Paper RF-15

A Simple Adjustment for Selection Bias Through Use of Propensity Scoring

Deanna Schreiber-Gregory, National University, Moorhead, MN

Abstract

Observational studies are vital to the exploration of health outcomes research. They allow researchers to estimate the effect of a treatment or behavior on a specific health outcome. This is something we would not be able to through an experimental procedure due to risk and ethical concerns. One concern of observational studies, however, is the fact that we cannot randomize participant placement into the treatment groups. This can result in the unwanted inclusion of a selection bias. One quick and easy way to adjust for a selection bias is through the utilization of a propensity score analysis through regression adjustment. In order to demonstrate how to do this, this presentation will seek to answer the question of how patients with a substance abuse/dependent diagnosis compare to the rest of the patient population in terms of status upon discharge. Data for this study was gathered through the National Hospital Discharge Survey, a nationwide project that collected data from acute care facilities in the United States until 2010. This presentation is designed for any level of statistician, SAS® programmer, or data analyst with an interest in controlling for selection bias, as well as for anyone who has an interest the explored topic.

Purpose

The purpose of this study was to explore whether there was an association between a primary diagnosis of substance abuse or dependency and discharge status as well as to serve as a case study analysis using a basic form of propensity scoring methodology. This topic was chosen based on the author's personal experiences working with individuals in both a short-term inpatient detoxification program and a long-term substance abuse treatment program. At this particular Midwest based hospital, a recent rise in both the severity and acuity of diagnoses in the patient population has been brought to the attention of nursing staff. This rise was also felt to be associated with the rise in emergency room admissions, greater length of required care, changes in how patients were paying for their treatment, and abrupt/inappropriate program discharges. A preliminary study employing a secondary analysis of data gathered from a nationally distributed survey was then proposed, implemented, and presented through this paper. The model for this study was set up to explore the association between a primary diagnosis of substance abuse or dependency and discharge status while controlling for gender, age, marital status, race, geographic region, principle form of payment, admission type, and days of care. Another factor that was taken into consideration was the effect that principle form of payment, days of care, and admission type would have on the discharge status of the patient. In order to address these effects, a propensity analysis was completed and propensity scores were included and controlled for in the model. These control variables were chosen based on a review of current literature and an assessment of risk behaviors and characteristics noted in the patient population of the aforementioned treatment program. Considering the sensitivity and complexity of substance abuse and dependency disorders compared to most other medical issues, we expect to see a significant difference in discharge status between these two groups (primary substance abuse or dependency diagnosis and other diagnoses). More specifically, we expect to see a higher rate of discharges that are against medical advice and to home (indicating a possible premature discharge) versus transfers (possibly a more appropriate alternative given the complex level of care needed) given the observed rates of inappropriate discharge and program incompletion in the aforementioned treatment program.

Introduction

Adults who struggle with substance-abuse issues have been noted by the medical population to report more illnesses and utilize acute and emergent health services at a much higher rate than the remainder of the population. In a study by Stein and colleagues (1993), observational evidence supported that these individuals were over two times more likely to use an emergency department and close to seven times more likely to be hospitalized than their non-substance abuse counterparts. Considering a similar study by O'Connor and colleagues which demonstrated that the healthcare needs that are associated with these types of diagnoses are often more complex, rigorous, and difficult, these types of hospitalizations usually require longer lengths of stay and often result in quite suboptimal outcomes. Further studies have also noted that during acute hospitalizations there are often delayed and inaccurate identifications of these addiction issues (Stein et al., 1996), low rates of completed medical care (Chan et al., 2004), low rates of successful referral to treatment by healthcare providers (Substance Abuse, 2004), and a high proportion of individuals who end up leaving the hospital against the medical advice of their provider (Bradley & Zarkin, 1997). Considering the results of the above studies and the general viewpoint of medical personnel, the problem of increased lengths of hospitalization and unsuccessful substance abuse treatment can no longer be overlooked as it must be addressed to not only assist in the overall health of individuals suffering from substance abuse disorders, but to assist in easing some pressure off of the medical system, thus freeing up more beds and services for other, more urgent, medical concerns.

The studies mentioned earlier primarily sought to describe and identify the type and course of treatments that individuals diagnosed with substance abuse disorders usually undergo. The current study seeks to contribute to this field of research by identifying to what extent and in what way a primary substance abuse or dependency diagnosis is associated with discharge status while controlling for demographic characteristics, admission type, principle form of payment, and days of care. In other words, this study seeks to bring to light the differences in substance abuse and dependency treatment, how current medical treatments are essentially inappropriate for this type of diagnosis, and to raise the question of whether many of the discharges that occur with this type of disorder as a primary diagnosis may be either inappropriate or premature; therefore, increasing the possibility of readmission and increased healthcare costs. This study also seeks to help identify some factors that could contribute to this discrepancy in discharge status and how these factors are represented among individuals with a primary substance abuse or dependency diagnosis versus some other type of diagnosis.

Introduction to Propensity Score Analysis

Randomized control trials (RCTs) measure the efficacy of treatment in controlled environments; however, this can often be restricted to subpopulations that limit generalizability of results. Observational studies, on the other hand, can evaluate treatment effectiveness in routine care settings or everyday use patterns. Considering this, a limitation of observational studies is the lack of treatment assignment. Non randomized groups usually differ in observed and unobserved characteristics causing selection bias when evaluating the effect of treatment.

Statistical techniques such as matching, stratification, and regression adjustment are commonly used to account for differences in treatment groups but may be limited if too few covariates are used in the adjustment process. The use of propensity score techniques avoids this limitation because it can summarize more or all of the covariate information into a single score. The question then becomes, what exactly is a propensity score? A propensity score is the conditional probability of being treated based on identified individual covariates. Rosenbaum and Rubin further demonstrate that propensity scores can account for imbalances in treatment groups and therefore provide a reduction bias by resembling randomization of subjects into treatment groups.

By using propensity scores to balance groups, traditional adjustment methods can better estimate treatment effect on outcomes while adjusting for covariates. One method proposed by D'Agostino was designed to adjust for the non-randomized treatment selection by using a propensity score method in conjunction with traditional regression techniques. This process can be performed using two steps, the first of which calculates propensity scores as the probability of patients being included in each treatment group based on pre-treatment observables. The specific aim of this step is to create a set of balanced treatment groups that simulate random treatment allocation. The second step utilizes the created propensity scores with ANCOVA as a more accurate estimate of treatment outcomes in order to study the possible covariate predictors in a more reliable environment.

Methods

Population Sample

A total of 135,238 patients included in the 2008, 2009, and 2010 collections of the National Hospital Discharge Survey (NHDS) data set and who had complete data for the identified variables were included within this study (NHDS, 2010; NHDS, 2009; NHDS, 2008). The individuals included in this study were randomly selected from participating hospitals in the United States of America with adherence to the NHDS data procurement and selection protocol. The gender distribution of the study population for the years 2008 – 2010 revealed that the slight majority of participants were female (59.93%), compared to the remaining 40.07% who were male. Throughout the study population and within the limitations that this study only looked at adults over the age of 18, there was a large age distribution with the mean age landing at about 57.57 years with a 20.43 year standard deviation. The geographic distribution of the study population revealed that slightly more participants were from the South (47.51%) followed by the MidWest (21.87%), Northeast (20.88%) and trailed by the West (9.75%). Additionally, the marital status distribution of this population revealed that the majority of patients included in this study indicated that they were married (48.66%), followed by single (27.08%), and trailed by previously married (ie: divorced, widowed, or separated) with 24.27%. Lastly, the racial distribution of the study population revealed that the majority of patients identified as Caucasian (75.09%), which was followed by those who identified as African American (17.06%), and trailed by the remainder of the population who indicated a racial identification as other than simply Caucasian or African American (7.85%).

Data

The database through which the chosen dataset was obtained is a continuously maintained project and joint effort of the Centers of Disease Control and Prevention (CDC) and the National Center for Health Statistics (NCHS). It is referred to as the National Hospital Discharge Survey (NHDS) and the information contained within it was gathered between the years of 1965 and 2010. The NHDS is a national probability survey that was designed to gather information on the characteristics of patients discharged from non-Federal short-stay hospitals in the United States of America. From the years of 1988 to 2007, the NHDS collected data from a sample of about 270,000 inpatient records obtained from a national sample of approximately 500 hospitals. From the years of 2008 to 2010, this sample size was reduced to 239 hospitals. Hospitals included in this survey were those with an average length of stay that was fewer than 30 days for all patients. This included general hospitals, children's hospitals, psychiatric hospitals, and residential settings that met the above requirements. Hospitals excluded from this survey included federal, military, and Department of Veterans Affairs (VA) hospitals, hospital units of institutions (such as prisons), and hospitals with fewer than six beds set aside for patient use.

Beginning in the year 1988, two types of data collection procedures were employed to gather the data used in this survey. The first method took the form of a manual system through which sample selection and medical transcription of hospital records to abstract forms were performed by the particular hospital's staff or directly by the staff of the U.S. Bureau of the Census on behalf of NCHS. The second method was an automated system through which the NCHS purchased electronic data files from a variety of sources, including commercial organizations, state data systems, hospitals, and hospital associations. Both of these data types contain information that relate to the personal characteristics of each patient involved in this study. The personal characteristics and demographics gathered include age, sex, race, ethnicity, marital status, and expected sources of payment. Administrative items gathered included information about admission and discharge dates (used to calculate length of stay) and discharge status. Medical items gathered included information about each patient's diagnoses and coded procedures (using the International Classification of Diseases, 9th Revision, Clinical Modification; aka. ICD-9-CM). Other considerations that were made include the use of random sampling and the exclusion of personal identifying information. Random sampling was conducting through the use of daily listing sheets provided by each hospital. By using these sheets, a systematic random sample of discharge records was selected for transmission or manual recording for the survey. However, since the survey sampled discharges and not patients, there is the possibility that the patients included in this survey could have multiple hospitalizations in a given year that may have inadvertently been included in the survey. As for personal identifying information, there were none included in this survey, so the patients seeking care at these hospitals could not be linked within the NHDS.

The NHDS survey itself consists of 4 separate and unique sections. The first section includes such patient identification information such as the particular hospital number, date of admission, date of discharge, and residence zip code. The second section includes patient demographic information such as date of birth, sex, ethnicity, race, and marital status. The third section includes administrative information such as type of admission, source of admission, expected source of payment, and the status of the patient upon their discharge. Finally, the fourth section includes medical information that was limited to up to 7 diagnoses and 4 procedures. The procedures and diagnosis codes used in this section were obtained and coded through the use of the International Classification of Diseases 9th Revision, Clinical Modification (ICD-9-CM). For this particular case study, the information included was gathered during the years 2008, 2009, and 2010.

The data used in this study was taken from individuals over the age of 18 who had complete data for the identified variables. The data was then split into two groups, one group in which the included individuals had a primary diagnosis of substance abuse or dependency (one of the pre-identified ICD-9 codes were used in the first discharge diagnosis slot for that patient) and included gender (sex), marital status, region, race, admission type, discharge status, principle form of payment, and length of hospital stay. The target variable was substance abuse or dependency diagnosis and the treatment variable was discharge status. The variables of gender, marital status, region, race, principle form of payment, and days of care were included as controls and to rule out any possible confounding interactions. A propensity score adjustment was also performed for discharge status using admission type, days of care, and principle form of payment as factors and included in the final model as an additional covariate. Applicable data adjustments are covered in detail in the statistics section.

Previous Uses

In reviewing published articles using the National Hospital Discharge Survey (NHDS), there appeared to be a significant lack of research exploring discharges and admissions for individuals diagnosed with substance abuse disorders. A good number of studies used this data set to explore cardiovascular difficulties and discharge status, demographic characteristics, and admission sources. However, one study was identified that used substance abuse diagnosis as an important criteria for inclusion in the study model. The goal of this study was to provide an analysis of inpatient drug abuse cases using the 1979-1985 administrations of the National Hospital Discharge Survey (Gfroerer et al., 1988). This study found that compared to other patients in the inpatient setting, individuals who presented with a diagnosis of

drug abuse were more likely to be male, between the ages of 15-44, and had a racial identification other than Caucasian (Gfroerer et al., 1988). The results of this study contributed to the decision to control for gender, age, and racial identification in the analysis.

Statistics

A predefined group of variables from the National Hospital Discharge Survey (NHDS) were used in this analysis. The continuous days of care variable was split at the point of 7 days to symbolize one week of care and recoded as either (0) for less than one week of care or (1) to represent a length of stay of one week or more. This split point was chosen based on an assumed average length of stay identified by nursing staff at the hospital setting sponsoring the exploration of this analysis. The admission type variable included in the administrative information section of the NHDS survey was used to determine whether the patient's admittance was emergency, trauma, urgent, or elective. A new variable was then created in which the admission type was coded as (1) for emergency or trauma, (2) for urgent, or (3) for elective. The discharge status variable included in the same section was used to determine whether a patient left against medical advice (AMA), was discharged routinely to home, or was transferred to another facility for continued care. A new variable was then created in which a code of (1) was given to an AMA discharge, (2) was given to a discharge that was routine or to home, and (3) was given to a discharge that resulted in a transfer. All other discharge statuses were excluded from this variable given that they did not indicate an actual form of discharge, but rather indicated varying levels of missing information or death. Lastly, we also chose principle source of payment from the administrative section in order to get an idea of how the patient was paying for their treatment which could provide some insight as to why a particular status of discharge was more likely. To represent principle source of payment, a new variable was created in which values were pulled from the principle source of payment variable in the NHDS and recoded to (1) for worker's compensation or government payment (signifying a specific course and compliance in order for the particular entity to complete payments), (2) for Medicare or Medicaid (signifying a level of disability or extreme need that needed to be met in order to receive this type of support), (3) for private insurance (signifying any common insurance either through work, self, parent, or spouse that was used for payment), (4) for self-pay, and (5) for some other form of payment not specified above.

A number of variables from the patient characteristics/demographics form were also included to help shed some light onto the characteristics of individuals with a continuous or episodic substance abuse or dependence diagnosis. This list of variables includes: sex, race, geographical region, and marital status. The continuous age variable was left as continuous given the large variance in ages for the target population. The categorical sex variable was left as is and only renamed as gender for clarity in results interpretation. The categorical region variable was also left as is with no need for adjustment. The race variable was recoded with a value of (1) for Caucasian racial identification, (2) for African American racial identification, and (3) for a positive identification with any other race (including codes 3, 4, 5, and 6; but excluding code 9 which indicates that a race was not specified). This recoding was based on the total number of representatives for each race and the need to maintain data completeness (some of the races did not have representatives for each group). Lastly, the categorical marital status variable was recoded as (1) for married or (2) for single, and (3) for previously married (which includes all other marital statuses except for 9 which was used for anyone who did not indicate a marital status).

The diagnosis codes used for continuous and episodic substance abuse and there representations are as follows: 30500 – Nondependent Alcohol Abuse Unspecified Drinking Behavior; 30501 – Nondependent Alcohol Abuse Continuous Drinking Behavior; 30502 – Nondependent Alcohol Abuse Episodic Drinking Behavior; 30520 – Nondependent Cannabis Abuse Unspecified Use; 30521 – Nondependent Cannabis Abuse Continuous Use; 30522 – Nondependent Cannabis Abuse Episodic Use; 30530 – Nondependent Hallucinogen Abuse Unspecified Use; 30531 – Nondependent Hallucinogen Abuse Continuous Use; 30532 – Nondependent Hallucinogen Abuse Episodic Use; 30540 – Sedative, Hypnotic or Anxiolytic Abuse, Unspecified; 30541 – Sedative, Hypnotic or Anxiolytic Abuse, Continuous; 30542 – Sedative, Hypnotic or Anxiolytic Abuse, Episodic; 30550 – Nondependent Opioid Abuse Unspecified Use; 30551 – Nondependent Opioid Abuse Continuous Use; 30552 – Nondependent Opioid Abuse Episodic Use; 30560 – Nondependent Cocaine Abuse Unspecified Use; 30561 – Nondependent Cocaine Abuse Continuous Use; 30562 – Nondependent Cocaine Abuse Episodic Use; 30570 – Nondependent Amphetamine or Related Acting Sympathomimetic Abuse Unspecified Use; 30571 – Nondependent Amphetamine or Related Acting Sympathomimetic Abuse Continuous Use; 30572 – Nondependent Amphetamine or Related Acting Sympathomimetic Abuse Episodic Use; 30580 – Nondependent Antidepressant Type Abuse Unspecified Use; 30581 – Nondependent Antidepressant Type Abuse Continuous Use; 30582 – Nondependent Antidepressant Type Abuse Episodic Use; 30590 – Nondependent Other mixed or Unspecified Drug Abuse Unspecified Use; 30591 – Nondependent Other Mixed or Unspecified Drug Abuse Continuous Use; and 30592 – Nondependent Other Mixed or Unspecified Drug Abuse Episodic Use.

The diagnosis codes used for continuous and episodic substance dependence and their representations are as follows: 30460 – Other Specified Drug Dependence Unspecified Use; 30461 – Other Specified Drug Dependence Continuous Use; 30462 – Other Specified Drug Dependence Episodic Use; 30470 – Combinations of Opioid Type Drug With Any Other Drug Dependence Unspecified Use; 30471 – Combinations of Opioid Type Drug With Any Other Drug Dependence Continuous Use; 30472 – Combinations of Opioid Type Drug With Any Other Drug Dependence Episodic Use; 30480 – Combinations of Drug Dependence Excluding Opioid Type Drug Unspecified Use; 30481 – Combinations of Drug Dependence Excluding Opioid Type Drug Continuous Use; 30482 – Combinations of Drug Dependence Excluding Opioid Type Drug Episodic Use; 30490 – Unspecified Drug Dependence Unspecified Use; 30491 – Unspecified Drug Dependence Continuous Use; and 30492 – Unspecified Drug Dependence Episodic Use. SAS Studio is the statistical program chosen to carry out the proposed analysis.

A series of univariate analyses, with the inclusion and consideration of chi-square statistics, were used to examine the adjusted associations of a primary substance abuse or dependence diagnosis with discharge status, days of care, principle form of payment, admission type, gender, age, marital status, race, and geographic region. Multivariate logistic regression models were then employed to compare the adjusted odds of a primary substance abuse or dependence diagnosis in relation to discharge status while also controlling for admission type, days of care, principle form of payment, gender, age, marital status, race, and geographic region. A propensity analysis was also conducted in order to control for any natural probability that patients may have towards a certain discharge status given three factors: principle form of payment, admission type, and days of care. This propensity score was then included in the final model as an adjusted covariate.

Results

An analysis of the content which was conducted through use of SAS Studio statistical analysis software indicates that of the 479,332 NHDS 2008, 2009, and 2010 participants, 135,238 (about 28.21%) had complete data for this study. The demographic characteristics of this population are compared in Table 1 which was also produced using SAS Studio. Of the entire target population the following demographics were distributed as such: the mean age was 57.57 years with a standard deviation of 20.43 years, 59.93% were female, 40.07% were male, 20.88% were from the Northeast, 21.87% were from the Midwest, 47.51% were from the South, 9.75% were from the West, 75.09% identified as Caucasian, 17.06% identified as African American, 7.85% identified as another race, 48.66% were married, 27.08% were single (never married), and 24.27% were no longer married.

In review of the available data in Table1 we can see that there was a proportionately older population in the “Transferred” discharge group (mean age = 74.28) and a proportionately younger population in the “AMA” discharge group (mean age = 47.74), with the “Routine” discharged group (mean age = 54.39) landing about in the middle ($p<.0001$). There were also proportionately more females than expected in the Routine and Transferred discharge groups as well as more males than expected in the AMA discharge group ($p<.0001$). There appeared to be proportionately more individuals from the South and less individuals from the West than expected in the all of the discharge groups and proportionately more individuals from the Northeast than individuals in the Midwest in the AMA and Transferred discharge groups, whereas the opposite was true in the Routine discharge group with the MidWest having proportionately more individuals in that group than the Northeast ($p<.0001$). As for days of care, there appeared to a far greater number of individuals who stayed less than one week than those who stayed more than one week in each of the discharge groups, however, the proportion of individuals from the AMA discharge group who stayed less than 1 week was far greater than those individuals who were discharged routinely to home, and the number of individuals who stayed less than one week and were discharged to home was proportionately greater than those who transferred to another facility ($p<.0001$). For principle form of payment, there was a proportionately greater number of individuals in the Transferred discharge group whose bills were being covered by Medicare or Medicaid than expected and proportionally fewer individuals in the Transferred discharge group who were being covered by any other insurance type. In addition to this, there were proportionately more individuals whose medical expenses were being covered by private insurance in the routine discharge group and proportionately more individuals who were self-pay in the AMA discharge group ($p<.0001$). For admission type, there were proportionately more individuals who were admitted through an emergency or trauma in the AMA discharge group, proportionately less elective and urgent admissions ins the AMA discharge group, and proportionately more elective and urgent admissions in the routine discharge group ($p<.0001$). In terms of race, there were proportionately more Caucasians in the transferred discharge group, proportionately less African Americans and other races in the transferred discharge group, and proportionately more African Americans in the AMA discharge group ($p<.0001$). Lastly, for marital status, there were proportionately more individuals who were married in the routine discharge group, proportionately more individuals who were single in the transferred discharge group, and proportionately more individuals who were previously married in the AMA discharge group ($p<.0001$).

The next set of comparisons were produced through SAS Studio and compiled into Table2. In review of the available data for Table2 we can see that the population who did not present with a primary substance abuse or dependency diagnosis was significantly older than the population that did present with a substance abuse or dependency diagnosis ($p<.0001$). There were also proportionately more females who were not in the target group (the target group consisting of those who presented with a primary substance abuse or dependency diagnosis per the ICD-9 coding structure) and proportionately more males who were in the target group ($p<.0001$). For days of care, there were proportionately more individuals who presented with a primary diagnosis of substance abuse or dependency who also stayed less than one week and proportionately less of those same individuals who stayed more than one week ($p<.0001$). For principle form of payment, there were proportionately more individuals in the target group who paid for their treatment via self-pay, workman's compensation, government sources, or other means than expected and proportionately more individuals who did not present with the target diagnosis who paid for their treatment via Medicare, Medicaid, or private insurance ($p<.0001$). For discharge status, there were proportionately more individuals in the target diagnosis group who discharged against medical advice and proportionately more individuals who were not in the target group who were transferred to another facility to continue treatment ($p<.0001$). For admissions type, there were proportionately more individuals from the target diagnosis group who presented for admission through an emergency or trauma and proportionately more individuals who were not from the target diagnosis group who presented for admission via urgent or elective means ($p<.0001$). For race, there were proportionately more individuals who identified as Caucasian in the group that did not have the target diagnosis, and proportionately more individuals than expected who identified as African American and who did have the target diagnosis ($p=0.0026$). For marital status, there were proportionately more married and single individuals than expected who did not present with the target diagnosis and proportionately more previously married individuals than expected who did present with the target diagnosis ($p<.0001$). Lastly, when considering geographic region, our results indicated that there were proportionately more individuals than expected from the MidWest and Northeast who did not present with the target diagnosis and proportionately more individuals from the South and West who did present with the target diagnosis ($p=0.0007$).

Adjusted odds ratios were produced through use of SAS Studio for the unadjusted model exploring the association between primary substance abuse or dependency diagnosis and discharge status and the results are presented in Table3. According to this table, we can see that individuals who discharged against medical advice were over 10 times more likely to have discharged with a primary substance abuse or dependency diagnosis than those individuals who discharged to home (OR=10.427, 95% CI = 8.257-13.168). The opposite was true when it came to those who transferred to another facility in that these individuals were almost 89% less likely to have discharged with a primary substance abuse or dependency diagnosis than those individuals who had discharged to home (OR=0.109, 95% CI = 0.060-0.199).

Adjusted odds ratios for an adjusted model exploring the association between primary substance abuse or dependency diagnosis and discharge status while controlling for race, gender, age, geographic region, marital status, principle form of payment, admission type, and days of care was then produced through use of SAS Studio and presented in Table4. According to this table we can see that older individuals were about 2% less likely than younger individuals to discharge with a primary diagnosis of substance abuse or dependency (OR=0.982, 95% CI = 0.974-0.990). As for gender, we see that females were about 42% less likely than males to discharge with a primary diagnosis of substance abuse or dependency (OR=0.576, 95% CI = 0.447-0.743). For days of care, we see that individuals who stayed in the hospital for one week or more were 71% less likely to discharge with a primary diagnosis of substance abuse or dependency than those who stayed for less than one week (OR=0.289, 95% CI = 0.168-0.496). For principle form of payment, we found that individuals who paid for their care through using Medicare, Medicaid, workman's compensation, or government resources were equally as likely to have discharged with a primary diagnosis of substance abuse or dependency as those individuals who had paid for their care through use of private insurance as indicative of the fact that the 95% confidence intervals for their odds ratio scores overlapped the value of 1, meaning that the interval within which the actual probability of their target diagnosis lied included the likelihood of equal probability (OR=0.793, 95% CI = 0.559-1.126 for Medicare/Medicaid; OR=1.934, 95% CI = 0.914-4.091 for workman's compensation/government support). However, individuals who paid for their care through other means were almost 7 times more likely to have a primary diagnosis of substance abuse or dependency upon discharge (OR=6.789, 95% CI = 4.435-10.390), and individuals who paid for their care through self-pay were over 4 times as likely to discharge with a primary diagnosis of substance abuse or dependency (OR=4.433, 95% CI = 3.121-6.298) than those individuals who paid for their medical service with private insurance. As for discharge status, we can see that by adjusting the model and controlling for the additional variables has significantly impacted the degree of odds ratios produced. This is apparent in that we now see that individuals who discharged against medical advice are now only about 4 times as likely to discharge with a primary substance abuse or dependency diagnosis (OR=3.667, 95% CI = 2.537-5.301) and those individuals who were transferred to another facility for additional care are now 93% less likely to discharge with a primary substance abuse or dependency diagnosis (OR=0.071, 95% CI = 0.017-0.286) than those individuals who discharged to home. For admission type, we see that individuals who were admitted through an emergency or trauma were over 5 times as likely to discharge with a primary diagnosis of substance abuse or dependency (OR=5.484, 95% CI = 3.172-9.481) while individuals who were admitted through urgent means were equally as likely to discharge with a substance abuse or

dependency diagnosis (OR=1.096, 95% CI = 0.486-2.474) than those individuals had an elective admission. For marital status, the resulting odds ratio scores indicated that individuals who were single were almost three times as likely as those who were married to discharge with a primary substance abuse diagnosis (OR=2.686, 95% CI = 1.940-3.270) while those who were previously married were about 51% more likely to present with a substance abuse diagnosis (OR=1.513, 95% CI = 1.010-2.265) than those who were currently married. As for race, there was not a significant odds discrepancy between African American (OR=0.740, 95% CI = 0.544-1.007) or any other race (OR=0.984, 95% CI = 0.680-1.422) and Caucasian racial identification. This was also the case for geographic region, in which there was not a significant odds discrepancy between the MidWest and the Northeast (OR=0.914, 95% CI = 0.462-1.809), the South (OR=1.548, 95% CI = 0.878-2.730), and the West (OR=1.659, 95% CI = 0.840-3.279).

A multivariate logistic regression analysis was then conducted to test whether or not and to what extent discharge status while controlling for admission type, principle form of payment, days of care, age, gender, discharge status, marital status, race, and geographical region helped explain the variation in whether a patient was discharged with a primary diagnosis of substance abuse or dependency. The analysis was run using SAS Studio and consisted of 135,238 total observations. At first glance, the analysis indicated that the convergence criterion was satisfied for this model and the interpretation of the results could then be implemented. The Wald Chi-Square value was 684.0687 with a p-value of <.0001 indicating that the defined model did significantly contribute to the explanation of whether a patient was likely to discharged with a primary diagnosis of substance abuse or dependency. While reviewing the Type 3 Analysis of Effects section, it is also apparent that eight of the nine variables included in this model significantly contributed to this effect. The only variable that did not significantly contribute to the effect of the model was racial identification (chi-square=4.5389; p-value=0.1034). The rest of the variables did significantly contribute to the model. The statistics for these effects are as such: age in years (chi-square=18.4121; p-value=<.0001), principle form of payment (chi-square=149.6017; p-value=<.0001), geographical region (chi-square=8.0247; p-value=0.0455), gender (chi-square=21.3461; p-value=<.0001), discharge status (chi-square=59.0775; p-value=<.0001), days of care (chi-square=18.5728; p-value=<.0001), admission type (chi-square=62.0726; p-value=<.0001), and marital status (chi-square=36.3210; p-value=<.0001).

Even though the above analysis proved to be significant, the author still feels that a propensity adjustment for discharge status is necessary in order to be sure that any confounding effects of this propensity are ruled out. A multivariate logistic regression analysis was then conducted to test whether or not and to what extent admission type, principle form of payment, and days of care helped explain the variation in discharge status. Through this analysis, subsequent propensity scores were also produced and outputted into a variable with intent to be used as a covariate adjustment in the final model. For the record, adjusted odds ratio scores were also produced and are presented in Table5, however, implications of these odds ratio scores will not be covered in this paper.

Adjusted odds ratios for an adjusted model exploring the association between primary substance abuse or dependency diagnosis and discharge status while controlling for race, gender, age, geographic region, marital status, principle form of payment, admission type, days of care, and the propensity adjustment for discharge status was then produced through use of SAS Studio and presented in Table6. According to this table we can see that older individuals were still about 2% less likely than younger individuals to discharge with a primary diagnosis of substance abuse or dependency (OR=0.982, 95% CI = 0.977-0.988). As for gender, we see that females were about 45% less likely than males to discharge with a primary diagnosis of substance abuse or dependency (OR=0.547, 95% CI = 0.457-0.656). For days of care, we see that individuals who stayed in the hospital for one week or more were 70% less likely to discharge with a primary diagnosis of substance abuse or dependency than those who stayed for less than one week (OR=0.304, 95% CI = 0.207-0.446). For principle form of payment, we found that individuals who paid for their care through using Medicare or Medicaid were equally as likely to have discharged with a primary diagnosis of substance abuse or dependency as those individuals who had paid for their care through use of private insurance as indicative of the fact that the 95% confidence intervals for their odds ratio scores overlapped the value of 1, meaning that the interval within which the actual probability of their target diagnosis lied included the likelihood of equal probability (OR=0.793, 95% CI = 0.559-1.126). As for those who received workman's compensation or government assistance with their expenses, these individuals were almost 90% more likely to have a primary diagnosis of substance abuse or dependency than those who paid with private insurance (OR=1.894, 95% CI = 1.115-3.219). However, individuals who paid for their care through other means were still almost 7 times more likely to have a primary diagnosis of substance abuse or dependency upon discharge (OR=6.874, 95% CI = 5.087- 9.290), and individuals who paid for their care through self-pay were over 4 times as likely to discharge with a primary diagnosis of substance abuse or dependency (OR=4.295, 95% CI = 3.350-5.505) than those individuals who paid for their medical service with private insurance. As for discharge status, we can see that individuals who discharged against medical advice are now only about 4 times as likely to discharge with a primary substance abuse or dependency diagnosis (OR=3.574, 95% CI = 2.755-4.635) and those individuals who were transferred to another facility for additional care are now 91% less likely to discharge with a primary substance abuse or dependency diagnosis (OR=0.087, 95% CI = 0.032-0.233) than those individuals who discharged to home. For admission type, we see that individuals who were admitted through an emergency or trauma were over 5 times as likely to discharge with a primary diagnosis of substance abuse or dependency (OR=5.689, 95%

CI = 3.860-8.384) while individuals who were admitted through urgent means were equally as likely to discharge with a substance abuse or dependency diagnosis (OR=1.005 95% CI = 0.565-1.788) than those individuals had an elective admission. For marital status, the resulting odds ratio scores indicated that individuals who were single were almost three times as likely as those who were married to discharge with a primary substance abuse diagnosis (OR=2.686, 95% CI = 2.134-3.382) while those who were previously married were about 51% more likely to present with a substance abuse diagnosis (OR=1.513, 95% CI = 1.137-2.012) than those who were currently married. As for race, odds ratio estimates indicated that African Americans were about 29% less likely to have a discharge diagnosis of a substance abuse or dependency disorder (OR=0.715, 95% CI = 0.575-0.890), while individuals who identified as some other race did not differ significantly from Caucasians as to whether they ended up having a primary discharge diagnosis of a substance abuse or dependency disorder (OR=0.936, 95% CI = 0.721-1.216). For marital status, individuals who were single were almost 3 times as likely to discharge with a primary diagnosis of a substance abuse or dependence disorder (OR=2.686, 95% CI = 2.134-3.3382) and individuals who had previously been married but were no longer married were about 51% more likely to discharge with a primary diagnosis of a substance abuse disorder than individuals who were married (OR=1.513, 95% CI = 1.137-2.012). When considering geographic region, individuals who were from the Northeast equally as likely as those individuals from the MidWest to discharge with a primary diagnosis of a substance abuse disorder (OR=0.914, 95% CI = 0.564-1.481), whereas individuals from the South were about 55% more likely (OR=1.548, 95% CI = 1.036-2.312) and individuals from the West were about 66% more likely (OR=1.659, 95% CI = 1.025-2.686) than individuals from the MidWest to discharge with a primary diagnosis of substance abuse or dependency. Lastly, when considering the adjusted propensity score, we found that the score itself did not provide any insight as to if a higher or lower score was more likely to indicate a discharge with a primary diagnosis of substance abuse or dependency (OR=0.996, 95% CI = 0.823-1.204).

A multivariate logistic regression analysis was then conducted to test whether or not and to what extent discharge status while controlling for admission type, principle form of payment, days of care, age, gender, discharge status, marital status, race, geographical region, and propensity adjustment for discharge status helped explain the variation in whether a patient was discharged with a primary diagnosis of substance abuse or dependency. The analysis was run using SAS Studio and consisted of 135,238 total observations. At first glance, the analysis indicated that the convergence criterion was satisfied for this model and the interpretation of the results could then be implemented. The Wald Chi-Square value was 1368.1449 with a p-value of <.0001 indicating that the defined model did significantly contribute to the explanation of whether a patient was likely to discharged with a primary diagnosis of substance abuse or dependency. While reviewing the Type 3 Analysis of Effects section, it is also apparent that nine of the ten variables included in this model significantly contributed to this effect. Keeping in mind that in the previous model, the only variable that did not significantly contribute to the effect of the model was racial identification, in the propensity adjusted model, the only variable that did not significantly contribute to the effect of the model was the adjusted propensity score (chi-square=0.0019; p-value=0.9653). The rest of the variables did significantly contribute to the model (including racial identification). The statistics for these effects are as such: racial identification (chi-square=9.0778; p-value=0.0107), age in years (chi-square=36.8241; p-value=<.0001), principle form of payment (chi-square=297.9490; p-value=<.0001), geographical region (chi-square=16.0494; p-value=0.0011), gender (chi-square=42.6922; p-value=<.0001), discharge status (chi-square=118.1552; p-value=<.0001), days of care (chi-square=37.1168; p-value=<.0001), admission type (chi-square=124.1174; p-value=<.0001), and marital status (chi-square=72.6422; p-value=<.0001).

In order to compare the effectiveness and fit of the multivariate logistic regression model before propensity adjustment inclusion to the multivariate logistic regression model after propensity adjustment inclusion, we need to look at the model fit statistics and r-square values produced by each model. In reviewing these statistics we can see that the original model, which sought to explore the association between primary substance abuse or dependence diagnosis and discharge status while controlling for admission type, principle form of payment, days of care, gender, age, marital status, race, and geographic region, had the following model fit statistics:

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	4022.221	3299.242
SC	4032.036	3485.723
-2 Log L	4020.221	3261.242

The first model had a max-rescaled r-square value of 0.1911 (max-rescaled r-square was used in the place of the default Cox-Snell r-square produced by SAS on account of upper-level boundary issues identified in the Cox-Snell

calculations; the max-rescaled r-square is SAS's solution to this problem) indicating that about 19% of the variance seen in the outcome variable of whether or not the primary discharge diagnosis was substance abuse or dependency could be explained by the defined model. The second model, which sought to explore the association between primary substance abuse or dependence diagnosis and discharge status while controlling for admission type, principle form of payment, days of care, gender, age, marital status, race, geographic region, and propensity score adjustment had these model fit statistics:

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	8042.442	6562.482
SC	8052.950	6772.641
-2 Log L	8040.442	6522.482

The second model also had a max-rescaled r-square value of 0.1911 indicating that about 19% of the variance seen in the outcome variable of whether or not the primary discharge diagnosis was substance abuse or dependency could be explained by the defined model (which included propensity adjustments). In comparing these two sets of statistics, we see that model fit statistics intercepts for all three calculations were higher in the second model than in the first, indicating that the first model was actually a better fit than the second. We can also see that the calculated r-square values for each model indicated that they predicted the same percentage of variability in the target variable. Given this information as well as the fact that both models proved to significantly explain the variance in having a primary discharge diagnosis of a substance abuse or dependence disorder, the propensity adjustment was not necessary to include in the end.

Discussion

The original purpose of the National Hospital Discharge Survey (NHDS) was to provide a response to the need for a comprehensive database that provided information on the characteristics of inpatients discharged from non-Federal short-stay hospitals within the United States of America. This dataset was developed between the years of 1965 and 2010 and continues to be maintained as part of a joint effort between the Centers of Disease Control and Prevention (CDC) and the National Center for Health Statistics (NCHS).

Limitations and Strengths

This dataset, like any other, has both limitations as well as strengths to its application and reference. For example, the total number of patients whose data was available for use and appropriate for this study between the years of 2008 and 2010 was only 135,238 individuals. Though this number may seem quite large and could be viewed as sufficient, when comparing this number to the total number of individuals hospitalized in a year within the borders of the United States of America one can quickly see how it can be difficult to effectively generalize the results of such a small sample to such a large target population. However, this number is still significantly more appropriate than what we could have otherwise obtained from a much smaller data pool. Considering this, we must point out one strength of this data set, which lies in the fact that the data pulled for this study was randomly extracted from a wide variety of hospitals and through two separate sources. This is a strength as it provides a much more varied sample that is more appropriately representative of the American population. However, another limitation is seen when considering the variety of variables that were obtained. So many of the variables obtained were interesting, however, there were many more variables that were left out that could have been included for a more robust sample (for example: various living conditions to which the patient is discharged, the patient's personal feelings of readiness upon discharge, financial status, insurance status and willingness to provide payment, relevant discharge delays, etc.). Another limitation lies in the admittance of the NHDS that since personal identifying information was not included and since the data and patients were pulled at random by the different hospitals, this data set could contain multiple entries or hospitalizations by the same patient. This could also mean that a patient seen at one hospital through which his/her information was pulled could have also been treated and pulled from another hospital later that year thus resulting in duplications within the overall dataset. In conclusion, though the NHDS has many limitations, it also has strengths. Both should be considered in the interpretation of results as well as in the development of future studies.

Detailed Variable Explanation

In consideration of the limitations of the NHDS dataset, the author would like make a more detailed note on the rationale behind the choice of identifying the number seven as the split point of the continuous variable of days of care as well as provide some preliminary results of an ad hoc data review and suggestions for future directions concerning this particular issue that could help to also address some of the limitations of the dataset. As a general rule of Prairie Saint John's Psychiatric and Addictions Care Hospital of Fargo, North Dakota, the author's place of employment, when asked what the average length of stay for a detoxing patient is the nursing and admissions staff are instructed to state that it is about 7 days give or take two days and completely dependent on the medical and therapeutic needs of the patient. While digging further into the rationale behind this informal policy, the author discovered a public access proposal from Prairie Saint John's Psychiatric and Addictions Care Hospital to erect a sister inpatient rehabilitation facility in the metropolitan area of Minneapolis/Saint Paul, Minnesota which stated that the planned average length of stay for this facility was 9.3 days (Minnesota Department of Health, 2008). Further research revealed from a similar drug and alcohol rehabilitation facility in Florida that, depending on the severity of the patient's drug/alcohol use, the key detoxification period is generally the beginning 5-10 days of physical detox which coincides directly with the timespan during which the patient is hospitalized before being admitted into a residential or outpatient rehabilitation program (Rosier, 2011). Given these numbers, the previously quoted 7 days appears to be an acceptable expectation. In exploration of this assumption, an ad hoc analysis of mean days of care was completed, compiled into Table7, and reviewed. This analysis revealed that the mean length of stay for the overall patient population included in this study was about 4.73 days and was only about 2.77 days for the target population (individuals with a primary substance abuse or dependency diagnosis). This result reveals that further exploration into mean days of care, discharge status, reason for discharge, future care plans, severity of diagnosis, specific type and cause of admission, and other defining patient characteristics would be worthy additions for future research. This result also leads us to consider more aspects of the dataset's limitations (especially in relation to discharge status) that could either use adjustment for future administrations or could redirect interested researchers in the direction of a more appropriate dataset. The questions that are raised in response to these limitations are as follows: 1) if a discharge occurred against medical advice, what was the reason for the discharge; 2) what was the chief complaint of the patient upon admission; especially if that admission was through the emergency room, 3) what medications were prescribed and was there a discrepancy between prescribed medications and requested medications; and 4) significant findings of lab results upon admission and discharge. Consideration of these limitations/questions, among others, would be important additions to future research into this topic.

Conclusion

Recent studies have found an association between substance use and emergency room use, hospitalization, and poor treatment outcomes (Stein et al., 1993; Substance Abuse, 2004). The question posed in this study took this association a step further to explore whether discharge status, while controlling for key variables, was significantly associated with a primary substance abuse or dependence diagnosis. A propensity analysis was also conducted in order to control for any natural propensities towards a specific discharge status due to admission type, days of care, and principle form of payment. A number of significant demographic and risk factors that could help explain or contribute to this association were also identified and included within the analysis.

Based on this secondary analysis of the 2010 NHDS data used in this study, a primary diagnosis of a substance abuse or dependence disorder was significantly associated with discharge status when controlling for gender, age, admission type, race, marital status, days of care, principle form of payment, geographic region, and the propensity for discharge status based on admission type, principle form of payment, and days of care. Since this study was successful in supporting this association, further study into the specifics behind it are recommended. Another recommendation would be to explore other outcome associations with unsuccessful discharge and re-admission due to a primary diagnosis of substance abuse or dependency. These recommendations are not only important for the success and increased quality of life for the patient with a primary diagnosis of substance abuse or dependency, but also the success and increased quality of care for the hospital.

References

Bradley CJ, Zarkin GA. *An inpatient profile of patients with a substance abuse diagnosis in Maryland.* J Subst Abuse Treat. 1997; 14:155–62.

Centers for Disease Control and Prevention (CDC). *National Hospital Discharge Survey 2010 Data.* Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2012.

Centers for Disease Control and Prevention (CDC). *National Hospital Discharge Survey 2010 Public Use Data File Documentation.* Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2012.

Chan AC, Palepu A, Guh DP. *HIV-positive injection drug users who leave the hospital against medical advice: the mitigating role of methadone and social support.* J Acquir Immune Defic Syndr. 2004; 35:56–9.

Gfroerer, JC, Adams, EH, Moien, M. *Drug abuse discharges form non-federal short-stay hospitals*. Am J Public Health. 1988; 78:1559-1562.

Minnesota Department of Health. (2008, February). *Minnesota Hospital Public Interest Review: Proposal for a Specialty Psychiatric Hospital in Woodbury, Minnesota*. Retrieved from http://www.health.state.mn.us/divs/hpsc/hep/publications/legislative/psjreport.pdf.

O'Connor PG, Samet JH, Stein MD. *Management of hospitalized intravenous dug users: role of the internist*. Am J Med. 1994; 96:551–8.

Rosier, R. (2011, December). *What is the average stay in detox and drug rehab?* Retrieved from http://blog.palmpartners.com/what-is-the-average-stay-in-detox-and-drug-rehab/.

Stein MD, O'Sullivan PS, Ellis P, Perrin H, Wartenberg A. *Utilization of medical services by drug abusers in detoxification.* J Subst Abuse. 1993; 5:187–93.

Stein MD, Wilinson J, Berglas N, O'Sullivan P. *Prevalence and detection of illicit drug disorders among hospitalized patients.* Am J Drug Alcohol Abuse. 1996; 22:463–71.

Substance Abuse and Mental Health Services Administration, Office of Applied Studies. 2004. *Treatment Episode Data Set (TEDS):* 1992–2002.

Contact Information

Your comments, questions, and suggestions are valued and encouraged. Contact the author at:

Deanna Schreiber-Gregory, BS
Masters Student
Health and Life Science Analytics
National University
E-mail: d.n.schreibergregory@gmail.com

TABLE 1. Univariate Associations of Characteristics of 135,238 Patients Who Participated in the 2008, 2009, or 2010 NHDS Study by Discharge Status.

Variable	Population *N*(%) or μ(σ)	Against Medical Advice *n*(%) or M(s) (N= 3677)	Routine - Home *n*(%) or M(s) (N= 225295)	Transferred *n*(%) or M(s) (N= 44986)	*p* value*
Age in years	57.57 (20.43)	47.74 (16.41)	54.39 (19.89)	74.28 (14.12)	<.0001
Gender					
Female	164187 (59.93)	1411 (38.37)	135834 (60.29)	26942 (59.89)	<.0001
Male	109771 (40.07)	2266 (61.63)	89461 (39.71)	18044 (40.11)	
Days of Care					
Less than 1 Week	222238 (81.12)	3398 (92.41)	191404 (84.96)	27436 (60.99)	<.0001
1 Week or More	51720 (18.88)	279 (7.59)	33891 (15.04)	17550 (39.01)	
Primary Payment Form					
Medicare-Medicaid	162361 (59.26)	2144 (58.31)	121415 (53.89)	38802 (86.25)	<.0001
Other	3788 (1.38)	95 (2.58)	3476 (1.54)	217 (0.48)	
Private Insurance	88020 (32.13)	683 (18.57)	82262 (36.51)	5075 (11.28)	
Self-Pay	14753 (5.39)	697 (18.96)	13529 (6.01)	527 (1.17)	
Workers Comp – Govn't	5036 (1.84)	58 (1.58)	4613 (2.05)	365 (0.81)	
Admission Type					
Emergency - Trauma	154475 (56.39)	3121 (84.88)	120112 (53.31)	31242 (69.45)	<.0001
Elective	75270 (27.48)	293 (7.97)	66891 (26.69)	8086 (17.97)	
Urgent	44213 (16.14)	263 (7.15)	38292 (17.00)	5658 (12.58)	
Race					
African American	46741 (17.06)	1009 (27.44)	39161 (17.38)	6571 (14.61)	<.0001
Other	21512 (7.85)	294 (8.00)	19159 (8.50)	2059 (4.58)	
Caucasian	205705 (75.09)	2374 (64.56)	166975 (74.11)	36356 (80.82)	
Marital Status					
Married	65801 (48.66)	551 (31.70)	57559 (51.74)	7691 (34.58)	<.0001
Single	36616 (27.08)	377 (21.69)	25823 (23.21)	10461 (46.83)	
Other	32821 (24.27)	810 (46.61)	27874 (25.05)	4137 (18.60)	
Region					
MidWest	59907 (21.87)	599 (16.29)	49088 (21.79)	10220 (22.72)	<.0001
Northeast	57207 (20.88)	980 (26.65)	44688 (19.84)	11539 (25.65)	
South	130146 (47.51)	1839 (50.01)	108356 (48.10)	19951 (44.35)	
West	26698 (9.75)	259 (7.04)	23163 (10.28)	3276 (7.28)	

* *p* values based on Pearson chi-square test of association.

TABLE 2. Univariate Associations of Characteristics of 135,238 Patients Who Participated in the 2008, 2009, or 2010 NHDS Study by Substance Use Diagnosis

Variable	Population *N*(%) or μ(σ)	No Substance Abuse/Dependence Diagnosis *n*(%) or M(s) (N= 3677)	Substance Abuse/Dependence Diagnosis *n*(%) or M(s) (N= 225295)	*p* value*
Age in years	57.57 (20.43)	57.60 (20.42)	42.91 (13.91)	<.0001
Gender				
Female	164187 (59.93)	163983 (59.99)	204 (34.06)	<.0001
Male	109771 (40.07)	109376 (40.01)	395 (65.94)	
Days of Care				
Less than 1 Week	222238 (81.12)	221680 (81.09)	558 (93.16)	<.0001
1 Week or More	51720 (18.88)	51679 (18.91)	41 (6.84)	
Primary Payment Form				
Medicare-Medicaid	162361 (59.26)	162144 (59.32)	217 (36.23)	
Other	3788 (1.38)	3739 (1.37)	49 (8.18)	
Private Insurance	88020 (32.13)	87905 (32.16)	115 (19.20)	<.0001
Self-Pay	14753 (5.39)	14561 (5.33)	192 (32.05)	
Workers Comp – Govn't	5036 (1.84)	5010 (1.83)	26 (4.34)	
Discharge Status				
Against Medical Advice (AMA)	3677 (1.34)	3593 (1.31)	84 (14.02)	
Routine - Home	225295 (82.24)	224791 (82.23)	504 (84.14)	<.0001
Transferred	44986 (16.42)	44975 (16.45)	11 (1.84)	
Admission Type				
Emergency - Trauma	154475 (56.39)	153966 (56.32)	509 (84.97)	
Elective	75270 (27.48)	75226 (27.52)	44 (7.35)	<.0001
Urgent	44213 (16.14)	44167 (16.16)	46 (7.68)	
Race				
African American	46741 (17.06)	46609 (17.05)	132 (22.04)	
Other	21512 (7.85)	21460 (7.85)	52 (8.68)	0.0026
Caucasian	205705 (75.09)	205290 (75.10)	415 (69.28)	
Marital Status				
Married	65801 (48.66)	65739 (48.71)	62 (22.14)	
Single	36616 (27.08)	36574 (27.10)	42 (15.00)	<.0001
Other	32821 (24.27)	32645 (24.19)	176 (62.86)	
Region				
MidWest	59907 (21.87)	59797 (21.87)	110 (18.36)	
Northeast	57207 (20.88)	57110 (20.89)	97 (16.19)	0.0007
South	130146 (47.51)	129824 (47.49)	322 (53.76)	
West	26698 (9.75)	26628 (9.74)	70 (11.69)	

* *p* values based on Pearson chi-square test of association.

TABLE 3. Multivariate Regression Analysis Comparing Substance Abuse / Dependence Diagnosis By Discharge Status of 135,238 Patients Who Participated in the 2008, 2009, or 2010 NHDS Study With No Adjustments.

Variable	No Substance Abuse/Dependence Diagnosis *n*(%) or M(s) (N= 3677)	Substance Abuse/Dependence Diagnosis n(%) or M(s) (N= 225295)	Odds Ratio (OR)*	Confidence Interval (CI) for Odds Ratio (OR)
Discharge Status				
Against Medical Advice (AMA)	3593 (1.31)	84 (14.02)	10.427	8.257 - 13.168
Routine - Home	224791 (82.23)	504 (84.14)	-----	-----
Transferred	44975 (16.45)	11 (1.84)	0.109	0.060 - 0.199

* Probability modeled was a positive substance abuse/dependence diagnosis

TABLE 4. Multivariate Regression Analysis Comparing Substance Abuse / Dependence Diagnosis By Discharge Status of 135,238 Patients Who Participated in the 2008, 2009, or 2010 NHDS Study Adjusting for Age, Admission Type, Primary Form of Payment, Days of Care, Gender, Marital Status, Region, and Race.

Variable	No Substance Abuse/Dependence Diagnosis *n*(%) or M(s) (N= 3677)	Substance Abuse/Dependence Diagnosis n(%) or M(s) (N= 225295)	Odds Ratio (OR)*	Confidence Interval (CI) for Odds Ratio (OR)
Age in years	57.60 (20.42)	42.91 (13.91)	0.982	0.974 - 0.990
Gender				
Female	163983 (59.99)	204 (34.06)	0.576	0.447 – 0.743
Male	109376 (40.01)	395 (65.94)	----	----
Days of Care				
Less than 1 Week	221680 (81.09)	558 (93.16)	----	----
1 Week or More	51679 (18.91)	41 (6.84)	0.289	0.168 – 0.496
Primary Payment Form				
Medicare-Medicaid	162144 (59.32)	217 (36.23)	0.793	0.559 – 1.126
Other	3739 (1.37)	49 (8.18)	6.789	4.435 – 10.390
Private Insurance	87905 (32.16)	115 (19.20)	----	----
Self-Pay	14561 (5.33)	192 (32.05)	4.433	3.121 – 6.298
Workers Comp – Govn't	5010 (1.83)	26 (4.34)	1.934	0.914 – 4.091
Discharge Status				
Against Medical Advice (AMA)	3593 (1.31)	84 (14.02)	3.667	2.537 – 5.301
Routine - Home	224791 (82.23)	504 (84.14)	----	----
Transferred	44975 (16.45)	11 (1.84)	0.071	0.017 – 0.286
Admission Type				
Emergency - Trauma	153966 (56.32)	509 (84.97)	5.484	3.172 – 9.481
Elective	75226 (27.52)	44 (7.35)	----	----
Urgent	44167 (16.16)	46 (7.68)	1.096	0.486 – 2.474
Race				
African American	46609 (17.05)	132 (22.04)	0.740	0.544 – 1.007
Other	21460 (7.85)	52 (8.68)	0.984	0.680 – 1.422
Caucasian	205290 (75.10)	415 (69.28)	----	----
Marital Status				
Married	65739 (48.71)	62 (22.14)	----	----
Single	36574 (27.10)	42 (15.00)	2.686	1.940 – 3.270
Other	32645 (24.19)	176 (62.86)	1.513	1.010 – 2.265
Region				
MidWest	59797 (21.87)	110 (18.36)	----	----
Northeast	57110 (20.89)	97 (16.19)	0.914	0.462 – 1.809
South	129824 (47.49)	322 (53.76)	1.548	0.878 – 2.730
West	26628 (9.74)	70 (11.69)	1.659	0.840 – 3.279

* Probability modeled was a positive substance abuse/dependence diagnosis

TABLE 5. Propensity Analysis of Discharge Status Adjusting for Admission Type, Days of Care, and Primary Form of Payment of 135,238 Patients Who Participated in the 2008, 2009, or 2010 NHDS Study.

Variable	Against Medical Advice *n*(%) or M(s) (N= 3677)	Routine - Home *n*(%) or M(s) (N= 225295)	Transferred *n*(%) or M(s) (N= 44986)	Odds Ratio (OR)*	Confidence Interval (CI) for Odds Ratio (OR)
Days of Care					
Less than 1 Week	3398 (92.41)	191404 (84.96)	27436 (60.99)	----	----
1 Week or More	279 (7.59)	33891 (15.04)	17550 (39.01)	0.348	0.337 – 0.359
Primary Payment Form					
Medicare-Medicaid	2144 (58.31)	121415 (53.89)	38802 (86.25)	0.307	0.295 – 0.318
Other	95 (2.58)	3476 (1.54)	217 (0.48)	1.481	1.274 – 1.722
Private Insurance	683 (18.57)	82262 (36.51)	5075 (11.28)	----	----
Self-Pay	697 (18.96)	13529 (6.01)	527 (1.17)	2.409	2.191 – 2.649
Workers Comp – Govn't	58 (1.58)	4613 (2.05)	365 (0.81)	0.839	0.735 – 0.957
Admission Type					
Emergency - Trauma	3121 (84.88)	120112 (53.31)	31242 (69.45)	0.695	0.671 – 0.720
Elective	293 (7.97)	66891 (26.69)	8086 (17.97)	----	----
Urgent	263 (7.15)	38292 (17.00)	5658 (12.58)	0.957	0.911 – 1.005

* Probability modeled was a positive substance abuse/dependence diagnosis

TABLE 6. Multivariate Regression Analysis Comparing Substance Use / Abuse Diagnosis By Discharge Status of 135,238 Patients Who Participated in the 2008, 2009, or 2010 NHDS Study Adjusting for Age, Admission Type, Primary Form of Payment, Days of Care, Gender, Marital Status, Region, Race, and Propensity Score

Variable	No Substance Abuse/Dependence Diagnosis *n*(%) or M(s) (N= 3677)	Substance Abuse/Dependence Diagnosis n(%) or M(s) (N= 225295)	Odds Ratio (OR)*	Confidence Interval (CI) for Odds Ratio (OR)
Age in years	57.60 (20.42)	42.91 (13.91)	0.982	0.977 - 0.988
Gender				
Female	163983 (59.99)	204 (34.06)	0.547	0.457 – 0.656
Male	109376 (40.01)	395 (65.94)	----	----
Days of Care				
Less than 1 Week	221680 (81.09)	558 (93.16)	----	----
1 Week or More	51679 (18.91)	41 (6.84)	0.304	0.207 – 0.446
Primary Payment Form				
Medicare-Medicaid	162144 (59.32)	217 (36.23)	1.000	0.772 – 1.294
Other	3739 (1.37)	49 (8.18)	6.874	5.087 – 9.290
Private Insurance	87905 (32.16)	115 (19.20)	----	----
Self-Pay	14561 (5.33)	192 (32.05)	4.295	3.350 – 5.505
Workers Comp – Govn't	5010 (1.83)	26 (4.34)	1.894	1.115 – 3.219
Discharge Status				
Against Medical Advice (AMA)	3593 (1.31)	84 (14.02)	3.574	2.755 – 4.635
Routine - Home	224791 (82.23)	504 (84.14)	----	----
Transferred	44975 (16.45)	11 (1.84)	0.087	0.032 – 0.233
Admission Type				
Emergency - Trauma	153966 (56.32)	509 (84.97)	5.689	3.860 – 8.384
Elective	75226 (27.52)	44 (7.35)	----	----
Urgent	44167 (16.16)	46 (7.68)	1.005	0.565 – 1.788
Race				
African American	46609 (17.05)	132 (22.04)	0.715	0.575 – 0.890
Other	21460 (7.85)	52 (8.68)	0.936	0.721 – 1.216
Caucasian	205290 (75.10)	415 (69.28)	----	----
Marital Status				
Married	65739 (48.71)	62 (22.14)	----	----
Single	36574 (27.10)	42 (15.00)	2.686	2.134 – 3.382
Other	32645 (24.19)	176 (62.86)	1.513	1.137 – 2.012
Region				
MidWest	59797 (21.87)	110 (18.36)	----	----
Northeast	57110 (20.89)	97 (16.19)	0.914	0.564 – 1.481
South	129824 (47.49)	322 (53.76)	1.548	1.036 – 2.312
West	26628 (9.74)	70 (11.69)	1.659	1.025 – 2.686
Propensity Score	0.42 (0.42)	0.46 (0.44)	0.996	0.823 – 1.204

* Probability modeled was a positive substance abuse/dependence diagnosis

TABLE 7. Means Associations of Days of Care of 135,238 Patients Who Participated in the 2008, 2009, or 2010 NHDS Study by Substance Use Diagnosis

Variable	Population *N*(%) or μ(σ)	No Substance Abuse/Dependence Diagnosis *n*(%) or M(s) (N= 3677)	Substance Abuse/Dependence Diagnosis *n*(%) or M(s) (N= 225295)
Days of Care	4.73 (6.79)	4.74 (6.80)	2.77 (2.85)

Significant Project Code Excerpts:

```
/*************************************************************************
*****************************************
We are conducting a study of substance abuse/dependence and discharge status
from acute care facilities and wish to
investigate whether there is an association between substance abuse/dependence
diagnosis and discharge status.

We believe that conducting an adjusted analysis is appropriate though you are
concerned that other factors need to be
matched on or otherwise taken into account differently than in an adjusted
analysis. You feel that the propensity of
"treatment" (Discharge status [AMA, Transferred to long-term, Discharged home])
assignment is conditional on at least two
observed baseline characteristics.

Age restricted to 18 and older, so that we are only looking at this trend in
adults.

Factors:
              1) Type of Admission (Emergency/Trauma, Urgent, Elective)
              2) Length of Stay (<7 days or 7 days and greater)
              3) Principal source of payment (Workers Comp/Govnt,
Medicare/Medicaid, Private Insurance, Self-Pay, Other)

Covariates Considered:
              1) Gender (female, male)
              2) Race (all types excluding uknown, not answered)
              3) Geographic Region (Midwest, Northeast, West, South)
              4) Marital Status (Married, Single, Other)
              5) Age in years (continuous variable)

*************************************************************************
*****************************************/


LIBNAME NHDS '/folders/myshortcuts/NHDS/NHDS/' ;

data NHDS.NHDS10;

infile '/folders/myshortcuts/NHDS/NHDS/NHDS10.pu.txt';

    input  surveyyear 1-2   Newborn 3   Ageunits 4   ageyears 5-6   sex 7
race 8   marital 9
dischargeMonth 10-11   dischargestatus 12    dayscare 13-16    LOS 17    region
18  numbbeds 19
hospowner 20    Analysisweight 21-25   twodigitssurveyyear 26-27    dx1 $ 28-32
dx2 $ 33-37   dx3 $ 38-42
```

```
dx4 $ 43-47  dx5 $ 48-52   dx6 $ 53-57    dx7 $ 58-62  dx8 $ 63-67   dx9 $ 68-
72   dx10 $ 73-77   dx11 $ 78-82   dx12 $ 83-87
dx13 $ 88-92   dx14 $ 93-97     dx15 $ 98-102    proc1 $ 103-106    proc2 $
107-110    proc3 $ 111-114    proc4 $ 115-118
proc5 $ 119-122   proc6 $ 123-126   proc7 $ 127-130   proc8 $ 131-134
prisourcepayment 135-136  secrourcepayment  137-138
drg 139-141   admisstype 142  admisssource 143-144  admissdxs $ 145-149;

run;

data NHDS.NHDS09;

infile '/folders/myshortcuts/NHDS/NHDS/NHDS09.pu.txt';

    input  surveyyear 1-2   Newborn 3   Ageunits 4   ageyears 5-6   sex 7
race 8   marital 9
dischargeMonth 10-11   dischargestatus 12    dayscare 13-16    LOS 17    region
18  numbbeds 19
hospowner 20    Analysisweight 21-25   twodigitssurveyyear 26-27    dx1 $ 28-32
dx2 $ 33-37   dx3 $ 38-42
dx4 $ 43-47  dx5 $ 48-52   dx6 $ 53-57    dx7 $ 58-62
proc1 $ 63-66    proc2 $ 67-70    proc3 $ 71-74    proc4 $ 75-78
prisourcepayment 79-80  secrourcepayment  81-82
drg 83-85   admisstype 86  admisssource 87-88  admissdxs $ 89-93;

run;

data NHDS.NHDS08;

infile '/folders/myshortcuts/NHDS/NHDS/NHDS08.pu.txt';

    input  surveyyear 1-2   Newborn 3   Ageunits 4   ageyears 5-6   sex 7
race 8   marital 9
dischargeMonth 10-11   dischargestatus 12    dayscare 13-16    LOS 17    region
18  numbbeds 19
hospowner 20    Analysisweight 21-25   twodigitssurveyyear 26-27    dx1 $ 28-32
dx2 $ 33-37   dx3 $ 38-42
dx4 $ 43-47  dx5 $ 48-52   dx6 $ 53-57    dx7 $ 58-62
proc1 $ 63-66    proc2 $ 67-70    proc3 $ 71-74    proc4 $ 75-78
prisourcepayment 79-80  secrourcepayment  81-82
drg 83-85   admisstype 86  admisssource 87-88  admissdxs $ 89-93;

run;

data NHDS08;
  set NHDS.NHDS08;
run;

data NHDS09;
  set NHDS.NHDS09;
run;

data NHDS10;
  set NHDS.NHDS10 (drop= dx8 dx9 dx10 dx11 dx12 dx13 dx14 dx15 proc5 proc6
proc7 proc8);
run;

/* This section details process to append the datasets for the years 2008,
2009, 2010 */
/* The ending data set contains all 3 and call it nhds200820092010 */

proc append base=NHDS08 data=NHDS09;
 run;
```

```
proc append base=NHDS08 data=NHDS10;
 run;

Data nhds200820092010;
  set NHDS08;

    /* Days of Care */
       format sevendayscare_cat $20.;
              if dayscare < 7 then sevendayscare= 0 ;
                 else if dayscare >= 7 then sevendayscare= 1 ;
              if dayscare < 7 then sevendayscare_cat= 'Less than 1 Week' ;
                 else if dayscare >= 7 then sevendayscare_cat = 'More than 1
Week' ;


       /* Race */
       format new_race_cat $20.;
              if race = 1 then new_race = 1;
                     else if race=2 then new_race = 2;
                     else if race~=1 and race~=2 and race~=9 then new_race = 3;
              if race = 1 then new_race_cat = 'White';
                     else if race=2 then new_race_cat = 'African American';
                     else if race~=1 and race~=2 and race~=9 then new_race_cat =
'Other';

       /* Payment Source */
       format pripayment_cat $20.;
              if prisourcepayment=01 or prisourcepayment=04 then pripayment = 1;
                     else if prisourcepayment=02 or prisourcepayment=03 then
pripayment = 2;
                     else if prisourcepayment=05 or prisourcepayment=06 or
prisourcepayment=07 then pripayment = 3;
                     else if prisourcepayment=08 then pripayment = 4;
                     else if prisourcepayment=09 or prisourcepayment=10 then
pripayment = 5;
              if prisourcepayment=01 or prisourcepayment=04 then pripayment_cat
= 'Workers Comp - Govnt ';
                     else if prisourcepayment=02 or prisourcepayment=03 then
pripayment_cat = 'Medicare - Medicaid';
                     else if prisourcepayment=05 or prisourcepayment=06 or
prisourcepayment=07  then pripayment_cat= 'Private Insurance';
                     else if prisourcepayment=08 then pripayment_cat= 'Self-
Pay';
                     else if prisourcepayment=09 or prisourcepayment=10 then
pripayment_cat= 'Other';

       /* Admission Type*/
       format admission_cat $20.;
              if admisstype=1 or admisstype=5 then admission=1;
                     if admisstype=2 then admission=2;
                     else if admisstype=3 then admission=3;
              if admisstype=1 or admisstype=5 then admission_cat='ER - Trauma';
                     if admisstype=2 then admission_cat='Urgent';
                     else if admisstype=3 then admission_cat='Elective';

       /* Region */
       format region_cat $20.;
              if region = 1 then region_cat = 'NorthEast';
                     else if region = 2 then region_cat = 'MidWest';
                     else if region = 3 then region_cat = 'South';
                     else if region = 4 then region_cat = 'West';
```

```
       /* Gender */
       format gender_cat $20.;
              if sex=1 then gender_num=1;
                     else if sex=2 then gender_num=2;
              if sex=1 then gender_cat='Male';
                     else if sex=2 then gender_cat='Female';

       /* Marital Status */
       format married_cat $20.;
              if marital=1 then married=1;
              if marital=2 then married=2;
                     else if marital~=1 and marital~=2 and marital~=9 then
married=3;
              if marital=1 then married_cat='Married';
              if marital=2 then married_cat='Single';
                     else if marital~=1 and marital~=2 and marital~=9 then
married_cat='Other';

       /* Discharge Status */
       format discharge_cat $20.;
              if dischargestatus=2 then discharge=1;
              if dischargestatus=1 then discharge=2;
                     else if dischargestatus=3 or dischargestatus=4 then
discharge=3;
              if dischargestatus=2 then discharge_cat='AMA';
              if dischargestatus=1 then discharge_cat='Routine - Home';
                     else if dischargestatus=3 or dischargestatus=4 then
discharge_cat='Transferred';

       /* Diagnosis of Substance Abuse/Dependence Disorder */
    array d(1) dx1;
       sub_abuse=0;
       do i=1;
           if (substr(d(i),1,5) in ('30520' '30521' '30522' '30530' '30531'
'30532'
                                                '30540' '30541' '30542' '30550'
'30551' '30552'
                                                '30560' '30561' '30562' '30570'
'30571' '30572'
                                                '30580' '30581' '30582' '30590'
'30591' '30592'
                                                '30500' '30501' '30502' '30460'
'30461' '30462'
                                                '30470' '30471' '30472' '30480'
'30481' '30482'
                                                '30490' '30491' '30492')) then
sub_abuse=1;
          end;
       format sub_abuse_cat $20.;
              if sub_abuse=0 then sub_abuse_cat='No';
              else if sub_abuse=1 then sub_abuse_cat='Yes';
run;

proc contents data=nhds200820092010;
run;

proc freq data=nhds200820092010;
  tables ( sevendayscare_cat pripayment_cat new_race_cat admission_cat
region_cat
                gender_cat discharge_cat married_cat) * sub_abuse_cat / chisq;
run;
```

```
data newnhds200820092010 (keep =   sevendayscare sevendayscare_cat pripayment
pripayment_cat new_race new_race_cat
                                                    admission admission_cat
region region_cat gender_num gender_cat discharge
                                                    discharge_cat sub_abuse
sub_abuse_cat married married_cat ageyears);
  set nhds200820092010 (where=  (  (sevendayscare in (0,1)) and (pripayment in
(1,2,3,4,5)) and (new_race in (1,2,3)) and
                                                    (admission in (1,2,3))
and (region in (1,2,3,4)) and (gender_num in (1,2)) and (discharge in (1,2,3))
                                                    and (sub_abuse in
(0,1)) and (ageyears>=18) ));
run;

proc freq data=newnhds200820092010;
  tables ( sevendayscare_cat pripayment_cat new_race_cat admission_cat
region_cat
                gender_cat married_cat) * discharge_cat / chisq;
run;

proc means data=newnhds200820092010;
       var ageyears;
run;

proc means data=newnhds200820092010;
       var ageyears;
       class discharge_cat;
       title "Mean Years of Age by Substance Use Diagnosis";
run;

proc sort data=newnhds200820092010;
       by discharge;
run;

proc reg data=newnhds200820092010;
       model discharge = ageyears;
run;

proc freq data=newnhds200820092010;
  tables ( sevendayscare_cat pripayment_cat new_race_cat admission_cat
region_cat
                gender_cat discharge_cat married_cat) * sub_abuse_cat / chisq;
run;

proc means data=newnhds200820092010;
       var ageyears;
       class sub_abuse_cat;
       title "Mean Years of Age by Substance Use Diagnosis";
run;

proc sort data=newnhds200820092010;
       by sub_abuse_cat;
run;

proc ttest data=newnhds200820092010;
       var ageyears;
       class sub_abuse_cat;
run;

proc logistic data = newnhds200820092010;
class sub_abuse_cat(ref='No')  discharge_cat (ref='Routine - Home')    /
param=ref;
model sub_abuse_cat = discharge_cat;
```

```
title 'Substance Abuse/Dependence Diagnosis by Discharge Status, Unadjusted';
run;

proc logistic data = newnhds200820092010;
class   sub_abuse_cat(ref='No')  discharge_cat(ref='Routine - Home')
admission_cat (ref='Elective') gender_cat(ref='Male')
             new_race_cat (ref='White') sevendayscare_cat (ref='Less than 1
Week') pripayment_cat (ref='Private Insurance')
             region_cat (ref='MidWest') married_cat (ref='Married') /
param=ref;
model sub_abuse_cat = discharge_cat admission_cat gender_cat new_race_cat
sevendayscare_cat pripayment_cat region_cat married_cat ageyears / rsq;
title 'Substance Abuse/Dependence Diagnosis by Discharge Status, Adjusted';
run;

/* This section details process to output the propensity for group selection */

proc logistic data = newnhds200820092010;
class   discharge_cat (ref='Routine - Home') admission_cat (ref='Elective')
gender_cat(ref='Male')
             new_race_cat (ref='White') sevendayscare_cat (ref='Less than 1
Week') pripayment_cat (ref='Private Insurance')
             region_cat (ref='MidWest') married_cat (ref='Married')  /
param=ref;
model discharge_cat =  admission_cat sevendayscare_cat pripayment_cat;
OUTPUT OUT=NHDS.AllPropen prob=prob; /*Output the propensity for group
selection (Discharge Status selection) */
title 'Propensity Scores for Discharge Status';
run;


proc contents data=NHDS.AllPropen;
run;


proc logistic data = NHDS.AllPropen;
class   sub_abuse_cat (ref='No')  discharge_cat(ref='Routine - Home')
admission_cat (ref='Elective') gender_cat(ref='Male')
             new_race_cat (ref='White') sevendayscare_cat (ref='Less than 1
Week') pripayment_cat (ref='Private Insurance')
             region_cat (ref='MidWest') married_cat (ref='Married')  /
param=ref;
model sub_abuse_cat = discharge_cat admission_cat gender_cat new_race_cat
sevendayscare_cat pripayment_cat region_cat married_cat ageyears prob / rsq;
title 'Propensity Scores Adjusted';
run;

proc means data=NHDS.AllPropen;
       var prob;
       title "Mean Propensity Score";
run;

proc means data=NHDS.AllPropen;
       var prob;
       class sub_abuse_cat;
       title "Mean Propensity Score by Substance Use Diagnosis";
run;

proc means data=nhds200820092010;
       var dayscare;
       title "Exploration of Days of Care";
run;
```

```
proc means data=nhds200820092010;
       var dayscare;
       class sub_abuse_cat;
run;
```