Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

__________________________  _________________________
Calli A. Taylor               Date
Prevalent and incident homelessness among a prospective cohort of black and white non-Hispanic men who have sex with men in Atlanta, GA

By

Calli A. Taylor

Master of Public Health

Department of Epidemiology

Travis Sanchez, DVM, MPH

Faculty Thesis Advisor
Prevalent and incident homelessness among a prospective cohort of black and white non-Hispanic men who have sex with men in Atlanta, GA

By

Calli A. Taylor

B.A., University of Tennessee, 2012

Faculty Thesis Advisor:

Travis Sanchez, DVM, MPH

An abstract of

a thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Public Health

in Epidemiology

2016
Abstract

Prevalent and incident homelessness among a prospective cohort of black and white non-Hispanic men who have sex with men in Atlanta, GA

By
Calli A. Taylor

Background: Homelessness is an endemic issue in the United States that disproportionately impacts certain populations, including African Americans and men who have sex with men (MSM). Herein we report on the prevalence and incidence of homelessness among men who have sex with men in Atlanta.

Methods: Data from the InvolveMENt study, a prospective cohort study of black and white Atlanta MSM that took place between June 2010 and March 2014, were used for this study. Baseline data (n=803) were used for prevalence analyses, while prospective data from HIV-negative MSM (n=556) were used for descriptive analyses and to estimate homelessness incidence rates. Significant predictors of homelessness incidence were determined using proportional hazards modeling.

Results: Among the baseline cohort, 93 (11.6%) MSM reported any prevalent homelessness, and black MSM were more likely to report prevalent homelessness than white MSM (15.2% and 6.9%, chi-square p=0.0003). Prevalent homelessness was higher among HIV-positive participants than among HIV-negative participants at baseline (17.4% and 9.0%, chi-square p=0.0006). Among the prospective cohort, there were 42 MSM who had incidence of homelessness, a rate of 5.9 per 100 person-years. Compared to white MSM, black MSM were twice as likely to have become homeless (RR: 2.1, 95% CI: 1.1, 4.1). Depressive symptoms (RR: 2.9, 95% CI: 1.6, 5.4), and age (RR: 2.4, 95% CI: 1.3, 4.5) were also found to be associated with incident homelessness.

Conclusions: Our results provided previously unreported homelessness prevalence and incidence measures of MSM in Atlanta, allowing for an examination of predictive factors of incident homelessness and providing a foundation for further research on this topic. Race, age, and depressive symptoms were found to be predictors of incident homelessness among this population. The estimates reported in this study emphasize the need for preventative programs and services focused on African American MSM, young MSM, and those with mental health disorders.
Prevalent and incident homelessness among a prospective cohort of black and white non-Hispanic men who have sex with men in Atlanta, GA

By

Calli A. Taylor

B.A., University of Tennessee, 2012

Faculty Thesis Advisor:

Travis Sanchez, DVM, MPH

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2016
Table of Contents

Introduction .............................................................................................................................. 1

Methods ................................................................................................................................. 3

Results ................................................................................................................................. 7

Discussion ............................................................................................................................ 9

Tables

Table 1. Prevalent homelessness at baseline, by HIV status, among black and white non-Hispanic MSM followed in an incidence cohort, Atlanta 2010-2014 .......... 14

Table 2. Baseline characteristics of 556 black and white non-Hispanic MSM followed in an incidence cohort, Atlanta, 2010-2014 ........................................... 15

Table 3. Predictors of incident homelessness in 556 black and white non-Hispanic MSM followed in an incidence cohort, Atlanta, 2010-2014 ......................... 16

Table 4. Adjusted black-white hazard ratios for homelessness from multivariable models, in an incidence cohort of 556 black and white non-Hispanic MSM, Atlanta, 2010-2014 ................................................................................. 18

Figures

Figure 1 Kaplan Meier plot illustrating incident homelessness in 556 black and white non-Hispanic MSM followed in an incidence cohort, Atlanta, 2010-2014 ................................................................................. 19

Figure 2 Kaplan Meier plot illustrating incident homelessness, by race, in 556 black and white non-Hispanic MSM followed in an incidence cohort, Atlanta, 2010-2014 ................................................................................. 19

Figure 3 Kaplan Meier plot illustrating incident homelessness, by age, in 556 black and white non-Hispanic MSM followed in an incidence cohort, Atlanta, 2010-2014 ................................................................................. 20

Figure 4 Kaplan Meier plot illustrating incident homelessness, by depressive symptoms, in 556 black and white non-Hispanic MSM followed in an incidence cohort, Atlanta, 2010-2014 ................................................................................. 20

Figure 5 Kaplan Meier plot illustrating incident homelessness, by race and age, in 556 black and white non-Hispanic MSM followed in an incidence cohort, Atlanta, 2010-2014 ................................................................................. 21
Introduction

Homelessness is an endemic issue in the United States that disproportionately impacts certain populations. As reported in the 2014 Point-in-Time (PIT) estimates, released by the Department of Housing and Urban Development, there were 578,424 sheltered and unsheltered homeless people in the United States.\[1\] In the report it was also estimated that there were 16,521 homeless people in Georgia, of whom 50.3% were unsheltered.\[1\] The 2013 metro Atlanta Tri-Jurisdictional PIT count found there were 6,664 homeless people in the metro Atlanta area, with 5,571 of those within the limits of the city of Atlanta.\[2\] In the same count it was found that a disproportionate 82% of the homeless individuals in the metro Atlanta area were male.\[2\] On a national level, disparities in race as well as sex are seen, with African Americans comprising 47% of the homeless population in 2007.\[3, 4\] In comparison, according to the U.S. Census Bureau, 13.3% of the national population in 2015 was comprised of African Americans.\[5\]

It is well documented that the homeless are often at high risk for chronic and acute health problems, including HIV.\[6\] Estimates from 2006 indicate that 3.4% of homeless people in the U.S. were HIV-positive, as compared to 0.4% of the general population.\[7\] It has also been reported that approximately 16% of the single, adult homeless population nationally has some form of severe, chronic mental illness.\[6\] In addition, the homeless population in the U.S. is more likely than the general population to be affected by all chronic health problems, aside from obesity, strokes, and cancer.\[8\] Homeless persons frequently do not have adequate access to health care resources, and so are unable to prevent or address these health problems if they do occur.\[8, 30\]

Disparities are also seen in regards to sexual orientation and homelessness in the United States. It has been reported that sexual minority youth are more likely to become homeless than are youth who identify as heterosexual, and it has been estimated that 6% to 40% of homeless
youth identify as lesbian, gay, bisexual, transgender, and questioning (LGBTQ).\textsuperscript{9, 10, 11, 12, 13} In comparison, reports indicate that LGBTQ individuals comprise approximately 2.3% to 3.8% of the general population of the U.S.\textsuperscript{14, 15}

One study compared the prevalence of current homelessness among sexual minority adolescents to that of their heterosexual counterparts in Massachusetts, and found that the sexual minority participants were between 4 and 13 times more likely to report homelessness.\textsuperscript{16} Another study, that investigated HIV seroprevalence among homeless and marginally housed people in San Francisco, found that 19.2% of the adult males included in the study identified as men who have sex with men (MSM).\textsuperscript{17} A recent report from an ongoing study of young sexual minority men in New York found that 12.7% reported unstable sleeping conditions, and 18.2% reported unstable living conditions.\textsuperscript{18} The same study reported that non-Hispanic black participants were more likely to report unstable housing, and strong associations were found between unstable housing and history of arrest and adverse childhood experiences.\textsuperscript{18} While there is research available regarding sexual minority homelessness, the majority of the available information is focused on prevalence of homelessness, and correlates or risks associated with homelessness among this population.

There are several factors that are commonly accepted to be associated with homelessness in the United States, including history of arrest, adverse childhood experiences, mental illness, and substance abuse. Research indicates that youth and adult homeless are more likely to have experienced adverse childhood events, including physical, emotional, and sexual abuse and neglect.\textsuperscript{9, 19} A number of studies have also shown an association between negative early life experiences, drug abuse, and homelessness among young MSM.\textsuperscript{9, 20, 21} Examinations of risk factors for homelessness have indicated that mental health problems and early childhood
adverse experiences are predictors of homelessness, and that history of arrest and drug and alcohol misuse are strongly associated with initiation and duration of homelessness.\textsuperscript{[21, 22, 23]}

While there is research available regarding MSM in Atlanta, and there is basic information about the homeless population in the city, there have been no reports regarding the intersection of these populations. In addition, there are very few reports of incidence of homelessness among sexual minorities in the United States. Public health professionals and policy makers require a more comprehensive picture of homelessness among MSM to plan programs and design health interventions for this at-risk population. Consequently, there is a public health need to better understand prevalence and incidence of homelessness among MSM in Atlanta. Research has shown that there are disparities in sex, race, and sexual orientation among homeless populations in the U.S., and these factors suggest that the risk of homelessness for African American MSM may be elevated as compared to the general population. Herein we report on the prevalence and incidence of homelessness, overall and by race and age group, among men who have sex with men in Atlanta.

**Methods**

**Study Description**

This study was conducted using data from the InvolveMENt study, a recently completed, prospective cohort study.\textsuperscript{[24, 25, 26]} The Emory University-led study was comprised of a computer-based questionnaire and HIV/STI testing of black and white metro Atlanta MSM, and has current Emory University IRB approval. Venue sampling and convenience sampling through Facebook were used to recruit black and white MSM for the study. A variety of venues throughout Atlanta where MSM congregate were randomly chosen, and at the venues study staff systematically
approached men to administer screening questions. Facebook sampling was achieved through paid advertisements directed towards men who were possibly eligible for the study.

Recruitment took place between June 2010 and December 2012 during 605 sampling events at 94 individual venues. Eligibility requirements included being male at birth, being capable of completing the study instruments in English, self-reported white or black race, currently living in the metropolitan Atlanta area, having at least one male sex partner in the last three months, being 18 years of age or older, and providing at least two means of contact. Exclusion criteria included reporting Hispanic or Latino ethnicity, having plans to move out of Atlanta within the next two years, being in a mutually monogamous relationship with a man, or currently participating in any HIV prevention research study.

Participants completed a computer assisted self-interview questionnaire at the baseline visit that included questions pertaining to demographics, psychosocial scales, community characteristics, individual-level HIV-related behaviors, and a dyadic inventory of the five most recent sex partners. Participants were also screened for HIV antibodies at each visit. At the baseline visit, participants were tested for syphilis, chlamydia, and gonorrhea. In addition, all participants were screened for biological markers of marijuana, cocaine, morphine, amphetamines, and methamphetamine at baseline. All participants who were confirmed to be HIV-negative at the baseline visit were then followed prospectively for two years to observe HIV and STI incidence. Those eligible returned for visits at 3, 6, 12, 18, and 24 months after the baseline visit, and the follow-up visits were conducted essentially the same as the baseline visit. Study participation ended after the 24-month visit or after HIV seroconversion.

**Measures**

This study was conducted using variables of interest from both the baseline and prospective datasets from the InvolveMENt study. The dependent variable, homelessness, was assessed at
the baseline visit as well as at each follow-up visit using two questions. The definition of homelessness used for this study was included in the first question; “In the past 12 months, have you been homeless at any time? By homeless, I mean you were living on the street, in a shelter, a Single Room Occupancy hotel (SRO), temporarily staying with friends or relatives, or living in a car?”. If this question was answered affirmatively, the second question, “Are you currently homeless”, was then asked. For the purposes of this study, historical prevalence is defined as homelessness within the 12 months prior to the baseline visit, while current prevalence was reported by participants at the baseline visit.

An event of homelessness was considered incident the first time a participant reported homelessness at a follow-up visit in addition to reporting no homelessness at the preceding visit. While several participants with incident homelessness reported multiple, separate events of homelessness throughout the follow-up time, for the purposes of this study participants were allowed only one event. Homelessness events were considered separate if the participant reported homelessness, and at a subsequent visit reported no longer being homeless, then later reported homelessness again.

Participants who reported incident homelessness at a follow-up visit were considered to have become homeless at the midpoint between that visit and the previous visit. Person-time for participants with an incident event was started at the date of enrollment into the study and ended at the incidence event. Person-time for participants who never reported homelessness was the difference between the date of the final study visit and the date of enrollment. For those who reported prevalent homelessness at baseline, contributed person-time was started after the first follow-up visit at which the participant reported no homelessness. If homelessness data was missing or reported as a “don’t know” at a visit, the participant was censored at the midpoint between that visit and the prior visit.
The following independent variables were all assessed at the baseline visit: race, age group, prior arrests, childhood abuse, clinically significant depressive symptoms, problem drinking, and self-reported illicit drug use. For the present analysis we did not include any independent variables from the follow-up visits. Age at baseline was categorized into two groups, 18-24 years of age, and 25 years of age and older. Prior arrests were defined as any arrest, drug arrests included, in the 12 months preceding the baseline visit. Childhood abuse included any prior physical and sexual abuse, and was reported at baseline. Clinically significant depressive symptoms were defined using a shortened version of the Center for Epidemiologic Studies Depression Scale (CESD-10). The CESD-10 was scored on a four-point Likert scale, and a higher score indicates higher agreement with depressive symptoms. The 10 items were summed to determine a score from 0 to 30, and a total score of 10 or greater is indicative of clinically significant depressive symptoms. The CAGE questionnaire, a 4-question validated screening tool, was used as an indicator for a probable alcohol problem. Participants were asked to consider how they felt and behaved over their whole life when answering the questionnaire, and a total score of 2 or more indicated a likely problem with alcohol. Illicit drug use was assessed by self-reported non-injection and injection drug use in the previous 12 months, and included any use of non-prescribed crystal meth, crack cocaine, cocaine, downers, painkillers, hallucinogens, ecstasy, special k, GHB, heroin, marijuana, and poppers.

**Analyses**

The baseline dataset, comprised of both HIV-positive and negative MSM, was analyzed to briefly describe prevalent homelessness by HIV status among black and white MSM in Atlanta. For the descriptive analyses of the prospective dataset, which includes HIV-negative participants only, chi-square or Fisher’s exact test were used to compare characteristics of black and white MSM that are commonly associated with homelessness.
Bivariate incidence-density rates, rate ratios, and exact 95% confidence intervals were computed for each independent variable to determine which were preliminarily associated with incidence homelessness among this population. Cox proportional hazards modeling was then used to determine which factors were predictors of incidence of homelessness among MSM in Atlanta. The adjusted hazard ratio for incident homelessness was estimated first, and a backwards-selection approach was used to arrive at the final model. Significant predictive factors, at a p-value of <0.05, were retained in the final model. For all models, adjusted hazard ratios and 95% CI were estimated. The Kaplan-Meier analysis method was employed to examine cumulative incidence of homelessness overall, and to further evaluate the independent variables that were determined to be significant predictors of incident homelessness among this population. Kaplan-Meier failure curves of overall homelessness incidence, stratified by race, by age at the baseline visit, by presence of depressive symptoms, and by race and age were produced. Differences in failure curves were evaluated by the log-rank test. Analysis for this study was performed using SAS 9.4.

Results

Of 19,931 men who were approached during recruitment for the InvolveMENt study, 8,983 were screened, and 2,144 were eligible. A total of 6,092 men clicked on the Facebook advertisement, 1,360 were screened, and 184 were eligible. Of the 1,010 black MSM and 713 white MSM who were offered participation in the study, 454 black MSM and 349 white MSM came to the baseline visit and consented to be enrolled in the study. Participants were followed for up to 24 months, and follow-up ended in March 2014. The final baseline dataset includes 803 men, while the prospective dataset is limited to 556 HIV-negative men.

Prevalent Homelessness
Of the 803 HIV-negative and positive black and white MSM who were enrolled in the baseline study, 93 (11.6%) reported any prevalent homelessness [Table 1]. Black MSM were significantly more likely than white MSM to report historical prevalent homelessness. Of the 803 enrolled MSM, 556 (69%) were HIV-negative, and 50 (9%) of those MSM reported any prevalence of homelessness at the baseline visit [Table 1]. Among the HIV-negative cohort that was followed prospectively, black MSM were again significantly more likely to report historical prevalent homelessness. Of the 247 HIV-positive MSM at baseline, 43 (17.4%) reported any prevalent homelessness, and there were no significant differences by race among that group. However, the percentage of prevalent homelessness among the 247 HIV-positive MSM (17.4%) was much higher than among the 556 HIV-negative MSM (9%) [Table 1].

**Incident Homelessness**

Of the 556 MSM (255 black, 301 white) in the prospective cohort, black MSM were, on average, younger. Among black MSM, 49.8% were 18-24 years of age at baseline, as compared to 33.2% of white MSM [Table 2]. There was a total of 710 person-years (PY) of follow-up time: 294.5 PY were contributed by black MSM, and 415.6 PY were contributed by white MSM [Table 3]. Of the follow-up time by age, 259.1 PY were contributed by MSM 18-24 years old at baseline, and 450.9 PY were contributed by MSM 25 years and older [Table 3].

There were 42 (IR: 5.9/100PY) reports of incident homelessness observed, with 27 (64.3%) reports among black MSM (IR: 9.2/100PY), and 15 (35.7%) reports among white MSM (IR: 3.6/100PY), yielding an incidence rate ratio of 2.5 (95% CI: 1.4, 4.9) [Table 3]. Of the 42 participants that reported incident homelessness, 10 (24%) had multiple homeless events: 2 (20%) among white MSM and 8 (80%) among black MSM. When considering age and race together, black MSM 18-24 years old had the highest incidence rate of homelessness (IR: 10.8, 95% CI: 6.0, 17.8). Compared to white MSM 25 years and older, black MSM 18-24 years old were
more than six times as likely to have become homeless (RR: 6.4, 95% CI: 2.4, 19.6). Incident homelessness was also associated with childhood abuse (RR: 2.0, 95% CI: 1.1, 3.7) and symptoms of depression (RR: 2.9, 95% CI: 1.6, 5.4) [Table 3]. Illicit drug use, problem drinking, and prior arrest were not found to be associated with homelessness incidence.

The final adjusted Cox proportional-hazards model included race, age, and depressive symptoms as significant predictors of incident homelessness [Table 4]. Compared to white MSM, black MSM were twice as likely to have become homeless (hazard ratio [HR]: 2.1, 95% CI: 1.1, 4.1). MSM 18-24 years old were more than twice as likely to report incident homelessness than MSM 25 years and older (HR: 2.4, 95% CI: 1.3, 4.5). Participants with depressive symptoms were nearly three times as likely to have become homeless compared to those without depressive symptoms (HR: 2.9, 95% CI: 1.6, 5.4). The interaction between race and age was also examined, but was not included in the final model as it was not found to be a significant predictive factor.

Stratified Kaplan-Meier failure estimate figures were then produced for the variables that were found to be significant predictors of incident homelessness from the proportional-hazards analysis. Proportion of homelessness from the figures was significantly different by race, by age group, by presence of depressive symptoms, and by race and age [Figures 1-5]. While the interaction of race and age was not found to be significant in the Cox proportional-hazards models, it was included in the Kaplan-Meier analysis due to the strong effect of the interaction that was observed in the bivariate analysis.

**Discussion**

There is little known about homelessness among MSM in the United States, but it has been well documented that certain populations are disproportionately represented among the homeless
throughout the country. This study confirmed the racial disparities seen within homeless populations in the U.S., and it also provided information on the characteristics and most commonly reported contributing factors of homelessness among MSM in Atlanta.

Among both the baseline and prospective cohorts, black MSM were more likely to report any prevalent homelessness than white MSM. Prevalence of homelessness was found to be higher among HIV positive participants (17.4%) than among HIV negative participants (9%). The incidence of homelessness among MSM in Atlanta was 5.9/100PY, which suggests that approximately 6% of MSM in Atlanta become newly homeless each year. In addition, significant racial disparities in homelessness incidence were observed, with 9.2% incidence among black MSM, in comparison to 3.6% for white MSM.

While available research regarding homelessness among MSM populations in the U.S. is sparse, the findings of this study are mostly consistent with estimates from the available research on young LGBTQ populations and homelessness more broadly. The Krause et al. study from New York reported similar historical prevalence estimates to those in this study (18.2% and 11.4%, respectively), while the Corliss et al. study from Massachusetts reported higher estimates of current homelessness prevalence among gay/lesbian participants than in this study (25% and 2.4%, respectively). Multiple studies, on both a national and regional scale, found young sexual minorities to be a group that is overrepresented within the homeless population, with approximately 6% to 40% of homeless youth identifying as LGBTQ. Findings from this study are consistent with those estimates, with 9% to 17.4% of the baseline cohort reporting any prevalent homelessness, depending on HIV status. The observed racial differences among MSM who reported homelessness in this study are also consistent with previous studies. A cross-sectional study found that African Americans comprised approximately 47% of the total
U.S. homeless population in 2007, which is a similar, but less extreme disparity, than we observed in this study, with black MSM reporting 64.3% of incident homelessness. [3, 4]

While the prevalence estimates from this study are consistent with the findings from many previous reports, we were able to make a rather unique contribution to the existing body of research with this study. Due to a variety of factors, the vast majority of previous studies related to LGBTQ homelessness have reported only prevalence measures, while this study was able to provide both prevalence and incidence estimates. Incidence measures of homelessness in general are useful indicators of risk, but this study was able to provide incidence estimates of homelessness among a population that is already at risk for many other health concerns. This incidence data allowed us to include an element of temporality, and to examine potential predictive factors of incident homelessness rather than solely reporting correlation. In addition, this study provides a foundation for further research on homelessness among MSM. This report demonstrates that incidence research on homelessness in this population is feasible, and shows that more specifically designed future studies could examine this issue in greater depth.

The unadjusted incidence results of this study indicate that among this population of MSM in Atlanta, race, age, depressive symptoms, and childhood abuse are associated with incidence of homelessness. When considering the adjusted model, race, age, and symptoms of depression were significant predictors of homelessness incidence. While prior arrest, illicit drug use, and problem drinking are known correlates of homelessness, they were not found to be significantly associated with incident homelessness among this study population.

This study was subject to some limitations. The censoring of HIV seroconverted men from further follow-up as part of the InvolveMENt study design may result in an underestimate of incidence of homelessness in this study. This is also a possibility as prevalence of HIV is higher among homeless populations in the U.S., and we found that prevalence of homelessness at
baseline was higher among those with HIV infection in this study as well.[6, 7] Additionally, the substance use variables were likely impacted by misclassification, as an analysis of the self-reported baseline measures indicated systematic underreporting of substance use by black participants.[25] Despite using systematic sampling methods, the cohort study was not representative of all black and white MSM in Atlanta. This study conservatively allowed only one event of incident homelessness per participant during the analysis, but 24% of participants reported multiple incident events. A more complex analysis, allowing participants to contribute multiple events or to be censored for missed visits and later continue contributing person-time, is recommended for subsequent studies. As this was a preliminary study, only factors that were previously reported as correlates of homelessness were included, and further research should explore other factors among this population of MSM in Atlanta. Moreover, this study used baseline characteristics as these measures were consistent across participants, and subsequent analyses should consider including time varying factors.

As demonstrated in this study, MSM in Atlanta may be at higher risk for homelessness than is the general population. Incidence of homelessness among this study population was 5.9/100PY, and there were clear racial and age disparities in homelessness incidence observed among MSM in Atlanta. In addition, symptoms of depression, race, and age were found to be significant predictors of incident homelessness. This information, which is elucidating the intersection of two at-risk populations, could aid public health professionals and policy makers in planning programs and health interventions tailored towards this specific population.

The homeless populations in the United States are at higher risk for many health problems, including HIV, and they frequently lack adequate access to health care.[6, 8, 30] Furthermore, MSM nationally and in Atlanta have been shown to be at elevated risk for certain health problems, and are also disproportionately affected by HIV.[7, 24, 31] In this study population
of MSM in Atlanta, prevalent homelessness was higher among those living with HIV infection, indicating a need for programs and services targeted towards ensuring housing stability for those with HIV. In addition, younger age and symptoms of depression were both predictors of incident homelessness among this population. These factors indicate that younger MSM may have a lack of support and resources, which places them at higher risk of housing instability, and that those with depression may be more likely to engage in behaviors and risks that increase their likelihood of becoming homeless. Additional programs and preventative services should be created with a focus on at-risk youth, and on those with mental health disorders.

African Americans nationally are at higher risk of homelessness, and racial disparities in prevalence and incidence of homelessness were observed in this study population as well.\[^3,4\] While there are many factors that may be associated with racial disparities in homelessness, the findings from this study confirm the need for programs and services focused specifically on African Americans as a population at risk for homelessness in Atlanta. Furthermore, sexual minority populations in the U.S. have been reported to be at higher risk of homelessness than their heterosexual counterparts, and the prevalence and incidence estimates of homelessness among MSM in this study are consistent with those findings.\[^9,10,12,13,16,17,18\] Support for sexual minority populations, including MSM in Atlanta, would be greatly beneficial to help prevent adverse public health outcomes, such as homelessness, among these at-risk populations.
### Table 1. Prevalent homelessness by baseline HIV status and race, among black and white non-Hispanic MSM followed in an incidence cohort, Atlanta 2010-2014

<table>
<thead>
<tr>
<th>Prevalent Homelessness, all MSM, at baseline</th>
<th>Overall (N=803) % (N)</th>
<th>Black (n=454) % (n)</th>
<th>White (n=349) % (n)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All prevalence</td>
<td>11.6 (93/803)</td>
<td>15.2 (69/454)</td>
<td>6.9 (24/349)</td>
<td>0.0003</td>
</tr>
<tr>
<td>Historical prevalence, previous 12 months</td>
<td>11.4 (91/798)</td>
<td>14.9 (67/451)</td>
<td>6.9 (24/347)</td>
<td>0.0005</td>
</tr>
<tr>
<td>Current prevalence, at baseline visit</td>
<td>2.4 (19/796)</td>
<td>3.8 (17/449)</td>
<td>0.6 (2/347)</td>
<td>0.003&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prevalent Homelessness, HIV- MSM, at baseline</th>
<th>Overall (N=556) % (N)</th>
<th>Black (n=255) % (n)</th>
<th>White (n=301) % (n)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All prevalence</td>
<td>9.0 (50/556)</td>
<td>14.1 (36/255)</td>
<td>4.7 (14/301)</td>
<td>0.0001</td>
</tr>
<tr>
<td>Historical prevalence, previous 12 months</td>
<td>8.8 (49/554)</td>
<td>13.7 (35/255)</td>
<td>4.7 (14/299)</td>
<td>0.0002</td>
</tr>
<tr>
<td>Current prevalence, at baseline visit</td>
<td>1.3 (7/550)</td>
<td>2.0 (5/251)</td>
<td>0.7 (2/299)</td>
<td>0.25&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prevalent Homelessness, HIV+ MSM, at baseline</th>
<th>Overall (N=247) % (N)</th>
<th>Black (n=199) % (n)</th>
<th>White (n=48) % (n)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All prevalence</td>
<td>17.4 (43/247)</td>
<td>16.6 (33/199)</td>
<td>20.8 (10/48)</td>
<td>0.49</td>
</tr>
<tr>
<td>Historical prevalence, previous 12 months</td>
<td>17.2 (42/244)</td>
<td>16.3 (32/196)</td>
<td>20.8 (10/48)</td>
<td>0.46</td>
</tr>
<tr>
<td>Current prevalence, at baseline visit</td>
<td>4.9 (12/246)</td>
<td>6.1 (12/198)</td>
<td>0.0 (0/48)</td>
<td>0.07&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup> From Fisher’s exact test
**Table 2. Baseline characteristics of 556 black and white non-Hispanic MSM followed in an incidence cohort, Atlanta, 2010-2014**

<table>
<thead>
<tr>
<th></th>
<th>Overall (N=556)</th>
<th>Black (n=255)</th>
<th>White (n=301)</th>
<th>p-value a</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age, years b</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>40.8 (227)</td>
<td>49.8 (127)</td>
<td>33.2 (100)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>≥25</td>
<td>59.1 (329)</td>
<td>50.2 (128)</td>
<td>66.8 (201)</td>
<td></td>
</tr>
<tr>
<td><strong>Illicit Drug Use, previous 12 months c</strong></td>
<td>40.1 (223)</td>
<td>27.1 (69)</td>
<td>51.1 (154)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td><strong>Problem Drinking, ever d</strong></td>
<td>14.5 (75/517)</td>
<td>9.1 (21/231)</td>
<td>18.9 (54/286)</td>
<td>0.002</td>
</tr>
<tr>
<td><strong>Childhood abuse, ever e</strong></td>
<td>38.9 (216)</td>
<td>43.5 (111)</td>
<td>34.9 (105)</td>
<td>0.04</td>
</tr>
<tr>
<td><strong>Arrested, previous 12 months f</strong></td>
<td>12.2 (68)</td>
<td>14.1 (36)</td>
<td>10.6 (32)</td>
<td>0.21</td>
</tr>
<tr>
<td><strong>Depressive Symptoms, past week g</strong></td>
<td>26.4 (141/535)</td>
<td>26.2 (64/244)</td>
<td>26.5 (77/291)</td>
<td>0.95</td>
</tr>
</tbody>
</table>

---

a From chi-square test for categorical variables and t-tests for continuous variables

b Age at baseline
c Illicit drug use included self-reported non-injection and injection drug use in the previous 12 months, and included any use of non-prescribed crystal meth, crack cocaine, cocaine, downers, painkillers, hallucinogens, ecstasy, special k, GHB, heroin, marijuana, and poppers
d Problem drinking was assessed using the CAGE questionnaire, reported at baseline, and considered lifetime feelings and behaviors
e Childhood abuse includes physical and sexual abuse, self-reported
f Any arrests, drug arrests included, in previous 12 months
g Clinically significant depressive symptoms were reported at baseline, based on feelings and behaviors of the past week, and assessed using the CESD-10 scale
Table 3. Predictors of incident homelessness in 556 black and white non-Hispanic MSM followed in an incidence cohort, Atlanta, 2010-2014

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Reported homelessness N (%)</th>
<th>Person-years (PY)</th>
<th>Incidence rate per 100 PY (95% CI)</th>
<th>Unadjusted incidence rate ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>42</td>
<td>710</td>
<td>5.9 (4.3, 7.9)</td>
<td>-- --</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>27 (64.3)</td>
<td>294.5</td>
<td>9.2 (6.0, 13.3)</td>
<td>2.5 (1.4, 4.9)</td>
</tr>
<tr>
<td>White</td>
<td>15 (35.7)</td>
<td>415.6</td>
<td>3.6 (2.0, 6.0)</td>
<td>ref.</td>
</tr>
<tr>
<td>Age, years, baseline</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-24</td>
<td>25 (59.5)</td>
<td>259.1</td>
<td>9.6 (6.2, 14.2)</td>
<td>2.6 (1.4, 4.8)</td>
</tr>
<tr>
<td>≥25</td>
<td>17 (40.5)</td>
<td>450.9</td>
<td>3.8 (2.2, 6.0)</td>
<td>ref.</td>
</tr>
<tr>
<td>Race and age</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black, 18-24</td>
<td>15 (35.7)</td>
<td>139.0</td>
<td>10.8 (6.0, 17.8)</td>
<td>6.4 (2.4, 19.6)</td>
</tr>
<tr>
<td>Black, 25+</td>
<td>12 (28.6)</td>
<td>155.5</td>
<td>7.7 (4.0, 13.5)</td>
<td>4.6 (1.6, 14.4)</td>
</tr>
<tr>
<td>White, 18-24</td>
<td>10 (23.8)</td>
<td>120.2</td>
<td>8.3 (4.0, 15.3)</td>
<td>4.9 (1.7, 15.9)</td>
</tr>
<tr>
<td>White, 25+</td>
<td>5 (11.9)</td>
<td>295.4</td>
<td>1.7 (0.5, 4.0)</td>
<td>ref.</td>
</tr>
<tr>
<td>Arrested, past 12 months, baseline a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>6 (14.3)</td>
<td>64.8</td>
<td>9.3 (3.4, 20.2)</td>
<td>1.7 (0.6, 3.8)</td>
</tr>
<tr>
<td>No</td>
<td>36 (85.7)</td>
<td>645.2</td>
<td>5.6 (3.9, 7.7)</td>
<td>ref.</td>
</tr>
<tr>
<td>Illicit drug use, past 12 months, baseline b</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>21 (50.0)</td>
<td>283</td>
<td>7.4 (4.6, 11.3)</td>
<td>1.5 (0.8, 2.8)</td>
</tr>
<tr>
<td>No</td>
<td>21 (50.0)</td>
<td>427</td>
<td>4.9 (3.0, 7.5)</td>
<td>ref.</td>
</tr>
<tr>
<td>Problem drinking, ever, baseline c,d</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>7 (18.4)</td>
<td>77.5</td>
<td>9.0 (3.6, 18.6)</td>
<td>1.7 (0.7, 3.7)</td>
</tr>
<tr>
<td>No</td>
<td>31 (81.6)</td>
<td>581.2</td>
<td>5.3 (3.6, 7.6)</td>
<td>ref.</td>
</tr>
<tr>
<td>Depressive symptoms, past week, baseline e,f</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>20 (48.8)</td>
<td>170.3</td>
<td>11.7 (7.2, 18.1)</td>
<td>2.9 (1.6, 5.4)</td>
</tr>
<tr>
<td>Childhood abuse, ever, baseline ⁶</td>
<td>No</td>
<td>Yes</td>
<td>ref.</td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>----</td>
<td>-----</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td></td>
<td>21 (51.2)</td>
<td>21 (50.0)</td>
<td>21 (50.0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>516.3</td>
<td>271.9</td>
<td>438.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.1 (2.5, 6.2)</td>
<td>9.6 (6.0, 14.7)</td>
<td>4.8 (3.0, 7.3)</td>
<td></td>
</tr>
</tbody>
</table>

* Any arrests, drug arrests included, in previous 12 months
  ⁵ I illicit drug use included self-reported non-injection and injection drug use in the previous 12 months, and included any use of non-prescribed crystal meth, crack cocaine, cocaine, downers, painkillers, hallucinogens, ecstasy, special k, GHB, heroin, marijuana, and poppers
  ⁶ Problem drinking was assessed using the CAGE questionnaire, reported at baseline, and considered lifetime feelings and behaviors
  ⁷ Problem drinking status missing for 4 participants with incident homelessness
  ⁸ Clinically significant depressive symptoms were reported at baseline, based on feelings and behaviors of the past week, and assessed using the CESD-10 scale
  ⁹ Depressive symptoms status missing for 1 participant with incident homelessness
  ⁶ Childhood abuse includes physical and sexual abuse, self-reported
Table 4. Adjusted hazard ratios for homelessness from multivariable models, in an incidence cohort of 556 black and white non-Hispanic MSM, Atlanta, 2010-2014

<table>
<thead>
<tr>
<th>Variables</th>
<th>Hazard Ratio</th>
<th>95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race, black vs. white</td>
<td>2.1</td>
<td>1.1 - 4.1</td>
<td>0.02</td>
</tr>
<tr>
<td>Age, 18-24 vs. 25+</td>
<td>2.4</td>
<td>1.3 - 4.5</td>
<td>0.008</td>
</tr>
<tr>
<td>Depressive symptoms</td>
<td>2.9</td>
<td>1.6 - 5.4</td>
<td>0.0007</td>
</tr>
</tbody>
</table>
Figure 1 Kaplan Meier plot illustrating incident homelessness in 556 black and white non-Hispanic MSM followed in an incidence cohort, Atlanta, 2010-2014

Figure 2 Kaplan Meier plot illustrating incident homelessness, by race, in 556 black and white non-Hispanic MSM followed in an incidence cohort, Atlanta, 2010-2014
Figure 3 Kaplan Meier plot illustrating incident homelessness, by age, in 556 black and white non-Hispanic MSM followed in an incidence cohort, Atlanta, 2010-2014

Figure 4 Kaplan Meier plot illustrating incident homelessness, by depressive symptoms, in 556 black and white non-Hispanic MSM followed in an incidence cohort, Atlanta, 2010-2014
Figure 5 Kaplan Meier plot illustrating incident homelessness, by race and age, in 556 black and white non-Hispanic MSM followed in an incidence cohort, Atlanta, 2010-2014
References


15. Gates GJ. *How many people are lesbian, gay, bisexual, and transgender?* The Williams Institute, 2011.


Appendix A: SAS Code

***********************************************************;
* Program: T:epiprojs\Sullivan_data_ident\C_Taylor_Inv\cat\thesis_data.sas7bdat *
* Date: 11/20/15 *
* Programmer: Calli Taylor *
* Purpose: This program uses the InvolveMENt datasets for thesis work. *
***********************************************************;

*** Run thesis_formats FIRST;
libname thesis 'T:epiprojs\Sullivan_data_ident\C_Taylor_Inv';

******
* Create temporary datasets from long stacked set to limit # of vars
and prepare sets for merging;
* Also format and recode covariates;
******;

data stacked;
set thesis.participants_bto24_stacked;

keep study_id visit homeless_p12m homeless_now homeless_p6m homeless1_int homeless2_int income poverty educ arrested ever arrested_p12m_final arrested_p6m drug_arrest abuse_child physabuse abuse_child sexabuse alcohol p6m drink_p12m cocaine_p12m cocaine_p6m crack_p12m crack_p6m crystalmeth_p12m crystalmeth_p6m downers_p12m downers_p6m ecstasy_p12m ecstasy_p6m ecstasy_p6m ghb_p12m ghb_p6m hallucinogen_p12m hallucinogen_p6m heroin_p12m heroin_p6m inject inject_p6 marijuana_p12m marijuana_p6m noninject_p12m noninject_p6
noninject_other_p12m painkillers_p12m painkillers_p6m poppers_p12m poppers_p6m specialk_p12m specialk_p6m cesd_hopeful cesd_happy cesd_hopefulp cesd_happyp cesd_lonely cesd_getgoing cesd_fearful cesd_effort cesd_depressed cesd_concentrate cesd_bother cesd_badsleep drink_critic drink_cut drink_morning drink_guilt;

format cocaine_p12m cocaine_p6m any_cocaine crack_p12m crack_p6m any_crack crystalmeth_p12m crystalmeth_p6m any_crystal downers_p12m downers_p6m any_downers ecstasy_p12m ecstasy_p6m ecstasy_p6m any_ecstasy ghb_p12m ghb_p6m any_ghb
if visit = 1 then homeless1_int = homeless_p12m;
if visit > 1 then homeless2_int = homeless_p6m;

if cocaine_p12m ge 1 then cocaine_p12m = 1;
if cocaine_p6m ge 1 then cocaine_p6m = 1;
if crack_p12m ge 1 then crack_p12m = 1;
if crack_p6m ge 1 then crack_p6m = 1;
if crystalmeth_p12m ge 1 then crystalmeth_p12m = 1;
if crystalmeth_p6m ge 1 then crystalmeth_p6m = 1;
if downers_p12m ge 1 then downers_p12m = 1;
if downers_p6m ge 1 then downers_p6m = 1;
if ecstasy_p12m ge 1 then ecstasy_p12m = 1;
if ecstasy_p6m ge 1 then ecstasy_p6m = 1;
if ghb_p12m ge 1 then ghb_p12m = 1;
if ghb_p6m ge 1 then ghb_p6m = 1;
if hallucinogen_p12m ge 1 then hallucinogen_p12m = 1;
if hallucinogen_p6m ge 1 then hallucinogen_p6m = 1;
if heroin_p12m ge 1 then heroin_p12m = 1;
if heroin_p6m ge 1 then heroin_p6m = 1;
if inject ge 1 then inject = 1;
if injectp6m ge 1 then injectp6m = 1;
if marijuana_p12m ge 1 then marijuana_p12m = 1;
if marijuana_p6m ge 1 then marijuana_p6m = 1;
if noninject_p12m ge 1 then noninject_p12m = 1;
if noninject_p6m ge 1 then noninject_p6m = 1;
if painkillers_p12m ge 1 then painkillers_p12m = 1;
if painkillers_p6m ge 1 then painkillers_p6m = 1;
if poppers_p12m ge 1 then poppers_p12m = 1;
if poppers_p6m ge 1 then poppers_p6m = 1;
if specialk_p12m ge 1 then specialk_p12m = 1;
if specialk_p6m ge 1 then specialk_p6m = 1;
*Recode scoring of positive CES-D items - reversed scoring;

```plaintext
if cesd_hopeful = 3 then cesd_hopefulp = 0;
if cesd_hopeful = 2 then cesd_hopefulp = 1;
if cesd_hopeful = 1 then cesd_hopefulp = 2;
if cesd_hopeful = 0 then cesd_hopefulp = 3;

if cesd_happy = 3 then cesd_happyp = 0;
if cesd_happy = 2 then cesd_happyp = 1;
if cesd_happy = 1 then cesd_happyp = 2;
if cesd_happy = 0 then cesd_happyp = 3;
```

run;

data stacked1;
set stacked;

keep study_id visit homeless_p12m homeless_now
homeless_p6m homeless1_int homeless2_int income
poverty educ arrested_ever arrested_p12m_final
arrested_p6m drug_arrest abuse_child_physabuse
abuse_child_sexabuse alcoholp6m drink_p12m
cocaine_p12m cocaine_p6m crack_p12m crack_p6m
crystalmeth_p12m crystalmeth_p6m dowers_p12m
dowers_p6m ecstasy_p12m ecstasy_p6m
ecstasy_p6m ghb_p12m ghb_p6m hallucinogen_p12m
hallucinogen_p6m heroin_p12m heroin_p6m inject
injectp6m marijuana_p12m marijuana_p6m
noninject_p12m noninject_p6
noninject_other_p12m painkillers_p12m
painkillers_p6m poppers_p12m poppers_p6m
specialk_p12m specialk_p6m cesd_hopeful
cesd_happy cesd_lonely cesd_getgoing
cesd_fearful cesd_effort cesd_depressed
cesd_concentrate cesd_bother cesd_badsleep
drink_critic drink_cut drink_morning
drink_guilt;

run;

data stacked2;
set stacked;

keep study_id visit homeless_p12m homeless_now
homeless_p6m homeless1_int homeless2_int income
poverty educ arrested_ever arrested_p12m_final
arrested_p6m drug_arrest abuse_child_physabuse
abuse_child_sexabuse alcoholp6m drink_p12m
cocaine_p12m cocaine_p6m crack_p12m crack_p6m
crystalmeth_p12m crystalmeth_p6m dowers_p12m
dowers_p6m ecstasy_p12m ecstasy_p6m
ecstasy_p6m ghb_p12m ghb_p6m hallucinogen_p12m
hallucinogen_p6m heroin_p12m heroin_p6m inject
injectp6m marijuana_p12m marijuana_p6m
noninject_p12m noninject_p6
noninject_other_p12m painkillers_p12m
painkillers_p6m poppers_p12m poppers_p6m
specialk_p12m specialk_p6m cesd_hopeful
cesd_happy cesd_lonely cesd_getgoing
cesd_fearful cesd_effort cesd_depressed
cesd_concentrate cesd_bother cesd_badsleep
drink_critic drink_cut drink_morning
drink_guilt;
proc sort data = stacked3;
   by study_id visit;
run;

data s1;
   set stacked;
   by study_id visit;
   if first.study_id then output;
run;

******
* To create vars that indicate visit of first reported homelessness;
 *** http://www.ats.ucla.edu/stat/sas/faq/enumerate.htm ***/
******;

run;
run;

data s2;
   set s1 (rename = (visit = first_event));
   keep study_id first_event;
run;

/*
proc print data = s2;
run;
*/

*** pl2m event;

proc sort data = stacked2;
   by study_id visit;
run;

data s3;
   set stacked2;
   by study_id visit;
     if first.study_id then output;
run;

data s4;
   set s3 (rename = (visit = first_1home));
   keep study_id first_1home;
run;

/*
proc print data = s4;
run;
*/

*** p6m event;

proc sort data = stacked3;
   by study_id visit;
run;

data s5;
   set stacked3;
   by study_id visit;
     if first.study_id then output;
run;

data s6;
   set s5 (rename = (visit = first_2home));
   keep study_id first_2home;
run;

/*
proc print data = s6;
run;
*/

******
* To merge the sets that indicate visit before homelessness;  
******;

```
proc sort data = s2;
  by study_id;
run;

proc sort data = s4;
  by study_id;
run;

proc sort data = s6;
  by study_id;
run;

data stacked_new;
  merge s2 s4 s6;
  by study_id;
run;

******

* Create a new var that indicates the first visit with reported  
  homelessness  
  Then make a var that indicates the visit prior to the first;  
  * Also make var that indicates whether a guy is incident  
    homeless;  
  * Then merge them into one set to later be merged into  
    master set;  
******;

proc sort data = stacked_new;
  by first_1home first_2home first_event;
run;

data s_first;
set stacked_new;
  if first_1home = 1 then first_visit = 1;
  else if first_event = 1 then first_visit = 1;
  else if first_2home le first_event
  then first_visit = first_2home;
  else if first_1home = . and first_event = .
  then first_visit = first_2home;

  if first_1home = 1 and first_event = . and
  first_2home = . then incident = 0;
  else if first_1home = 1 and first_event = 1 and
  first_2home = . then incident = 0;
  else if first_1home = . and first_event = 1 and
  first_2home = . then incident = 0;
  else if first_1home = 1 and first_event = 1 and
  first_2home = 2 and study_id ne 0202941
  then incident = 0;
  else incident = 1;
```
run;

data s_prior;
   set s_first;
       visit_prior = first_visit-1;
run;

proc sort data = s_first;
   by study_id;
run;

proc sort data = s_prior;
   by study_id;
run;

data s;
   merge s_first s_prior;
   by study_id;
run;

******
* Create a temporary dataset stat identical to permanent dataset status keeping only vars of interest;
******;

data stat;
   set thesis.status;
       keep study_id race_inc age_baseline dt_enroll last_visit_date part_dob daysenrolled daysenrolledstop daysenrolledvisit met_behav_crit prospective_pt baseline_hiv double_enroll last_visit;
run;

******
* To get the right n from the stat set;
******;

* Baseline positives included (n=803);

data stat_base;
   set stat;
       where double_enroll^=1 and met_behav_crit=1;
run;

* Prospective negative guys only (n=556);

data stat_pro_neg;
   set stat;
       where double_enroll^=1 and met_behav_crit=1 and prospective_pt=1 and baseline_hiv=1;
run;
* Process to make wide dataset -
  create datasets with one row per participant based on id, by
  visit,
  then change var names and merge all back together;

```
data baseline_survey;
  set stacked (rename =
    homeless_now = homeless_1now
    homeless_p12m = homeless_1
    income = income1
    poverty = poverty1
    educ = edu1
    arrested_ever = arrested_ever1
    arrested_p6m = arrested_p6m1
    arrested_p12m_final = arrested_p12m_final1
    drug_arrest = drug_arrest1
    abuse_child_physabuse = abuse_child_physabuse1
    abuse_child_sexabuse = abuse_child_sexabuse1
    alcoholp6m = alcoholp6m1
    drink_p12m = drink_p12m1
    cocaine_p12m = cocaine_p12m1
    cocaine_p6m = cocaine_p6m1
    crack_p12m = crack_p12m1
    crack_p6m = crack_p6m1
    crystalmeth_p12m = crystalmeth_p12m1
    crystalmeth_p6m = crystalmeth_p6m1
    downers_p12m = downers_p12m1
    downers_p6m = downers_p6m1
    ecstasy_p12m = ecstasy_p12m1
    ecstasy_p6m = ecstasy_p6m1
    ghb_p12m = ghb_p12m1
    ghb_p6m = ghb_p6m1
    hallucinogen_p12m = hallucinogen_p12m1
    hallucinogen_p6m = hallucinogen_p6m1
    heroin_p12m = heroin_p12m1
    heroin_p6m = heroin_p6m1
    inject = inject1
    injectp6m = injectp6m1
    marijuana_p6m = marijuana_p6m1
    marijuana_p12m = marijuana_p12m1
    noninject_p12m = noninject_p12m1
    noninject_p6m = noninject_p6m1
    noninject_other_p12m = noninject_other_p12m1
    painkillers_p12m = painkillers_p12m1
    painkillers_p6m = painkillers_p6m1
    poppers_p12m = poppers_p12m1
    poppers_p6m = poppers_p6m1
    specialk_p12m = specialk_p12m1
    specialk_p6m = specialk_p6m1
    cesd_hopefulp = cesd_hopeful1
    cesd_happyp = cesd_happy1
    cesd_lonely = cesd_lonely1
    cesd_getgoing = cesd_getgoing1
    cesd_fearful = cesd_fearful1
```
```
cesd_effort = cesd_effort1
cesd_depressed = cesd_depressed1
cesd_concentrate = cesd_concentrate1
cesd_bother = cesd_bother1
cesd_badsleep = cesd_badsleep1
drink_critic = drink_critic1
drink_cut = drink_cut1
drink_morning = drink_morning1
drink_guilt = drink_guilt1);

where visit = 1;
drop visit;
run;

data month3_survey ;
set stacked (rename = (homeless_now = homeless_2now
income = income2
poverty = poverty2
educ = educ2
arrested_ever = arrested_ever2
arrested_p6m = arrested_p6m2
arrested_p12m_final = arrested_p12m_final2
drug_arrest = drug_arrest2
abuse_child_physabuse = abuse_child_physabuse2
abuse_child_sexabuse = abuse_child_sexabuse2
alcoholp6m = alcoholp6m2
drink_p12m = drink_p12m2
cocaine_p12m = cocaine_p12m2
cocaine_p6m = cocaine_p6m2
crack_p12m = crack_p12m2
crack_p6m = crack_p6m2
crystalmeth_p12m = crystalmeth_p12m2
crystalmeth_p6m = crystalmeth_p6m2
downers_p12m = downers_p12m2
downers_p6m = downers_p6m2
ecstasy_p12m = ecstasy_p12m2
ecstasy_p6m = ecstasy_p6m2
ecstasy_p6m = ecstasy_p6m2
ghb_p12m = ghb_p12m2
ghb_p6m = ghb_p6m2
hallucinogen_p12m = hallucinogen_p12m2
hallucinogen_p6m = hallucinogen_p6m2
heroin_p12m = heroin_p12m2
heroin_p6m = heroin_p6m2
inject = inject2
injecp6m = injeckp6m2
marijuana_p6m = marijuana_p6m2
marijuana_p12m = marijuana_p12m2
noninject_p12m = noninject_p12m2
noninject_p6m = noninject_p6m2
noninject_other_p12m = noninject_other_p12m2
painkillers_p12m = painkillers_p12m2
painkillers_p6m = painkillers_p6m2
poppers_p12m = poppers_p12m2
poppers_p6m = poppers_p6m2
specialk_{p12m} = specialk_{p12m2}
specialk_{p6m} = specialk_{p6m2}
cesd_hopefulp = cesd_hopeful2
cesd_happyp = cesd_happy2
cesd_lonely = cesd_lonely2
cesd_getgoing = cesd_getgoing2
cesd_fearful = cesd_fearful2
cesd_effort = cesd_effort2
cesd_depressed = cesd_depressed2
cesd_concentrate = cesd_concentrate2
cesd_bother = cesd_bother2
cesd_badsleep = cesd_badsleep2
drink_critic = drink_critic2
drink_cut = drink_cut2
drink_morning = drink_morning2
drink_guilt = drink_guilt2));

where visit = 2;
drop visit;
run;
data month6_survey ;
set stacked (rename = (homeless_now = homeless_3now
income = income3
poverty = poverty3
educ = educ3
arrested_ever = arrested_ever3
arrested_p6m = arrested_p6m3
arrested_p12m_final = arrested_p12m_final3
drug_arrest = drug_arrest3
abuse_child_physabuse = abuse_child_physabuse3
abuse_child_sexabuse = abuse_child_sexabuse3
alcoholp6m = alcoholp6m3
drink_p12m = drink_p12m3
cocaine_p12m = cocaine_p12m3
cocaine_p6m = cocaine_p6m3
crack_p12m = crack_p12m3
crack_p6m = crack_p6m3
crystalmeth_p12m = crystalmeth_p12m3
crystalmeth_p6m = crystalmeth_p6m3
downers_p12m = downers_p12m3
downers_p6m = downers_p6m3
ecstasy_p12m = ecstasy_p12m3
ecstasy_p6m = ecstasy_p6m3
ecstasy_p6m = ecstasy_p6m3
ghb_p12m = ghb_p12m3
ghb_p6m = ghb_p6m3
hallucinogen_p12m = hallucinogen_p12m3
hallucinogen_p6m = hallucinogen_p6m3
heroin_p12m = heroin_p12m3
heroin_p6m = heroin_p6m3
inject = inject3
injectp6m = injectp6m3
marijuana_p6m = marijuana_p6m3
marijuana_p12m = marijuana_p12m3
noninject_p12m = noninject_p12m3


noninject_p6m = noninject_p6m3
noninject_other_p12m = noninject_other_p12m3
painkillers_p12m = painkillers_p12m3
painkillers_p6m = painkillers_p6m3
poppers_p12m = poppers_p12m3
poppers_p6m = poppers_p6m3
specialk_p12m = specialk_p12m3
specialk_p6m = specialk_p6m3
cesd_hopefulp = cesd_hopefulp3
cesd_happyp = cesd_happyp3
cesd_lonely = cesd_lonelyp3
cesd_getgoing = cesd_getgoingp3
cesd_fearful = cesd_fearfulp3
cesd_effort = cesd_effortp3
cesd_depressed = cesd_depressedp3
cesd_concentrate = cesd_concentratep3
cesd_badsleep = cesd_badsleepp3
drink_critic = drink_criticp3
drink_cut = drink_cutp3
drink_morning = drink_morningp3
drink_guilt = drink_guilt3);

where visit = 3;
drop visit;
run;
data month12_survey;
  set stacked (rename = (homeless_now = homeless_4now
homeless_p6m = homeless_4
income = income4
poverty = poverty4
educ = educ4
arrested_ever = arrested_ever4
arrested_p6m = arrested_p6m4
arrested_p12m_final = arrested_p12m_final4
drug_arrest = drug_arrest4
abuse_child_physabuse = abuse_child_physabuse4
abuse_child_sexabuse = abuse_child_sexabuse4
alcoholp6m = alcoholp6m4
drink_p12m = drink_p12m4
cocaine_p12m = cocaine_p12m4
cocaine_p6m = cocaine_p6m4
crack_p12m = crack_p12m4
crack_p6m = crack_p6m4
crystalmeth_p12m = crystalmeth_p12m4
crystalmeth_p6m = crystalmeth_p6m4
downers_p12m = downers_p12m4
downers_p6m = downers_p6m4
ecstasy_p12m = ecstasy_p12m4
ecstasy_p6m = ecstasy_p6m4
ecstasy_p6m = ecstasy_p6m4
ghb_p12m = ghb_p12m4
ghb_p6m = ghb_p6m4
hallucinogen_p12m = hallucinogen_p12m4
hallucinogen_p6m = hallucinogen_p6m4
heroin_p12m = heroin_p12m4)
heroin_p6m = heroin_p6m4
inject = inject4
injectp6m = injectp6m4
marijuana_p6m = marijuana_p6m4
marijuana_p12m = marijuana_p12m4
noninject_p12m = noninject_p12m4
noninject_p6m = noninject_p6m4
noninject_other_p12m = noninject_other_p12m4
painkillers_p12m = painkillers_p12m4
painkillers_p6m = painkillers_p6m4
poppers_p12m = poppers_p12m4
poppers_p6m = poppers_p6m4
specialk_p12m = specialk_p12m4
specialk_p6m = specialk_p6m4
cesd_hopefulp = cesd_hopefulp4
cesd_happy4 = cesd_happy4
cesd_lonely = cesd_lonely4
cesd_getgoing = cesd_getgoing4
cesd_fearful = cesd_fearful4
cesd_effort = cesd_effort4
cesd_depressed = cesd_depressed4
cesd_concentrate = cesd_concentrate4
cesd_bother = cesd_bother4
cesd_badsleep = cesd_badsleep4
drink_critic = drink_critic4
drink_cut = drink_cut4
drink_morning = drink_morning4
drink_guilt = drink_guilt4);

where visit = 4;
drop visit;
run;
data month18_survey;
set stacked (rename=(homeless_now = homeless_5now
homeless_p6m = homeless_5
income = income5
poverty = poverty5
educ = educ5
arrested_ever = arrested_ever5
arrested_p6m = arrested_p6m5
arrested_p12m_final = arrested_p12m_final5
drug_arrest = drug_arrest5
abuse_child_physabuse = abuse_child_physabuse5
abuse_child_sexabuse = abuse_child_sexabuse5
alcoholp6m = alcoholp6m5
drink_p12m = drink_p12m5
cocaine_p12m = cocaine_p12m5
cocaine_p6m = cocaine_p6m5
crack_p12m = crack_p12m5
crack_p6m = crack_p6m5
crystalmeth_p12m = crystalmeth_p12m5
crystalmeth_p6m = crystalmeth_p6m5
downers_p12m = downers_p12m5
downers_p6m = downers_p6m5
ecstasy_p12m = ecstasy_p12m5
ecstasy_p6m = ecstasy_p6m5

);
ecstasy_p6m = ecstasy_p6m5
gbh_p12m = ghb_p12m5
gbh_p6m = ghb_p6m5
hallucinogen_p12m = hallucinogen_p12m5
hallucinogen_p6m = hallucinogen_p6m5
heroin_p12m = heroin_p12m5
heroin_p6m = heroin_p6m5
injection = injection5
injection_p6m = injection_p6m5
marijuana_p6m = marijuana_p6m5
marijuana_p12m = marijuana_p12m5
noninjection_p12m = noninjection_p12m5
noninjection_p6m = noninjection_p6m5
noninjection_other_p12m = noninjection_other_p12m5
painkillers_p12m = painkillers_p12m5
painkillers_p6m = painkillers_p6m5
poppers_p12m = poppers_p12m5
poppers_p6m = poppers_p6m5
specialk_p12m = specialk_p12m5
specialk_p6m = specialk_p6m5
cesd_hopeful_p5 = cesd_hopeful_p5
cesd_happy_p5 = cesd_happy_p5
cesd_lonely = cesd_lonely_p5
cesd_getgoing = cesd_getgoing_p5
cesd_fearful = cesd_fearful_p5
cesd_effort = cesd_effort_p5
cesd_depressed = cesd_depressed_p5
cesd_concentrating = cesd_concentrating_p5
cesd_bother = cesd_bother_p5
cesd_bad_sleep = cesd_bad_sleep_p5
drink_critic = drink_critic_p5
drink_cut = drink_cut_p5
drink_morning = drink_morning_p5

where visit = 5;
drop visit;

run;

data month24_survey ;
set stacked (rename = (homeless_now = homeless_6now
homeless_p6m = homeless_6
income = income6
poverty = poverty6
educ = educ6
arrested_ever = arrested_ever6
arrested_p6m = arrested_p6m6
arrested_p12m_final = arrested_p12m_final6
drug_arrest = drug_arrest6
abuse_child_physabuse = abuse_child_physabuse6
abuse_child_sexabuse = abuse_child_sexabuse6
alcoholp6m = alcoholp6m6
drink_p12m = drink_p12m6
cocaine_p12m = cocaine_p12m6
cocaine_p6m = cocaine_p6m6
crack_p12m = crack_p12m6

crack_p6m = crack_p6m6
crystalmeth_p12m = crystalmeth_p12m6
crystalmeth_p6m = crystalmeth_p6m6
downers_p12m = downers_p12m6
downers_p6m = downers_p6m6
ecstasy_p12m = ecstasy_p12m6
ecstasy_p6m = ecstasy_p6m6
eccstacy_p6m = eccstacy_p6m6
ghb_p12m = ghb_p12m6
ghb_p6m = ghb_p6m6
hallucinogen_p12m = hallucinogen_p12m6
hallucinogen_p6m = hallucinogen_p6m6
heroin_p12m = heroin_p12m6
heroin_p6m = heroin_p6m6
inject = inject6
injectp6m = injectp6m6
marijuana_p6m = marijuana_p6m6
marijuana_p12m = marijuana_p12m6
noninject_p12m = noninject_p12m6
noninject_p6m = noninject_p6m6
noninject_other_p12m = noninject_other_p12m6
painkillers_p12m = painkillers_p12m6
painkillers_p6m = painkillers_p6m6
poppers_p12m = poppers_p12m6
poppers_p6m = poppers_p6m6
specialk_p12m = specialk_p12m6
specialk_p6m = specialk_p6m6
cesd_hopefulp = cesd_hopeful6
cesd_happyp = cesd_happy6
cesd_lonely = cesd_lonely6
cesd_getgoing = cesd_getgoing6
cesd_fearful = cesd_fearful6
cesd_effort = cesd_effort6
cesd_depressed = cesd_depressed6
cesd_concentrate = cesd_concentrate6
cesd_bother = cesd_bother6
cesd_badsleep = cesd_badsleep6
drink_critic = drink_critic6
drink_cut = drink_cut6
drink_morning = drink_morning6
drink_guilt = drink_guilt6)

where visit = 6;
drop visit;
run;

******
* To merge pulled apart stacked sets back into one wide set;
******;

proc sort data = baseline_survey;
    by study_id;
run;

proc sort data = month3_survey;
    by study_id;
run;

proc sort data = month6_survey;
    by study_id;
run;

proc sort data = month12_survey;
    by study_id;
run;

proc sort data = month18_survey;
    by study_id;
run;

proc sort data = month24_survey;
    by study_id;
run;

data stacked_wide;
    merge baseline_survey month3_survey month6_survey
        month12_survey month18_survey month24_survey;
    by study_id;
run;

*****
* To merge set (with visit indicator and incidence vars) with wide stacked;
*****;

proc sort data = stacked_wide;
    by study_id;
run;

proc sort data = s;
    by study_id;
run;

data stacks;
    merge stacked_wide s;
    by study_id;
run;

*****
To merge stat (stat_base set, n=803) and stacked_wide datasets;
* Set DKs to missings;
*****;

proc sort data = stacks;
    by study_id;
run;

proc sort data = stat_base;
    by study_id;
run;
data merged1;
  merge stacks (in=survey) stat_base (in=status);
  by study_id;
  if status;
    if homeless_1now eq 9 then homeless_1now = .;
    if homeless_1 eq 9 then homeless_1 = .;
    if homeless_2now eq 9 then homeless_2now = .;
    if homeless_2 eq 9 then homeless_2 = .;
    if homeless_3now eq 9 then homeless_3now = .;
    if homeless_3 eq 9 then homeless_3 = .;
    if homeless_4now eq 9 then homeless_4now = .;
    if homeless_4 eq 9 then homeless_4 = .;
    if homeless_5now eq 9 then homeless_5now = .;
    if homeless_5 eq 9 then homeless_5 = .;
    if homeless_6now eq 9 then homeless_6now = .;
    if homeless_6 eq 9 then homeless_6 = .;
  endif;
  if age_baseline le 24 then age_n = '18-24 yr';
  if age_baseline ge 25 then age_n = '25+ yr';
  any_arrestp12m = sum(arrested_p12m_final1, drug_arrest1);
  any_abuse = sum(abuse_child_physabuse1, abuse_child_sexabuse1);
  all_drug = sum(noninject_p12m1, inject1);
  if any_abuse ge 1 then any_abuse = 1;
  if any_arrestp12m ge 1 then any_arrestp12m = 1;
  if all_drug ge 1 then all_drug = 1;

*Create var for total response for CES-D scale;
  cesd_total = 0;
  cesd_total = cesd_bother1 + cesd_concentrate1 +
              cesd_depressed1 + cesd_effort1 + cesd_hopefull1 +
              cesd_fearfull + cesd_badsleep1 + cesd_happy1 +
              cesd_lonely1 + cesd_getgoing1;

*Create binary outcome for CES-D scale. The cutoff is 10;
  if cesd_total gt . and cesd_total lt 10 then cesd = 2;
  if cesd_total ge 10 then cesd = 1;

*Create var for total response for CAGE scale questions;
cage = 0;
if drink_cut1 = . then cage = .;
if drink_critic1 = . then cage = .;
if drink_guilt1 = . then cage = .;
if drink_morning1 = . then cage = .;
if drink_cut1 = 1 then cage = cage + 1;
if drink_critic1 = 1 then cage = cage + 1;
if drink_guilt1 = 1 then cage = cage + 1;
if drink_morning1 = 1 then cage = cage + 1;

*Create binary var for CAGE outcome. 3-4 indicates problematic alcohol use;

if cage gt . and cage le 2 then drink = 0;
if cage gt 2 then drink = 1;

if homeless_1 = 1 or homeless_lnow = 1 then prev_home = 1;
else prev_home = 0;

label age_n = "Age"
person time = "Person time in days"
pt_month = "Study Month"
race_inc = "Race"
cesd = "Depressive Symptoms"
drink = "Problem Drinking"
all_drug = "Illicit Drug Use"
any_arrestp12m = "Arrested"
any_abuse = "Childhood Abuse";

format drink all_drug any_arrestp12m any_abuse homeless_1
homeless_lnow prev_home
yaynay.
format cesd
cesdbi.

run;

*****
* To merge stat (stat_pro_neg set, n=556) and stacked_wide datasets;
* Set DKs to missings;
* Create summed covariate vars, and PT vars;
*****;

proc sort data = stacks;
   by study_id;
run;

proc sort data = stat_pro_neg;
   by study_id;
run;
data merged3;
  merge stacks (in=survey) stat_pro_neg (in=status);
    by study_id;
    if status;
    if homeless_1now eq 9 then homeless_1now = .;
    if homeless_1 eq 9 then homeless_1 = .;
    if homeless_2now eq 9 then homeless_2now = .;
    if homeless_2 eq 9 then homeless_2 = .;
    if homeless_3now eq 9 then homeless_3now = .;
    if homeless_3 eq 9 then homeless_3 = .;
    if homeless_4now eq 9 then homeless_4now = .;
    if homeless_4 eq 9 then homeless_4 = .;
    if homeless_5now eq 9 then homeless_5now = .;
    if homeless_5 eq 9 then homeless_5 = .;
    if homeless_6now eq 9 then homeless_6now = .;
    if homeless_6 eq 9 then homeless_6 = .;

  any_homeless = sum(homeless_1, homeless_1now, homeless_2,
      homeless_2now, homeless_3, homeless_3now,
      homeless_4, homeless_4now, homeless_5,
      homeless_5now, homeless_6, homeless_6now);

  anyp6m_arrest = sum(arrested_p6m1, arrested_p6m2,
         arrested_p6m3, arrested_p6m4, arrested_p6m5,
         arrested_p6m6);

  any_arrest = sum(arrested_ever1, arrested_p12m_final1,
      anyp6m_arrest, drug_arrest1);

  any_arrestp12m = sum(arrested_p12m_final1, drug_arrest1);

  any_abuse = sum(abuse_child_physabuse1,
      abuse_child_sexabuse1);

  any_alcohol = sum(drink_p12m1, alcoholp6m3, alcoholp6m4,
      alcoholp6m5, alcoholp6m6);

  any_cocaine = sum(cocaine_p12m1, cocaine_p6m1,
      cocaine_p6m2, cocaine_p6m3, cocaine_p6m4,
      cocaine_p6m5, cocaine_p6m6);

  any_crack = sum(crack_p12m1, crack_p6m2, crack_p6m3,
      crack_p6m4, crack_p6m5, crack_p6m6);

  any_crystal = sum(crystalmeth_p12m1, crystalmeth_p6m2,
      crystalmeth_p6m3, crystalmeth_p6m4,
      crystalmeth_p6m5, crystalmeth_p6m6);

  any_downers = sum(downers_p12m1, downers_p6m2,
      downers_p6m3, downers_p6m4, downers_p6m5,
      downers_p6m6);

  ecstasy_p6m2 = ecstasy_p6m2;
any_ecstasy = sum(ecstasy_p12m1, ecstasy_p6m2, ecstasy_p6m3, ecstasy_p6m4, ecstasy_p6m5, ecstasy_p6m6);

any_ghb = sum(ghb_p12m1, ghb_p6m2, ghb_p6m3, ghb_p6m4, ghb_p6m5, ghb_p6m6);

any_hallucinogen = sum(hallucinogen_p12m1, hallucinogen_p6m2, hallucinogen_p6m3, hallucinogen_p6m4, hallucinogen_p6m5, hallucinogen_p6m6);

any_heroin = sum(heroin_p12m1, heroin_p6m2, heroin_p6m3, heroin_p6m4, heroin_p6m5, heroin_p6m6);

any_inject = sum(inject1, injectp6m2, injectp6m3, injectp6m4, injectp6m5, injectp6m6);

any_marijuana = sum(marijuana_p12m1, marijuana_p6m2, marijuana_p6m3, marijuana_p6m4, marijuana_p6m5, marijuana_p6m6);

any_noninject = sum(noninject_p12m1, noninject_p6m2, noninject_p6m3, noninject_p6m4, noninject_p6m5, noninject_p6m6);

any_painkillers = sum(painkillers_p12m1, painkillers_p6m2, painkillers_p6m3, painkillers_p6m4, painkillers_p6m5, painkillers_p6m6);

any_poppers = sum(poppers_p12m1, poppers_p6m2, poppers_p6m3, poppers_p6m4, poppers_p6m5, poppers_p6m6);

any_specialk = sum(specialk_p12m1, specialk_p6m2, specialk_p6m3, specialk_p6m4, specialk_p6m5, specialk_p6m6);

any_noninjectp12m = sum(cocaine_p12m1, crack_p12m1, crystalmeth_p12m1, downers_p12m1, ecstasy_p12m1, ghb_p12m1, hallucinogen_p12m1, heroin_p12m1, marijuana_p12m1, painkillers_p12m1, poppers_p12m1, specialk_p12m1);

invol_noninjectp12m = sum(cocaine_p12m1, crack_p12m1, crystalmeth_p12m1, downers_p12m1, ecstasy_p12m1, ghb_p12m1, hallucinogen_p12m1, painkillers_p12m1, specialk_p12m1);

all_noninject_p12m = sum(noninject_p12m1, noninject_other_p12m);

all_drug = sum(noninject_p12m1, inject1);

if any_abuse ge 1 then any_abuse = 1;
if any_arrestpl2m ge 1 then any_arrestpl2m = 1;

if first_event = . and first_home = . and first_2home = . then incident = 0;

if all_drug ge 1 then all_drug = 1;

v1_dt = dt_enroll;
v2_dt = (dt_enroll + 91.25);
v3_dt = (dt_enroll + 182.5);
v4_dt = (dt_enroll + 365);
v5_dt = (dt_enroll + 547.5);
v6_dt = (dt_enroll + 730);

* Create var for total response for CES-D scale;

cesd_total = 0;
cesd_total = cesd_bother1 + cesd_concentratel + cesd_depressed1 + cesd_effortl + cesd_hopefull + cesd_fearfull + cesd_badsleepl + cesd_happy1 + cesd_lonely1 + cesd_getgoing1;

* Create binary outcome for CES-D scale. The cutoff is 10;

if cesd_total gt . and cesd_total lt 10 then cesd = 2;
if cesd_total ge 10 then cesd = 1;

* Create var for total response for CAGE scale questions;

cage = 0;
cage = cage + 1;
cage = cage + 1;
cage = cage + 1;

* Create binary var for CAGE outcome. 3-4 indicates problematic alcohol use;

run;

******
* Make set/var so I can subtract incident pt time from total pt calculations;
* Also make a var, time, that indicates pt at first incident event;
* Also also make a var that censors guys when they have missing
p6m homelessness data;
* Create var to correct start time (to start after they've answered no to homeless questions) for current prevalent guys;
* Create final person time var that calculates for everyone;
* Create age groups;
* Create format for drug covariates;

*****;

data incidence;
  set merged3;

  if homeless_2 = 1 then event_time = (dt_enroll + 45.625);
  else if (homeless_2 = 0 or homeless_2 = .) and homeless_2now = 1 then event_time = (dt_enroll + 91.25);
  else if (homeless_2 = 0 or homeless_2 = .) and (homeless_2now = 0 or homeless_2now = .) and homeless_3 = 1 then event_time = (dt_enroll + 136.875);
  else if (homeless_2 = 0 or homeless_2 = .) and (homeless_2now = 0 or homeless_2now = .) and (homeless_3 = 0 or homeless_3 = .) and homeless_3now = 1 then event_time = (dt_enroll + 182.5);
  else if (homeless_2 = 0 or homeless_2 = .) and (homeless_2now = 0 or homeless_2now = .) and (homeless_3 = 0 or homeless_3 = .) and (homeless_3now = 0 or homeless_3now = .) and homeless_4 = 1 then event_time = (dt_enroll + 273.75);
  else if (homeless_2 = 0 or homeless_2 = .) and (homeless_2now = 0 or homeless_2now = .) and (homeless_3 = 0 or homeless_3 = .) and (homeless_3now = 0 or homeless_3now = .) and (homeless_4 = 0 or homeless_4 = .) and homeless_4now = 1 then event_time = (dt_enroll + 365);
  else if (homeless_2 = 0 or homeless_2 = .) and (homeless_2now = 0 or homeless_2now = .) and (homeless_3 = 0 or homeless_3 = .) and (homeless_3now = 0 or homeless_3now = .) and (homeless_4 = 0 or homeless_4 = .) and (homeless_4now = 0 or homeless_4now = .) and homeless_5 = 1 then event_time = (dt_enroll + 456.25);
  else if (homeless_2 = 0 or homeless_2 = .) and (homeless_2now = 0 or homeless_2now = .) and (homeless_3 = 0 or homeless_3 = .) and (homeless_3now = 0 or homeless_3now = .) and (homeless_4 = 0 or homeless_4 = .) and (homeless_4now = 0 or homeless_4now = .) and (homeless_5 = 0 or homeless_5 = .) and homeless_5now = 1
then \( \text{event\_time} = (\text{dt\_enroll} + 547.5) \); 

```plaintext
else if (homeless_2 = 0 or homeless_2 = .) and 
(homeless_2now = 0 or homeless_2now = .) and 
(homeless_3 = 0 or homeless_3 = .) and (homeless_3now 
= 0 or homeless_3now = .) and (homeless_4 = 0 or 
homeless_4 = .) and (homeless_4now = 0 or 
homeless_4now = .) and (homeless_5 = 0 or homeless_5 
= .) and (homeless_5now = 0 or homeless_5now = .) and 
homeless_6 = 1 
then \( \text{event\_time} = (\text{dt\_enroll} + 638.75) \); 
```

```plaintext
else if (homeless_2 = 0 or homeless_2 = .) and 
(homeless_2now = 0 or homeless_2now = .) and 
(homeless_3 = 0 or homeless_3 = .) and (homeless_3now 
= 0 or homeless_3now = .) and (homeless_4 = 0 or 
homeless_4 = .) and (homeless_4now = 0 or 
homeless_4now = .) and (homeless_5 = 0 or homeless_5 
= .) and (homeless_5now = 0 or homeless_5now = .) and 
homeless_6 = 1 
then \( \text{event\_time} = (\text{dt\_enroll} + 730) \); 
```

* Fix event status for guys that look incident but are not;  
* Also create event status var;

```plaintext
if study_id = 0300361 then event = 0; 
else if study_id = 0300641 then event = 0; 
else if study_id = 0300731 then event = 0; 
else if incident = 1 then event = 1; 
else event = 0; 
```

* Fix start time for incident guys with prevalent homelessness;

```plaintext
if study_id = 0102491 
then start\_time = (\text{dt\_enroll} + 45.625); 
  * YYN.N.N.YY...; 
else if study_id = 0102521 
then start\_time = (\text{dt\_enroll} + 136.875); 
  * YYYN.N.N...; 
else if study_id = 0202471 
then start\_time = (\text{dt\_enroll} + 45.625); 
  * YYN.N....; 
else if study_id = 0202941 
then start\_time = (\text{dt\_enroll} + 91.25); 
  * YYNYY...N...; 
else if study_id = 0301531 
then start\_time = (\text{dt\_enroll} + 182.5); 
  * YYYYNN.N.N.; 
```

* Fix start time for guys with DKs;

```plaintext
else if study_id = 0100361 
then start\_time = (\text{dt\_enroll} + 45.625); 
  * DDN.N.N.N.N.; 
else if study_id = 0102241 
```

```plaintext
```
then start_time = (dt_enroll);
  * DN..........;
else if study_id = 0200201
  then start_time = (dt_enroll + 182.5);
  * NNN.N...N.D.;
else if study_id = 0200991
  then start_time = (dt_enroll + 182.5);
  * NNN.N.D.N...;
else if study_id = 0202911
  then start_time = (dt_enroll + 182.5);
  * NNN.N.N.D...;
else if study_id = 0400411
  then start_time = (dt_enroll);
  * DN..........;
else start_time = dt_enroll;

* Fix end time for incident guys with prevalent homelessness;

if study_id = 0102491
  then end_time = (dt_enroll + 456.25);
  * YYN.N.N.YY..;
else if study_id = 0102521
  then end_time = (last_visit_date);
  * YYYYN.N.N...;
else if study_id = 0202471
  then end_time = (last_visit_date);
  * YYN.N.N.....;
else if study_id = 0202941
  then end_time = (dt_enroll + 136.875);
  * YYYNY..N...;
else if study_id = 0301531
  then end_time = (last_visit_date);
  * YYYYYN.N.N.;

* Fix end time for guys with DKs;

else if study_id = 0100361
  then end_time = (last_visit_date);
  * DDN.N.N.N.N.;
else if study_id = 0102241
  then end_time = (last_visit_date);
  * DN..........;
else if study_id = 0200201
  then end_time = (dt_enroll + 273.75);
  * NNN.N...N.D.;
else if study_id = 0200991
  then end_time = (dt_enroll + 273.75);
  * NNN.N.D.N...;
else if study_id = 0202911
  then end_time = (dt_enroll + 456.25);
  * NNN.N.N.D...;
else if study_id = 0400411
  then end_time = (last_visit_date);
  * DN.........;
* Fix end time for guys with missings for homelessness var;

```plaintext
else if event = 1 then end_time = event_time;
else if homeless_2 = . then end_time = (dt_enroll + 45.625);
else if homeless_3 = . then end_time = (dt_enroll + 136.875);
else if homeless_4 = . then end_time = (dt_enroll + 273.75);
else if homeless_5 = . then end_time = (dt_enroll + 456.25);
else if homeless_6 = . then end_time = (dt_enroll + 638.75);
else end_time = last_visit_date;
```

* Create final person time var;
* Also fix person time for incident guys with missings;

```plaintext
person_time = end_time - start_time;
if study_id = 0100221 then person_time = 0;
    * NY....N....;
if study_id = 0300361 then person_time = 0;
    * YY..YY.....;
if study_id = 0300641 then person_time = 45.625;
    * NN..YNN.N.N.;
if study_id = 0300731 then person_time = 45.625;
    * YDYY..N.D...;
```

```plaintext
if age_baseline le 24 then age_n = '18-24 yr';
if age_baseline ge 25 then age_n = '25+ yr';

pt_month = person_time/30.42;
pt_years = person_time/365.25;
```

```plaintext
label event = "Proportion Homeless"
person_time = "Person time in days"
pt_month = "Study Month"
race_inc = "Race"
age_n = "Age"
cesd = "Depressive Symptoms"
drink = "Problem Drinking"
all_drug = "Illicit Drug Use"
any_arrestp12m = "Arrested"
any_abuse = "Childhood Abuse";
```

```plaintext
format drink all_drug any_arrestp12m any_abuse
    homeless_1 homeless_1now prev_home
    yaynay.;
format cesd
```
cesdbi.;
run;

******
* Merge sets to create master set, thesis;
******;
proc sort data = incidence;
  by study_id;
run;
proc sort data = merged3;
  by study_id;
run;
data thesis;
  merge incidence merged3;
    by study_id;
run;

******
* Make sure pt vars are working correctly
******;
/*
  proc freq data = thesis;
    tables homeless_1 homeless_1now homeless_2 homeless_2now
       homeless_3 homeless_3now homeless_4 homeless_4now
       homeless_5 homeless_5now homeless_6 homeless_6now;
    where study_id = 0100221;
  run;

  proc freq data = thesis;
    tables study_id;
    where person_time = .;
  run;

  proc freq data = thesis;
    tables curr_prev*event;
    where homeless_1 = 1;
  run;

  proc freq data = thesis;
    tables curr_prev*study_id;
    where curr_prev = 1;
  run;
proc freq data = thesis;
  tables miss_pt;
run;

proc means data = thesis;
  var pt_years;
  where pt_years ge 2;
  by study_id;
run;

proc means data = thesis;
  var person_time;
run;

proc freq data = thesis;
  tables homeless_1 homeless_2
       homeless_3 homeless_4
       homeless_5 homeless_6;
  where study_id = 0300641;
run;

proc freq data = thesis;
  where person_time = 879;
run;

proc freq data = thesis;
  tables dt_enroll event_time start_time end_time
       person_time;
  where study_id = 0100361;
run;

proc univariate data = thesis;
  var person_time;
  where event = 0;
run;

proc freq data = thesis;
  tables event;
run;

proc freq data = thesis;
  tables pt_month*person_time;
run;

/*
******************************************************************************
* Program: T:\epiprojs\Sullivan_data_ident\       *
* C_Taylor_Inv\Cat -     *
*******************************************************************************/
* \thesis_var_explore.sas7bdat *
* Date: 02/28/16 *
* Programmer: Calli Taylor *
* Purpose: This program uses the merged InvolveMENt datasets for thesis work (variable exploration).
* ******************************************************;

*** Run thesis_formats FIRST;
******************************************************;
* Call in thesis merged data *
******************************************************;

$include "T:\ep1projs\Sullivan_data_ident\C_Taylor_Inv\cat\thesis_data.sas"

Exploring vars of interest, checking #s;
******************************************************;

****** Arreests; ******;

*** Baseline guys only;

  proc freq data = merged1;
    tables arrested_ever1
         arrested_ever2
         arrested_ever3
         arrested_ever4
         arrested_ever5
         arrested_ever6;
  run;

  * Baseline ? only, 310 yes, 493 no, n=803;

  proc freq data = merged1;
    tables arrested_p12m_final1
         arrested_p12m_final2
         arrested_p12m_final3
         arrested_p12m_final4
         arrested_p12m_final5
         arrested_p12m_final6;
  run;

  * Baseline ? only, of the 310 ever arrested, 86 were in p12m
   86 yes, 716 no, 1 missing, n=803;

  proc freq data = merged1;
    tables arrested_p6m1
         arrested_p6m2
         arrested_p6m3
         arrested_p6m4
         arrested_p6m5
         arrested_p6m6;
  run;

  * Baseline ? only, of the 310 ever arrested, 86 were in p6m
   86 yes, 493 no, n=803;
arrested_p6m3
arrested_p6m4
arrested_p6m5
arrested_p6m6;
run;

proc freq data = merged1;
    tables anyp6m_arrest;
run;

* Follow up ? only, 46 yes, 492 no, 265 missing, n=803;

proc freq data = merged1;
    tables drug_arrest1
drug_arrest2
drug_arrest3
drug_arrest4
drug_arrest5
drug_arrest6;
run;

*Baseline ? only, 26 yes, 287 no, 490 missing, n=803;

proc freq data = merged1;
    tables any_arrest;
run;

* 335 ever arrested, 468 never, n=803;

*** Prospective guys only;

proc freq data = merged3;
    tables arrested_ever1
arrested_ever2
arrested_ever3
arrested_ever4
arrested_ever5
arrested_ever6;
run;

* Baseline ? only, 173 yes, 383 no, n=556;

proc freq data = merged3;
    tables arrested_p12m_final1
arrested_p12m_final2
arrested_p12m_final3
arrested_p12m_final4
arrested_p12m_final5
arrested_p12m_final6;
run;

* Baseline ? only, of the 173 ever arrested, 55 were in p12m
55 yes, 501 no, n=556;
proc freq data = merged3;
  tables arrested_p6m1
      arrested_p6m2
      arrested_p6m3
      arrested_p6m4
      arrested_p6m5
      arrested_p6m6;
run;

proc freq data = merged3;
  tables anyp6m_arrest;
run;

* Follow up ? only, 45 yes, 483 no, 28 missing, n=556;

proc freq data = merged3;
  tables drug_arrest1
      drug_arrest2
      drug_arrest3
      drug_arrest4
      drug_arrest5
      drug_arrest6;
run;

* Baseline ? only, 15 yes, 208 no, 333 missing, n=556;

proc freq data = merged3;
  tables any_arrest;
run;

* 208 ever arrested, 348 never, n=556;

*** Among homeless prospective guys;

proc freq data = merged3;
  tables any_arrest;
  where any_homeless ge 1;
run;

* 36 ever arrested, 41 never, n=77;

******
Abuse;
******;

* Baseline ? only, so only need 1st visit var;

*** Among baseline guys;

proc freq data = merged1;
  tables abuse_child_physabuse1 abuse_child_sexabuse1;
run;
**Drug Use;**

*** Alcohol;

* No V2 data, ? wasn't asked at that visit?;

    proc freq data = merged3;
        tables drink_p12m1;
    run;

    proc freq data = merged3;
        tables
    run;

    proc freq data = merged3;
        tables
    run;
alcoholp6m2
alcoholp6m3
alcoholp6m4
alcoholp6m5
alcoholp6m6;
run;

proc freq data = merged3;
  tables any_alcohol;
run;

* 536 yes, 20 no, n=556;

proc freq data = merged3;
  tables any_alcohol;
  where any_homeless ge 1;
run;

* 75 yes, 2 no, n=77;

*** Cocaine;

proc freq data = merged3;
  tables cocaine_p12m1;
run;

proc freq data = merged3;
  tables cocaine_p6m2
cocaine_p6m3
cocaine_p6m4
cocaine_p6m5
cocaine_p6m6;
run;

proc freq data = merged3;
  tables any_cocaine;
run;

* 125 yes, 152 no, 279 missing n=277;

proc freq data = merged3;
  tables any_cocaine;
  where any_homeless ge 1;
run;

* 20 yes, 21 no, 36 missing, n=41;

*** Crack;

proc freq data = merged3;
  tables crack_p12m1;
run;
``` Sas
proc freq data = merged3;
  tables crack_p6m2
  crack_p6m3
  crack_p6m4
  crack_p6m5
  crack_p6m6;
run;

proc freq data = merged3;
  tables any_crack;
run;

* 15 yes, 260 no, 281 missing, n=275;

proc freq data = merged3;
  tables any_crack;
  where any_homeless ge 1;
run;

*  5 yes, 35 no, 37 missing, n=40;

*** Crystal Meth;

proc freq data = merged3;
  tables crystalmeth_p12m1;
run;

* 22 yes, 190 no, 344 missing, n=212;

proc freq data = merged3;
  tables crystalmeth_p6m2
  crystalmeth_p6m3
  crystalmeth_p6m4
  crystalmeth_p6m5
  crystalmeth_p6m6;
run;

proc freq data = merged3;
  tables any_crystal;
run;

* 44 yes, 231 no, 281 missing, n=275;

*** Downers;

proc freq data = merged3;
  tables downers_p12m1;
run;

* 44 yes, 168 no, 344 missing, n=212;
```
**proc freq data = merged3;**
   **tables**
     downers_p6m2
     downers_p6m3
     downers_p6m4
     downers_p6m5
     downers_p6m6;
   **run;**

**proc freq data = merged3;**
   **tables** any_downers;
   **run;**

* 70 yes, 204 no, 282 missing, n=274;

*** Ecstasy;***

**proc freq data = merged3;**
   **tables** ecstasy_p12m1;
   **run;**

* 62 yes, 149 no, 345 missing, n=211;

**proc freq data = merged3;**
   **tables**
     ecstasy_p6m2
     ecstasy_p6m3
     ecstasy_p6m4
     ecstasy_p6m5
     ecstasy_p6m6;
   **run;**

**proc freq data = merged3;**
   **tables** any_ecstasy;
   **run;**

* 90 yes, 188 no, 278 missing, n=278;

*** GHB;***

**proc freq data = merged3;**
   **tables** ghb_p12m1;
   **run;**

* 19 yes, 190 no, 347 missing, n=209;

**proc freq data = merged3;**
   **tables**
     ghb_p6m2
     ghb_p6m3
     ghb_p6m4
     ghb_p6m5
ghb_p6m6;
run;
proc freq data = merged3;
    tables any_ghb;
run;

* 31 yes, 243 no, 282 missing, n=274;

*** Hallucinogen;

proc freq data = merged3;
    tables hallucinogen_p12m1;
run;

* 32 yes, 177 no, 347 missing, n=209;

proc freq data = merged3;
    tables
        hallucinogen_p6m2
        hallucinogen_p6m3
        hallucinogen_p6m4
        hallucinogen_p6m5
        hallucinogen_p6m6;
run;

proc freq data = merged3;
    tables any_hallucinogen;
run;

* 53 yes, 222 no, 281 missing, n=275;

*** Heroin;

proc freq data = merged3;
    tables heroin_p12m1;
run;

* 2 yes, 206 no, 348 missing, n=208;

proc freq data = merged3;
    tables
        heroin_p6m2
        heroin_p6m3
        heroin_p6m4
        heroin_p6m5
        heroin_p6m6;
run;

proc freq data = merged3;
    tables any_heroin;
run;
* 5 yes, 269 no, 282 missing, n=274;

*** Inject;

```r
proc freq data = merged3;
  tables inject1;
run;
```

* 4 yes, 551 no, 1 missing, n=555;

```r
proc freq data = merged3;
  tables injecnptp6m2 injecnptp6m3 injecnptp6m4 injecnptp6m5 injecnptp6m6;
run;
```

```r
proc freq data = merged3;
  tables any_inject;
run;
```

* 13 yes, 543 no, n=556;

*** Weed;

```r
proc freq data = merged3;
  tables marijuana_p12m1;
run;
```

* 203 yes, 21 no, 332 missing, n=224;

```r
proc freq data = merged3;
  tables marijuana_p6m2 marijuana_p6m3 marijuana_p6m4 marijuana_p6m5 marijuana_p6m6;
run;
```

```r
proc freq data = merged3;
  tables any_marijuana;
run;
```

* 266 yes, 28 no, 262 missing, n=294;

*** Non Inject;

* Unspecified drug use, just asked if any non inject drugs in pl2m;
proc freq data = merged3;
   tables noninject_p12m1;
run;

* 223 yes, 330 no, 3 missing, n=553;

proc freq data = merged3;
   tables noninject_p6m2
       noninject_p6m3
       noninject_p6m4
       noninject_p6m5
       noninject_p6m6;
run;

proc freq data = merged3;
   tables any_noninject;
run;

* 292 yes, 264 no, n=556;

proc freq data = merged3;
   tables noninject_other_p12m1;
run;

* 10 yes, 213 no, 333 missing, n=223;

*** Painkillers;

proc freq data = merged3;
   tables painkillers_p12m1;
run;

* 59 yes, 152 no, 345 missing, n=211;

proc freq data = merged3;
   tables painkillers_p6m2
       painkillers_p6m3
       painkillers_p6m4
       painkillers_p6m5
       painkillers_p6m6;
run;

proc freq data = merged3;
   tables any_painkillers;
run;

* 86 yes, 186 no, 284 missing, n=272;

*** Poppers;

proc freq data = merged3;
   tables poppers_p12m1;
run;

* 63 yes, 147 no, 346 missing, n=210;

    proc freq data = merged3;
    tables poppers_p6m2
             poppers_p6m3
             poppers_p6m4
             poppers_p6m5
             poppers_p6m6;
    run;

    proc freq data = merged3;
    tables any_poppers;
    run;

* 101 yes, 173 no, 282 missing, n=274;

*** Special K;

    proc freq data = merged3;
    tables specialk_p12m1;
    run;

* 15 yes, 193 no, 348 missing, n=208;

    proc freq data = merged3;
    tables specialk_p6m2
             specialk_p6m3
             specialk_p6m4
             specialk_p6m5
             specialk_p6m6;
    run;

    proc freq data = merged3;
    tables any_specialk;
    run;

* 29 yes, 243 no, 284 missing, n=272;

*** Various non inject drug vars;

    proc freq data = merged3;
    tables noninject_p12m1;
    run;

* 223 yes, 330 no, 3 missing, n=553;

    proc freq data = merged3;
    tables any_noninjectp12m;
run;

* 223 yes, 3 no, 330 missing, n=226;
   proc freq data = merged3;
   tables invol_noninjectp12m;
run;

* 144 yes, 74 no, 338 missing, n=218;
   proc freq data = merged3;
   tables all_noninject_p12m;
run;

* 223 yes, 330 no, 3 missing, n=553;

**** Look at patterns of missings;
   proc format;
      value miss . = 'Missing'
                   other = 'Ok';
   run;

   proc freq data = merged3;
      tables drink_p12m1*noninject_p12m1*inject1*marijuana_p12m1
            *any_abuse*any_arrestp12m*race_inc*age_baseline/list
            missing;
      format drink_p12m1 noninject_p12m1 inject1 marijuana_p12m1
            any_abuse any_arrestp12m race_inc age_baseline miss.;
      where incident = 1;
   run;

** Among all pro. homeless, 39 are missing marijuana_p12m1, 1 is also
   missing noninject_p12m1;
** Among incident homeless, 21 are missing marijuana_p12m1, 1 is also
   missing noninject_p12m1;

   proc freq data = merged3;
      tables homeless_1 homeless_1now homeless_2 homeless_2now
            homeless_3 homeless_3now homeless_4 homeless_4now
            homeless_5 homeless_5now homeless_6 homeless_6now;
   run;

*/
/*
******
* Problem drinking vars;
******;
* None of the problem drinking vars were asked at V2;
* All reported numbers are for baseline only;
**proc freq data = thesis;**
  tables drink_cut1
drink_cut3
drink_cut4
drink_cut5
drink_cut6;
run;

* 218 yes, 306 no, 32 missing, n=524;

**proc freq data = thesis;**
  tables drink_critic1
drink_critic3
drink_critic4
drink_critic5
drink_critic6;
run;

* 89 yes, 432 no, 35 missing, n=521;

**proc freq data = thesis;**
  tables drink_guilt1
drink_guilt3
drink_guilt4
drink_guilt5
drink_guilt6;
run;

* 150 yes, 369 no, 37 missing, n=519;

**proc freq data = thesis;**
  tables drink_morning1
drink_morning3
drink_morning4
drink_morning5
drink_morning6;
run;

* 69 yes, 450 no, 37 missing, n=519;

*****;

**proc freq data = thesis;**
  tables cage
drink;
run;

* 75 yes (3-4), 442 no (0-2), 39 missing, n=517;

******
* Depressive symptoms vars;******;
* None of the depressive symptoms vars were asked at V2;
* All reported numbers are for baseline only;

```latex
proc freq data = thesis;
  tables cesd_hopeful1
       cesd_hopeful3
       cesd_hopeful4
       cesd_hopeful5
       cesd_hopeful6;
run;

* 236 no, 3 missing, n=553;

proc freq data = thesis;
  tables cesd_happy1
       cesd_happy3
       cesd_happy4
       cesd_happy5
       cesd_happy6;
run;

* 273 no, 3 missing, n=553;

proc freq data = thesis;
  tables cesd_lonely1
       cesd_lonely3
       cesd_lonely4
       cesd_lonely5
       cesd_lonely6;
run;

* 196 no, 4 missing, n=552;

proc freq data = thesis;
  tables cesd_getgoing1
       cesd_getgoing3
       cesd_getgoing4
       cesd_getgoing5
       cesd_getgoing6;
run;

* 289 no, 4 missing, n=552;

proc freq data = thesis;
  tables cesd_fearful1
       cesd_fearful3
       cesd_fearful4
       cesd_fearful5
       cesd_fearful6;
run;

* 373 no, 3 missing, n=553;
```
proc freq data = thesis;
    tables cesd_effort1
    cesd_effort3
    cesd_effort4
    cesd_effort5
    cesd_effort6;
run;
* 229 no, 4 missing, n=552;

proc freq data = thesis;
    tables cesd_depressed1
    cesd_depressed3
    cesd_depressed4
    cesd_depressed5
    cesd_depressed6;
run;
* 326 no, 5 missing, n=551;

proc freq data = thesis;
    tables cesd_concentrate1
    cesd_concentrate3
    cesd_concentrate4
    cesd_concentrate5
    cesd_concentrate6;
run;
* 245 no, 5 missing, n=551;

proc freq data = thesis;
    tables cesd_bother1
    cesd_bother3
    cesd_bother4
    cesd_bother5
    cesd_bother6;
run;
* 355 no, 5 missing, n=551;

proc freq data = thesis;
    tables cesd_badsleep1
    cesd_badsleep3
    cesd_badsleep4
    cesd_badsleep5
    cesd_badsleep6;
run;
* 239 no, 4 missing, n=552;
****;
proc freq data = thesis;
    tables cesd_total
    cesd;
run;

* 141 yes (10+), 394 no (0-9), 21 missing, n=535;

***********************************************************************;
* Program: T:\epiprojs\Sullivan_data_ident\C_Taylor_Inv\cat\thesis_analyses.sas7bdat *
* Date: 04/13/16 *
* Programmer: Calli Taylor *
* Purpose: This program uses the InvolveMENt datasets for thesis analyses. *
***********************************************************************;

***********************************************************************;
* Call in thesis merged data *
***********************************************************************;

%include "T:\epiprojs\Sullivan_data_ident\C_Taylor_Inv\cat\thesis_data.sas";

*Run thesis_formats FIRST;

***************************************************************************************;
* It's analysis tiimmmeeee;
***************************************************************************************;

*****
* Final data exploration;
*****;

proc freq data = merged1;
    tables race_inc
    any_abuse
    any_arrestp12m
    all_drug
    drink
    cesd;
run;

proc freq data = thesis;
    tables race_inc
    any_abuse
    any_arrestp12m
    all_drug
    drink
run;
* Figure out best age groups;

```sas
proc univariate data = thesis;
  var age_baseline;
run;
```

```sas
proc sort data = thesis;
  by race_inc;
run;
```

```sas
proc boxplot data = thesis;
  plot age_baseline*race_inc;
run;
```

```sas
proc freq data = thesis;
  tables race_inc age_n;
run;
```

* Age_n, 18-24 and 25+;

```sas
proc univariate data = thesis;
  var pt_years pt_month;
run;
```

```sas
proc corr data = thesis plots (maxpoints = none) = matrix (histogram);
  var pt_years
    pt_month
    race_inc
    age_baseline
    any_abuse
    any_arrestp12m
    all_drug
    drink
cesd;
run;
```

******;
* Table 1, prevalence;
******;

* Need to include test stats for racial differences in prevalence;
* Overall (n=803), 56.5% black, 43.5% white;
* Pro. only (n=556), 46% black, 54% white;

```sas
proc freq data = merged1;
  tables race_inc;
run;
```
* Historical, current, and all prevalence by race, HIV- and HIV+ (n=803) at baseline;

```plaintext
proc freq data = merged1 order = formatted;
   tables race_inc*homeless_1/chisq;
run;

proc freq data = merged1 order = formatted;
   tables race_inc*homeless_1now/fisher;
run;

proc freq data = merged1 order = formatted;
   tables race_inc*prev_home/chisq;
run;
```

* Historical, current, and all prevalence by race, HIV- only (n=556) at baseline;

```plaintext
proc freq data = thesis order = formatted;
   tables race_inc*homeless_1/chisq;
run;

proc freq data = thesis order = formatted;
   tables race_inc*homeless_1now/fisher;
run;

proc freq data = thesis order = formatted;
   tables race_inc*prev_home/chisq;
run;
```

* Historical, current, and all prevalence by race, HIV+ only (n=247) at baseline;

```plaintext
proc freq data = merged1 order = formatted;
   tables race_inc*homeless_1/chisq;
where prospective_pt ne 1 or baseline_hiv ne 1;
run;

proc freq data = merged1 order = formatted;
   tables race_inc*homeless_1now/fisher;
where prospective_pt ne 1 or baseline_hiv ne 1;
run;

proc freq data = merged1 order = formatted;
   tables race_inc*prev_home/chisq;
where prospective_pt ne 1 or baseline_hiv ne 1;
run;
```

******;
* Table 2, baseline characteristics;
******;
proc freq data = thesis;
  tables race_inc*age_n/chisq;
run;

proc freq data = thesis order = formatted;
  tables race_inc*all_drug/chisq;
run;

proc freq data = thesis order = formatted;
  tables race_inc*drink/chisq;
run;

proc freq data = thesis order = formatted;
  tables race_inc*any_abuse/chisq;
run;

proc freq data = thesis order = formatted;
  tables race_inc*any_arrestp12m/chisq;
run;

proc freq data = thesis order = formatted;
  tables race_inc*cesd/chisq;
run;

********;
* Table 3, for incidence rates and rate ratios;
********;

proc freq data = thesis;
  tables event;
run;

proc univariate data = thesis;
  var person_time;
run;

proc univariate data = thesis;
  var pt_years;
run;

*** Race;

proc freq data = thesis;
  tables race_inc*event;
run;

proc univariate data = thesis;
  var pt_years;
class race_inc;
run;

*** Age;
proc freq data = thesis;
tables age_n*event;
run;

proc univariate data = thesis;
var pt_years;
class age_n;
run;

*** Race and age;
proc freq data = thesis;
tables race_inc*age_n;
where event = 1;
run;

proc univariate data = thesis;
var pt_years;
class race_inc age_n;
run;

*** Arrested;
proc freq data = thesis;
tables any_arrestp12m;
where event = 1;
run;

proc univariate data = thesis;
var pt_years;
class any_arrestp12m;
run;

*** Drug;
proc freq data = thesis;
tables all_drug;
where event = 1;
run;

proc univariate data = thesis;
var pt_years;
class all_drug;
run;

*** Drink;
*** Depression;

proc freq data = thesis;
  tables cesd;
  where event = 1;
run;

proc univariate data = thesis;
  var pt_years;
  class cesd;
run;

*** Abuse;

proc freq data = thesis;
  tables any_abuse;
  where event = 1;
run;

proc univariate data = thesis;
  var pt_years;
  class any_abuse;
run;

********;
* Survival analysis;
********;

**** Nonparametric, Kaplan Meier plots (Figures 1-4);

/*
* To find the sas template for survival analysis, saved it in thesis_KM_template,
  can edit it there before running each plot;
ods trace on;
ods graphics on;
proc lifetest data = thesis plots = survival (nocensor failure test);
  time pt_month*event(0);
run;
ods trace off;
proc template;
run;
proc template;
run;
*/

**** Alter/Run thesis_KM_template before each plot;

* Overall;
  
  proc lifetest data = thesis plots = survival (nocensor failure test);
    time pt_month*event(0);
  run;

* By race;
  
  proc lifetest data = thesis plots = survival (nocensor failure test);
    strata race_inc;
    time pt_month*event(0);
  run;

* By age;
  
  proc lifetest data = thesis plots = survival (nocensor failure test);
    strata age_n;
    time pt_month*event(0);
  run;

* By race and age;
  
  proc lifetest data = thesis plots = survival (nocensor failure test);
    strata age_n race_inc;
    time pt_month*event(0);
  run;

**** Cox proportional hazards models, Table 4;

* All covariates, race*age included, Model #1;
**proc phreg** data = thesis plots = survival;
  class race_inc age_n any_abuse any_arrest12m all_drug
drink cesd;
  model person_time*event(0) = race_inc age_n any_abuse
any_arrest12m all_drug drink cesd race_inc | age_n /
risklimits;

  hazardratio 'Effect of change in race by age' race_inc /at
  (age_n = all);
run;

* All covariates, no interaction terms, Model #2;*

**proc phreg** data = thesis plots = survival;
  class race_inc age_n any_abuse any_arrest12m all_drug
drink cesd;
  model person_time*event(0) = race_inc age_n any_abuse
any_arrest12m all_drug drink cesd / risklimits;
run;

* Dropped problem drinking due to amount of missings, Model #3;*

**proc phreg** data = thesis plots = survival;
  class race_inc age_n any_abuse any_arrest12m all_drug
cesd;
  model person_time*event(0) = race_inc age_n any_abuse
any_arrest12m all_drug cesd / risklimits;
run;

* Age, race, age | race, and depression;*

**proc phreg** data = thesis plots = survival;
  class race_inc age_n cesd;
  model person_time*event(0) = race_inc age_n cesd race_inc |
age_n / risklimits;
run;

* Age, race, and depression only, Model #4;*

**proc phreg** data = thesis plots = survival;
  class race_inc age_n cesd;
  model person_time*event(0) = race_inc age_n cesd / risklimits;
run;

* Race and age only, Model #5;*

**proc phreg** data = thesis plots = survival;
  class race_inc age_n;
  model person_time*event(0) = race_inc age_n / risklimits;
run;

* Race only, Model #6;

proc phreg data = thesis plots = survival;
  class race_inc;
  model person_time*event(0) = race_inc / risklimits;
run;

********************************************************************;
* Program: T:\epiprojs\Sullivan_data_ident\ 
  C_Taylor_Inv\Cat 
  \thesis_KM_template.sas7bdat *
* Date: 07/13/16 *
* Programmer: Calli Taylor (but mostly SAS) *
* Purpose: This program uses the SAS ods template *
  for KM failure curve graphs to alter the produced *
  plots for thesis work. *
********************************************************************;

******
* Kaplan Meier failure curve template; ******

proc template;
dynamic NStrata xName maxTime plotAtRisk plotCensored plotCL plotHW plotEP labelCL labelHW labelEP yMin xtickVals xtickValFitPol method StratumID classAtRisk plotTest GroupName Transparency rowWeights SecondTitle TestName pValue _byline_ _bytitle_ _byfootnote_;
BeginGraph;
if (NSTRATA=1)
  if (EXISTS(STRATUMID))
    entrytitle " " ;
  else
    entrytitle " " ;
endif;
if (PLOTATRISK=1) entrytitle "With Number of Subjects at Risk" / textattrs=GRAPHVALUETEXT;
endif;
layout overlay / xaxisopts=(shortlabel=XNAME offsetmin=.05 linearopts=(viewmax=MAXTIME tickvaluelist=XTICKVALS tickvaluefitpolicy=XTICKVALFITPOL))
  yaxisopts=(label="Proportion Homeless" shortlabel="Failure"
            linearopts=(viewmin=0 viewmax=.5 tickvaluelist=(0 .1 .2 .3 .4 .5)))
  ;
if (PLOTHW=1 AND PLOTEP=0)
bandplot LimitUpper=eval (1-HW_LCL) LimitLower=eval (1-
HW_UCL) x=TIME / displayTail=false modelname=
"Failure" fillattrs=GRAPHCONFIDENCE name="HW"
legendlabel=LABELHW;
endif;
if (PLOTHW=0 AND PLOTEP=1)
bandplot LimitUpper=eval (1-EP_LCL) LimitLower=eval (1-
EP_UCL) x=TIME / displayTail=false modelname=
"Failure" fillattrs=GRAPHCONFIDENCE name="EP"
legendlabel=LABELEP;
endif;
if (PLOTHW=1 AND PLOTEP=1)
bandplot LimitUpper=eval (1-HW_LCL) LimitLower=eval (1-
HW_UCL) x=TIME / displayTail=false modelname=
"Failure" fillattrs=GRAPHCONFIDENCE datatransparency=.55
name="HW" legendlabel=LABELHW;
bandplot LimitUpper=eval (1-EP_LCL) LimitLower=eval (1-
EP_UCL) x=TIME / displayTail=false modelname=
"Failure" fillattrs=GRAPHDATA2 datatransparency=.55
name="EP" legendlabel=LABELEP;
endif;
if (PLOTCL=1)
if (PLOTHW=1 OR PLOTEP=1)
bandplot LimitUpper=eval (1-SDF_LCL) LimitLower=eval (1-
SDF_UCL) x=TIME / displayTail=false modelname =
"Failure" display=(outline) outlineattrs=GRAPHPREDICTIONLIMITS
name="CL" legendlabel=LABELCL;
else
bandplot LimitUpper=eval (1-SDF_LCL) LimitLower=eval (1-
SDF_UCL) x=TIME / displayTail=false modelname =
"Failure" fillattrs=GRAPHCONFIDENCE name="CL"
legendlabel=LABELCL;
endif;
endif;
stepplot y=eval (1-SURVIVAL) x=TIME / name="Failure"
rolename=(_tip1=ATRISK _tip2=EVENT) tiplabel=(_y=
"Failure Probability" _tip1="Number at Risk"
_tip2="Observed Events") tip=(x y _tip1 _tip2)
legendlabel = "Failure";
if (PLOTCENSORED) scatterplot y=eval (1-CENSORED) x=TIME
/ tiplabel=(y="Failure Probability")
markerattrs=(symbol=plus) name="Censored" legendlabel="Censored";
endif;
if (PLOTCL=1 OR PLOTHW=1 OR PLOTEP=1)
discretelegend "Censored" "CL" "HW" "EP" /
location=outside halign=center;
else
if (PLOTCENSORED=1)
discretelegend "Censored" / location=inside autoalign=(topleft bottomright);
endif;
endif;
if (PLOTATRISK=1)
innermargin / align=bottom;
axistable x=TATRISK value=ATRISK / display=(label)
valueattrs=(size=7pt);
endinnermargin;
endif;
endlayout;
else
entrytitle " ";
if (EXISTS(SECONDTITLE))
entrytitle SECONDTITLE / textattrs=GRAPHVALUETEXT;
endif;
layout overlay / xaxisopts=(shortlabel=XNAME offsetmin=.05
linearopts=(viewmax=MAXTIME tickvaluelist=XTICKVALS
tickvaluefitpolicy=XTICKVALFITPOL))
yaxisopts=(label="Proportion Homeless" shortlabel="Failure"
linearopts=(viewmin=0 viewmax=.5 tickvaluelist=(0 .1
.2 .3 .4 .5)));
if (PLOTHW=1) bandplot LimitUpper=eval (1-HW_LCL)
LimitLower=eval (1-HW_UCL) x=TIME / displayTail=false
group=STRATUM index=STRATUMNUM modelname="Failure"
datatransparency=Transparency;
endif;
if (PLOTEP=1)
bandplot LimitUpper=eval (1-EP_LCL) LimitLower=eval
(1-EP_UCL) x=TIME / displayTail=false group=STRATUM
index=STRATUMNUM modelname="Failure"
datatransparency=Transparency;
endif;
if (PLOTCL=1)
if (PLOTHW=1 OR PLOTEP=1)
bandplot LimitUpper=eval (1-SDF_LCL) LimitLower=eval
(1-SDF_UCL) x=TIME / displayTail=false group=STRATUM
index=STRATUMNUM modelname="Failure"
display=(outline) outlineattrs=(pattern=ShortDash);
else
bandplot LimitUpper=eval (1-SDF_UCL) LimitLower=eval
(1-SDF_LCL) x=TIME / displayTail=false group=STRATUM
index=STRATUMNUM modelname="Failure"
datatransparency=Transparency;
endif;
endif;
stepplot y=eval (1-SURVIVAL) x=TIME / group=STRATUM
index=STRATUMNUM name="Failure"
rolename=(tip1=ATRISK _tip2=EVENT)
tiplabel=(y="Failure Probability" _tip1="Number at
Risk" _tip2="Observed Events")
tip=(x y _tip1 _tip2);
if (PLOTCENSORED=1)
scatterplot y=eval (1-CENSORED) x=TIME /
tiplabel=(y="Failure Probability") group=STRATUM
index=STRATUMNUM markerattrs=(symbol=plus);
endif;
if (PLOTATRISK=1)
innermargin / align=bottom;
axistable x=TATRISK value=ATRISK / display=(label)
valueattrs=(size=7pt) class=CLASSATRISK colorgroup
=CLASSATRISK;
endinnermargin;
```sas
endif;
DiscreteLegend "Failure" / title=GROUPNAME
location=outside;
if (PLOTCENSORED=1)
if (PLOTTEST)
layout gridded / rows=2 autoalign=(TOPLEFT BOTTOMRIGHT BOTTOM TOP) border=true BackgroundColor=
GraphWalls:Color Opaque=true;
entry "+ Censored";
if (PVALUE < .0001) entry TESTNAME " p " eval (PUT(PVALUE, PVALUE6.4));
  else
    entry TESTNAME " p=" eval (PUT(PVALUE, PVALUE6.4));
  endif;
endlayout;
else
layout gridded / rows=1 autoalign=(TOPLEFT BOTTOMRIGHT BOTTOM TOP) border=true BackgroundColor=
GraphWalls:Color Opaque=true;
entry "+ Censored";
endlayout;
endif;
else
  if (PLOTTEST=1)
    layout gridded / rows=1 autoalign=(TOPLEFT BOTTOMRIGHT BOTTOM TOP) border=true BackgroundColor=
GraphWalls:Color Opaque=true;
    if (PVALUE < .0001) entry TESTNAME " p " eval (PUT(PVALUE, PVALUE6.4));
      else
        entry TESTNAME " p=" eval (PUT(PVALUE, PVALUE6.4));
      endif;
    endlayout;
  endif;
endif;
endlayout;
if (_BYTITLE_)
  entrytitle _BYLINE_ / textattrs=GRAPHVALUETEXT;
else
  if (_BYFOOTNOTE_)
    entryfootnote halign=left _BYLINE_;
  endif;
endif;
EndGraph;
end;

******************************************************************************
*                        Involve[men]t Study                                 *
*                Formats and Labels for SAS Data sets                       *
******************************************************************************;
options pagesize=66 linesize=120 pageno=1 missing=' ' date FORMCHAR="|---|+|---+|=|/\<>";
* Programmer : Nicole Luisi
* Purpose : Program to set values and labels for all Involve\[men\]t SAS data sets.
* Program Location: T:\EpiProjs\MSM cohort\SAS\data_mgt
* Program Name: SQL_Formats.sas

* Set pathways;
libname library 'T:\epiprojs\Sullivan_data_ident\C_Taylor_Inv\cat';

proc format library = library;
*********************** VALUES ***********************;
* GENERAL ;
value cesdbi 1 = 'Yes'
2 = 'No'
. = 'Missing';

value yaynay 1 = '1 Yes'
0 = '2 No'
. = 'Missing';

value DRUG 0 = 'Didnt Use'
1 = 'Have Used';

value BIN 0 = 'No'
1 = 'Yes';

value YNF 0 = 'No'
1 = 'Yes' ;

value VISIT 1 = 'Baseline'
2 = 'Month 3'
3 = 'Month 6'
4 = 'Month 12'
5 = 'Month 18'
6 = 'Month 24'
11 = 'Baseline - REDO'
12 = 'Month 3 - REDO'
13 = 'Month 6 - REDO'
14 = 'Month 12 - REDO'
15 = 'Month 18 - REDO'
16 = 'Month 24 - REDO';

value RACE_INC 4 = 'White/Caucasian'
1 = 'Black/African American'
7 = 'Other';

value Binary_prefernot 1="Yes"
0 = "No"
2 = "Not Sure"
7 = "Prefer not to answer"
9 = "Don't know"

value arrested_jaildays
1 = "Under 30 days"
2 = "Over 30 days"
9 = "Don't know"

value drugfreq
0 = "Didn't use"
1 = "More than once a day"
2 = "Once a day"
3 = "More than once a week"
4 = "Once a week"
5 = "More than once a month"
6 = "Once a month"
7 = "Less than once a month"
8 = "Refused to Answer"

value LikertFive
5 = 'Always'
4 = 'Very Often'
3 = 'Sometimes'
2 = 'Rarely'
1 = 'Never'

value CESD
0 = 'Rarely or none of the time'
1 = 'Some or little of the time'
2 = 'Occasionally or a moderate amount of the time'
3 = 'Most or all of the time'

value noninject_often
1 = "More than once a day"
2 = "Once a day"
3 = "More than once a week"
4 = "Once a week"
5 = "More than once a month"
6 = "Once a month"
7 = "Less than once a month"
8 = "Refused to answer"