Factors Predicting Healthcare Access for Clients of Homeless Assistance Programs

Mao Hu

Background: Homeless persons have increased rates of illness and disability compared to housed persons, as well as reporting many barriers to healthcare access. However, healthcare access is not the same amongst all homeless persons. This study identifies factors that predict access to healthcare.

Sample: 3529 homeless and non-homeless clients of homeless assistance programs distributed across the United States, interviewed in the National Survey of Homeless Assistance Providers and Clients.

Outcome Measure: In this study, client's access to healthcare is measured by whether or not the client has been examined or treated by a medical professional in the last year (having a healthcare event). A related analysis is conducted about whether the client's last healthcare event was in an emergency room, inpatient care, or ambulatory care.

Results: The most statistically significant predictors of increased odds that a client has had a healthcare event in the last year are: having insurance; having more than 1 comorbid chronic illnesses; and having some combination of alcohol, drug, and mental health problems. Having insurance is also associated with reduced rates of emergency room care and increased rates of inpatient care.

Introduction

Homeless persons in the United States suffer from high rates of illness and disability and experience many barriers to healthcare access. A study on homeless men and women in Baltimore reports that on average, homeless men and women have 8 to 9 physical health problems. Homeless persons also suffer from high rates of comorbidity between mental illness and substance abuse (Breakey, 1989). Compounding these health issues are the difficulties homeless persons face when they need health care. One article describes life on the streets as a struggle where the homeless respond to immediate needs such as food, shelter, and safety before chronic illnesses. Not only is regular healthcare not a priority, but many homeless persons have no insurance and are socially isolated. (Levy, 2004)

These barriers to healthcare access for the homeless not only impact the lives of many people, they incur substantial costs for "safety net" hospitals and local governments. Homeless people unable to access or pay for healthcare may postpone treatment until their condition is so severe that expensive emergency room treatment is needed. In San Diego, 15 chronically homeless persons used $1.5 million dollars of medical services alone in 1.5 years. (Project 25, n.d.) Moreover, one study on public general hospitals in New York City discovered that hospital stays for homeless persons are both 4.5 days longer on average and more expensive than hospital stays of non-homeless patients. (Salit, Kuhn, Hartz, Vu, & Mosso, 1998).

A national study on healthcare access found that homeless persons with health insurance are more likely to use ambulatory and in-patient hospitalization care. The researchers suggest that improving insurance coverage may reduce excess morbidity in the homeless population as well as reduce costs associated with emergency treatment versus inpatient or ambulatory care. (Kushel, 2001) Regular access to healthcare may also help reach these objectives.

In this study, access to healthcare is measured by whether or not a person has been treated or examined by a medical professional in the past year (has had a healthcare event). Although the Kushel study examines only the strictly homeless, this study expands the sample population to all users of homeless assistance programs. The research objective is to identify the factors that are i) predictive for having had a self-reported healthcare event in the past year and ii) whether or not the client's last healthcare event was in an emergency room, inpatient, or ambulatory setting.

Data

The factors predictive for healthcare access were analyzed in a sample of 3529 clients of homeless assistance providers distributed throughout the United States. These clients included persons who are homeless, formerly homeless, or never homeless using services at organizations such as emergency shelters. Persons of all three types of housing status are likely to report chronic health conditions and have similar levels of reported acute infectious conditions. (Homelessness: Programs and the People They Serve, 1999).

The analyzed sample is a subset of a larger sample of clients from the National Survey of Homeless Assistance Providers and Clients, a 1996 study of "providers of homeless assistance and the characteristics of persons who use these services." This study collected information from 76 areas: the "28 largest metropolitan statistical areas," "24 randomly sampled small and medium metropolitan statistical areas," and "24 randomly selected groups of rural counties or parts of counties." Clients of homeless assistance programs were selected by first randomly selecting programs in the study areas, and then randomly selecting 4200 clients. Oral interviews were conducted with these clients regarding a large variety of characteristics, including demographics, income, health status, and others. All answers except those referring to the location of the interview and the observed sex of the client were self-reported. Moreover, no attempt to verify answers to questions about illnesses using medical records was made. (NSHAPC Design and Data Collection, 1999)

Table 1 summarizes 21 explanatory variables of interest, as well as the outcome variable *visityear,* indicating whether or not a client has had a healthcare event in the past year. Included are demographic characteristics, factors associated with housing, and factors associated with health and insurance status. The explanatory variables were selected as possibly predictive for *visityear*, based on previous studies. The variables *age, race infect, chronic,* and *insur* are presented with several categories collapsed into fewer categories, since initial regressions suggest the coefficients of those categories are similar.

After removing observations with missing values for the explanatory variables, the values for most of the explanatory variables in the reduced population of 3529 clients were similar in proportion to the original population. The notable exception is *urbrural* – in the sample without missing values there are no clients interviewed in rural locations, compared to 11.9% in the original sample.

Table 1: Descriptions and Descriptive Statistics for Explanatory and Outcome Variables

Outcome Variable		% of Sample €
Have received treatment from or examined by a medical professional in the past year		
visityear	No	24.74
	Yes	75.26
Explanatory Variables		**% of Sample €**
Age in years		
age	< 18	1.33
	18-44	32.64
	≥ 44	66.02
Sex as observed by interviewer		
sex	Male	66.14
	Female	33.86
Race		
race	Not Black	55.17
	Black	44.83
# of comorbid infectious illnesses ¥		
infect	0	72.80
	1	23.46
	≥ 2	3.74
# of comorbid chronic illnesses †		
chronic	0	47.97
	1	26.64
	2	13.11
	≥ 3	12.27
Living on the street in the past week		
street	No	79.63
	Yes	20.37
Housing status on day of interview		
homlss	Currently homeless	75.43
	Formerly homeless	15.90
	Never homeless	8.67
Is a veteran of the Armed Forces		
veteran	No	78.27
	Yes	21.73
Visited by an outreach worker in the past week		
outreach	No	93.26
	Yes	6.744

Explanatory Variable		% of Sample €
Alcohol, drug, and mental health problems in the past year		
adm	None of these problems	29.22
	Alcohol problem only	10.37
	Drug problem only	7.34
	Mental health problem only	15.39
	Alcohol and drug problems	10.43
	Alcohol and mental health problems	7.11
	Drug and mental health problems	7.06
Alcohol, drug, and mental health problems 13.09		
Visited a drop in center in the past week		
drop	No	79.60
	Yes	20.40
Has medical insurance £		
insur	No	50.88
	Yes	49.22
Level of education		
ed	Less than high school	38.06
	High school diploma or GED	33.15
	More than high school	28.79
Location of interview		
urbrural	Central city	83.65
	Suburban/Urban fringe area	16.35
Number of children		
child	0	84.92
	≥ 1	15.08
Had at least 1 food problem in the last 30 days ‡		
food	No	44.69
	Yes	45.31
Marital Status		
marital	Now married	8.13
	Widowed	4.64
	Divorced	24.62
	Seperated	12.99
	Never married, don't know, or refused to answer	49.62

€ All variables are categorical, thus the percentage of the sample in each category is reported, without s.d.
¥ Infectious illnesses such as chest infection, cough, pneumonia, tuberculosis, syphilis, etc.
† Chronic illnesses such as diabetes, anemia, heart disease, cancer, HIV, etc.
£ Types of insurance include Medicaid, VA medical insurance, private insurance, and other types of insurance
‡ Having a food problem is defined as being hungry and unable to eat due to lack of money

Figure 1: Final Model (1)

$$\log\left(\frac{P(visityear=1)}{P(visityear=0)}\right) = \beta_0 + \beta_1\text{age2} + \beta_2\text{age3} + \beta_3\text{race2} + \beta_4\text{sex} + \beta_5\text{infect1} + \beta_6\text{infect2} + \beta_7\text{chronic1} + \beta_8\text{chronic2} + \beta_9\text{homlss2} + \beta_{10}\text{homlss3} + \beta_{11}\text{street1} + \beta_{12}\text{adm1} + \beta_{13}\text{adm2} + \beta_{14}\text{adm3} + \beta_{15}\text{adm4} + \beta_{16}\text{adm5} + \beta_{17}\text{adm6} + \beta_{18}\text{adm7} + \beta_{19}\text{drp1} + \beta_{20}\text{insur}$$

Table 2: Factors Associated with Health Care Usage in the Past Year from the Reduced Model (n=3529)

Variable (name, description)		Unadjusted Rate, % ¥	Odds Ratio (95% CI) €	z-statistic	Drop in Deviance
Age					15.25
age	< 18	91.89	1		
	18-44	77.17	0.3 (0.1-0.8)	-2.07	
	≥ 45	73.99	0.2 (0.1-0.6)	-2.56	
Race & ethnicity					5.64
race	Not Black	73.75	1		
	Black	77.12	1.2 (1.0-1.5)	2.37	
Sex					25.63
sex	Male	70.69	1		
	Female	84.18	1.7 (1.4-2.0)	4.99	
# of comorbid infectious illnesses					13.67
infect	0	72.67	1		
	1	80.56	1.1 (0.9-1.4)	1.29	
	≥ 2	92.42	3.0 (1.6-6.2)	3.17	
# of comorbid chronic illnesses					73.49
chron	0	68.10	1		
	1	77.66	1.5 (1.3-1.9)	4.16	
	2	83.15	2.1 (1.6-2.9)	5.27	
	≥ 3	89.60	3.4 (2.4-4.9)	6.91	
Alcohol, drug, and mental health problems in the past year					42.53
adm	None of these problems	69.54	1		
	Alcohol problem only	63.11	0.9 (0.7-1.2)	-0.59	
	Drug problem only	77.61	1.8 (1.3-2.5)	3.30	
	Mental health problem only	84.16	1.8 (1.4-2.4)	4.09	
	Alcohol and drug problems	76.63	1.8 (1.3-2.4)	3.68	
	Alcohol and mental health problems	79.68	1.7 (1.2-2.4)	2.87	
	Drug and mental health problems	78.71	1.4 (1.0-2.0)	1.87	
	Alcohol, drug, and mental health problems	80.52	1.6 (1.2-2.1)	3.09	

Variable (name, description)		Unadjusted Rate, % ¥	Odds Ratio (95% CI) €	z-statistic	Drop in Deviance
Status on day of interview					23.67
homlss	Currently homeless	75.66	1		
	Formerly homeless	75.40	0.6 (0.5-0.8)	-3.58	
	Never homeless (other service users)	71.57	0.5 (0.4-0.7)	-4.07	
Living on the street in the past week					18.4
street	No	77.47	1		
	Yes	66.62	0.6 (0.5-0.8)	-4.31	
Visit to drop in center in the past week					10.09
drop	No	74.61			
	Yes	77.77	1.4 (1.1-1.7)	3.13	
Has medical insurance					112.67
insur	No	65.84			
	Yes	84.97	2.6 (2.2-3.1)	10.37	
Intercept Coefficient: *1.61 ,SE 0.55, z-statistic 2.95* ***Residual Deviance:*** *3526.1*		***Hosmer-Lemeshow Test*** *(70 bins) p-value: .74*			

¥ The unadjusted rate of a category of an explanatory variable is the percentage of the clients in that category that have *visityear=1*. For example, 91.5% of clients with *age* < 18 have *visityear=1*, and 77.2% of clients with *age* from 18-44 have *visityear=1*.
€ The odds ratio of one value of an explanatory variable is the multiplicative factor by which the predicted odds of *visityear=1* increases or decreases if the explanatory variable takes that value, all other explanatory variables being held constant. For example, if a client has *age* 18-44, the odds of *visityear*=1 is estimated to be 0.3 times less than if the client has *age* < 18, all other variables held constant.

Analyses

Logistic regression of visityear on explanatory variables

Figure 1 presents the logistic regression model relating explanatory variables to the log-odds of *visityear* (1). **Table 2** presents the results of this logistic regression model. There is evidence that the following explanatory variables are statistically significant: *insur (*drop in deviance 112.67), *infect (*drop in deviance 13.67)*, chron (*drop in deviance 73.49)*, adm (*drop in deviance 42.53)*, street (*drop in deviance 18.4)*, homlss (*drop in deviance 23.67)*, drop (*drop in deviance 10.09)*, sex (*drop in deviance 25.63)*,* and *age (*drop in deviance 15.25). Holding all other explanatory variables constant, a client with insurance is estimated to have 2.6 times greater odds (95% CI 2.2-3.1) of *visityear=1* than a client without insurance; a female client is estimated to have 1.7 times greater odds (95% CI 1.4-2.0) of *visityear=1* than a male client; and a client with 2 or more chronic illnesses is estimated to have 3.0 times greater odds (95% CI 1.6-6.2) of visityear = 1 than a client without any chronic illnesses.

The predictive performance of model (1) is similar to other models with interaction terms and the expanded model including all explanatory variables under consideration. **Figure 2** describes a model including an interaction term between *age* and *sex* and a model containing all the explanatory variables under consideration. **Table 4** summarizes statistics comparing models (1) (2) and (3). The Hosmer-Lemeshow test indicates that model (2) and model (3) have superior predictive performance than model (1); however this test is unreliable because all the explanatory variables are categorical (Reiter, Logistic Regression, 2013). Indeed, the confusion matrices with threshold at 0.7 show that all the models have roughly the same predictive ability. Moreover, the area under the receiver operator characteristic (ROC) curves[1] for models (1) (2) and (3) indicate that all the models have better performance than random guessing (ROC curve areas for all models > .5). **Figure 3** displays these ROC curves.

For model (1), the linear regression assumption that explanatory variables are not linear combinations of the others is valid (none of the variables have correlation greater than .9 or less than -.9). Residuals show no reason for concern with regards to model assumptions. Additionally, plots of leverages and Cook's distances show that no cases have large influence on the model.

---

[1] ***Method: Receiver Operator Characteristic (ROC) Curves***

ROC curves are a graphical representation of the accuracy of a model, created by plotting the sensitivity of the model against one minus its specificity **Invalid source specified.**. Sensitivity is the number of cases correctly predicted as *visityear=1* over the total number of cases with *visityear=1*; specificity is the number of cases correctly predicted as *visityear=0* over the total number of cases with *visityear=0*. These statistics are calculated from confusion matrices of the model at different cutoffs, examples of which are displayed in **Table 4.** Plotting many of these sensitivity, 1-specificity pairs creates the ROC curve. A completely random guess about the outcome of the cases would fall on the diagonal line bisecting the graph. Thus, models with ROC curves with area > 0.5 are believed to be more accurate than random guessing. In R, the package pROC is used to plot these curves, using the function roc. (Xavier Robin, 2011)

Table 3: Sensitivity, 1-Specificity Pairs for Model (1)

	Confusion Matrix			Sensitivity	1-Specificity
A	cutoff = 0.5	visityear=0	visityear=1	2535/(2535+121) = .954	1-132/(132+741) =0.849
	predicted=0	132	121		
	predicted=1	741	2535		
B	cutoff = 0.7	visityear=0	visityear=1	1981/(1981+675) = .746	1-515/(515+358) = 0.41
	predicted=0	515	675		
	predicted=1	358	1981		
C	cutoff = 0.9	visityear=0	visityear=1	2057/(2057+599) = .744	1-827/(827+46) = .052
	predicted=0	827	2057		
	predicted=1	46	599		

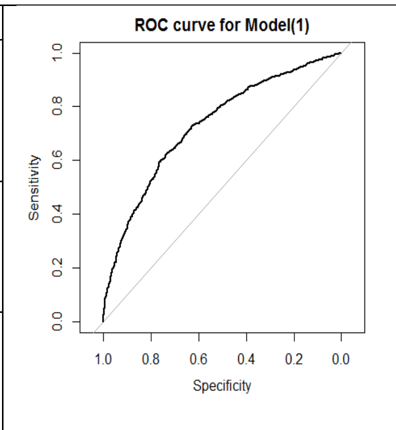


Figure 2: Description of Alternative Models

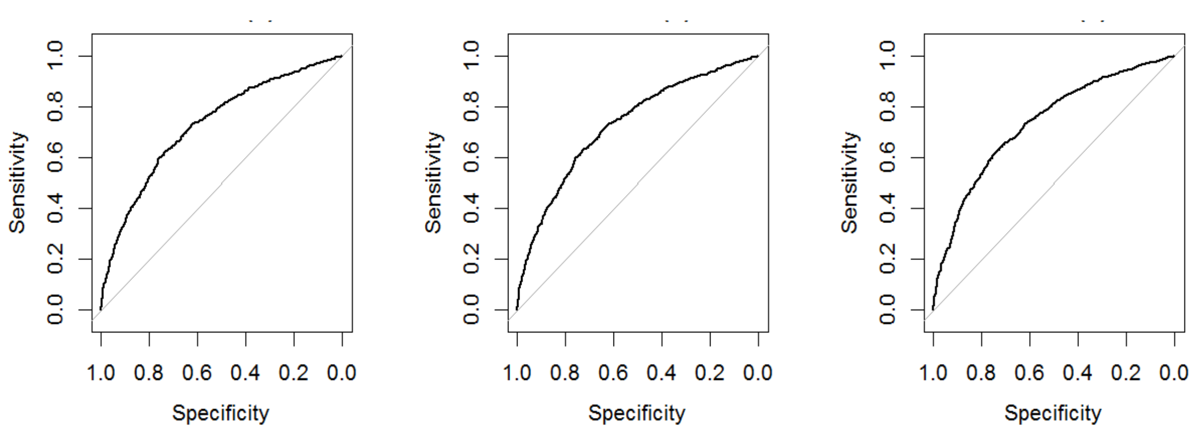
Interaction Model (2)	Expanded Model (3)
$\log\left(\frac{P(visityear=1)}{P(visityear=0)}\right) = \beta_0 + \beta_1\text{age2} + \beta_2\text{age3} + \beta_3\text{race2} + \beta_4\text{sex} + \beta_5\text{infect1} + \beta_6\text{infect2} + \beta_7\text{chronic1} + \beta_8\text{chronic2} + \beta_9\text{homlss2} + \beta_{10}\text{homlss3} + \beta_{11}\text{street1} + \beta_{12}\text{adm1} + \beta_{13}\text{adm2} + \beta_{14}\text{adm3} + \beta_{15}\text{adm4} + \beta_{16}\text{adm5} + \beta_{17}\text{adm6} + \beta_{18}\text{adm7} + \beta_{19}\text{drp1} + \beta_{20}\text{insur} + \beta_{21}\text{age2} * \text{sex} + \beta_{22}\text{age3} * \text{sex}$	Includes all explanatory variables under consideration, without interaction terms.

*The interaction term age*sex is added to model (2) because it is expected that the effect of sex on the log-odds would differ by age. For example, women of childbearing age (18-44) may be more likely to have had a medical care event in the past year.*

Table 4: Statistics Comparing Goodness-of-Fit between Reduced and Expanded Models

	Hosmer-Lemeshow (70 bins) p-value	Confusion matrix (threshold =.7)			Area under ROC Curve
Final Model (1)	.47		visityear=0	visityear=1	0.7301
		predicted=0	515	675	
		predicted=1	358	1981	
Interaction Model (2)	0.77		visityear=0	visityear=1	0.7302
		predicted=0	519	672	
		predicted=1	354	1984	
Expanded Model (3)	0.75		visityear=0	visityear=1	0.7361
		predicted=0	522	678	
		predicted=1	351	1978	

Figure 3: ROC Curves of Models (1), (2), and (3)



Multinomial logistic regression[2] *of healthcare event location on* **age, sex, street,** ***and*** **insur**

Table 5 summarizes 3 explanatory variables predicting one nominal outcome variable, *caretype.* The sample size is smaller than the logistic regression sample (n=3501 vs. n=3529), but since the difference is small, it is believed that the sample characteristics are similar. Statistics evaluating model fit are unreliable when both the model includes many categorical variables and some arrangements of explanatory variables include few cases. (Agresti, Contingency Tables, 2007) Therefore, a model including only three highly predictive variables is reported.

Table 6 summarizes a multinomial regression of *caretype* on the 3 explanatory variables. There is evidence that *sex*, *street,* and *insur* are predictive of *caretype* (p-values: $< 3\text{x}10^{-5}$, .001, $< 1.7\text{x}10^{-7}$ respectively). A client that is female is estimated to have 0.59 (95% CI 0.42-0.84) times less odds of *caretype=inpatient* vs *caretype=emergency room* than a client that is male. A client that has *street=1* is

---

[2] ***Method: Multinomial Logistic Regression***

Multinomial logistic regression is a generalization of logistic regression. The outcome variable, *caretype*, can take three values, *caretype*={emergency room, inpatient care, and ambulatory care}. Let Y_1 = # of clients whose last healthcare event was in an emergency room, Y_2 = # of clients whose last healthcare event was in inpatient care, Y_3 = # of clients whose last healthcare event was in ambulatory care. Y = {Y_1, Y_2, Y_3} is distributed as Multinomial (n,p) where p = { π_1, π_2, π_3 } is a vector of probabilities. (Reiter, Generalized Linear Models, 2013) (Agresti, Multicategory Logit Models, 2007)

$$(1)\ \log\left(\frac{\pi_{i2}}{\pi_{i2}}\right) = \beta_{02} + \beta_{12}sex_i + \beta_{22}street_i + \beta_{32}insur_i$$

$$(2)\ \log\left(\frac{\pi_{i2}}{\pi_{i2}}\right) = \beta_{02} + \beta_{12}sex_i + \beta_{22}street_i + \beta_{32}insur_i$$

One category, *caretype*=emergency room, is set as the baseline. We evaluate 2 logistic equations; one predicting the log-odds of caretype=inpatient vs. the baseline and the other predicting the log-odds of caretype=ambulatory vs. the baseline for some observation x_i = {sex_i + $street_i$ + $insur_i$}. Equations (1) and (2) display the model equations, where π_{i1} = P(*caretype=emergency room*|x_i), π_{i2} = P(*caretype=inpatient*|x_i), π_{i3} = P(*caretype=ambulatory*|x_i). The R package *nnet* provides a function *multinom* which fits both of these equations simultaneously. (Venables, 2002)

Model-checking is conducted here using likelihood ratio tests of multinomial models and chi-square test. Likelihood ratio tests are used here to compare two models, one with more terms designated B_i than the other. Under H_0: B_i = 0, the likelihood ratio statistic $D = -2\ln\left(\frac{l_0}{l_1}\right) \sim \chi^2(d.f. = d_1 - d_2)$, where l_0 is the maximum likelihood under H_0, l_1 is the maximum likelihood under no restrictions, d_1 is the number of free parameters in the bigger model, and d_2 is the number of free parameters in the smaller model. (Agresti, Building and Applying Logistic Regression Models, 2002) (Agresti, Introduction, 2002) (Agresti, Contingency Tables, 2007) (Agresti, Multicategory Logit Models, 2007) The R function anova calls this test for multinomial models. Alternatively, the "deviance" reported by the multinom function is -2ln(l(x)). (Package 'nnet', 2013) (Thomson, 2009)

Agresti also describes the use of the scoring statistic X^2. This statistic is calculated by $Q = \sum\frac{(observed - expected)^2}{expected} \sim \chi^2(d.f. = \#logits - \#of\ model\ parameters)$, where # logits refers to the # of π_{ij} calculated for every arrangement of the explanatory variables. (Agresti, Building and Applying Logistic Regression Models, 2002) (Agresti, Multicategory Logit Models, 2007)

Independence of irrelevant alternatives is an assumption that is made while using multinomial logistic regression. This assumption is not tested in this analysis, but may be tested with a Hausman diagnostic test. (Kwak, 2002) estimated to have 0.65 (95% CI 0.51-0.82) times fewer odds of *caretype=ambulatory* vs *caretype=emergency room* than a client that does not have *street=1*. Having insurance is associated with increased odds of having *caretype=inpatient* vs. *caretype=emergency room* and increased odds of having *caretype=ambulatory* vs. *caretype=emergency room*. There is little evidence that the model does not fit the data (p-value from the X^2 test statistic: 0.66). (Agresti, Building and Applying Logistic Regression Models, 2002) **Table 7** displays expected probabilities for different arrangements of the explanatory variables.

Table 5: Explanatory and Outcome Variables for Multinomial Logistic Regression

Outcome Variable		% of Sample €	Explanatory Variables		% of Sample €
The location of the client's last healthcare event			Sex		
caretype	Emergency room	13.17		Male	66.10
	Inpatient care	8.40		Female	33.90
	Ambulatory care	78.43	Living on the street in the past week		
				No	79.72
				Yes	20.28
			Has medical insurance £		
				No	50.50
				Yes	49.50

€ All variables are categorical, thus the percentage of the sample in each category is reported, without s.d.
£ Types of insurance include Medicaid, VA medical insurance, private insurance, and other types of insurance

Table 6: Summary of Multinomial Regression of caretype *on* sex, street, *and* insur *(n=3501)*

$\frac{\pi(inpatient)}{\pi(emergency\ room)}$†				$\frac{\pi(ambulatory)}{\pi(emergency\ room)}$†			
Variable	Odds Ratio	95% CI	p-value‡	Variable	Odds Ratio	95% CI	p-value‡
sex=Male	1			sex=Male	1		
sex=Female	0.59	0.42-0.84	< $3x10^{-5}$	sex=Female	1.14	0.91-1.42	< $3x10^{-5}$
street=No	1			street=No	1		
street=Yes	0.81	0.57-1.15	.001	street=Yes	0.65	0.51-0.82	.001
insur=No	1			insur=No	1		
insur=Yes	2.83	1.75-3.24	< $1.7x10^{-7}$	insur=Yes	1.47	1.19-1.82	< $1.7x10^{-7}$
Intercept coefficient: -0.68 (SE .13)				*Intercept coefficient: 1.67 (SE .084)*			

† π(inpatient) signifies the probability that *caretype=inpatient.*
‡ p-value comes from likelihood ratio tests of multinomial models testing $H_0: \beta_i = 0$, where β_i is the coefficient of the explanatory variable x_i in that row
p-value from X^2 test: 0.66

***Table 7: Expected Probabilities for* caretype**

sex	*street*	*insur*	*Location of client's last healthcare event*		
			caretype = emergency	*caretype = inpatient*	*caretype = ambulatory*
Male	No	No	0.15	0.07	0.78
	No	Yes	0.1	0.12	0.78
	Yes	No	0.21	0.08	0.71
	Yes	Yes	0.14	0.14	0.72
Female	No	No	0.14	0.04	0.82
	No	Yes	0.09	0.07	0.84
	Yes	No	0.19	0.05	0.76
	Yes	Yes	0.14	0.08	0.79

Discussion

The most statistically significant predictors for whether or not a client has had a healthcare event in the past year are having insurance, having 1 or more chronic illnesses, having a combination drug, alcohol, or mental health problems in the last year, the sex of the client, the housing status of the client, and whether or not the client has lived on the street in the past week.

Having insurance is predictive for lower probabilities of emergency care and higher probabilities of inpatient care. Being female is predictive for lower rates of both emergency and inpatient care. Living on the street is predictive for higher rates of emergency and inpatient care and lower rates of ambulatory care.

Reliability of observations

All variables except *sex* and *urbrural* are self-reported. No medical conditions or healthcare visits have been corroborated with medical records. The outcome variables *visityear* and *caretype* might not reflect the general pattern of usage for each client: each is an assessment of healthcare access at one particular instance in time. Similarly*, drop* and *outreach* do not reflect how clients receive support over time.

***Relationships between explanatory variables and outcome variable* visityear**

Since this study is observational, no causal relationship can be drawn between the explanatory variables and the outcome variable. However, there is a strong theoretical basis for a causal relationship. 95% of clients have income less than $1200 a month, and may not be able to pay for healthcare otherwise.

It is more difficult to speculate about hypothetical causal relationships between the other explanatory variables and *visityear*. In particular, since *chronic* and *infect* are self-reported, it is unclear whether or not clients seek healthcare as a result of their illnesses, whether they are able to report many illnesses due to having had a healthcare event, or neither.

Differences between homeless, formerly homeless, and other service users

Although homeless, formerly homeless, and other service users have similar unadjusted rates of *visityear=1*, a formerly homeless client is estimated to have odds of *visityear=1* 0.6 times less (95% CI 0.5-0.8) than a currently homeless client. Similarly, a client who has never been homeless is estimated to have odds of *visityear=1* 0.5 times less (95% CI 0.4-0.7) than a currently homeless client. However, formerly homeless and other services users have higher rates of *insur=1* than the currently homeless (66.67% and 68.30% vs. 43.35%, respectively).

Independence of Irrelevant Alternatives in Multinomial Model

An assumption made in multinomial regression necessary to solve for the probabilities is that the odds ratio between two categories does not change with the inclusion or exclusion of a third category. (Kwak, 2002) This assumption has not been tested. A priori, this assumption may be unreasonable (having access to inpatient care may change the odds ratio between emergency care and ambulatory care).

Inference to populations

The sample excludes all clients whose interviews were conducted in rural locations, therefore no inference to clients in those areas can be made. Since the clients were sampled only semi-randomly, inference to clients in central urban, fringe urban, or suburban areas is speculative.

References

Agresti, A. (2002). Building and Applying Logistic Regression Models. In A. Agresti, *An Introduction to Categorical Data Anlaysis* (p. 146). John Wiley & Sons, Inc.

Agresti, A. (2002). Introduction. In A. Agresti, *An Introduction to Categorical Data Analysis* (pp. 11-12). John Wiley & Sons, Inc.

Agresti, A. (2002). Introduction. In A. Agresti, *An Introduction to Categorical Data Analysis* (pp. 11-12). John Wiley & Sons, Inc.

Agresti, A. (2007). Contingency Tables. In A. Alan, *An Introduction to Categorical Data Analysis* (p. 40). John Wiley & Sons, Inc.

Agresti, A. (2007). Multicategory Logit Models. In A. Alan, *An Introduction to Categorical Data Analysis* (pp. 173-175). John Wiley & Sons, Inc.

Agresti, A. (n.d.). *R AND S-PLUS EXAMPLES.* Retrieved 4 18, 2013, from stat.ufl.edu: http://www.stat.ufl.edu/~aa/cda/R_web.pdf

Breakey, W. (1989). Health and Mental Health Problems of Homeless Men and Women in Baltimore. *Journal of the American Medical Association, 262*(10), 1352-1356.

Homelessness: Programs and the People They Serve. (1999, December). (Department of Housing and Urban Development) Retrieved 4 4, 2013, from Comparing Currently with Formerly Homeless: http://www.huduser.org/publications/homeless/homelessness/ch_3.html

Kushel, M. E. (2001). Factors Associated with the Health Care Utilization of Homeless Persons. *Journal of the American Medical Association, 285*(2), 200-206.

Kwak, C. A.-M. (2002). Methods: Multinomial Logistic Regression. *Nursing Research, 51*(6), 404-410.

Levy, B. O. (2004). Health Care for Homeless Persons. *New England Journal of Medicine, 350*(23), 2329-2332.

(1999). *NSHAPC Design and Data Collection.* Department of Housing and Urban Devleopment. Retrieved 4 4, 2013, from http://www.huduser.org/portal/publications/pdf/home_tech/tchap-02.pdf

O'Toole, T. A.-M. (2003). Health Care of Homeless Veterans. *Journal of General Internal Medicine, 18*, 921-933.

Package 'nnet'. (2013, March 20). Retrieved April 12, 2013, from cran.r-project.org: http://cran.r-project.org/web/packages/nnet/nnet.pdf

Project 25. (n.d.). Retrieved 4 22, 2013, from Ending Chronic Homelessness in San Diego: http://homeagainsd.org/our-progress/project-25

Reiter, J. (2013, 4 24). Generalized Linear Models.

Reiter, J. (2013, March 19). Logistic Regression.

Salit, S. A., Kuhn, E. M., Hartz, A. J., Vu, J. M., & Mosso, A. L. (1998). Hospitalization costs associated with homelessness in New York City. *The New England Journal of Medicine, 338*(24), 1734-1740.

Thomson, L. (2009). *R (and S-PLUS) Manual to Accompany Agresti's Categorical Data Analysis 2009 2nd edition.* Retrieved 4 20, 2013, from comcast.net: https://home.comcast.net/~lthompson221/Splusdiscrete2.pdf

Venables, W. N. (2002). *Modern Applied Statistics with S. Fourth Edition. .* New York: Springer.

Xavier Robin, N. T.-C. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics, 12*, 77.