; pcured2 = 0.1296;
                                    end;
RUN;

/*11. Puts data in format needed for Proc MDC */
DATA thesis.nis_2014_additions_subgroup_t1 (keep= KEY_NIS Decision
Treatment_Option PDeath PSemi_cured PCured Severity Treatment);
      set thesis.nis_2014_additions_subgroup_t;
      array dvec{2} pdeath1 - pdeath2;
      array svec{2} psemi1 - psemi2;
      array cvec{2} pcured1 - pcured2;
```

```
      retain pid 0;
      pid + 1;
      do i = 1 to 2;
            Treatment_Option = i;
            PDeath = dvec{i};
            PSemi_cured = svec{i};
            PCured = cvec{i};
            Decision = (Treatment = i);
            output;
      end;
RUN;

/*12. Creates Discrete Choice Model*/
/*Run 1: Conditional Logit with PDeath*/
PROC MDC data=thesis.nis_2014_additions_subgroup_t1;
      model Decision = PDeath PSemi_cured PCured /
                  type=clogit
                  nchoice=2
                  optmethod=qn
                  covest=hessian;
      id KEY_NIS;
      output out=thesis.probdata pred=p;
RUN;

/*Run 2: Conditional Logit without PDeath*/
PROC MDC data=thesis.nis_2014_additions_subgroup_t1;
      model Decision = PSemi_cured PCured /
                  type=clogit
                  nchoice=2
                  optmethod=qn
                  covest=hessian;
      id KEY_NIS;
      output out=thesis.probdata pred=p;
RUN;

/*Run 3: Multinomial Probit without PDeath*/
PROC MDC data=thesis.nis_2014_additions_subgroup_t1;
      model Decision = PSemi_cured PCured /
                  type= mprobit
                  nchoice=2
                  optmethod=qn
                  covest=hessian;
      id KEY_NIS;
      output out=thesis.probdata pred=p;
RUN;

/*13. Forecasts Choice Probabilities*/
PROC PRINT data=thesis.probdata;
   var Treatment Decision p PDeath PSemi_cured PCured;
   id KEY_NIS;
RUN;
```

Appendix E

HCUP Data Partners

Thank you to the HCUP Data Partners.

Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, District of Columbia, Florida, Georgia, Hawaii, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, Wyoming

BIBLIOGRAPHY

2014 Introduction to the NIS. Healthcare Cost and Utilization Project (HCUP). (2016 December). Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/db/nation/nis/NIS_Introduction_2014.jsp.

2016 National Healthcare Quality and Disparities Report. (2018 June). Agency for Healthcare Research and Quality, Rockville, MD. Retrieved from: http://www.ahrq.gov/research/findings/nhqrdr/nhqdr16/index.html

American Academy of Otolaryngology - Head and Neck Surgery (AAO-HNS). (2009, July 22). Tonsillectomy Facts in the U.S.: From ENT Doctors [Press Release]. Retrieved from: https://www.entnet.org/content/tonsillectomy-facts-us-ent-doctors

Allen, S. (2018). 2019 Global health care outlook: Shaping the future. A report by Deloitte LLP.

Alzheimer's Association. (2019). What Is Alzheimer's? Retrieved from https://www.alz.org/alzheimers-dementia/what-is-alzheimers

Barnes, K., Isgur, B., and Judy, R. (2018). Medical cost trend: behind the numbers 2019. A report by PwC's Health Research Institute, PwC.

Bernstein SL, Verghese V, Leung W, Lunney AT, Perez I (2003). Development and validation of a new index to measure emergency department crowding. Acad Emerg Med. 10(9):938-942.

De Bekker-Grob, E. W. (2009). Discrete Choice Experiments in Health Care: Theory and

Applications. J.E. Jurriaanse Stichting, the Department of Public Health Erasmus MC, and the Erasmus University Rotterdam. Retrieved from: https://core.ac.uk/download/pdf/18513128.pdf

Devlin, N. and Parkin, D. (2004). Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. Health Econ., 13: 437-452. doi:10.1002/hec.864

DiMatteo, M., Haskard, K., & Williams, S. (2007). Health Beliefs, Disease Severity, and Patient Adherence: A Meta-Analysis. Medical Care, 45(6), 521-528. Retrieved from: http://www.jstor.org/stable/40221467

Di Saverio, S., Tugnoli, G., Catena, F., Ansaloni, L., and Naidoo, N. (2014). Chapter 2 Management of the Polytrauma Patient, *Trauma Surgery (Vol. 1) Trauma Management, Trauma Critical Care, Orthopaedic Trauma and Neuro-Trauma* (17-25). Milan: Springer.

Emergency Severity Index (ESI): A Triage Tool for Emergency Departments. Content last reviewed October 2014. Agency for Healthcare Research and Quality, Rockville, MD.

Escher, M., Perneger, T. V., & Chevrolet, J. C. (2004). National questionnaire survey on what influences doctors' decisions about admission to intensive care. BMJ (Clinical research ed.), 329(7463), 425.

Evans, M. (2018, November 14). To Curb Wasteful Health Spending, Walmart to Send Employees Traveling for Spine Surgery. The Wall Street Journal. Retrieved from: https://www.wsj.com/articles/to-curb-wasteful-health-spending-walmart-to-send-employees-traveling-for-spine-surgery-1542205164

Fichera, Eleonora & Banks, James & Sutton, Matt. (2014). Health behaviours and the patient

doctor interaction: The double moral hazard problem. Manchester Economics Discussion Paper Series.

Froom, J., Culpepper, L., Green, L., de Melker, R., Grob, P., Heeren, T., and van Balen, F. (2001). A cross-national study of acute otitis media: risk factors, severity, and treatment at initial visit. Report from the International Primary Care Network (IPCN) and the Ambulatory Sentinel Practice Network (ASPN). Journal of the American Board of Family Medicine, November 1, 2001 Vol. 14 No. 6 406-417.

Greene, W. H., and Hensher, D. A. (2009). Modeling Ordered Choices. Department of Economics, Stern School of Business, New York University, New York, NY. Retrieved from: http://pages.stern.nyu.edu/~wgreene/DiscreteChoice/Readings/OrderedChoiceSurvey.pdf

Goode, D. (n.d.). What is the hypothalamo-pituitary axis? Retrieved from http://www.endocrinesurgeon.co.uk/index.php/what-is-the-hypothalamo-pituitary-axis

Gordon, R., Burrill, S., and Chang, C. (2018). Volume- to value-based care: physicians are willing to manage cost but lack data and tools. A report by the Deloitte Center for Health Solutions, Deloitte LLP.

Hajjaj, F. M., Salek, M. S., Basra, M. K., & Finlay, A. Y. (2010). Non-clinical influences on clinical decision-making: a major challenge to evidence-based practice. Journal of the Royal Society of Medicine, 103(5), 178-87.

HCUP National Inpatient Sample (NIS). Healthcare Cost and Utilization Project (HCUP). (2012). Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/nisoverview.jsp

HCUP NIS Description of Data Elements. Healthcare Cost and Utilization Project (HCUP).

(September 2008). Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/db/vars/dxn/nisnote.jsp.

HCUP Supplemental Files for Revisit Analyses. Healthcare Cost and Utilization Project (HCUP). (November 2018). Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/toolssoftware/revisit/revisit.jsp.

Holmes, W. F., MacGregor, E. A., Sawyer, J. P. and Lipton, R. B. (2001). Information About Migraine Disability Influences Physicians' Perceptions of Illness Severity and Treatment Needs. Headache: The Journal of Head and Face Pain, 41: 343-350. doi:10.1046/j.1526-4610.2001.111006343.

Institute of Medicine (US) Roundtable on Value & Science-Driven Health Care; Yong PL, Olsen LA, McGinnis JM, editors. Value in Health Care: Accounting for Cost, Quality, Safety, Outcomes, and Innovation. Washington (DC): National Academies Press (US); 2010. 2, Stakeholder Perspectives on Value. Available from: https://www.ncbi.nlm.nih.gov/books/NBK50926/

Lundin, D. (2000). Moral hazard in physician prescription behavior. Journal of Health Economics, Volume 19, Issue 5, Pages 639-662, ISSN 0167-6296, https://doi.org/10.1016/S0167-6296(00)00033-3.

Lyu, H., Xu, T., Brotman, D., Mayer-Blackwell, B., Cooper, M., Daniel, M., Wick, E. C., Saini, V., Brownlee, S., … Makary, M. A. (2017). Overtreatment in the United States. PloS one, 12(9), e0181970. doi:10.1371/journal.pone.0181970

Mayo Clinic. (2018 June 30). Parkinson's disease. Retrieved from https://www.mayoclinic.org/diseases-conditions/parkinsons-disease/symptoms-causes/syc-20376055

NICE (2017 October 09). Head Injury Pathway. Retrieved from: https://pathways.nice.org.uk/pathways/head-injury

NICE (2014 October 23). Head Injury Quality Standard. Retrieved from: https://www.nice.org.uk/guidance/qs74

NICE (2014 January 22). Head Injury: Assessment and Early Management. Retrieved from: https://www.nice.org.uk/guidance/CG176

Risdall, J., and Menon, D. (2011 January 27). *Traumatic brain injury*. Philosophical Transactions of the Royal Society B: Biological Sciences Volume 366, Issue 1562. Retrieved from https://doi.org/10.1098/rstb.2010.0230

Ritz, D., Althauser, C., and Wilson, K. (2014). Connecting Health Information Systems for Better Health: Leveraging interoperability standards to link patient, provider, payor, and policymaker data. Seattle, WA: PATH and Joint Learning Network for Universal Health Coverage. Retrieved from: https://jln1.pressbooks.com/chapter/3-introducing-the-key-stakeholders-patients-providers-payors-and-policymakers-the-four-ps/

Russo, A., and Steiner, C. (March 2007). *Hospital Admissions for Traumatic Brain Injuries, 2004*. HCUP Statistical Brief #27. Agency for Healthcare Research and Quality, Rockville, MD. Retrieved from https://www.hcup-us.ahrq.gov/reports/statbriefs/sb27.jsp

Seabury, S., Bognar, K., Xu, Y., Huber, C., Commerford, R., and Tayama, D. (2017 September). Regional disparities in the quality of stroke care. The American Journal of Emergency Medicine, Volume 35, Issue 9, Pages 1234-1239, ISSN 0735-6757. Retrieved from: https://doi.org/10.1016/j.ajem.2017.03.046.

Solish, N. , Bertucci, V. , Dansereau, A. , Hong, H. C., Lynde, C. , Lupin, M. , Smith, K. C. and

Storwick, G. (2007). A Comprehensive Approach to the Recognition, Diagnosis, and Severity-Based Treatment of Focal Hyperhidrosis: Recommendations of the Canadian Hyperhidrosis Advisory Committee. Dermatologic Surgery, 33: 908-923. doi:10.1111/j.1524-4725.2007.33192.

Taylor, D., Bateman, E., Boulet, L-P., Boushey, H., Busse, W., Casale, T., Chanez, P., Enright, P., Gibson, P., de Jongste, J., Kerstjens, H., Lazarus, S., Levy, M., O'Byrne, P., Partridge, M., Pavord, I., Sears, M., Sterk, P., Stoloff, S., Szefler, S., Sullivan, S., Thomas, M., Wenzel, S., and Reddel, H. (2008). A New Perspective on Concepts of Asthma Severity and Control. European Respiratory Journal 2008 32: 545-554. doi:10.1183/09031936.00155307

Thomas, S., Bachman, D., Khuperkar, A., and Vadnerkar, G. (2014). Dig deep: Impacts and implications of rising out-of-pocket health care costs. A report by the Deloitte Center for Health Solutions, Deloitte LLP.

Ubel, P., DeKay, M., Baron, J., & Asch, D. (1996). Cost-Effectiveness Analysis in a Setting of Budget Constraints — Is It Equitable? Massachusetts Medical Society, N Engl J Med 1996; 334:1174-1177. doi: 10.1056/NEJM199605023341807

Vos P., and Diaz-Arrastia R. (2015). Traumatic Brain Injury. Hoboken, NJ: John Wiley & Sons.

Young, W. W., Kohler, S., & Kowalski, J. (1994). PMC Patient Severity Scale: derivation and validation. *Health Services Research, 29*(3), 367–390.

Kelly Tamburo

EDUCATION & AWARDS

The Pennsylvania State University | *Schreyer Honors College* *University Park, PA*
B.S. Risk Management, Actuarial Science option | *Smeal College of Business* *Class of 2019*
B.S. Economics | *College of Liberal Arts* | Honors Economics Program
Minor in Statistics & Mathematics | Two-Piece in Organizational Leadership
Passed SOA Exam MFE/3, FM/2, & P/1 | *Earned* SOA Exam SRM/6 & VEEs (Economics, Corp. Finance, Applied Statistics)

University of Sussex *Brighton, England*
International Summer School *Summer 2016*

Skills: Microsoft Office Products, VBA, R, SAS, SQL

- MICASU Scholarship Recipient
- Roy C. Buck Scholarship Recipient
- Anthony Buzzelli Sapphire Scholarship Recipient
- W & E Leitzell Scholarship Recipient
- 2018 International Case Competition @ Maastricht Competitor
- 1st Place 2018 Wharton Undergraduate Case Competition
- 1st Place 2017 Deloitte Case Competition (PSU)
- 2017 Deloitte Case Competition National Competitor (DU)
- Highest Ranked North American Team 2018 SDS Case Competition (New Zealand)

PROFESSIONAL EXPERIENCE

Deloitte Consulting **New York, NY**
Human Capital Summer Scholar *June 2018 – Aug 2018*

- Worked on a team of 4 to analyze a Fortune 100 financial services company's internal financial reporting for consistency & accuracy
- Validated embedded value model through single-cell testing by critically analyzing modeling techniques in comparison to formal documentation for more than 4 different retirement & insurance products
- Evaluated the landscape of a Fortune 100 company's pension plan for a hedge fund with over $10 billion in assets under management
- Shadowed a team of 3 that strategized efficient provider contracting and network development for a health plan to create a new Medicare company that leapfrogs the competition in the Florida market

Highmark Health (Blue Cross Blue Shield of Pittsburgh) **Pittsburgh, PA**
Actuarial Intern *May 2017 – Aug 2017*

- Applied analytical skills by extracting data, automating processes, & calculating relative costs between facilities for use in product & network development, pricing, & manual claim development
- Developed ranking metrics to exhaustively evaluate 6 relative cost methodologies & performed a disruption analysis to determine impact of implementation on potential savings & future business opportunities
- Investigated number discrepancies between forecasting reports & improved efficiency of current forecasting processes by reducing calculation time by 95%
- Crafted visually-appealing deliverable & summarized complex actuarial concepts into high-level key takeaways on summer project to present in front of senior management

LEADERSHIP EXPERIENCE

Nittany Consulting Group **University Park, PA**
Executive Board | Director of Corporate & Alumni Outreach *Dec 2017 – Dec 2018*

- Leveraged alumni & corporate relations to increase engagement by developing targeted networking or outreach events

Engagement Manager *Jan 2018 – May 2018*

- Managed a team of 3 to strategize ways to increase market share, decrease costs, and improve reputation of a locally-owned restaurant

Senior Consultant | Consultant Training Program (CTP) *Sept 2016 – Present*

- Produced deliverables for client to identify the problem through data analysis and offer potential solutions
- Selected from 130 applicants to participate in 12-week program to improve critical thinking & presentation skills (CTP)

Sapphire Leadership Academic Program **University Park, PA**
Executive Board | Leadership Development Captain | Symposium Chair *Jan 2016 – Dec 2017*

- Managed a team of 2 to implement a comprehensive 4-year leadership development program for 200 students
- Utilized time-management & marketing strategies to organize general body meetings & annual college-wide leadership symposium

Distinguished Leader *Aug 2015 – Present*

- Represented top 5% of students in Smeal's incoming class & placed into Sapphire-specific curriculum
- Completed 16+ hours of community service, professional development, & leadership development per semester

Actuarial Science Club **University Park, PA**
Executive Board | Community Affairs Liaison *Aug 2017 – Dec 2017*

- Restructured and streamlined executive board from 15 direct reports to a team-based structure with no more than 4 direct reports
- Spearheaded critical analysis of organizational inefficiencies such as community engagement and administrative tasks
- Developed strategic plan to improve internal inefficiencies & utilized change management techniques to create buy-in

ADDITIONAL EXPERIENCE

Morgan Academic Center, *Tutor | Introductory Microeconomics* *Aug 2016 – Present*
Smeal Allocation Board, *Board Member* *Sept 2016 – Present*
THON Committee, *Second-In-Command | Pass Leader | Revamp Chair* *Oct 2015 – Feb 2017*
Department of Economics, *Teaching Assistant | Introductory Microeconomics* *Jan 2016 – May 2016*