LENGTH OF STAY FOR SUBSTANCE ABUSE TREATMENT

**ISM 6137 – Statistical Data Mining**

**Final Project – Fall 2022**



**Syed Asjid Raza**

**Akanksha Kottakota**

**Kshitiz Kharel**

**Zakeer Ahmed Shaik**

Table of Contents

[Length of Stay– Analysis Report 3](#_Toc118735891)

[Executive Summary 3](#_Toc118735892)

[Problem Definition and Significance 3](#_Toc118735893)

[Prior Literature 4](#_Toc118735894)

[Data Source, Preparation and Feature Engineering 5](#_Toc118735895)

[Explanation of key variables 6](#_Toc118735896)

[Descriptive Analytics and Visualizations 7](#_Toc118735897)

[Variable Exclusion & Initial Predictor Table 8](#_Toc118735898)

[Predictor Table 8](#_Toc118735899)

[Statistical Models for Analysis 9](#_Toc118735900)

[Model 1 – Base Model 10](#_Toc118735901)

[Model 2 – Poisson 10](#_Toc118735902)

[Model 3 – Multilevel 10](#_Toc118735903)

[Choice of model: 10](#_Toc118735904)

[Interpretation of the marginal effects 10](#_Toc118735905)

[Model 3 : Multi Level 11](#_Toc118735906)

[Assumptions: 12](#_Toc118735907)

[Recommendations: 12](#_Toc118735908)

[References 12](#_Toc118735909)

[Abstract 13](#_Toc118735910)

[Model Summaries 14](#_Toc118735911)

[Stargazer Output for Comparison 16](#_Toc118735912)

[Prior Work & Relevance 21](#_Toc118735913)

[Links to Data Source & Dictionary 24](#_Toc118735914)

[R-Code 24](#_Toc118735915)

# Length of Stay– Analysis Report

## Executive Summary

The phrase "substance abuse" refers to a pattern of drug or substance usage that results in serious issues or distress. It could lead to substance abuse-related legal issues or persistent substance use those harms friendships, family ties, or both. Addiction to illegal substances like marijuana, heroin, cocaine, or methamphetamine is referred to as substance abuse, which is a recognized medical psychological problem. Furthermore, it may involve abusing substances that are legal, including alcohol, cigarettes, or prescription drugs. We aim to study the length of stay of patients addicted to substances.

The dataset used for this analysis is from the TEDS system. The Treatment Episode Data Set (TEDS) system is a repository of treatment data periodically collected by different states in the US to monitor their respective substance use treatment systems.

This dataset comprises of demographic, clinical, and substance use characteristics of admissions to alcohol or drug treatment of 258664 patients in facilities that report to state administrative data systems.

This report will include the steps performed prior to analysis such as data cleaning, feature selection, feature engineering as well as steps performed in order to perform multilevel analysis.

As the dependent variable is count of number of days in the facility and the visual graph of data distribution, we understand that our data belongs to a Poisson Distribution.

## Problem Definition and Significance

Drug Addition, also known as substance, use disorder, is considered a disease that affects a person's brain and behavior. This can lead to an inability to control the use of a legal or illegal drug or medicine. Alcohol, marijuana, cocaine, opioids, and heroin are considered as some of the most addictive ones. The addict usually does not care about the side effects of the drug they administer, making this a dangerous way of living. We also know that

According to the National Survey on Drug Use and Health (NSDUH), 19.7 million American adults battled a substance use disorder in 2017. 38% of the adults were also battling illicit drug use, data also pointed to 1 in 8 adults having both a drug and alcohol problem. Additionally, $740 billion cost was endured by the American society in annually lost workspace productivity, healthcare expenses, and crime related costs. Finally, 40% - 60% of the substance abusers are expected to relapse

Most of the hospitals, insurance companies and federal health agencies understand the concept of length of stay. Our dataset supports this perspective, so our primary target is to identify the length of stay for a patient that is undergoing substance abuse treatment. Hence, we will be looking at the major factors that primarily affect length of stay

## Prior Literature

In order to better understand our problem at hand we went through different prior publications which dealt with analysis of substance abuse and hospital length of stay. After examining multiple such case scenarios, eight papers were chosen as our primary references from which we understand the following:

* Importance must be given to identifying and addressing emotional/mental disorders.

(Battjes RJ,… 2004)

* Facility-level factors such as type of treatment, influence length of stay for patients with substance abuse disorders (Gifford E,…2008)
* Psychological and behavioral measures are important when looking at substance abuse(Barenholtz,..2020)
* Alcohol increases chances of endangerment within teens. Teens with alcohol comorbidities had lower lengths of stay than teens without alcohol comorbidities, although teens with drug comorbidities have higher charges and lengths of stay than teens without drug comorbidities (Peek-(Asa C,… 2011)
* Sexuality and social media have profound effect on substance abuse and length of stay (Ovalle,…. 2021)
* Severity, Frequency of use of substance, and insurance are crucial factors to affect length of stay prediction (Baek H,…2018)
* Multivariate models partly explained variance in hospital stay, suggesting the importance of pre-admission and post-discharge factors, including the healthcare environment, the availability of primary and secondary care resources, and the threshold for hospital admission.(S [.P. Wright](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorRaw=Wright%2C+SP),[D.](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorRaw=Verouhis%2C+D) 2003)
* Racial-ethnic differences in referral source, diagnosis, and length of stay in substance abuse treatment were examined. African Americans had longer stays than Hispanics and whites. (Miriam D,… 2011)

## Data Source, Preparation and Feature Engineering

All the data used in the analysis has been sourced from Substance Abuse & Mental Health Data Archive and has been concatenated from years 2015-2019. The original dataset contained 7,928,437 observations of 76 variables. After diligent feature selection and dropping missing values, 200,000 observations with 21 features were examined in the study. The data preparation and feature engineering performed on the dataset are mentioned below:

1. Target Variable, Length of Stay (LOS) up to 30 has been considered in the study.

* The structure of LOS for up to 30 days is continuous and after 30 days the column is divided into categories of different bins.
* To achieve consistency in our target variable, LOS up to 30 days has been considered.

1. Patients whose treatment is completed has only been considered.

* There are different reasons for discharge of an individual involved in the substance abuse treatment.
  + The different reasons are treatment completed
  + dropped out of treatment
  + terminated by facility
  + transferred to another treatment program or facility
  + incarcerated and death.

1. Age of First use for Primary and Secondary substances has been converted to binary category

* The structure of age of first use is categorical with different bins.
* The first age column has been converted to two binary factors, 0 being the first use age below 18 and 1 being the first use age above 18, to examine the LOS differences in adults and nonadults.

1. Concatenated part-time and full-time as employed in employment column.

* The data contains 4 categories for employment. The categories are:
  + part-time, full-time, unemployed, and not in the labor force.
  + Part-time and full-time employment has been concatenated into one.

1. Categorizing patients as insured and uninsured on Health insurance column

* There are different categories for Health Insurance
  + Private Insurance, Medicaid, Medicare and individual without any insurance.
  + The data, for the purpose of our study, has been split into two categories, i.e. People with Insurance which includes Private, Medicaid and Medicare and People with no Insurance.

1. Categorized substances with low proportion into others

* Substances such as Alcohol, Cocaine, Marijuana, Heroin, Methamphetamine, opiates, and synthetics have the majority proportion on the substance use.
* Other drugs such as Non-prescription methadone, PCP, hallucinogens, benzodiazepines, barbiturates, over-the-counter medications, etc. contains about 3.5% of total observations, so, these drugs have been converted into “others”.

## Explanation of key variables

* **Substance use at admission and discharge(primary):** identify the client's primary substance use at admission and discharge.
* **Substance use at admission and discharge(secondary):** identify the client's secondary substance use at admission and discharge.
* **Age at admission:** Calculated from date of birth and date of admission.
* **Gender:** This field identifies the client's biological sex.
* **Employment status at Discharge:** This field identifies the client’s employment status at discharge. (Employed, unemployed and not in labor force)
* **Type of treatment/service setting**: This field describes the type of treatment service or treatment setting in which the client is placed.
* **Living arrangements at admission:** Identifies whether the client is homeless, a dependent (living with parents or in a supervised setting) or living independently on his or her own at the time of discharge.
* **Previous substance uses treatment episodes:** Indicates if the client has received in any substance use treatment program.
* **Route of administration (primary and secondary):** This field identifies the usual route of administration of the corresponding substance identified in Substance Use
* **Frequency of use at admission (primary and secondary):** Specifies the frequency of use of the corresponding substance identified
* **Age at first use (primary and secondary):** For alcohol use, this is the age of first intoxication. For substances other than alcohol, this field identifies the age at which the client first used the corresponding substance identified in Substance Use.
* **Co-occurring mental and substance use disorders**: This field indicates whether the client has co-occurring mental, and substance use disorders
* **Health insurance:** This field indicates whether the patient has health insurance or not.
* **Census state FIPS code:** The state that patient belongs to.
* **Year of discharge:** Year of client's discharge from substance use treatment.

## Descriptive Analytics and Visualizations

Target Variable distribution:

Chart, histogram

Description automatically generated

Visualization of the important predictor variables:

Chart, bar chart

Description automatically generatedGraphical user interface, text, table

Description automatically generated

Most patients undergo Detox,24-hour, free standing residential treatment when compared to other forms of treatment.

Bar chart, square

Description automatically generated

We can see that there are almost equal numbers of adult and non-adult patients who are victims of substance abuse.

Chart, bar chart

Description automatically generatedA picture containing table

Description automatically generated

It is visible that most patients use oral and injection as the route of administration for the substance.

## Variable Exclusion & Initial Predictor Table

Initially in the dataset we started off with 33 variables out of 76 that we believed held the highest importance. We then started to find that 13 of the narrowed down variables had many missing values, some were missing over 50% data. Other variables did not supply a significant change in Length of stay so they were also excluded. Our final dataset then ended up with 21 variables including Length of stay.

## Predictor Table

|  |  |  |
| --- | --- | --- |
| Variables | Effect On LOS | Rationale |
| AGE | + | As age increases, the length of stay increases as older patients take more time to recover after substance abuse. |
| GENDER | +/- | Females might take more time to recover after a substance abuse. |
| EMPLOY\_D | - | If the patients are employed at the time they are discharged, they might be more motivated to complete their treatment, this decreasing length of stay. |
| SLIVARAG | +/- | If the patient lives under parent supervision, they might not be able to use excessive drugs, thus decreasing length of stay |
| STFIPS | ? | Different states might have different drug laws and different facilities which affect the length of stay |
| SERVICES\_D | ? | The services provided to patients will have different effects on the length of stay |
| NOPRIOR | ? | Having prior treatment, might decrease the length of stay of current treatment. |
| SUB1 | ? | Drugs such as Opioids and Methamphetamine might have adverse effects on patients than alcohol, thus increasing the length of stay. |
| SUB1\_D | ? | Patients might choose to stop pursuing their primary drug but might still be open to other various forms of drugs. |
| ROUTE1 | ? | Drugs administered through injection might have more effect than the rest. |
| FREQ1 | + | Higher the number of times the drug is used, the higher that affects their length of stay. |
| FRSTUSE1 | + | Recent first use of drug might cause shorter length of stay. |
| SUB2 | ? | Drugs such as Opioids and Methamphetamine might have adverse effects on patients than alcohol, thus increasing the length of stay. |
| SUB2\_D | ? | Patients might choose to stop pursuing their primary drug, but might still be open to other various forms of drugs |
| ROUTE2 | ? | Drugs administered through injection might have more effect than the rest. |
| FREQ2 | + | The higher the number of times the drug is used, the higher that affects their length of stay. |
| FRSTUSE2 | + | Recent first use of drug might cause shorter length of stay. |
| PSYPROB | + | Having co-occurring mental and substance abuse disorders increases length of stay. |
| DISYR | ? | The availability of drugs might vary in different years, because of stricter laws. |
| HLTHINS | + | If the patient has health insurance, they might prefer to go through the treatment, thus increasing length of stay. |

## Statistical Models for Analysis

The section discusses the models built for analysis, please refer to the appendix for stargazer and summary outputs

## Model 1 – Base Model

basic=glm(LOS ~ AGE + GENDER  + EMPLOY + LIVARAG + SERVICES + SUB\_PR + SUB\_PR\_D + ROUTE\_PR + FREQ\_PR +FRSTUSE\_PR+ PSYPROB  + HLTHINS ,data = d\_1, family=poisson(link=log))

This is a basic model with only primary substances included. We chose this model to read the LOS against age, gender, employment status, living arrangement, health insurance, mental health, and primary substance use.

## Model 2 – Poisson

m2 <- glm(LOS ~ AGE + GENDER  + EMPLOY + LIVARAG + SERVICES + NOPRIOR + SUB\_PR + SUB\_PR\_D + ROUTE\_PR + FREQ\_PR  + FRSTUSE\_PR + SUB\_SD + SUB\_SD\_D + ROUTE\_SD + FREQ\_SD  + FRSTUSE\_SD + PSYPROB  + HLTHINS, data = d\_1, family=poisson(link=log))

This is our second model, which uses all the factors we want to study using the Poisson model without including the effects of states on LOS. We used Poisson as our target variable is count data and has no negative and decimal values.

## Model 3 – Multilevel

m1 <- glmer(LOS ~ AGE + GENDER  + EMPLOY + LIVARAG + SERVICES + NOPRIOR + SUB\_PR + SUB\_PR\_D + ROUTE\_PR + FREQ\_PR  + FRSTUSE\_PR + SUB\_SD + SUB\_SD\_D + ROUTE\_SD + FREQ\_SD  + FRSTUSE\_SD + PSYPROB  + HLTHINS + (1 | STFIPS) + (1 | DISYR, data = d\_1\_sample, family=poisson(link=log))

We created a subset with 50,000 rows from entire dataset to run the multi-level model. This is a multi-level model with all the factors we want to study. It has a level on STFIPS to study the differences between states and another level on DISYR to study the differences between years. We can’t use states and years at the same level because they are not the same type of data.

## Choice of model:

We subset the data and built a model upon it to understand the common variables between different years. Using years and states on levels makes more sense as different states have different laws on drugs and alcohol, and laws keep changing year by year in so many states. Reading States and years let us know the impact of each state or each year separately.

## Interpretation of the marginal effects

Note: we have used color coding in marginal effects tables to indicate direction of effect. Blue indicates top three favorable (negative) effects and Orange indicates top three unfavorable (positive) effects. Green denotes the effects of key variables of interest.

### Model 3 : Multi Level

**Type of Service**: The longest length of stay is for Rehab short term (30 days or less) which is 100.2% longer stay than detox 24-hour hospital inpatient.

**Secondary Substance at admission**: Patients whose secondary substance is Marijuana have 15.6 % longer stay when compared to people whose secondary substance is alcohol.

**Age at admission:**  People between 15-17 years of age have 15.4% longer length of stay in the treatment facility compared to the people aged between 12 and 14.

**Living Arrangement:** People in dependent living have 7.1 % higher length of stay compared to homeless people.

**Health Insurance**: Uninsured people have 5.4% shorter stay at the treatment facility compared to people with insurance.

**Type of Service**: Ambulatory detoxification has the least amount of length of stay which is 23.1% lower when compared to detox 24-hour hospital inpatient.

**Secondary substance route of administration:**  People whose route of administration of Secondary substance is oral have the highest length of stay and whose route of administration is smoking have 10.2% lower length of stay compared to people whose route of administration is oral.

**Primary substance frequency of use:** People who had some use of substance have a 9.8% lower length of stay compared to people who had no use in the past month.

|  |  |  |
| --- | --- | --- |
| **Variable** | **β** | **Interpretation Value (%)** |
| Type of Service: Rehab Short Term (less than 30 days) | 1.024 | 102.40% |
| Secondary Substance Use at Admission: Marijuana | 0.156 | 15.60% |
| Age at admission:15-17 years | 0.154 | 15.40% |
| Living Arrangement: Dependent Living | 0.071 | 7.10% |
| Health Insurance: Uninsured | -0.054 | -5.40% |
| Type of Service: Ambulatory Detoxification | -0.231 | -23.10% |
| Secondary substance Route of Administration: Smoking | -0.102 | -10.20% |
| Frequency of Use: Some use | -0.098 | -9.80% |

## Assumptions:

As we are using a GLMER model, we do not need to check for normality, homoscedasticity, and linearity. Since all the variables considered are factors, multicollinearity cannot be tested.

## Recommendations:

* We see that patients who are homeless have lower Length of Stay, which might be because they can’t afford it. Since it is necessary to be treated after substance abuse, we recommend that cheaper facilities as well as government funded facilities must be made available to ensure everyone has access to treatment.
* It is also clear that uninsured patients have a lower length of stay, this might also be because they cant afford it or can get admitted, thus the government should subsidize health insurance policies for people that cannot afford proper living arrangements
* We can see that people with co-occurring substance and mental disorders have longer Length of Stay, we recommend that therapy counselling sessions must be provided to them to help in the recovery process.
* We see that people below 18 years of age are also consuming alcohol and it has adverse effects on them. As most underage drinking is done in private residences, there should be a neighborhood check on houses occupied primarily by teenagers by the HOA

## References

<https://americanaddictioncenters.org/rehab-guide/addiction-statistics>

<https://www.mayoclinic.org/diseases-conditions/drug-addiction/symptoms-causes/syc-20365112>

Citations:

Baek H, Cho M, Kim S, Hwang H, Song M, Yoo S. Analysis of length of hospital stay using electronic health records: A statistical and data mining approach. PLoS One. 2018 Apr 13;13(4): e0195901. doi: 10.1371/journal.pone.0195901. PMID: 29652932; PMCID: PMC5898738.

Battjes RJ, Gordon MS, O'Grady KE, Kinlock TW. Predicting retention of adolescents in substance abuse treatment. Addict Behav. 2004 Jul;29(5):1021-7. doi: 10.1016/j.addbeh.2004.02.054. PMID: 15219352.

Editorial Staff. (2022, October 21). *Alcohol and Drug Abuse Statistics (facts about addiction)*. American Addiction Centers. Retrieved November 3, 2022, from https://americanaddictioncenters.org/rehab-guide/addiction-statistics

Gifford E, Foster EM. Provider-level effects on psychiatric inpatient length of stay for youth with mental health and substance abuse disorders. Med Care. 2008 Mar;46(3):240-6. doi: 10.1097/MLR.0b013e318158aee7. PMID: 18388838.

Mayo Foundation for Medical Education and Research. (2022, October 4). *Drug addiction (substance use disorder)*. Mayo Clinic. Retrieved November 3, 2022, from

Ovalle A, Goldstein O, Kachuee M, Wu E, Hong C, Holloway I, Sarrafzadeh M Leveraging Social Media Activity and Machine Learning for HIV and Substance Abuse Risk Assessment: Development and Validation Study J Med Internet Res 2021;23(4): e22042 DOI: 10.2196/22042

Peek-Asa C, Yang J, Ramirez M, Hamann C, Cheng G. Factors affecting hospital charges and length of stay from teenage motor vehicle crash-related hospitalizations among United States teenagers, 2002-2007. Accid Anal Prev. 2011 May;43(3):595-600. doi: 10.1016/j.aap.2010.07.019. Epub 2010 Dec 18. PMID: 21376843.

Wright SP, Verouhis D, Gamble G, Swedberg K, Sharpe N, Doughty RN. Factors influencing the length of hospital stay of patients with heart failure. Eur J Heart Fail. 2003 Mar;5(2):201-9. doi: 10.1016/s1388-9842(02)00201-5. PMID: 12644013.

## Abstract

Parting thoughts:

We tried to run three models to find our length of stay, Poisson was our first choice, but we realized while doing our project that we need to do a multi-level model. The issue we primarily had we that was we could not run the whole dataset; thus we were left to use a sample of the data, which did not run after 50,000 samples. Our goal next time would be to run the whole dataset to get more accurate results.

## Model Summaries

BASE MODEL:

**Call:**

**glm(formula = LOS ~ AGE + GENDER + EMPLOY + LIVARAG + SERVICES +**

**+SUB\_PR + SUB\_PR\_D + ROUTE\_PR + FREQ\_PR + FRSTUSE\_PR + PSYPROB +**

**HLTHINS, family = poisson(link = log), data = d\_1)**

**Deviance Residuals:**

**Min       1Q   Median       3Q      Max**

**-7.2575  -1.6585  -0.4851   1.0216   8.4092**

**Coefficients:**

**Estimate Std. Error  z value Pr(>|z|)**

**(Intercept)                 1.9839343  0.0170760  116.182  < 2e-16 \*\*\***

**AGE15–17 years              0.0685267  0.0164963    4.154 3.27e-05 \*\*\***

**AGE18–20 years             -0.0763030  0.0161946   -4.712 2.46e-06 \*\*\***

**AGE21–24 years             -0.1180118  0.0160276   -7.363 1.80e-13 \*\*\***

**AGE25–29 years             -0.1286357  0.0159903   -8.045 8.65e-16 \*\*\***

**AGE30–34 years             -0.1297621  0.0159943   -8.113 4.94e-16 \*\*\***

**AGE35–39 years             -0.1311321  0.0160108   -8.190 2.61e-16 \*\*\***

**AGE40–44 years             -0.1592621  0.0160559   -9.919  < 2e-16 \*\*\***

**AGE45–49 years             -0.1726497  0.0160775  -10.739  < 2e-16 \*\*\***

**AGE50–54 years             -0.2227212  0.0161066  -13.828  < 2e-16 \*\*\***

**AGE55–64 years             -0.2414739  0.0161223  -14.978  < 2e-16 \*\*\***

**AGE65 years and older      -0.3986000  0.0181706  -21.937  < 2e-16 \*\*\***

**GENDERMale                 -0.0059653  0.0013077   -4.562 5.07e-06 \*\*\***

**EMPLOYNot in labor force    0.1098321  0.0018628   58.960  < 2e-16 \*\*\***

**EMPLOYUnemployed            0.0050616  0.0017857    2.835 0.004589 \*\***

**LIVARAG2                    0.0713697  0.0020441   34.915  < 2e-16 \*\*\***

**LIVARAG3                    0.0968705  0.0015573   62.204  < 2e-16 \*\*\***

**SERVICES2                  -0.0959974  0.0052958  -18.127  < 2e-16 \*\*\***

**SERVICES3                   0.7180525  0.0109193   65.760  < 2e-16 \*\*\***

**SERVICES4                   1.0999637  0.0052767  208.457  < 2e-16 \*\*\***

**SERVICES5                   0.9169198  0.0062047  147.779  < 2e-16 \*\*\***

**SERVICES6                   1.0113343  0.0056521  178.932  < 2e-16 \*\*\***

**SERVICES7                   0.6265383  0.0056541  110.812  < 2e-16 \*\*\***

**SERVICES8                  -0.0495658  0.0085957   -5.766 8.10e-09 \*\*\***

**SUB\_PRCocaine               0.0413824  0.0066101    6.260 3.84e-10 \*\*\***

**SUB\_PRHeroine              -0.0934462  0.0045947  -20.338  < 2e-16 \*\*\***

**SUB\_PRMarijuana             0.1028942  0.0073549   13.990  < 2e-16 \*\*\***

**SUB\_PRMethamphetamine       0.0378639  0.0092920    4.075 4.60e-05 \*\*\***

**SUB\_PROpiates\_Synthetics   -0.0211160  0.0056553   -3.734 0.000189 \*\*\***

**SUB\_PROthers                0.0052429  0.0060588    0.865 0.386854**

**SUB\_PR\_DCocaine             0.2314622  0.0063154   36.650  < 2e-16 \*\*\***

**SUB\_PR\_DHeroine             0.2372464  0.0037749   62.849  < 2e-16 \*\*\***

**SUB\_PR\_DMarijuana           0.0034223  0.0069464    0.493 0.622242**

**SUB\_PR\_DMethamphetamine     0.0805948  0.0090020    8.953  < 2e-16 \*\*\***

**SUB\_PR\_DNone               -0.0037112  0.0031098   -1.193 0.232711**

**SUB\_PR\_DOpiates\_Synthetics  0.1283281  0.0058841   21.809  < 2e-16 \*\*\***

**SUB\_PR\_DOthers              0.1730249  0.0066284   26.104  < 2e-16 \*\*\***

**ROUTE\_PR2                   0.0111952  0.0033948    3.298 0.000975 \*\*\***

**ROUTE\_PR3                   0.0410532  0.0032354   12.689  < 2e-16 \*\*\***

**ROUTE\_PR4                   0.0372078  0.0031458   11.828  < 2e-16 \*\*\***

**ROUTE\_PR5                  -0.0481001  0.0075511   -6.370 1.89e-10 \*\*\***

**FREQ\_PR2                   -0.1678260  0.0022336  -75.136  < 2e-16 \*\*\***

**FREQ\_PR3                   -0.0001649  0.0020402   -0.081 0.935599**

**FRSTUSE\_PR1                -0.0170691  0.0014252  -11.977  < 2e-16 \*\*\***

**PSYPROB2                   -0.0569423  0.0012485  -45.608  < 2e-16 \*\*\***

**HLTHINSUNINSURED           -0.1372993  0.0013185 -104.131  < 2e-16 \*\*\***

**---**

**Signif. codes:  0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

**(Dispersion parameter for poisson family taken to be 1)**

**Null deviance: 2125993  on 258663  degrees of freedom**

**Residual deviance: 1138526  on 258618  degrees of freedom**

**AIC: 2116806**

**Number of Fisher Scoring iterations: 5**

**M2 MODEL:**

**Call:**

**glm(formula = LOS ~ AGE + GENDER + EMPLOY + LIVARAG + SERVICES +**

**NOPRIOR + SUB\_PR + SUB\_PR\_D + ROUTE\_PR + FREQ\_PR + FRSTUSE\_PR +**

**SUB\_SD + SUB\_SD\_D + ROUTE\_SD + FREQ\_SD + FRSTUSE\_SD + PSYPROB +**

**HLTHINS, family = poisson(link = log), data = d\_1)**

**Deviance Residuals:**

**Min       1Q   Median       3Q      Max**

**-7.4268  -1.5317  -0.5051   1.0044   8.2381**

**Coefficients:**

**Estimate Std. Error  z value Pr(>|z|)**

**(Intercept)                 2.036189   0.017205  118.352  < 2e-16 \*\*\***

**AGE15–17 years              0.069919   0.016499    4.238 2.26e-05 \*\*\***

**AGE18–20 years             -0.086804   0.016209   -5.355 8.54e-08 \*\*\***

**AGE21–24 years             -0.136066   0.016047   -8.479  < 2e-16 \*\*\***

**AGE25–29 years             -0.148887   0.016014   -9.297  < 2e-16 \*\*\***

**AGE30–34 years             -0.152168   0.016021   -9.498  < 2e-16 \*\*\***

**AGE35–39 years             -0.154121   0.016039   -9.609  < 2e-16 \*\*\***

**AGE40–44 years             -0.182558   0.016085  -11.349  < 2e-16 \*\*\***

**AGE45–49 years             -0.193757   0.016107  -12.029  < 2e-16 \*\*\***

**AGE50–54 years             -0.231877   0.016137  -14.369  < 2e-16 \*\*\***

**AGE55–64 years             -0.243707   0.016153  -15.087  < 2e-16 \*\*\***

**AGE65 years and older      -0.376943   0.018198  -20.713  < 2e-16 \*\*\***

**GENDERMale                 -0.011110   0.001315   -8.449  < 2e-16 \*\*\***

**EMPLOYNot in labor force    0.098124   0.001872   52.426  < 2e-16 \*\*\***

**EMPLOYUnemployed            0.014330   0.001794    7.989 1.36e-15 \*\*\***

**LIVARAG2                    0.057835   0.002047   28.254  < 2e-16 \*\*\***

**LIVARAG3                    0.081972   0.001564   52.418  < 2e-16 \*\*\***

**SERVICES2                  -0.072518   0.005306  -13.667  < 2e-16 \*\*\***

**SERVICES3                   0.715660   0.010927   65.497  < 2e-16 \*\*\***

**SERVICES4                   1.097104   0.005284  207.631  < 2e-16 \*\*\***

**SERVICES5                   0.924875   0.006213  148.866  < 2e-16 \*\*\***

**SERVICES6                   1.004023   0.005663  177.295  < 2e-16 \*\*\***

**SERVICES7                   0.616444   0.005673  108.670  < 2e-16 \*\*\***

**SERVICES8                  -0.042752   0.008628   -4.955 7.22e-07 \*\*\***

**NOPRIOR1                   -0.005116   0.001449   -3.531 0.000414 \*\*\***

**SUB\_PRCocaine               0.024626   0.006786    3.629 0.000284 \*\*\***

**SUB\_PRHeroine              -0.104399   0.004866  -21.456  < 2e-16 \*\*\***

**SUB\_PRMarijuana             0.091855   0.007680   11.960  < 2e-16 \*\*\***

**SUB\_PRMethamphetamine       0.014156   0.009569    1.479 0.139016**

**SUB\_PROpiates\_Synthetics   -0.026655   0.005801   -4.595 4.33e-06 \*\*\***

**SUB\_PROthers               -0.008209   0.006154   -1.334 0.182239**

**SUB\_PR\_DCocaine             0.208470   0.006537   31.891  < 2e-16 \*\*\***

**SUB\_PR\_DHeroine             0.209253   0.004107   50.955  < 2e-16 \*\*\***

**SUB\_PR\_DMarijuana          -0.014542   0.007313   -1.989 0.046755 \***

**SUB\_PR\_DMethamphetamine     0.070829   0.009295    7.621 2.53e-14 \*\*\***

**SUB\_PR\_DNone               -0.020974   0.005937   -3.533 0.000411 \*\*\***

**SUB\_PR\_DOpiates\_Synthetics  0.130365   0.006040   21.582  < 2e-16 \*\*\***

**SUB\_PR\_DOthers              0.150115   0.006728   22.311  < 2e-16 \*\*\***

**ROUTE\_PR2                   0.037755   0.003413   11.062  < 2e-16 \*\*\***

**ROUTE\_PR3                   0.049542   0.003258   15.206  < 2e-16 \*\*\***

**ROUTE\_PR4                   0.044242   0.003183   13.898  < 2e-16 \*\*\***

**ROUTE\_PR5                  -0.029617   0.007755   -3.819 0.000134 \*\*\***

**FREQ\_PR2                   -0.133196   0.002684  -49.634  < 2e-16 \*\*\***

**FREQ\_PR3                   -0.004483   0.002648   -1.693 0.090414 .**

**FRSTUSE\_PR1                -0.022932   0.001497  -15.318  < 2e-16 \*\*\***

**SUB\_SDCocaine               0.180926   0.005072   35.671  < 2e-16 \*\*\***

**SUB\_SDHeroine               0.072202   0.006358   11.356  < 2e-16 \*\*\***

**SUB\_SDMarijuana             0.291138   0.005433   53.589  < 2e-16 \*\*\***

**SUB\_SDMethamphetamine       0.116747   0.008414   13.875  < 2e-16 \*\*\***

**SUB\_SDOpiates\_Synthetics    0.020130   0.005670    3.551 0.000384 \*\*\***

**SUB\_SDOthers               -0.037815   0.004769   -7.929 2.21e-15 \*\*\***

**SUB\_SD\_DCocaine             0.132154   0.004967   26.604  < 2e-16 \*\*\***

**SUB\_SD\_DHeroine             0.145400   0.006431   22.610  < 2e-16 \*\*\***

**SUB\_SD\_DMarijuana          -0.035319   0.005270   -6.702 2.05e-11 \*\*\***

**SUB\_SD\_DMethamphetamine     0.008222   0.008381    0.981 0.326535**

**SUB\_SD\_DNone                0.029096   0.006136    4.742 2.12e-06 \*\*\***

**SUB\_SD\_DOpiates\_Synthetics  0.062946   0.006134   10.261  < 2e-16 \*\*\***

**SUB\_SD\_DOthers             -0.039058   0.005209   -7.499 6.45e-14 \*\*\***

**ROUTE\_SD2                  -0.269989   0.002491 -108.388  < 2e-16 \*\*\***

**ROUTE\_SD3                  -0.153077   0.002733  -56.004  < 2e-16 \*\*\***

**ROUTE\_SD4                  -0.167516   0.002819  -59.420  < 2e-16 \*\*\***

**ROUTE\_SD5                  -0.222509   0.007548  -29.480  < 2e-16 \*\*\***

**FREQ\_SD2                   -0.032136   0.002342  -13.723  < 2e-16 \*\*\***

**FREQ\_SD3                   -0.007297   0.002417   -3.019 0.002535 \*\***

**FRSTUSE\_SD1                 0.032500   0.001470   22.106  < 2e-16 \*\*\***

**PSYPROB2                   -0.049524   0.001253  -39.537  < 2e-16 \*\*\***

**HLTHINSUNINSURED           -0.115886   0.001332  -86.984  < 2e-16 \*\*\***

**---**

**Signif. codes:  0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

**(Dispersion parameter for poisson family taken to be 1)**

**Null deviance: 2125993  on 258663  degrees of freedom**

**Residual deviance: 1106990  on 258597  degrees of freedom**

**AIC: 2085312**

**Number of Fisher Scoring iterations: 5**

## Stargazer Output for Comparison

**===========================================================================**

**Dependent variable:**

**------------------------------------------------**

**LOS**

**Poisson            generalized linear**

**mixed-effects**

**(1)            (2)              (3)**

**---------------------------------------------------------------------------**

**AGE15–17 years                0.069\*\*\*       0.070\*\*\*         0.154\*\*\***

**(0.016)        (0.016)          (0.039)**

**AGE18–20 years               -0.076\*\*\*      -0.087\*\*\*          0.033**

**(0.016)        (0.016)          (0.039)**

**AGE21–24 years               -0.118\*\*\*      -0.136\*\*\*          -0.005**

**(0.016)        (0.016)          (0.038)**

**AGE25–29 years               -0.129\*\*\*      -0.149\*\*\*          -0.033**

**(0.016)        (0.016)          (0.038)**

**AGE30–34 years               -0.130\*\*\*      -0.152\*\*\*          -0.034**

**(0.016)        (0.016)          (0.038)**

**AGE35–39 years               -0.131\*\*\*      -0.154\*\*\*          -0.041**

**(0.016)        (0.016)          (0.038)**

**AGE40–44 years               -0.159\*\*\*      -0.183\*\*\*          -0.059**

**(0.016)        (0.016)          (0.038)**

**AGE45–49 years               -0.173\*\*\*      -0.194\*\*\*          -0.035**

**(0.016)        (0.016)          (0.039)**

**AGE50–54 years               -0.223\*\*\*      -0.232\*\*\*         -0.090\*\***

**(0.016)        (0.016)          (0.039)**

**AGE55–64 years               -0.241\*\*\*      -0.244\*\*\*         -0.084\*\***

**(0.016)        (0.016)          (0.039)**

**AGE65 years and older        -0.399\*\*\*      -0.377\*\*\*          -0.031**

**(0.018)        (0.018)          (0.043)**

**GENDERMale                   -0.006\*\*\*      -0.011\*\*\*        -0.015\*\*\***

**(0.001)        (0.001)          (0.003)**

**EMPLOYNot in labor force      0.110\*\*\*       0.098\*\*\*         0.027\*\*\***

**(0.002)        (0.002)          (0.004)**

**EMPLOYUnemployed              0.005\*\*\*       0.014\*\*\*         0.048\*\*\***

**(0.002)        (0.002)          (0.004)**

**LIVARAG2                      0.071\*\*\*       0.058\*\*\*         0.071\*\*\***

**(0.002)        (0.002)          (0.005)**

**LIVARAG3                      0.097\*\*\*       0.082\*\*\*         0.059\*\*\***

**(0.002)        (0.002)          (0.004)**

**SERVICES2                    -0.096\*\*\*      -0.073\*\*\*        -0.101\*\*\***

**(0.005)        (0.005)          (0.014)**

**SERVICES3                     0.718\*\*\*       0.716\*\*\*         0.465\*\*\***

**(0.011)        (0.011)          (0.028)**

**SERVICES4                     1.100\*\*\*       1.097\*\*\*         1.024\*\*\***

**(0.005)        (0.005)          (0.014)**

**SERVICES5                     0.917\*\*\*       0.925\*\*\*         0.980\*\*\***

**(0.006)        (0.006)          (0.016)**

**SERVICES6                     1.011\*\*\*       1.004\*\*\*         0.972\*\*\***

**(0.006)        (0.006)          (0.014)**

**SERVICES7                     0.627\*\*\*       0.616\*\*\*         0.692\*\*\***

**(0.006)        (0.006)          (0.014)**

**SERVICES8                    -0.050\*\*\*      -0.043\*\*\*        -0.231\*\*\***

**(0.009)        (0.009)          (0.021)**

**NOPRIOR1                                    -0.005\*\*\*        -0.037\*\*\***

**(0.001)          (0.004)**

**SUB\_PRCocaine                 0.041\*\*\*       0.025\*\*\*         0.047\*\*\***

**(0.007)        (0.007)          (0.016)**

**SUB\_PRHeroine                -0.093\*\*\*      -0.104\*\*\*          -0.010**

**(0.005)        (0.005)          (0.011)**

**SUB\_PRMarijuana               0.103\*\*\*       0.092\*\*\*         0.087\*\*\***

**(0.007)        (0.008)          (0.017)**

**SUB\_PRMethamphetamine         0.038\*\*\*        0.014            -0.002**

**(0.009)        (0.010)          (0.022)**

**SUB\_PROpiates\_Synthetics     -0.021\*\*\*      -0.027\*\*\*         -0.026\*\***

**(0.006)        (0.006)          (0.013)**

**SUB\_PROthers                   0.005          -0.008           0.003**

**(0.006)        (0.006)          (0.014)**

**SUB\_PR\_DCocaine               0.231\*\*\*       0.208\*\*\*         0.153\*\*\***

**(0.006)        (0.007)          (0.015)**

**SUB\_PR\_DHeroine               0.237\*\*\*       0.209\*\*\*         0.102\*\*\***

**(0.004)        (0.004)          (0.010)**

**SUB\_PR\_DMarijuana              0.003         -0.015\*\*        -0.045\*\*\***

**(0.007)        (0.007)          (0.017)**

**SUB\_PR\_DMethamphetamine       0.081\*\*\*       0.071\*\*\*         0.149\*\*\***

**(0.009)        (0.009)          (0.021)**

**SUB\_PR\_DNone                   -0.004       -0.021\*\*\*         0.151\*\*\***

**(0.003)        (0.006)          (0.017)**

**SUB\_PR\_DOpiates\_Synthetics    0.128\*\*\*       0.130\*\*\*         0.105\*\*\***

**(0.006)        (0.006)          (0.014)**

**SUB\_PR\_DOthers                0.173\*\*\*       0.150\*\*\*         0.123\*\*\***

**(0.007)        (0.007)          (0.015)**

**ROUTE\_PR2                     0.011\*\*\*       0.038\*\*\*         0.025\*\*\***

**(0.003)        (0.003)          (0.008)**

**ROUTE\_PR3                     0.041\*\*\*       0.050\*\*\*          0.002**

**(0.003)        (0.003)          (0.007)**

**ROUTE\_PR4                     0.037\*\*\*       0.044\*\*\*          -0.010**

**(0.003)        (0.003)          (0.007)**

**ROUTE\_PR5                    -0.048\*\*\*      -0.030\*\*\*          0.001**

**(0.008)        (0.008)          (0.018)**

**FREQ\_PR2                     -0.168\*\*\*      -0.133\*\*\*        -0.098\*\*\***

**(0.002)        (0.003)          (0.006)**

**FREQ\_PR3                      -0.0002        -0.004\*         -0.082\*\*\***

**(0.002)        (0.003)          (0.006)**

**FRSTUSE\_PR1                  -0.017\*\*\*      -0.023\*\*\*        -0.014\*\*\***

**(0.001)        (0.001)          (0.003)**

**SUB\_SDCocaine                                0.181\*\*\*         0.082\*\*\***

**(0.005)          (0.012)**

**SUB\_SDHeroine                                0.072\*\*\*         0.065\*\*\***

**(0.006)          (0.014)**

**SUB\_SDMarijuana                              0.291\*\*\*         0.156\*\*\***

**(0.005)          (0.012)**

**SUB\_SDMethamphetamine                        0.117\*\*\*          0.034\***

**(0.008)          (0.019)**

**SUB\_SDOpiates\_Synthetics                     0.020\*\*\*         0.042\*\*\***

**(0.006)          (0.013)**

**SUB\_SDOthers                                -0.038\*\*\*          -0.011**

**(0.005)          (0.011)**

**SUB\_SD\_DCocaine                              0.132\*\*\*         0.097\*\*\***

**(0.005)          (0.011)**

**SUB\_SD\_DHeroine                              0.145\*\*\*         0.081\*\*\***

**(0.006)          (0.015)**

**SUB\_SD\_DMarijuana                           -0.035\*\*\*         -0.023\*\***

**(0.005)          (0.012)**

**SUB\_SD\_DMethamphetamine                       0.008           0.125\*\*\***

**(0.008)          (0.019)**

**SUB\_SD\_DNone                                 0.029\*\*\*          -0.001**

**(0.006)          (0.014)**

**SUB\_SD\_DOpiates\_Synthetics                   0.063\*\*\*         0.042\*\*\***

**(0.006)          (0.014)**

**SUB\_SD\_DOthers                              -0.039\*\*\*          0.011**

**(0.005)          (0.012)**

**ROUTE\_SD2                                   -0.270\*\*\*        -0.102\*\*\***

**(0.002)          (0.006)**

**ROUTE\_SD3                                   -0.153\*\*\*        -0.058\*\*\***

**(0.003)          (0.006)**

**ROUTE\_SD4                                   -0.168\*\*\*        -0.067\*\*\***

**(0.003)          (0.006)**

**ROUTE\_SD5                                   -0.223\*\*\*        -0.083\*\*\***

**(0.008)          (0.018)**

**FREQ\_SD2                                    -0.032\*\*\*        -0.037\*\*\***

**(0.002)          (0.005)**

**FREQ\_SD3                                    -0.007\*\*\*        -0.057\*\*\***

**(0.002)          (0.006)**

**FRSTUSE\_SD1                                  0.032\*\*\*         0.013\*\*\***

**(0.001)          (0.003)**

**PSYPROB2                     -0.057\*\*\*      -0.050\*\*\*        -0.062\*\*\***

**(0.001)        (0.001)          (0.003)**

**HLTHINSUNINSURED             -0.137\*\*\*      -0.116\*\*\*        -0.054\*\*\***

**(0.001)        (0.001)          (0.003)**

**Constant                      1.984\*\*\*       2.036\*\*\*         1.937\*\*\***

**(0.017)        (0.017)          (0.066)**

**---------------------------------------------------------------------------**

**Observations                  258,664        258,664           50,000**

**Log Likelihood             -1,058,357.000 -1,042,589.000    -184,833.700**

**Akaike Inf. Crit.          2,116,806.000  2,085,312.000     369,805.400**

**Bayesian Inf. Crit.                                         370,414.000**

**===========================================================================**

**Note:                                           \*p<0.1; \*\*p<0.05; \*\*\*p<0.01**

**Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [**

**glmerMod]**

**Family: poisson  ( log )**

**Formula: LOS ~ AGE + GENDER + EMPLOY + LIVARAG + SERVICES + NOPRIOR +**

**SUB\_PR + SUB\_PR\_D + ROUTE\_PR + FREQ\_PR + FRSTUSE\_PR + SUB\_SD +**

**SUB\_SD\_D + ROUTE\_SD + FREQ\_SD + FRSTUSE\_SD + PSYPROB + HLTHINS +**

**(1 | STFIPS) + (1 | DISYR)**

**Data: d\_1\_sample**

**AIC       BIC    logLik  deviance  df.resid**

**369805.4  370414.0 -184833.7  369667.4     49931**

**Scaled residuals:**

**Min      1Q  Median      3Q     Max**

**-6.1432 -1.1558 -0.4856  0.8443 16.2938**

**Random effects:**

**Groups Name        Variance  Std.Dev.**

**STFIPS (Intercept) 0.1032579 0.32134**

**DISYR  (Intercept) 0.0003529 0.01879**

**Number of obs: 50000, groups:  STFIPS, 40; DISYR, 5**

**Fixed effects:**

**Estimate Std. Error z value Pr(>|z|)**

**(Intercept)                 1.937059   0.066369  29.186  < 2e-16 \*\*\***

**AGE15–17 years              0.154441   0.039483   3.912 9.17e-05 \*\*\***

**AGE18–20 years              0.032788   0.038755   0.846 0.397526**

**AGE21–24 years             -0.005326   0.038413  -0.139 0.889733**

**AGE25–29 years             -0.032992   0.038343  -0.860 0.389545**

**AGE30–34 years             -0.033969   0.038345  -0.886 0.375681**

**AGE35–39 years             -0.040801   0.038390  -1.063 0.287872**

**AGE40–44 years             -0.059337   0.038487  -1.542 0.123139**

**AGE45–49 years             -0.035069   0.038536  -0.910 0.362809**

**AGE50–54 years             -0.090062   0.038604  -2.333 0.019652 \***

**AGE55–64 years             -0.084310   0.038652  -2.181 0.029163 \***

**AGE65 years and older      -0.030896   0.043384  -0.712 0.476368**

**GENDERMale                 -0.014887   0.003021  -4.927 8.33e-07 \*\*\***

**EMPLOYNot in labor force    0.026766   0.004429   6.043 1.51e-09 \*\*\***

**EMPLOYUnemployed            0.048332   0.004155  11.632  < 2e-16 \*\*\***

**LIVARAG2                    0.071255   0.004803  14.836  < 2e-16 \*\*\***

**LIVARAG3                    0.058560   0.003591  16.306  < 2e-16 \*\*\***

**SERVICES2                  -0.100843   0.013762  -7.327 2.35e-13 \*\*\***

**SERVICES3                   0.464784   0.028143  16.515  < 2e-16 \*\*\***

**SERVICES4                   1.023715   0.013559  75.499  < 2e-16 \*\*\***

**SERVICES5                   0.979546   0.015596  62.806  < 2e-16 \*\*\***

**SERVICES6                   0.972460   0.014288  68.062  < 2e-16 \*\*\***

**SERVICES7                   0.692477   0.014237  48.638  < 2e-16 \*\*\***

**SERVICES8                  -0.231308   0.021477 -10.770  < 2e-16 \*\*\***

**NOPRIOR1                   -0.036958   0.003516 -10.510  < 2e-16 \*\*\***

**SUB\_PRCocaine               0.046773   0.015944   2.934 0.003351 \*\***

**SUB\_PRHeroine              -0.009511   0.011143  -0.854 0.393344**

**SUB\_PRMarijuana             0.086635   0.017398   4.980 6.37e-07 \*\*\***

**SUB\_PRMethamphetamine      -0.002140   0.021980  -0.097 0.922436**

**SUB\_PROpiates\_Synthetics   -0.026261   0.013335  -1.969 0.048913 \***

**SUB\_PROthers                0.003166   0.014165   0.223 0.823147**

**SUB\_PR\_DCocaine             0.153063   0.015477   9.889  < 2e-16 \*\*\***

**SUB\_PR\_DHeroine             0.101541   0.009517  10.669  < 2e-16 \*\*\***

**SUB\_PR\_DMarijuana          -0.044722   0.016611  -2.692 0.007096 \*\***

**SUB\_PR\_DMethamphetamine     0.148993   0.021376   6.970 3.16e-12 \*\*\***

**SUB\_PR\_DNone                0.151195   0.017109   8.837  < 2e-16 \*\*\***

**SUB\_PR\_DOpiates\_Synthetics  0.104562   0.013912   7.516 5.66e-14 \*\*\***

**SUB\_PR\_DOthers              0.123314   0.015491   7.960 1.71e-15 \*\*\***

**ROUTE\_PR2                   0.024571   0.007748   3.171 0.001517 \*\***

**ROUTE\_PR3                   0.002053   0.007475   0.275 0.783600**

**ROUTE\_PR4                  -0.009981   0.007273  -1.372 0.169957**

**ROUTE\_PR5                   0.001346   0.018024   0.075 0.940466**

**FREQ\_PR2                   -0.097687   0.006171 -15.829  < 2e-16 \*\*\***

**FREQ\_PR3                   -0.081638   0.006106 -13.371  < 2e-16 \*\*\***

**FRSTUSE\_PR1                -0.014421   0.003421  -4.216 2.49e-05 \*\*\***

**SUB\_SDCocaine               0.081785   0.011651   7.020 2.22e-12 \*\*\***

**SUB\_SDHeroine               0.064508   0.014458   4.462 8.13e-06 \*\*\***

**SUB\_SDMarijuana             0.155931   0.012300  12.677  < 2e-16 \*\*\***

**SUB\_SDMethamphetamine       0.034332   0.018980   1.809 0.070472 .**

**SUB\_SDOpiates\_Synthetics    0.042279   0.012728   3.322 0.000895 \*\*\***

**SUB\_SDOthers               -0.010658   0.010880  -0.980 0.327315**

**SUB\_SD\_DCocaine             0.096789   0.011376   8.508  < 2e-16 \*\*\***

**SUB\_SD\_DHeroine             0.080859   0.014594   5.541 3.01e-08 \*\*\***

**SUB\_SD\_DMarijuana          -0.023377   0.011806  -1.980 0.047681 \***

**SUB\_SD\_DMethamphetamine     0.125306   0.018823   6.657 2.79e-11 \*\*\***

**SUB\_SD\_DNone               -0.001433   0.014376  -0.100 0.920610**

**SUB\_SD\_DOpiates\_Synthetics  0.041688   0.013820   3.017 0.002557 \*\***

**SUB\_SD\_DOthers              0.011007   0.011840   0.930 0.352582**

**ROUTE\_SD2                  -0.101532   0.005904 -17.198  < 2e-16 \*\*\***

**ROUTE\_SD3                  -0.057595   0.006278  -9.174  < 2e-16 \*\*\***

**ROUTE\_SD4                  -0.066957   0.006491 -10.315  < 2e-16 \*\*\***

**ROUTE\_SD5                  -0.082778   0.017510  -4.728 2.27e-06 \*\*\***

**FREQ\_SD2                   -0.037473   0.005423  -6.910 4.86e-12 \*\*\***

**FREQ\_SD3                   -0.056715   0.005614 -10.103  < 2e-16 \*\*\***

**FRSTUSE\_SD1                 0.013407   0.003356   3.995 6.46e-05 \*\*\***

**PSYPROB2                   -0.062450   0.002980 -20.957  < 2e-16 \*\*\***

**HLTHINSUNINSURED           -0.054060   0.003459 -15.630  < 2e-16 \*\*\***

**---**

**Signif. codes:  0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

## Prior Work & Relevance

|  |  |  |  |
| --- | --- | --- | --- |
| Article | Cause | Intervention | Other Notable Points |
| Robert J Battjes, Michael S Gordon, Kevin E O'Grady, Timothy W Kinlock, Predicting retention of adolescents in substance abuse treatment, Addictive Behaviors, Volume 29, Issue 5, 2004.https://www.sciencedirect.com/science/article/pii/S0306460304000668 | Demographic characteristics - substance-use involvement, deviant behavior such as aggressive behaviors, physical force, criminal activity, consequences of substance use | -Give importance to identifying and addressing emotional problems  -Pressure to enter the treatment is often short-lived, so, initial pressure is not sufficient to retain clients in treatment. | Only low severed substance use youths were studied, findings not to be generalized with more severe disorders |
| Gifford E, Foster EM. Provider-level effects on psychiatric inpatient length of stay for youth with mental health and substance abuse disorders. Med Care. 2008 Mar;46(3):240-6. doi: 10.1097/MLR.0b013e318158aee7. PMID: 18388838 | -Facility level factors which include:  Type of hospital  Experience of the doctors  Facilities available | About half of the variation in LOS is explained by facility-level factors. Given the vulnerable nature of youth who are in need of inpatient psychiatric care, it may be particularly important to monitor provider-level processes and outcomes | Several studies identified region of the state or area as being a salient factor that is external to the child in predicting LOS |
| Barenholtz, Elan, et al. “Machine-Learning Approaches to Substance-Abuse Research: Emerging Trends and Their Implications.” Current Opinion in Psychiatry, vol. 33, no. 4, 2020, pp.33442, <https://doi.org/10.1097/YCO.0000000000000611.> | A multidimensional dataset that included EEG measures, genomic data [alcohol-related single nucleotide polymorphisms (SNPs)], and histories of alcohol use disorder, including longitudinal data from an early age | Multiple different ML models were used to discover emerging trends such as artificial neural networks (ANN), random forest, and gradient boosting machine (GBM). On a social media dataset, they applied Deep convolutional neural network (CNN’s) which predicted with the highest accuracy trends on posted pictures for future substance abusers. SVM models that combined both EEG and SNP features achieved higher classification performance for later abuse | A general observed trend is that some of the best performances were for concurrent physiological and behavioral measures for predicting current substance use, although even those models did not always generalize beyond their training data |
| Peek-Asa, Yang, J., Ramirez, M., Hamann, C., & Cheng, G. (2011). Factors affecting hospital charges and length of stay from teenage motor vehicle crash-related hospitalizations among United States teenagers, 2002–2007. Accident Analysis and Prevention, 43(3), 595–600. <https://doi.org/10.1016/j.aap.2010.07.019> | This article studies the length of stay in hospital because of other health related issues too.  Motor vehicle crashes are the leading cause of death for teenagers in the United and lead to considerable medical costs and LOS | Alcohol and drug use at the time of the crash was not known. Teens with alcohol comorbidities had lower charges and lengths of stay than teens without alcohol comorbidities, although teens with drug comorbidities have higher charges and lengths of stay than teens without drug comorbidities | Of hospitalized teens in our sample, 5.8% were diagnosed with an alcohol abuse comorbidity and 4.4% were diagnosed with a drug abuse comorbidity. |
| Ovalle, Anaelia, et al. “Leveraging Social Media Activity and Machine Learning for HIV and Substance Abuse Risk Assessment: Development and Validation Study.” Journal of Medical Internet Research, vol. 23, no. 4, 2021, pp. e22042–e22042, https://doi.org/10.2196/22042 | HIV and other sexual diseases are related to an increase in substance abuse.  Social media and dating apps contribute to the increase in drug use by making it normal. Trigger words used online could also encourage substance abuse | Standardized counts of each tokenized word and text-summary features, such as message length, were used as simple features to predict drug use and sexually transmitted diseases, this showed how effective social media is in normalizing drug abuse. There is a positive correlation between drug use and sexual diseases, with HIV more prevalent in methamphetamine users | Creating the data collector required custom handling of each data source. For instance, Facebook's policy change called for a completely different approach to data collecting in the middle of the study. To provide a scalable solution, the data collection platform should be flexible enough to adjust to a new collection regime due to circumstances outside of the study's control. |
| Baek H, Cho M, Kim S, Hwang H, Song M, Yoo S. Analysis of length of hospital stay using electronic health records: A statistical and data mining approach. PLoS One. 2018 Apr 13;13(4):e0195901. doi: 10.1371/journal.pone.0195901. PMID: 29652932; PMCID: PMC5898738. | This article shows the data mining approach of Length of hospital stay analysis.  Transfer, discharge delay time, operation frequency, frequency of diagnosis, severity, bed grade, and insurance type affect the most. | The department of rehabilitation medicine (RH) had the highest average LOS at 15.9 days. Of all the conditions diagnosed over 250 times, diagnoses of I63.8 (cerebral infarction, middle cerebral artery), I63.9 (infarction of middle cerebral artery territory) and I21.9 (myocardial infarction) were associated with the longest average hospital stay and high standard deviation.  **Analysis Items**  Performance analysis for LOS.  LOS analysis in accordance with diagnosis  Analysis for long-term hospitalization patients  LOS analysis in terms of transfer patterns Deriving correlated factors on LOS | 55% of inpatients were discharged within 4 days.  Patients were analyzed according to the following three categories: descriptive and exploratory analysis, process pattern analysis using process mining techniques, and statistical analysis and pre- diction of LOS. |
| Factors influencing the length of hospital stay of patients with heart failure[S.P. Wright](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorRaw=Wright%2C+SP),[D. Verouhis](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorRaw=Verouhis%2C+D),[G. Gamble](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorRaw=Gamble%2C+G),[K. Swedberg](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorRaw=Swedberg%2C+K),[N. Sharpe](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorRaw=Sharpe%2C+N),[R.N. Doughty](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorRaw=Doughty%2C+RN) | Determinants of length of hospital stay for patients with Heart Failure include ethnicity  medical comorbidity  disease severity  clinical presentation  in-patient treatment in-hospital progress | Multivariate models only partly explained variance in hospital stay, suggesting the importance of pre-admission and post-discharge factors, including the healthcare environment, the availability of primary and secondary care resources, and the threshold for hospital admission. | Factors independently associated with length of stay in the top quartile (>10 days) on logistic regression included the presence of oedema at admission, change in weight during stay, duration of treatment with iv diuretic, the development of renal impairment, concurrent respiratory problems requiring specific treatment, and social problems requiring intervention. |
| Miriam Delphin-Rittmon, Ph.D., Raquel Andres-Hyman, Ph.D., Elizabeth H Flanagan, Ph.D., Jose Ortiz, M.S., M.B.A., Mona M Amer, Ph.D., and Larry Davidson, Ph.D. Racial-Ethnic Differences in Referral Source, Diagnosis, and Length of Stay in Inpatient Substance Abuse Treatment <https://doi.org/10.1176/appi.ps.201100322>  , Ph.D.,  Raquel Andres-Hyman  , Ph.D., | Referral source  - Diagnosis at the time of admission such as alcohol use disorder diagnosis, drug use disorder diagnosis, personality disorder diagnosis, personality disorder diagnosis  - number of days spent in inpatient substance use treatment | Identification of racial-ethnic differences in substance abuse treatment can provide the foundation to achieve equity in care and treatment | findings are from a single behavioral health system over a limited time period. Information not unavailable for the nativity of the ethnic groups. |

## Links to Data Source & Dictionary

Dataset: <https://www.datafiles.samhsa.gov/dataset/treatment-episode-data-set-admissions-2019-teds-2019-ds0001>

Codebook: https://www.datafiles.samhsa.gov/sites/default/files/field-uploads-protected/studies/TEDS-D-2019/TEDS-D-2019-datasets/TEDS-D-2019-DS0001/TEDS-D-2019-DS0001-info/TEDS-D-2019-DS0001-info-codebook\_V1.pdf

## R-Code

**library(car)**

**library(rio)**

**d\_2019 = import('tedsd\_puf\_2019.csv')**

**d\_2018 = import('tedsd\_puf\_2018.csv')**

**d\_2017 = import('tedsd\_puf\_2017.csv')**

**d\_2016 = import('tedsd\_2016\_puf.csv')**

**d\_2015 = import('tedsd\_2015\_puf.csv')**

**#Treatment Completed only, Reason 1 is treatment completed.**

**d\_2019 = subset(d\_2019, REASON == 1)**

**d\_2018 = subset(d\_2018, REASON == 1)**

**d\_2017 = subset(d\_2017, REASON == 1)**

**d\_2016 = subset(d\_2016, REASON == 1)**

**d\_2015 = subset(d\_2015, REASON == 1)**

**#Columns that we are using.**

**d\_2019= d\_2019[,c("AGE","GENDER","EMPLOY\_D","LIVARAG",**

**"STFIPS","SERVICES\_D",**

**"LOS","NOPRIOR","SUB1","SUB1\_D","ROUTE1","FREQ1",**

**"FRSTUSE1","SUB2","SUB2\_D","ROUTE2","FREQ2","FRSTUSE2",**

**"PSYPROB","DISYR","HLTHINS" )]**

**d\_2018= d\_2018[,c("AGE","GENDER","EMPLOY\_D","LIVARAG",**

**"STFIPS","SERVICES\_D",**

**"LOS","NOPRIOR","SUB1","SUB1\_D","ROUTE1","FREQ1",**

**"FRSTUSE1","SUB2","SUB2\_D","ROUTE2","FREQ2","FRSTUSE2",**

**"PSYPROB","DISYR","HLTHINS")]**

**d\_2017= d\_2017[,c("AGE","GENDER","EMPLOY\_D","LIVARAG",**

**"STFIPS","SERVICES\_D",**

**"LOS","NOPRIOR","SUB1","SUB1\_D","ROUTE1","FREQ1",**

**"FRSTUSE1","SUB2","SUB2\_D","ROUTE2","FREQ2","FRSTUSE2",**

**"PSYPROB","DISYR","HLTHINS" )]**

**d\_2016= d\_2016[,c("AGE","GENDER","EMPLOY\_D","LIVARAG",**

**"STFIPS","SERVICES\_D",**

**"LOS","NOPRIOR","SUB1","SUB1\_D","ROUTE1","FREQ1",**

**"FRSTUSE1","SUB2","SUB2\_D","ROUTE2","FREQ2","FRSTUSE2",**

**"PSYPROB","DISYR","HLTHINS" )]**

**d\_2015= d\_2015[,c("AGE","GENDER","EMPLOY\_D","LIVARAG",**

**"STFIPS","SERVICES\_D",**

**"LOS","NOPRIOR","SUB1","SUB1\_D","ROUTE1","FREQ1",**

**"FRSTUSE1","SUB2","SUB2\_D","ROUTE2","FREQ2","FRSTUSE2",**

**"PSYPROB","DISYR","HLTHINS" )]**

**d= rbind(d\_2015, d\_2016, d\_2017, d\_2018, d\_2019)**

**d\_1=subset(d,LOS<31)**

**d\_1=subset(d\_1,GENDER!='-9')**

**d\_1=subset(d\_1,EMPLOY\_D!='-9')**

**d\_1=subset(d\_1,LIVARAG!='-9')**

**d\_1=subset(d\_1,SERVICES\_D!='-9')**

**d\_1=subset(d\_1,NOPRIOR!='-9')**

**d\_1=subset(d\_1,SUB1!='-9')**

**d\_1=subset(d\_1,SUB1\_D!='-9')**

**d\_1=subset(d\_1,ROUTE1!='-9')**

**d\_1=subset(d\_1,FREQ1!='-9')**

**d\_1=subset(d\_1,FRSTUSE1!='-9')**

**d\_1=subset(d\_1,SUB2!='-9')**

**d\_1=subset(d\_1,SUB2\_D!='-9')**

**d\_1=subset(d\_1,ROUTE2!='-9')**

**d\_1=subset(d\_1,FREQ2!='-9')**

**d\_1=subset(d\_1,FRSTUSE2!='-9')**

**d\_1=subset(d\_1,PSYPROB!='-9')**

**d\_1=subset(d\_1,HLTHINS!='-9')**

**d\_1=subset(d\_1,SUB1!='1')**

**d\_1=subset(d\_1,SUB2!='1')**

**str(d\_1)**

**d\_1$SUB1 = ifelse(d\_1$SUB1 == 2,"Alcholol",ifelse(d\_1$SUB1 == 3,"Cocaine",**

**ifelse(d\_1$SUB1 == 4, "Marijuana", ifelse(d\_1$SUB1 == 5, "Heroine",**

**ifelse(d\_1$SUB1 ==7, "Opiates\_Synthetics",ifelse(d\_1$SUB1 == 10,"Methamphetamine",**

**"Others"))))))**

**d\_1$SUB2 = ifelse(d\_1$SUB2 == 2,"Alcholol",ifelse(d\_1$SUB2 == 3,"Cocaine",**

**ifelse(d\_1$SUB2 == 4, "Marijuana", ifelse(d\_1$SUB2 == 5, "Heroine",**

**ifelse(d\_1$SUB2 ==7, "Opiates\_Synthetics",ifelse(d\_1$SUB2 == 10,"Methamphetamine",**

**"Others"))))))**

**d\_1$SUB1\_D = ifelse(d\_1$SUB1\_D == 2,"Alcholol",ifelse(d\_1$SUB1\_D == 3,"Cocaine",**

**ifelse(d\_1$SUB1\_D == 4, "Marijuana", ifelse(d\_1$SUB1\_D == 5, "Heroine",**

**ifelse(d\_1$SUB1\_D ==7, "Opiates\_Synthetics",ifelse(d\_1$SUB1\_D == 10,"Methamphetamine",ifelse(d\_1$SUB1\_D == 1,"None",**

**"Others")))))))**

**d\_1$SUB2\_D = ifelse(d\_1$SUB2\_D == 2,"Alcholol",ifelse(d\_1$SUB2\_D == 3,"Cocaine",**

**ifelse(d\_1$SUB2\_D == 4, "Marijuana", ifelse(d\_1$SUB2\_D == 5, "Heroine",**

**ifelse(d\_1$SUB2\_D ==7, "Opiates\_Synthetics",ifelse(d\_1$SUB2\_D == 10,"Methamphetamine",ifelse(d\_1$SUB2\_D == 1,"None",**

**"Others")))))))**

**#FirstUse Age into binary(non adults and adults)**

**#FRSTUSE1**

**d\_1$FRSTUSE1 = ifelse(d\_1$FRSTUSE1 <= 3,0,1)**

**#FRSTUSE2**

**d\_1$FRSTUSE2 = ifelse(d\_1$FRSTUSE2 <= 3,0,1)**

**d\_1$AGE = ifelse(d\_1$AGE == 1,'12â“14 years',**

**ifelse(d\_1$AGE == 2,'15â“17 years',**

**ifelse(d\_1$AGE == 3,'18â“20 years',**

**ifelse(d\_1$AGE ==4,'21â“24 years',**

**ifelse(d\_1$AGE==5,'25â“29 years',**

**ifelse(d\_1$AGE==6,'30â“34 years',**

**ifelse(d\_1$AGE==7,'35â“39 years',**

**ifelse(d\_1$AGE==8,'40â“44 years',**

**ifelse(d\_1$AGE==9,'45â“49 years',**

**ifelse(d\_1$AGE==10,'50â“54 years',**

**ifelse(d\_1$AGE==11,'55â“64 years',**

**ifelse(d\_1$AGE==12,'65 years and older',NA)**

**)))))))))))**

**d\_1$GENDER=ifelse(d\_1$GENDER==1,'Male','Female')**

**d\_1$EMPLOY\_D <- ifelse(d\_1$EMPLOY\_D ==1, "Employed",**

**ifelse(d\_1$EMPLOY\_D==2, "Employed",**

**ifelse(d\_1$EMPLOY\_D ==3, "Unemployed",**

**ifelse(d\_1$EMPLOY\_D==4, "Not in labor force", NA))))**

**d\_1$HLTHINS=ifelse(d\_1$HLTHINS<=3,'INSURED','UNINSURED')**

**#Factorizing the columns**

**cols\_to\_be\_factorized = c("AGE","GENDER","EMPLOY\_D","LIVARAG",**

**"STFIPS","SERVICES\_D",**

**"NOPRIOR","SUB1","SUB1\_D","ROUTE1","FREQ1",**

**"FRSTUSE1","SUB2","SUB2\_D","ROUTE2","FREQ2","FRSTUSE2",**

**"PSYPROB","DISYR","HLTHINS" )**

**d\_1[cols\_to\_be\_factorized] = lapply(d\_1[cols\_to\_be\_factorized],factor)**

**str(d\_1)**

**colnames(d\_1)[which(names(d\_1) == 'EMPLOY\_D')] <- 'EMPLOY'**

**colnames(d\_1)[which(names(d\_1) == 'SERVICES\_D')] <- 'SERVICES'**

**colnames(d\_1)[which(names(d\_1) == 'SUB1')] <- 'SUB\_PR'**

**colnames(d\_1)[which(names(d\_1) == 'SUB1\_D')] <- 'SUB\_PR\_D'**

**colnames(d\_1)[which(names(d\_1) == 'ROUTE1')] <- 'ROUTE\_PR'**

**colnames(d\_1)[which(names(d\_1) == 'FREQ1')] <- 'FREQ\_PR'**

**colnames(d\_1)[which(names(d\_1) == 'FRSTUSE1')] <- 'FRSTUSE\_PR'**

**colnames(d\_1)[which(names(d\_1) == 'SUB2')] <- 'SUB\_SD'**

**colnames(d\_1)[which(names(d\_1) == 'SUB2\_D')] <- 'SUB\_SD\_D'**

**colnames(d\_1)[which(names(d\_1) == 'ROUTE2')] <- 'ROUTE\_SD'**

**colnames(d\_1)[which(names(d\_1) == 'FREQ2')] <- 'FREQ\_SD'**

**colnames(d\_1)[which(names(d\_1) == 'FRSTUSE2')] <- 'FRSTUSE\_SD'**

**d\_1\_sample = d\_1[sample(1:nrow(d\_1),50000,**

**replace=FALSE),]**

**library(lme4)**

**m1 <- glmer(LOS ~ AGE + GENDER + EMPLOY + LIVARAG + SERVICES +**

**NOPRIOR + SUB\_PR + SUB\_PR\_D + ROUTE\_PR + FREQ\_PR +**

**FRSTUSE\_PR + SUB\_SD + SUB\_SD\_D + ROUTE\_SD + FREQ\_SD + FRSTUSE\_SD**

**+ PSYPROB + HLTHINS + (1 | STFIPS) + (1 | DISYR)**

**,data = d\_1\_sample, family=poisson(link=log))**

**summary(m1)**

**ranef(m1)**

**m2 <- glm(LOS ~ AGE + GENDER + EMPLOY + LIVARAG + SERVICES +**

**NOPRIOR + SUB\_PR + SUB\_PR\_D + ROUTE\_PR + FREQ\_PR +**

**FRSTUSE\_PR + SUB\_SD + SUB\_SD\_D + ROUTE\_SD + FREQ\_SD + FRSTUSE\_SD**

**+ PSYPROB + HLTHINS**

**,data = d\_1, family=poisson(link=log))**

**summary(m2)**

**basic=glm(LOS ~ AGE + GENDER + EMPLOY + LIVARAG + SERVICES +**

**+ SUB\_PR + SUB\_PR\_D + ROUTE\_PR + FREQ\_PR +FRSTUSE\_PR+**

**PSYPROB + HLTHINS ,data = d\_1, family=poisson(link=log))**

**summary(basic)**

**library(stargazer)**

**stargazer(basic,m2,m1,type='text', Single=TRUE)**