

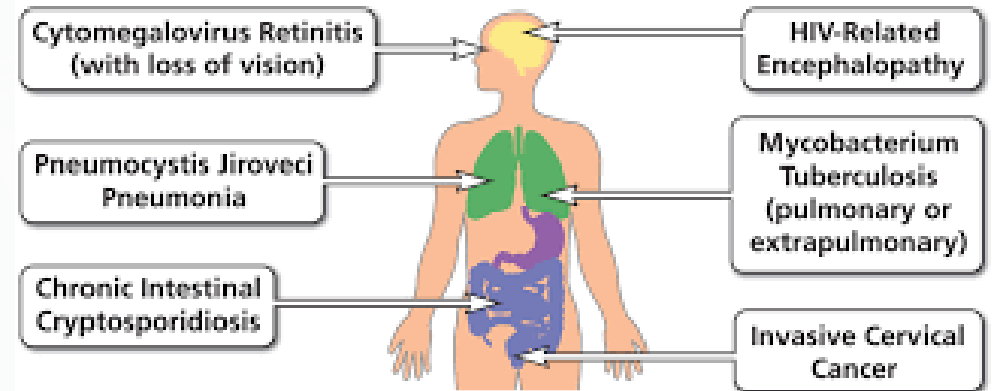
Prevalence of AIDS Defining illness in HIV Patients with Substance Abuse

432 Project 2

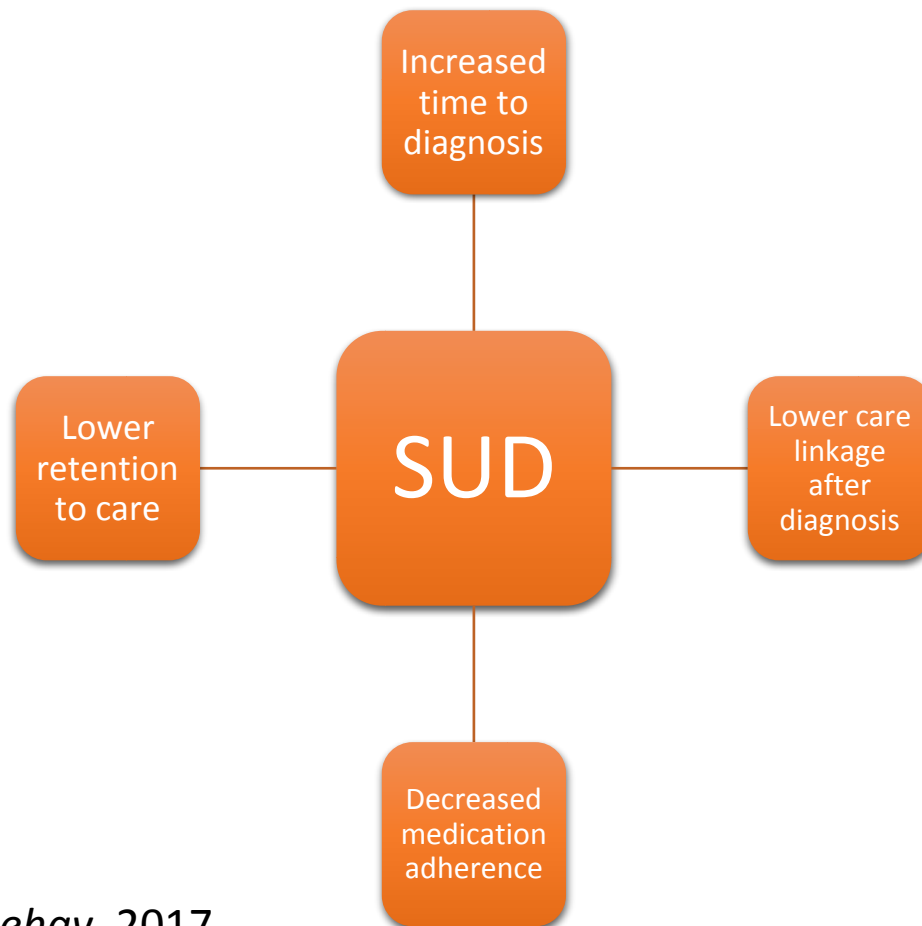
Lindsay Petrenchik

May 7, 2021

Examples of AIDS-Defining Conditions

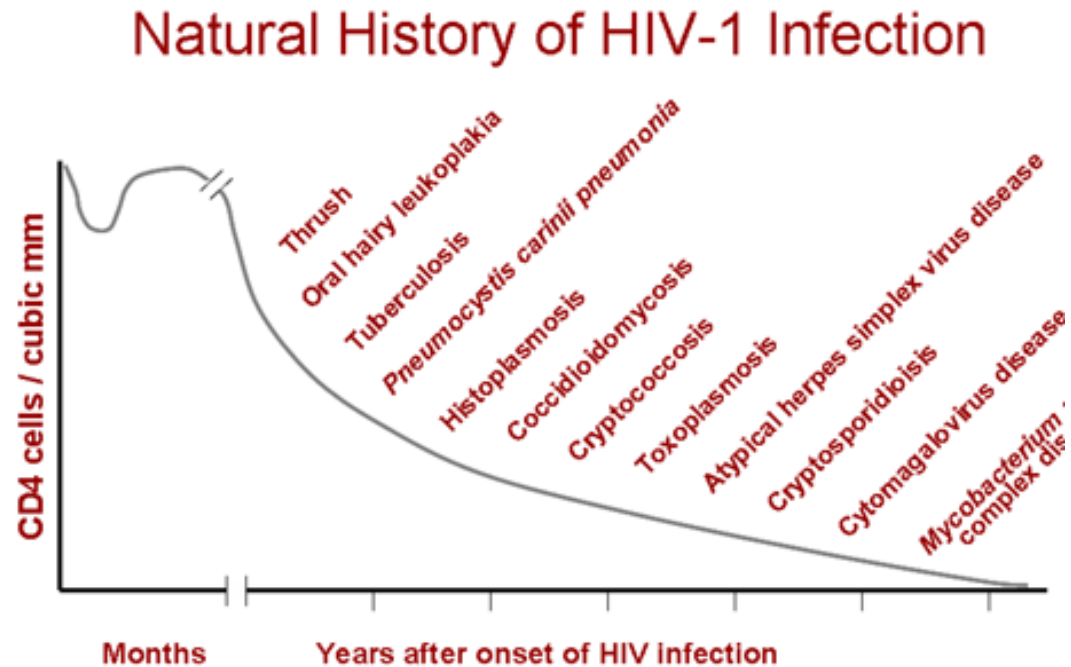


Substance Use Disorder (SUD) in HIV



- 48% (21-71%) of people living with HIV (PLWH) are estimated to have a SUD
- SUD adversely affects the continuum of care leading to increased risk of disease progression

HIV Disease progression



- As HIV progresses, the immune system is weakened, and patients are at an increased risk of opportunistic infections (OIs) /AIDS defining illness

Previous Literature

- Previous cohort studies indicate that there is an association between OIs and SUD in PLWH

Study (year)	Design, duration	Population	SUD definition	Findings
Cook et al. (2008)	Prospective cohort (1996-2004)	Women enrolled in 6 research centers (N=1686)	Crack cocaine	HR=1.65
Lucas et al. (2006)	Prospective cohort (1998-2003)	Baltimore, Maryland (N=1851)	Heroin or cocaine	OR = 1.6 (1.2–2.3)
Anastos et al. (2005)	Prospective cohort (1995-2003)	Women after initiating HAART (N=961)	Crack, cocaine, heroin	HR=1.49

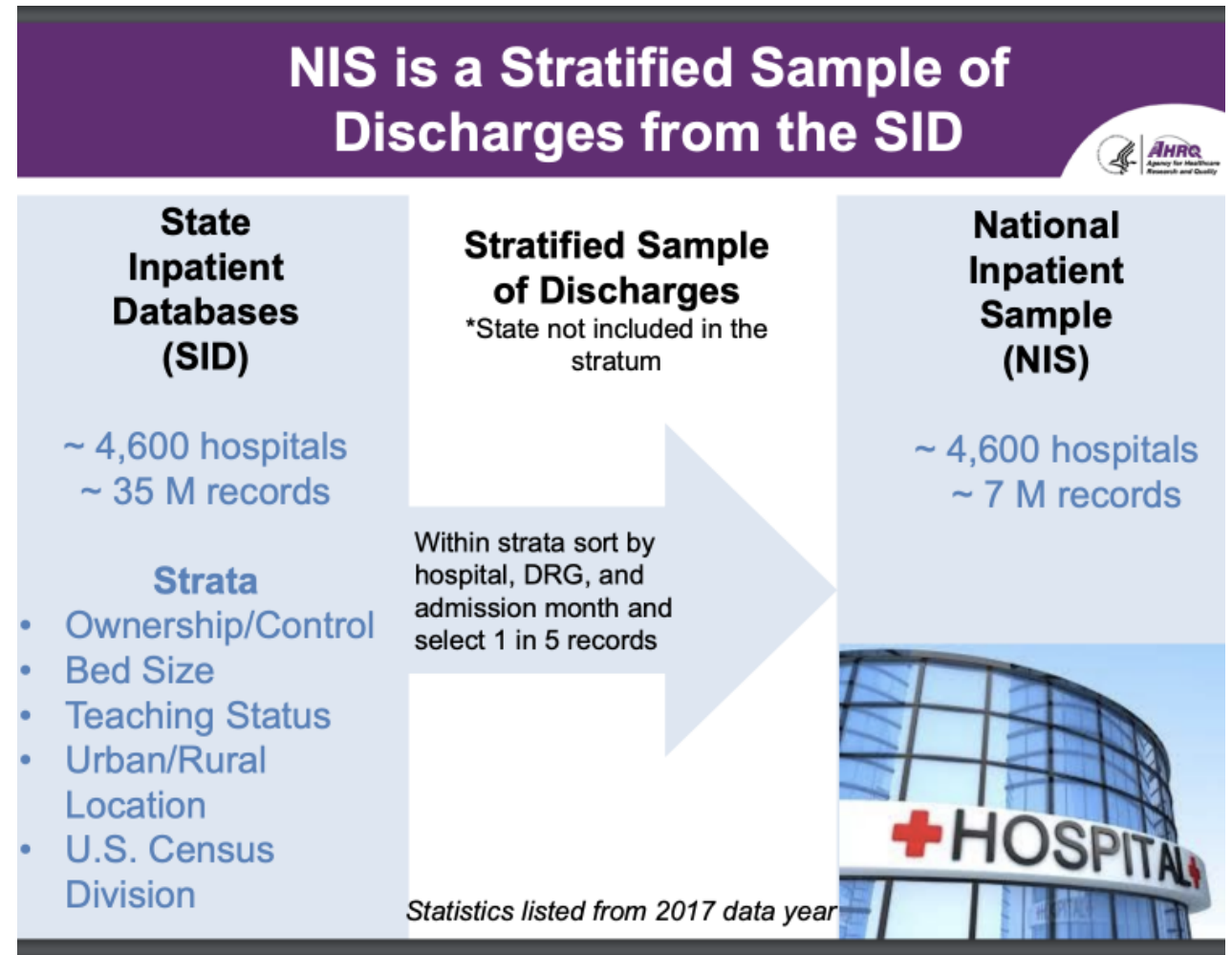
Research Questions

In PLWH with SUD vs no SUD
what is the difference in..

- AIDS defining illnesses
- Length of hospital stay

Data Source

- Healthcare Cost and Utilization Project, National (Nationwide) Inpatient Sample (HCUP-NIS)
- a 20-percent stratified sample of all discharges from 48 United States hospitals



Exposure & Outcomes

Exposure

- **Substance Use Disorder:** abuse of alcohol (F10), opioids, sedatives, hypnotics, anxiolytics (F13), cocaine (F14), other stimulants (F15), hallucinogens (F16), inhalants (F18), or other psychoactive substances/multiple drug use (F19)

Outcomes

- **CDC defined AIDS defining illness (Opportunistic infections):** candidiasis of the esophagus (B37.81), bronchi, trachea, or lungs (B371); invasive cervical cancer (C53); coccidiomycosis (B38); cryptococcosis (B45); cryptosporidiosis (A07.2); cytomegalovirus disease or CMV (B25); histoplasmosis (B39); isosporiasis (A07.3); Kaposi sarcoma (C46); Burkitt's, immunoblastic, Hodgkin's, and Non-Hodgkin's lymphoma (Burkitt's, immunoblastic); mycobacterium avium complex (A31.2, A31.8); mycobacterium tuberculosis (A15); pneumocystis pneumonia (B59); recurrent pneumonia (Z87.01); progressive multifocal leukoencephalopathy (A81.2), salmonella septicemia (A02.1) and toxoplasmosis of brain (B58.2).
- **LOS:** discharge date-admission date (days)

7 Covariates

Quantitative

- Age

Binary

- Sex

Multi categorical

- Hospital division (5-cat)
- Insurance (5-cat)
- Race (5-cat)
- Population density (6-cat)
- Median Income based on zip (4-cat)

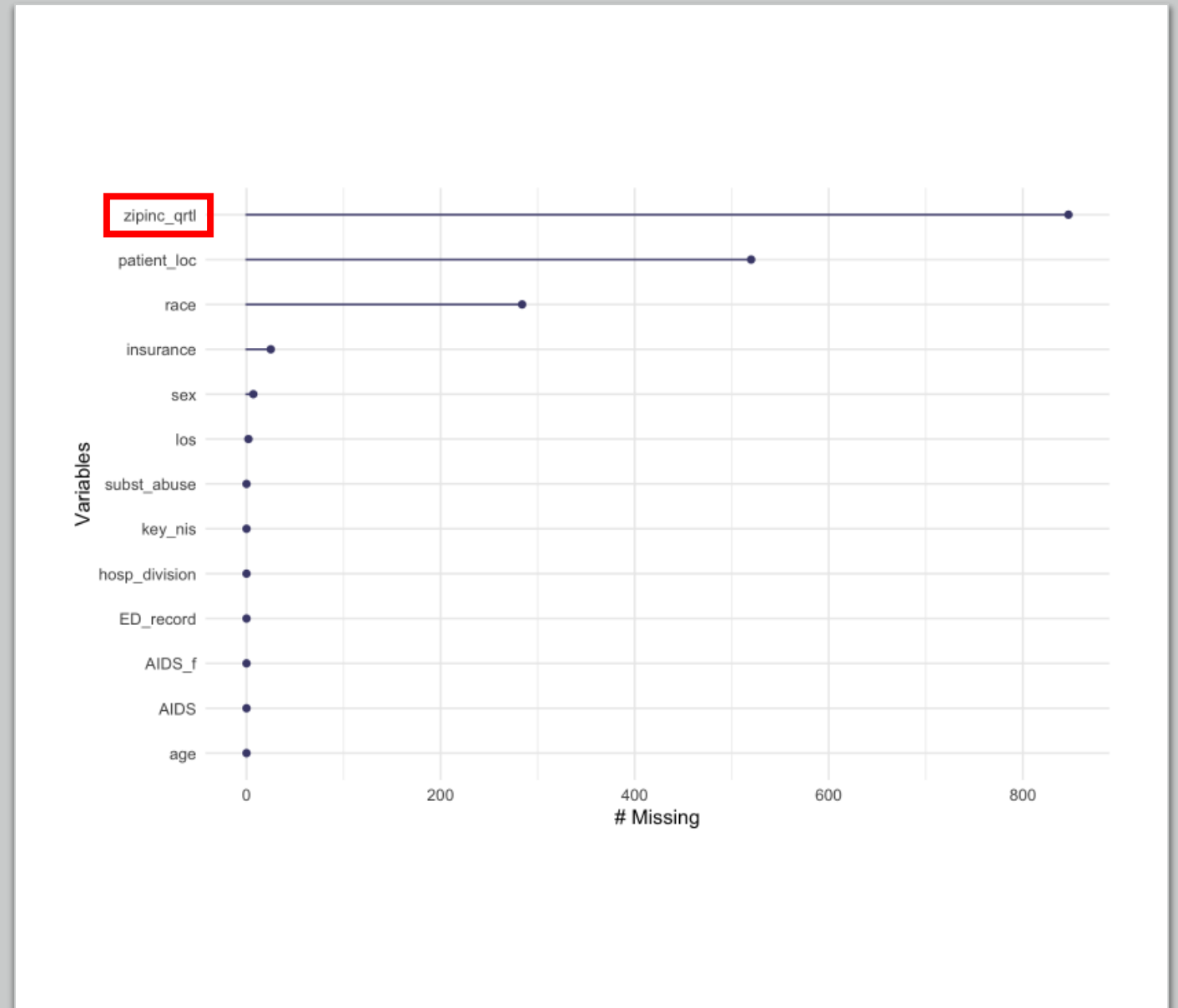
Population (N=24,118)

- Eligibility: age ≥ 20
- Equally balanced
 - SUD: 11,791
 - No SUD: 12,327
- Mostly black middle-aged men from a metro area (>1 million people) in the South Atlantic region with a median income based on zip code of <48 k

	Stratified by subst_abuse		p	test
	yes	no		
n	11791	12327		
hosp_division (%)			<0.001	
South_Atlantic	3392 (28.8)	3945 (32.0)		
Northeast	2944 (25.0)	2696 (21.9)		
South	1956 (16.6)	2326 (18.9)		
West	2001 (17.0)	2005 (16.3)		
Midwest	1498 (12.7)	1355 (11.0)		
age (mean (SD))	48.22 (11.58)	50.69 (13.62)	<0.001	
sex = female (%)	3413 (29.0)	3852 (31.3)	<0.001	
insurance (%)			<0.001	
Medicaid	5773 (49.0)	3736 (30.3)		
Medicare	3595 (30.5)	4895 (39.7)		
Private	1185 (10.1)	2469 (20.0)		
Self_pay	892 (7.6)	861 (7.0)		
Other	330 (2.8)	357 (2.9)		
race (%)			<0.001	
Black	5990 (51.4)	6212 (51.0)		
White	3534 (30.3)	3278 (26.9)		
Hispanic	1590 (13.6)	2042 (16.8)		
Other	402 (3.4)	445 (3.7)		
Asian	65 (0.6)	159 (1.3)		
NativeA	73 (0.6)	44 (0.4)		
patient_loc (%)			<0.001	
Central	6453 (56.7)	6466 (52.9)		
Fringe	1642 (14.4)	2401 (19.7)		
metro>250K	1990 (17.5)	1924 (15.8)		
metro>50K	604 (5.3)	672 (5.5)		
micro	440 (3.9)	472 (3.9)		
Other	258 (2.3)	277 (2.3)		
zipinc_qrtl (%)			<0.001	
<48K	6028 (53.7)	5369 (44.6)		
48-61K	2499 (22.3)	2909 (24.2)		
61-82K	1669 (14.9)	2171 (18.0)		
82K+	1032 (9.2)	1595 (13.2)		

Missing Data

- Max missing 3.5% (zipinc_qrtl)
- Assumed MNAR
- Multiple Imputation with 'mice' (m=4)



A photograph of various medical supplies on a wooden surface. In the foreground, there are several green and white capsules and white pills. A silver spoon holds a pile of white powder. To the right, a yellow syringe with markings is visible. In the background, there are more white pills in a small container and a white packet. The overall scene suggests a medical or pharmaceutical setting.

Analysis 1

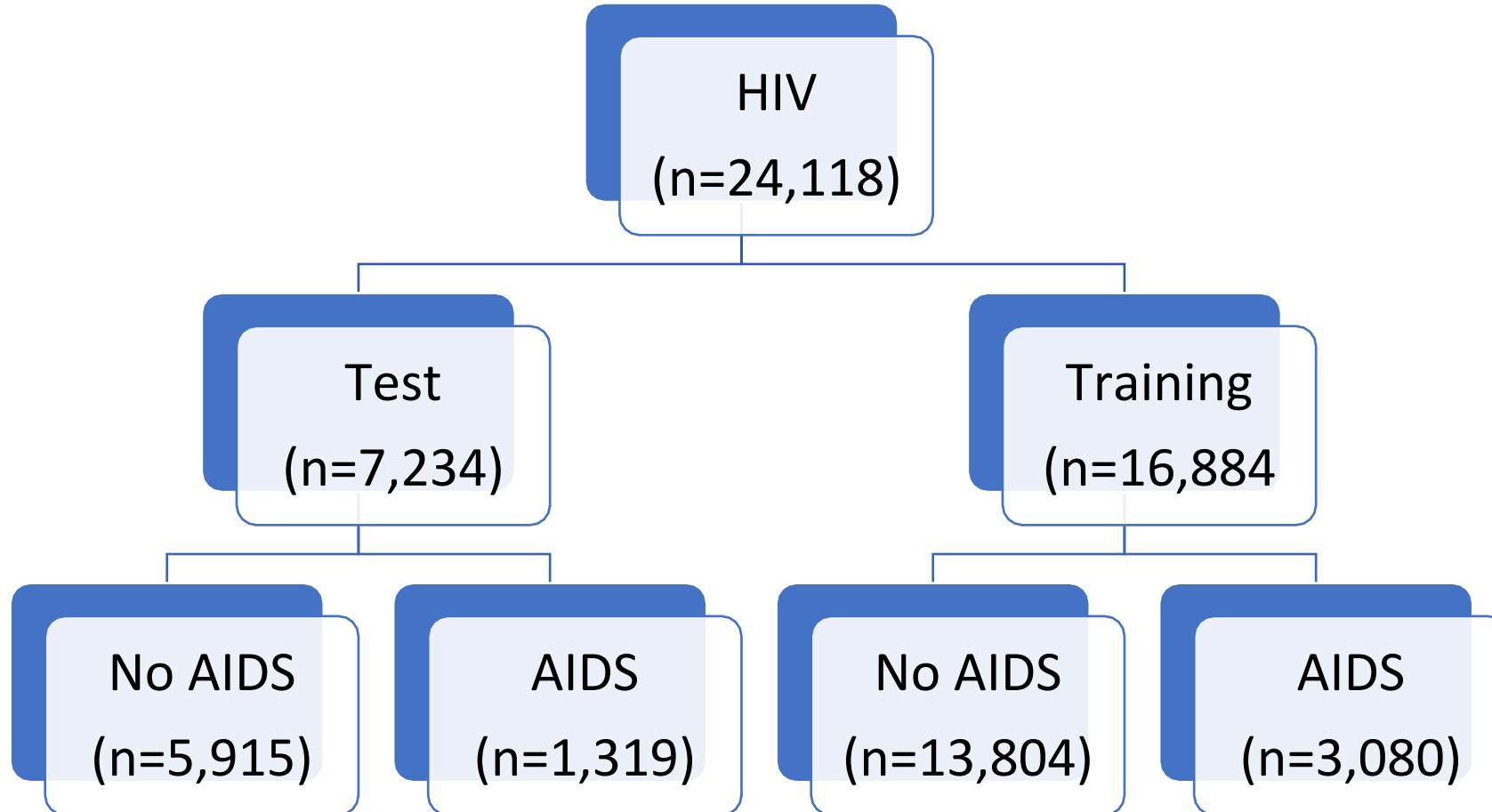
AIDS illness in PLWH with SUD vs without

Exposure and Outcome

- Lower burden of AIDS illnesses in non-SUD group
 - $OR_{unadjusted}=0.785$

	AIDS illness	No AIDS illness	Total
SUD	1929 (16.4%)	9862 (83.6%)	11,791
No SUD	2470 (20.0%)	9857 (80.0%)	12,327
Total	4399 (18.2%)	19,719 (81.8%)	24,118 (100.0%)

Split 70/30 with **18.2%** in each sample with the outcome



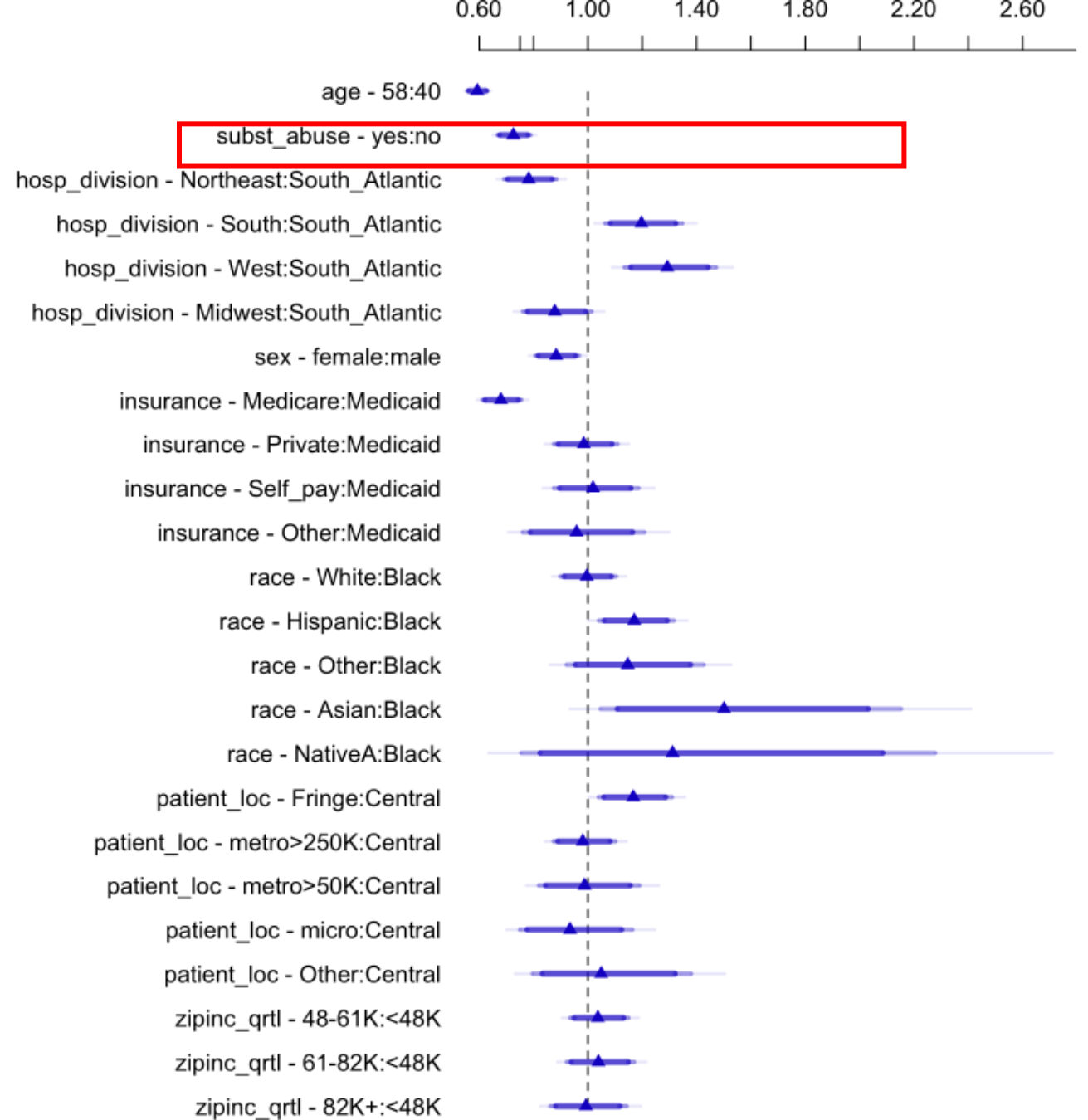
Logistic Regression Adjustment (Model 1)

$$\log \frac{\Pr(AIDS)}{1 - \Pr(AIDS)} = \beta_0 + \beta_{\text{subst_abuse}}\mathbf{x} + \beta_{\text{hosp_division}}\mathbf{x} + \beta_{\text{age}}\mathbf{x} \\ + \beta_{\text{sex}}\mathbf{x} + \beta_{\text{insurance}}\mathbf{x} + \beta_{\text{race}}\mathbf{x} + \beta_{\text{patient_loc}}\mathbf{x} + \beta_{\text{zipinc_quartile}}\mathbf{x}$$

8 predictors were chosen based on previous literature and reasoning

Effect Size

- Holding all other predictors constant, model 1 predicts that in PLWH, the **odds of AIDS** in those with **SUD** is **0.726** (95% CI 0.668, 0.789) the odds of those without SUD (**N=16,884**)
- Other predictors with meaningful ORs
 - Hospital division*
 - Medicare vs Medicaid*
 - Asian vs black*
 - Native American vs black
- Lack of effect with income



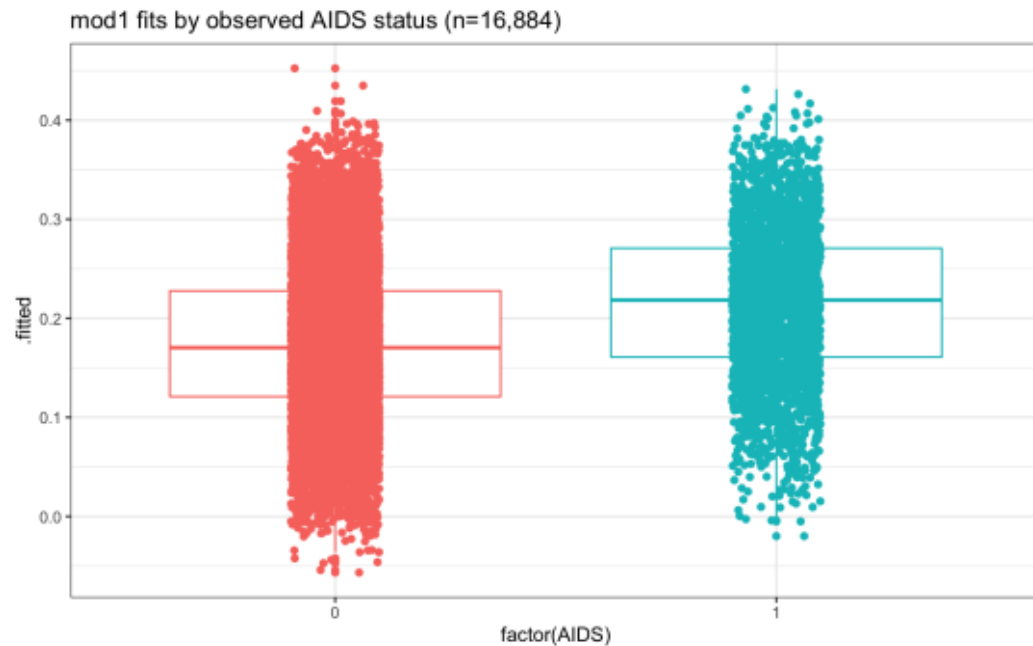
But how good was the model fit...

Not good.

Fit Statistic	Index	Corrected
R2	0.066	0.061
C	0.650	0.645

Confusion Matrix Results

Not good.



Confusion Matrix Statistic	Result
Sensitivity	25%
Specificity	88%
Positive Predictive Value	32%

*Based on the plot, chose a cut off of 0.27

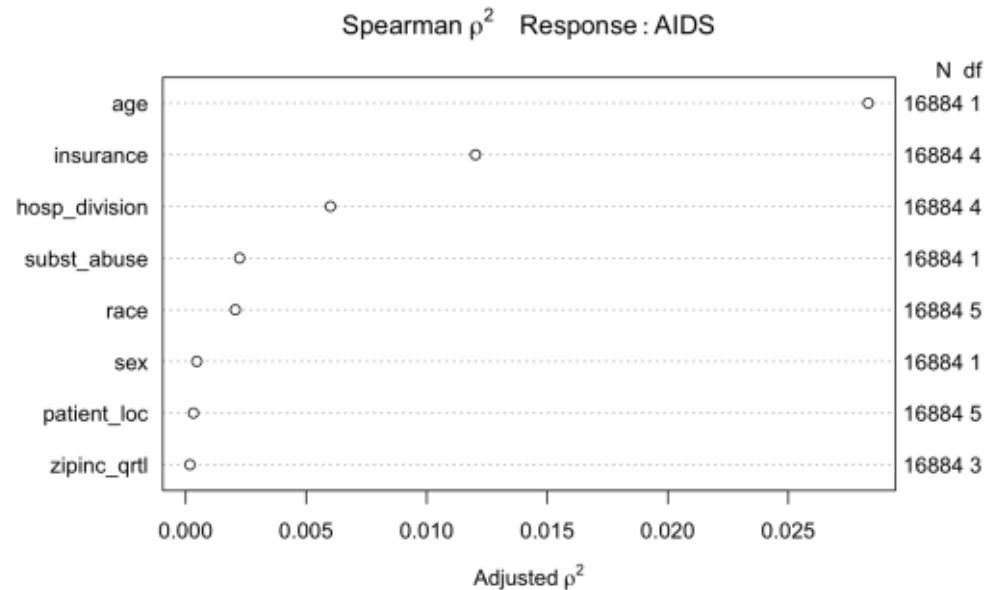


Plan B!

Can I make it better with nonlinear terms?

Logistic Regression Models Built

ANOVA: compared to model 1, all reduced the lack of fit



Model	Description	DF
1	All linear terms	24
1c	Model 1 but rcs(age,4)	28
1d	Model 1 + age%ia%insurance	28
1b	1c + 1d	30

Comparison of Logistic Regression Models

Model	AIC	BIC	Index R2	Index C	Corrected R2	Corrected C	Validated Accuracy	Validated Kappa
1	15135	15336	0.066	0.6496	0.0614	0.64495	0.769	0.152
1b	15118	15366	0.069	0.6521	0.0630	0.64625	0.762	0.164
1c	15113	15329	0.069	0.6519	0.0634	0.64635	0.762	0.163
1d	15137	15369	0.066	0.6499	0.0623	0.6458	0.767	0.154

- **In sample statistics:** all models were almost equally poorly fit ($r^2=0.07$, $C=0.65$)
- **Validated statistics:** model 1c negligibly better
- **Accuracy in training sample:** model 1
- **Kappa in training sample:** model 1b

A person in a hospital bed, seen from behind, looking out a window. The room is dimly lit, with a sink and counter visible on the left and medical equipment on the right. The person is wearing a light blue hospital gown. The window has white curtains. The overall mood is somber and contemplative.

Analysis 2

Length of hospital stay in those with SUD vs without

Variables

Exposure

- SUD

Outcome

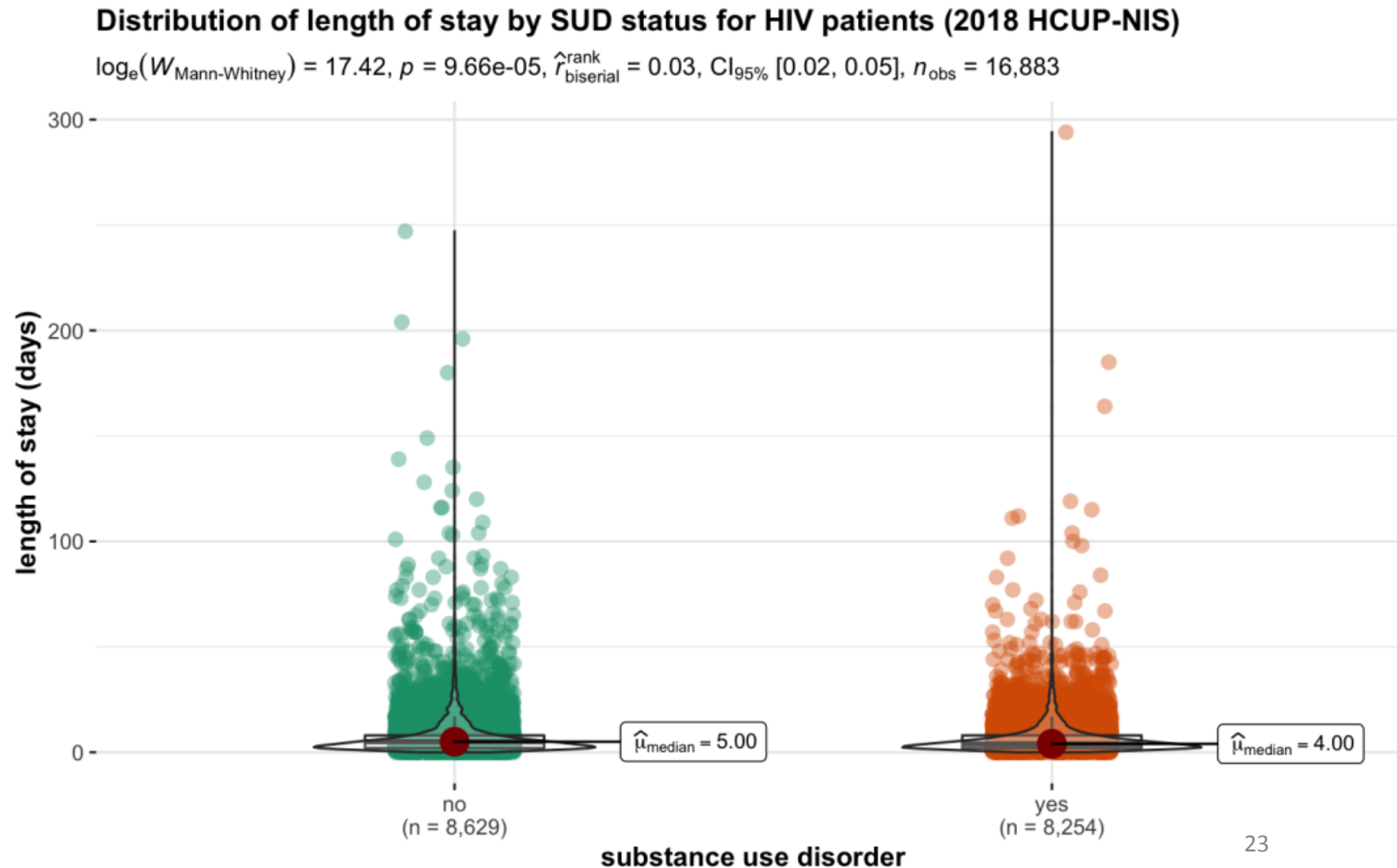
- LOS

Covariates

- Same as Analysis 1, but switched sex for AIDS

Distribution of length of stay by SUD

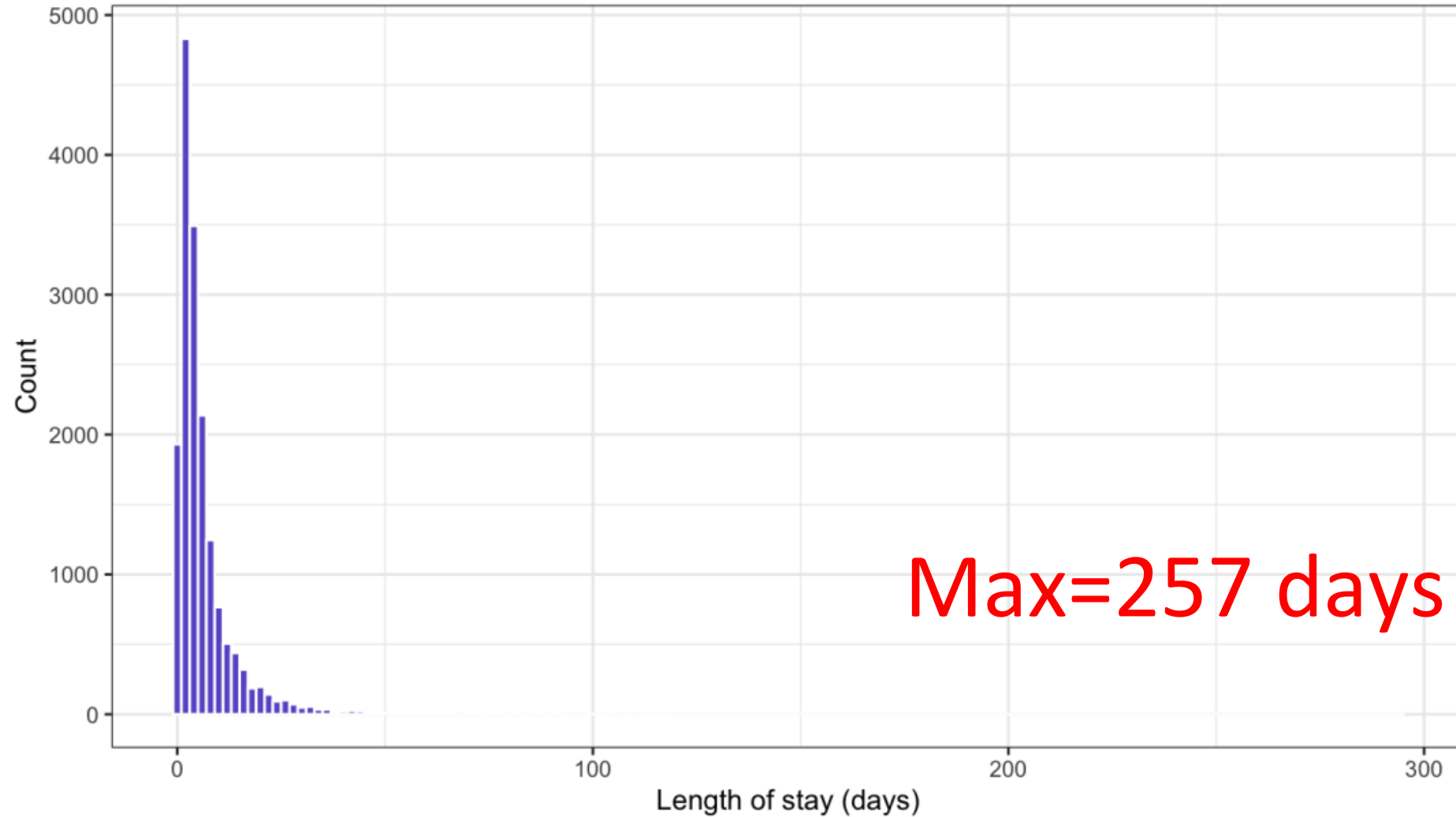
- Justifies Poisson/NB
 - All positive
 - Not normal
- Most los <50 days
 - Range: 0 to 247
- Higher in non SUD group



Lots of 1's

Distribution of length of inpatient stay of HIV patients

24081 observations from NIS 2018 data



Analysis

- Single imputation
- Data Split
 - Strata: SUD, AIDS
- Compared 6 models

Poisson (n=16,883)

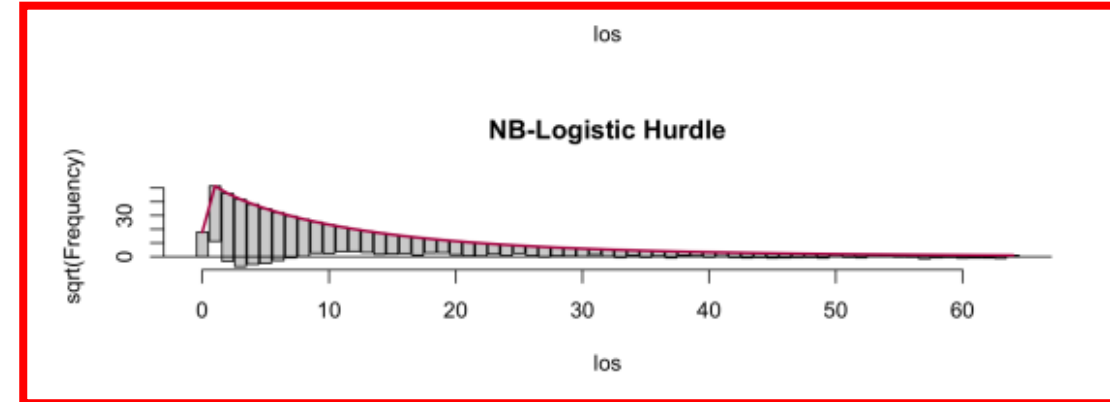
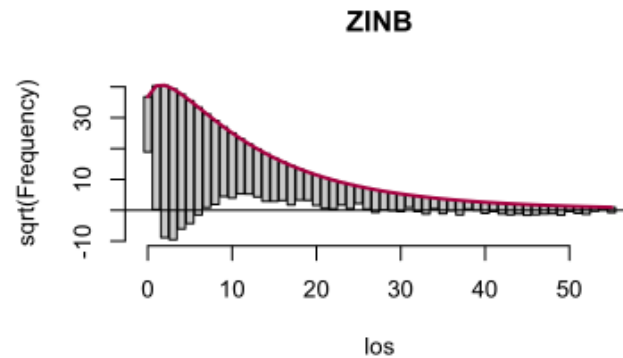
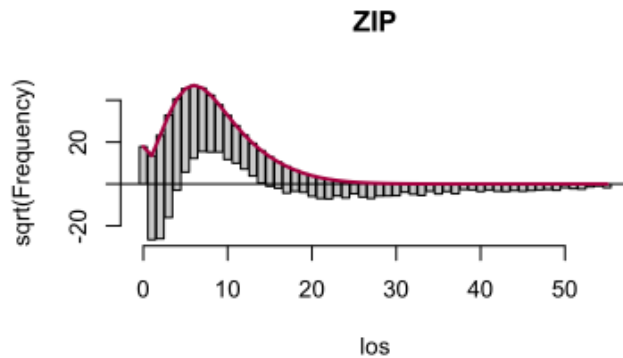
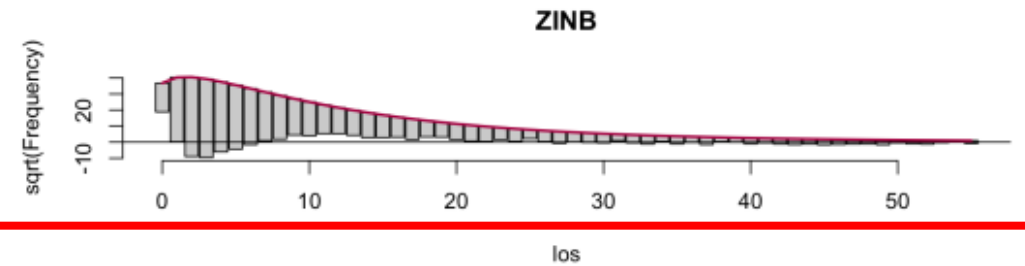
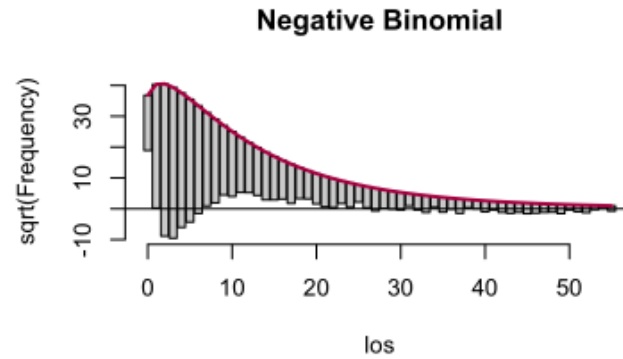
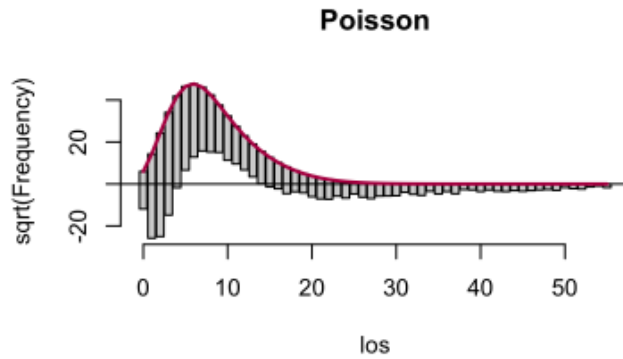
- ZIP
- Hurdle

Negative Binomial (n=16,883)

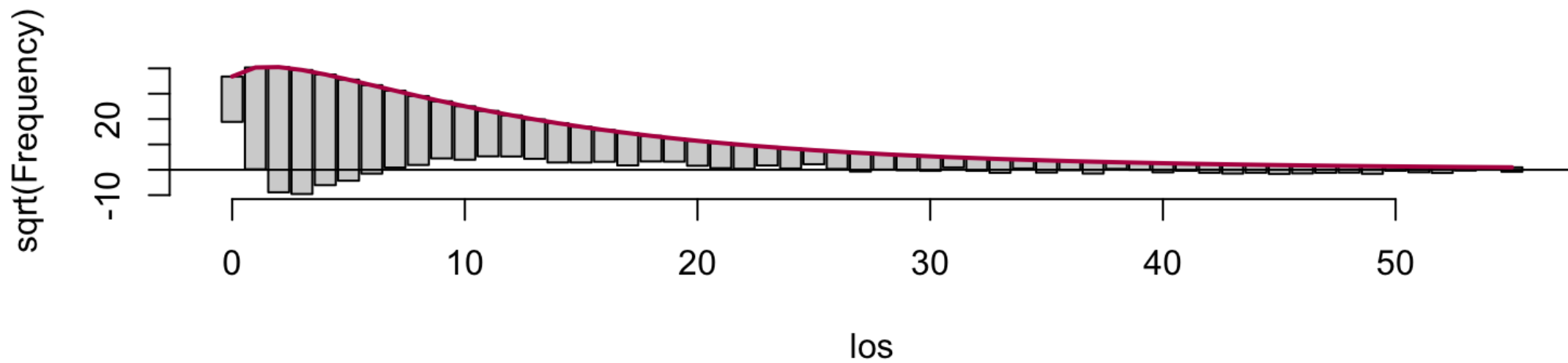
- ZINB
- Hurdle

Best Rootogram for NB-based models

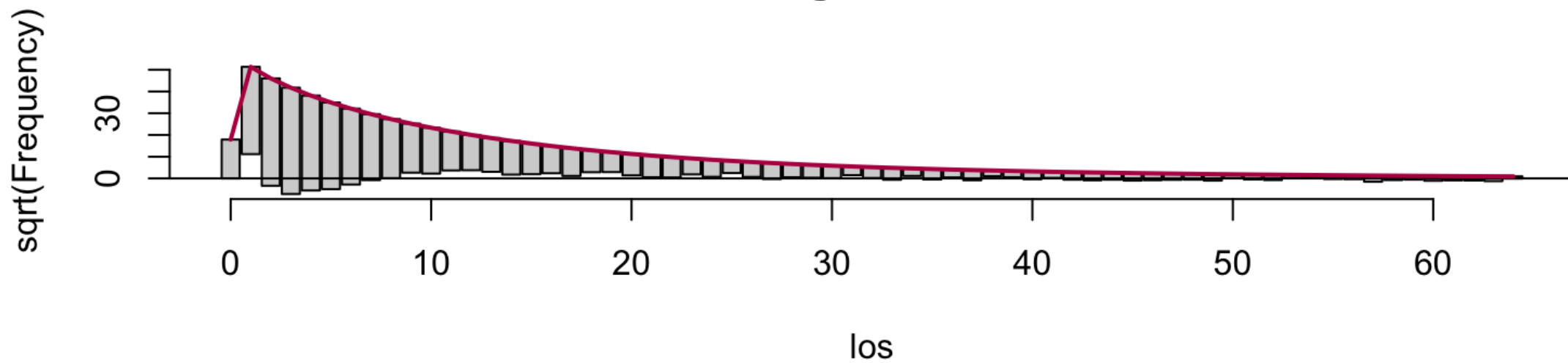
- NB hurdle model was the best



ZINB



NB-Logistic Hurdle



Validation in Test Sample (N=7,198)

.metric	POIS	NB	ZIP	ZINB	hurdlePOIS	hurdleNB
rsq	0.040	0.040	0.041	0.042	0.041	0.042
rmse	10.446	10.446	8.966	8.965	8.967	8.965
mae	5.348	5.347	5.014	5.008	5.014	5.008

- Validated fit statistics best for ZINB and hurdleNB
 - Vuong test ZINB vs NB-hurdle: $P < 0.000$
- Not a good model: low r^2 and high mae

Coefficients

Holding all else constant, hurdleNB predicts that in PLWH (N=16,883), *log(IoS)* for:

- **SUD vs no SUD:** 0.11 lower (0.9 days, 95% CI 0.86,0.93)
- **AIDS vs no AIDS:** 0.59 higher (1.8, 95% CI 1.83,1.9)

Count model coefficients (truncated negbin with log link):

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	1.7133477	0.0479004	35.769	< 2e-16	***
subst_abuseyes	-0.1106580	0.0183972	-6.015	1.80e-09	***
AIDS_fyes	0.5936846	0.0235180	25.244	< 2e-16	***
age	0.0044841	0.0007756	5.781	7.41e-09	***
raceWhite	-0.0253442	0.0220174	-1.151	0.249691	
raceHispanic	-0.0793630	0.0271422	-2.924	0.003456	**
raceOther	0.1498052	0.0497040	3.014	0.002579	**
raceAsian	0.0390066	0.0920889	0.424	0.671876	
raceNativeA	0.2792220	0.1248266	2.237	0.025294	*
hosp_divisionNortheast	0.0586923	0.0254488	2.306	0.021095	*
hosp_divisionSouth	-0.0152584	0.0270947	-0.563	0.573332	
hosp_divisionWest	-0.1029889	0.0290843	-3.541	0.000399	***
hosp_divisionMidwest	-0.1849104	0.0307075	-6.022	1.73e-09	***
zipinc_qrtl48-61K	0.0055749	0.0227359	0.245	0.806298	
zipinc_qrtl61-82K	0.0356644	0.0260498	1.369	0.170974	
zipinc_qrtl82K+	0.0257215	0.0303426	0.848	0.396603	
insuranceMedicare	-0.1367474	0.0222882	-6.135	8.49e-10	***
insurancePrivate	-0.1589493	0.0275364	-5.772	7.82e-09	***
insuranceSelf_pay	-0.1467868	0.0373464	-3.930	8.48e-05	***
insuranceOther	0.0560034	0.0553886	1.011	0.311969	
ED_recordyes	-0.1613901	0.0222716	-7.246	4.28e-13	***

Conclusion

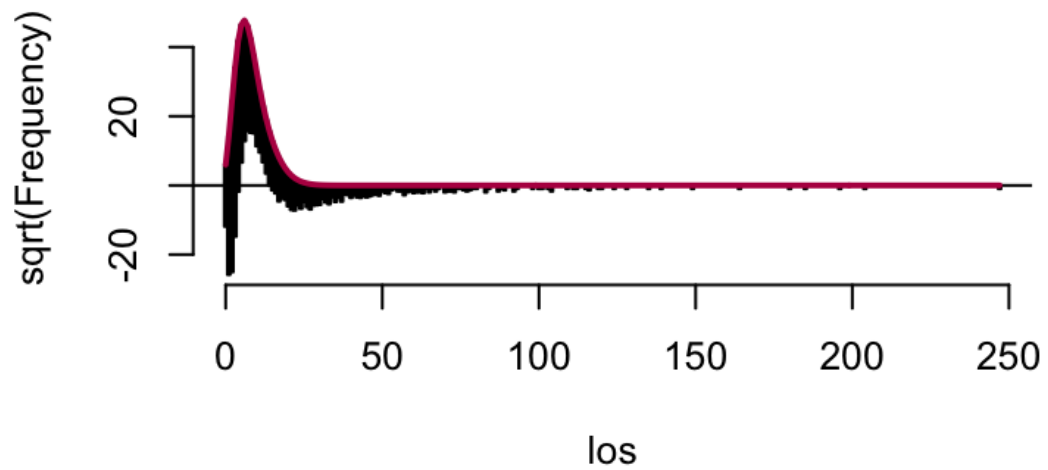
- In U.S. hospitalized PLWH in 2018 (N=24,118), **SUD** is associated **with lower burden of AIDS defining illness & lower LOS**
- **Relationships inconsistent with:**
 - AIDS: previous literature (3 cohort studies)
 - Los: general population
- **Limitations**
 - Data collected for reimbursement (i.e all diagnoses may not be captured)
 - Important covariates unavailable (OI prophylaxis, labs, meds)
 - Broad exposure definition → next step narrow to stimulants
 - Did not adjust for hospital characteristics (los analysis) or relevant comorbidities → PS match on those next

Appendix

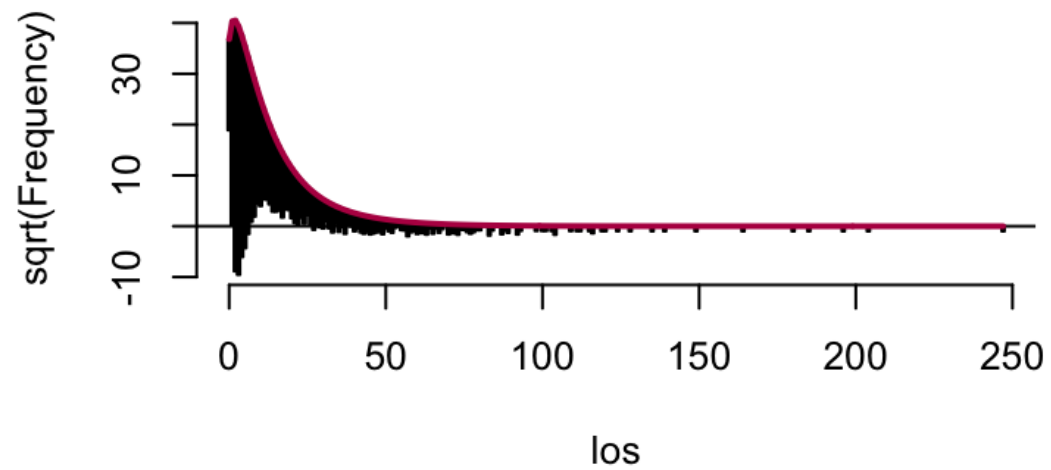
- Code book

Variable	Type	Description
key_nis	character	key_nis identifier
subst_abuse	binary	main predictor whether or not somebody has of substance use disorder. Patients were classified as having a history of SUD if they had an ICD-10 code for abuse of alcohol (F10), opioids, sedatives, hypnotics, anxiolytics (F13), cocaine (F14), other stimulants (F15), hallucinogens (F16), inhalants (F18), or other psychoactive substances/multiple drug use (F19) (yes/no)
age	quant	age in years.
sex	binary	male, female.
race	5-cat	Black, White, Hispanic, Other, Asian, Native American
hosp_division	5-cat	Northeast, Midwest, South, South_Atlantic, West
zipinc_qrtl	4-cat	Median household income for patient's ZIP Code (based on current year). Values include <48K, 48-61K, 61-82K, 82K+.
insurance	5-cat	expected primary payer (Medicare, Medicaid, private insurance, self pay, other)
patient_loc	6-cat	Patient Location ("Central" counties of metro areas of >=1 million population, "Fringe" counties of metro areas of >=1 million population, Counties in metro areas of 250,000-999,999 population, Counties in metro areas of 50,000-249,999 population, Micropolitan counties, Not metropolitan or micropolitan counties)
ED_record	binary	records that have evidence of emergency department (ED) services reported on the HCUP record (yes/no)

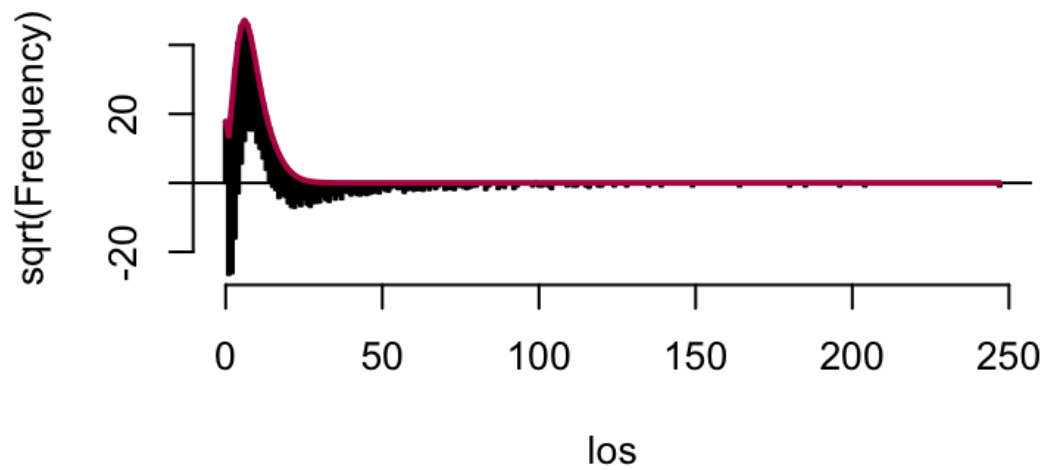
Poisson



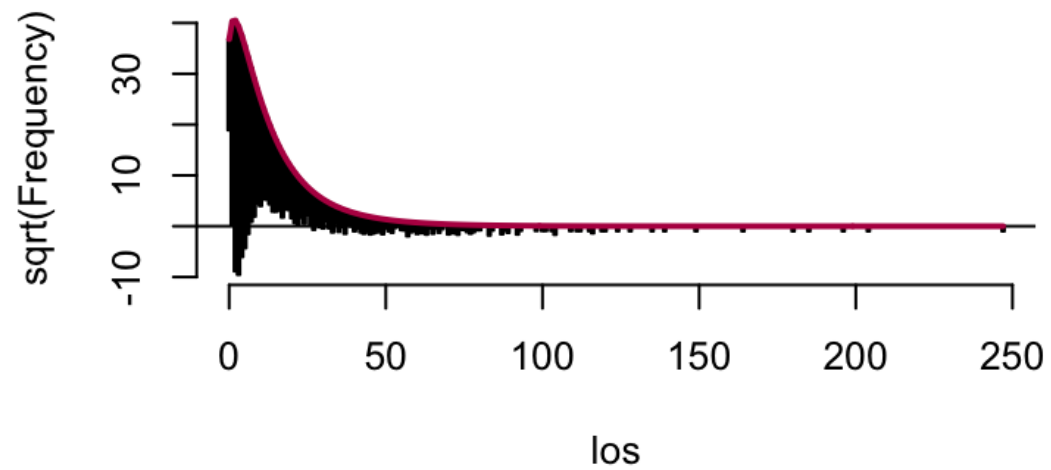
Negative Binomial



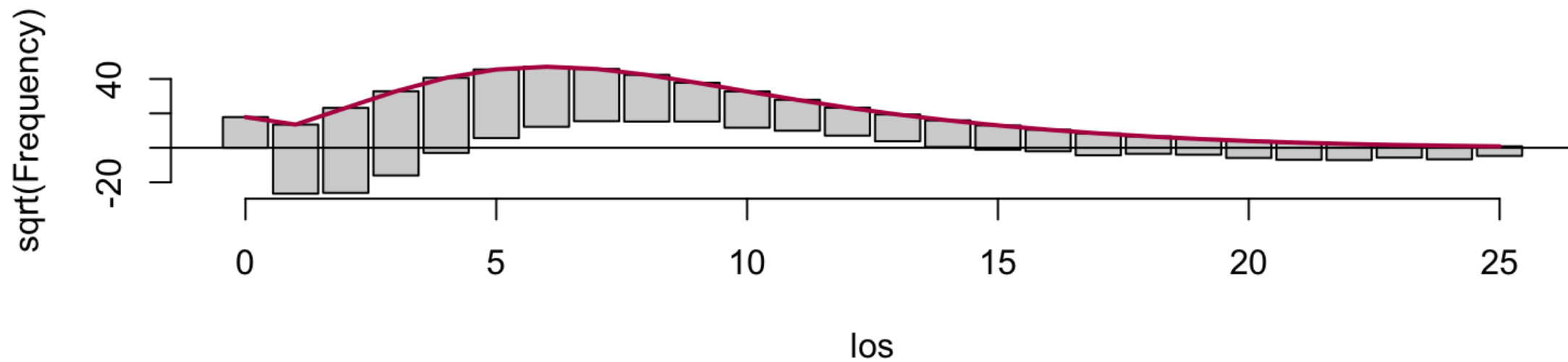
ZIP



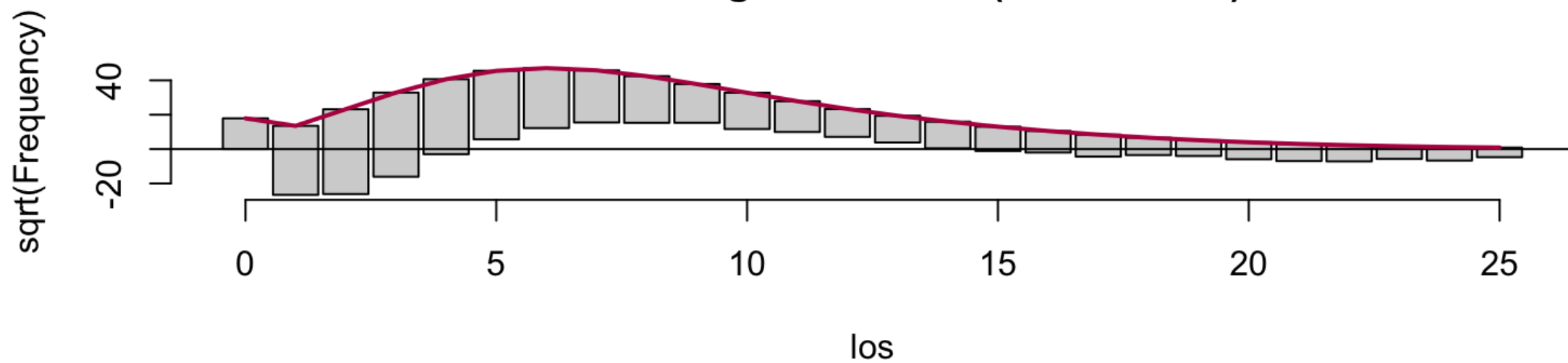
ZINB



ZIP



Poisson-Logistic Hurdle (hurdlePOIS)



Coefficient	Los estimate (days)	P value
subst_abuseyes	-0.9	1.80e-09
AIDS_fyes	1.81	< 2e-16
age	1	7.41e-09
raceWhite	-0.97	0.249691
raceHispanic	-0.92	0.003456
raceOther	1.16	0.002579
raceAsian	1.04	0.671876
raceNativeA	1.32	0.025294
hosp_divisionNortheast	1.06	0.021095
hosp_divisionSouth	-0.98	0.573332
hosp_divisionWest	-0.9	0.000399
hosp_divisionMidwest	-0.83	1.73e-09
zipinc_qrtl48-61K	1.01	0.806298
zipinc_qrtl61-82K	1.04	0.170974
zipinc_qrtl82K+	1.03	0.396603
insuranceMedicare	-0.87	8.49e-10
insurancePrivate	-0.85	7.82e-09
insuranceSelf_pay	-0.86	8.48e-05
insuranceOther	1.06	0.311969

Count Model

$$\text{Log}(Ios) = \beta_0 + \beta_{\text{subst_abuse}}\mathbf{x} + \beta_{\text{AIDS}}\mathbf{x} + \beta_{\text{hosp_division}}\mathbf{x} + \beta_{\text{age}}\mathbf{x} \\ + \beta_{\text{ED_record}}\mathbf{x} + \beta_{\text{insurance}}\mathbf{x} + \beta_{\text{race}}\mathbf{x} + \beta_{\text{zipinc_quartile}}\mathbf{x}$$

8 predictors were chosen based on previous literature and reasoning

Analysis

1. Single imputation
2. Split into training & testing
 1. Strata= SUD, AIDS
3. Fit & compared 6 different models (poisson, negative binomial, ZIP, ZINB, hurdle based)
 1. Rootograms
 2. Fit statistics (in and out of sample)
4. Used the chosen model to produce estimates for each predictor

Analysis Flow

