Emil Kosa Jr., *Cloverleaf Confusion*, 1950s. Courtesy of The Hilbert Collection.

Early Intervention to Prevent Persistent Homelessness

Predictive Models for Identifying Unemployed Workers and Young Adults who become Persistently Homeless

March 2019

ECONOMIC ROUNDTABLE

Knowledge for the Greater Good

Early Intervention to Prevent Persistent Homelessness

Predictive Models for Identifying Unemployed Workers and Young Adults who become Persistently Homeless

March 2019

Economic Roundtable
Halil Toros
Daniel Flaming
Patrick Burns

Underwritten by
The John Randolph Haynes and Dora Haynes Foundation

Report available at: www.economicrt.org

This report can be downloaded from the Economic Roundtable web site:
www.economicrt.org

Follow us on Twitter @EconomicRT
Like us on Facebook.com/EconomicRT

Table of Contents

Unsheltered homeless in the Great Depression: San Gabriel Canyon squatters. Security Pacific National Bank Collection, courtesy of the Los Angeles Public Library.

Executive Summary

Thousands of persistently homeless Americans are turning sidewalks of U.S. cities into camps for internally displaced persons. In major west coast metropolitan areas, the number of long-term homeless needing housing far exceeds the available housing supply, making it difficult to move persistently homeless individuals off of the streets. One of the most promising approaches to reducing these numbers lies in early identification and quick, effective intervention to help those most likely to become persistently homeless.

This report presents two predictive screening models for intervening early to help individuals who would otherwise become persistently homeless. The first tool identifies the eight percent of low-wage workers who become persistently homeless after losing their jobs. The second tool identifies the eight percent of youth receiving public assistance who become persistently homeless in the first three years of adulthood.

A majority of people entering homelessness over the course of a year make rapid exits, often with little help, but roughly two-fifths become persistently homeless. These screening tools are valuable for identifying individuals who will remain stuck in homelessness. The screening tools are in the public domain. The factors used in the tools (parameter estimates) are shown in *Tables A-4* and *A-7* in the *Appendix*.

It is reasonable to use the tools in metropolitan areas throughout the United States, based on the large and broadly representative study population used to develop the tools, and the scarcity of comparable information for other regions. The study population includes nearly everyone who was homeless during fifteen years, a total of over one million people, in the most populous county in the United States.

The tools can be reconfigured to use locally available data and still retain a high level of accuracy, provided that key attributes of individuals are addressed. This includes demographic characteristics, homeless and employment histories, and use of services provided by the health, behavioral health, social service, and justice systems.

Using Predictive Analytic Models to Guide Homeless Interventions

Because it is hard to differentiate newly homeless individuals who will make rapid exits from those who will remain stuck in homelessness, the prevailing service delivery model calls for "progressive engagement." Progressively more help is given as individuals remain homeless longer. If individuals become chronically homeless, they are offered permanent supportive housing, if units are available.

Progressive engagement is pragmatic, but it gives rise to two problems:

1. The longer people are homeless, the worse their problems become, making it more difficult and expensive to stably house them.

2. The flow of people into long-term homelessness is not reduced, so there is growing demand for the most expensive homeless intervention, permanent supportive housing, and meeting this demand is challenging.

Predictive analytic models can distinguish accurately between different types of homelessness and predict future outcomes. This can improve both the efficiency and effectiveness of homeless interventions. For example:

1. Predictive models can provide a fair, objective system for prioritizing who gets to be housed based on likely duration of homelessness or public costs in future years.
2. Predictive models can identify newly homeless individuals who are likely to become persistently homeless so they can be targeted for early interventions that will help them escape homelessness with less distress and public cost.

In addition to housing chronically homeless individuals, the most promising strategy for combating homelessness is to have tools for differentiating level of need among newly homeless individuals and to intervene early with intensive help for individuals who are likely to become persistently homeless. The pay-offs are:

1. Interventions can be provided that are specifically tailored to meeting the needs of discrete high-risk subpopulations.
2. Small reductions of the flow of people into chronic homelessness will have a large impact on reducing the number of people who are chronically homeless.
3. Individuals will have far less social, economic, legal, and medical damage in their lives, making it more feasible and less costly to help them become stably housed.

Historic pictures are used to illustrate this report. These pictures make two points. First, homelessness is not new in Los Angeles, what is new is the number of people living without shelter. Second, we have responded successfully to homelessness in the past by providing housing and jobs.

Workers Who Lose Their Jobs and Become Persistently Homeless

All low-wage workers face some level of risk that they will become persistently homeless if they lose their jobs, but this risk is disproportionately high for workers who are African American, male and single. It is important that screening to identify unemployed workers who are likely to become persistently homeless be carried out in ways that effectively reach these groups with especially high-risks.

Over the course of homelessness, the widening monthly cost gap for local public services and higher cumulative costs for workers who become persistently homeless provides financial justification for a comprehensive package of re-employment services to avoid higher long-term costs. The predictive analytic tool described in this report can target high-risk workers for a package of re-employment services as soon as they lose their jobs, and often before they become homeless.

Some high-risk workers have barriers to employment resulting from substance abuse and involvement in the criminal justice system. This indicates that some need behavioral health services to overcome substance abuse problems as well as legal services to expunge or lessen their criminal justice records.

One quarter of high-risk workers are part of a family unit and one third are homeless before they lose their jobs. This indicates that some workers need affordable child care and many workers need affordable transitional housing.

Almost one third of workers who become persistently homeless have held down jobs despite having limiting physical or mental conditions. These disabilities become much more frequent during post-unemployment homelessness. These workers are likely to have better employment and job retention prospects if they receive health care support in treating and managing their conditions. These conditions most frequently involve back, joint and arthritic problems. These workers will benefit from finding work in occupations that are less physically demanding than their previous jobs.

Workers who become persistently homeless often have histories of job turnover, under-employment and low earnings. This indicates that many high-risk workers need education and training that will enable them to compete for better jobs. They may also need temporary housing and wage subsidies to encourage employers to give them an opportunity to demonstrate their capabilities.

Young Adults Who Become Persistently Homeless

The young adult screening tool is designed to identify the eight percent of young adults receiving public benefits who will become persistently homeless within three years.

Youth who become persistently homeless are far more likely to be solitary and not connected to a family unit. Youth who experienced homelessness in their six years preceding adulthood were more than three times as likely to be homeless as young adults than those who had not previously been homeless. The risk of persistent homelessness is especially high for:

- African American youth

- Youth who have been in the foster care system
- Youth who were homeless as children
- Youth who are homeless when they enter adulthood
- Youth who have been incarcerated

It is important that screening to identify young adults who are likely to become persistently homeless be carried out in ways that effectively reach these groups with especially high-risks.

Substance abuse problems increase the likelihood of justice system encounters and are much more prevalent among youth who are persistently homeless. Many high-risk young adults need behavioral health services to overcome substance abuse problems, and some need legal services to expunge or lessen their criminal justice records.

Only five percent of the young adult population spent time in the foster care system, but 13 percent of those who were persistently homeless had been in foster care.

The enactment of California Assembly Bill 12 in 2012 extended foster care services until youth are 21 years old. It has improved but not eliminated the problem of youth homelessness. Youth who were eligible for extended foster care services under AB 12 had better outcomes – 16 percent of these youth experienced persistent homelessness compared to 24 percent of older foster youth who emancipated into adulthood when they were 18 years old, before the bill took effect.

Disabilities emerged rapidly among young adults who were homeless – a quarter of persistently homeless youth had persistent disabilities at the end of the three-year study window. The largest share of these disabilities were for mental conditions. Effective early intervention for young adults who are on a path toward persistent homelessness can reduce the rapid emergence of long-term physical and mental disabilities that result from continued homelessness.

Persistently homeless youth have higher employment rates but lower earnings than their peers who are not stuck in homelessness. This demonstrates a strong drive to earn enough money to pay for housing but little success in obtaining sustaining employment. Many high-risk young adults need human capital investments in the form of education and training that will enable them to compete for better jobs. They may also need wage subsidies to encourage employers to give them an opportunity to demonstrate their capabilities.

Young adults who become persistently homeless often have histories of social disconnection, high levels of effort to find employment but low earnings, and behavioral health needs. This indicates that many high-risk young adults need education and training that will enable them to compete for better jobs, and behavioral health services. They may also need affordable housing and wage subsidies to encourage employers to give them an opportunity to demonstrate their capabilities.

Public Costs

Individuals who become persistently homeless use more public services and have far higher public costs than their peers who do not become homeless. These costs are ongoing and increase as individuals become older.

Health care costs were five times higher for persistently homeless workers and four times higher for persistently homeless youth than for their counterparts who did not become homeless.

Justice system costs were nine times higher for persistently homeless workers and seven times higher for persistently homeless youth than for their counterparts who did not become homeless.

Using predictive screening tools to identify high-risk individuals and intervene early before they become persistently homeless can help them avoid hardship and help the public avoid continuing high costs from ongoing, intensive and increasing use of local services.

Conclusions

Both predictive models are very accurate and particularly strong when using high probability cutoff levels for targeting high-risk individuals. A key strength of the models is that the accuracy of predictions was validated using three years of post-prediction data. Another key strength is that the models are transparent and identify distinctive attributes of high-cost individuals. The results confirm that local public costs for targeted individuals are likely to be high and to increase over time.

The tools are particularly useful for prioritizing unemployed workers and young adults for services because each individual who is screened is given a probability of becoming persistently homeless. Prioritizing individuals for access to early, comprehensive interventions is important because the resources that are most effective for preventing homelessness, including subsidized housing and employment, are scarce in relation to the demand for those resources.

The purpose of the models is to target individuals for additional help. Unlike models used to predict credit rating or justice system outcomes that have punitive consequences, the consequences for individuals targeted by these models are beneficial.

The optimal probability cutoff level for individuals who will be targeted for services is not simply an empirical decision but is influenced by resource availability and longer term cost avoidance. Greater program capacity for helping unemployed workers obtain new jobs and for helping young adults make a successful transition into adulthood can increase the percent targeted for help. Longer term public cost avoidance also should be considered in deciding on funding levels for delivery of these targeted services.

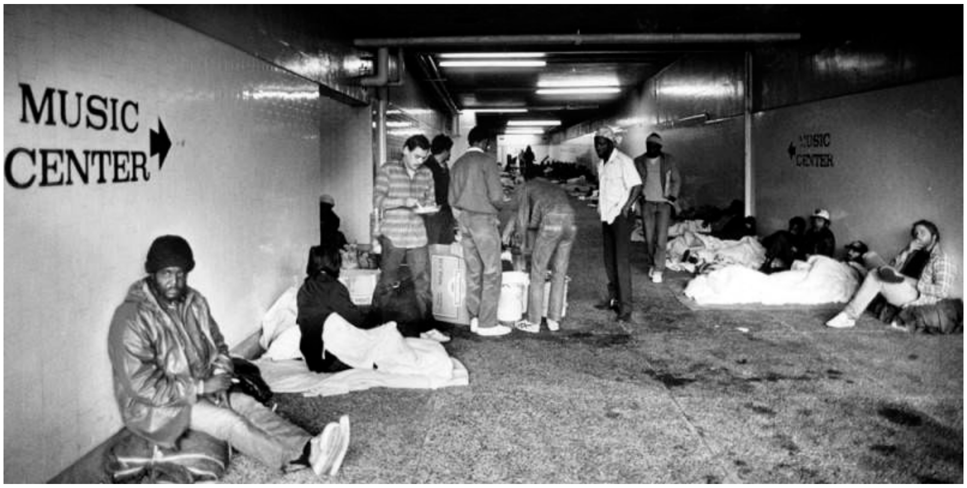
Both models are system-based tools. Depending on the population targeted, they require information about healthcare, justice system involvement, foster care, employment, homeless history, and demographics that is most readily available from the records of public agencies. Cooperation of public agencies is valuable for providing the data required for the tools.

Because of the level of effort required to obtain and integrate the necessary data, the most efficient use of the tools is regular, ongoing system-wide screening of linked records. Screening clients individually is a fallback option. By using either system-wide or person-by-person screening to predict how likely each person is to become persistently homeless, it is possible to prioritize individuals for access to the scarce supply of housing and services.

Because the tools do not correctly identify all high-risk individuals, the screening process should include an option to override the probability score based on the judgment of service providers. Allowing overrides permits service providers to adapt to changing populations and conditions and to be responsive to unique circumstances.

The descriptive information in this report and the factors used in the predictive models provide extensive information about the characteristics and needs of individuals who become persistently homeless. This information identifies needs that should be addressed but it does not define the program models for addressing those needs. Programs should be structured using evidence-based findings about best practices for helping unemployed workers obtain sustaining employment and helping high-risk young adults make a successful transition to adulthood.

The strong validation results for these models show that it is possible to develop many other predictive models that will target other distinct homeless populations for specific types of interventions. Each model is likely to target only a narrow segment of the overall homeless population because discrete population groups with distinctive attributes are needed to produce accurate predictive results. An updated typology of homelessness that breaks out distinct homeless trajectories will be valuable for mapping the full range of groups that should be targeted for interventions that will minimize the harm, cost and duration of homelessness.



Homeless in Civic Center tunnel. By Anne Knudsen, Herald Examiner Collection, 1986. Courtesy of Los Angeles Public Library.

Population Overview

Overview

There is a solution to every individual's problems but there are no mass solutions. Differing durations of homelessness point to differing barriers to becoming stably housed and differing solutions. A large population enters homelessness over the course of a year, but only a minority confronts barriers to escaping homelessness so severe that they remain homeless more than a year.

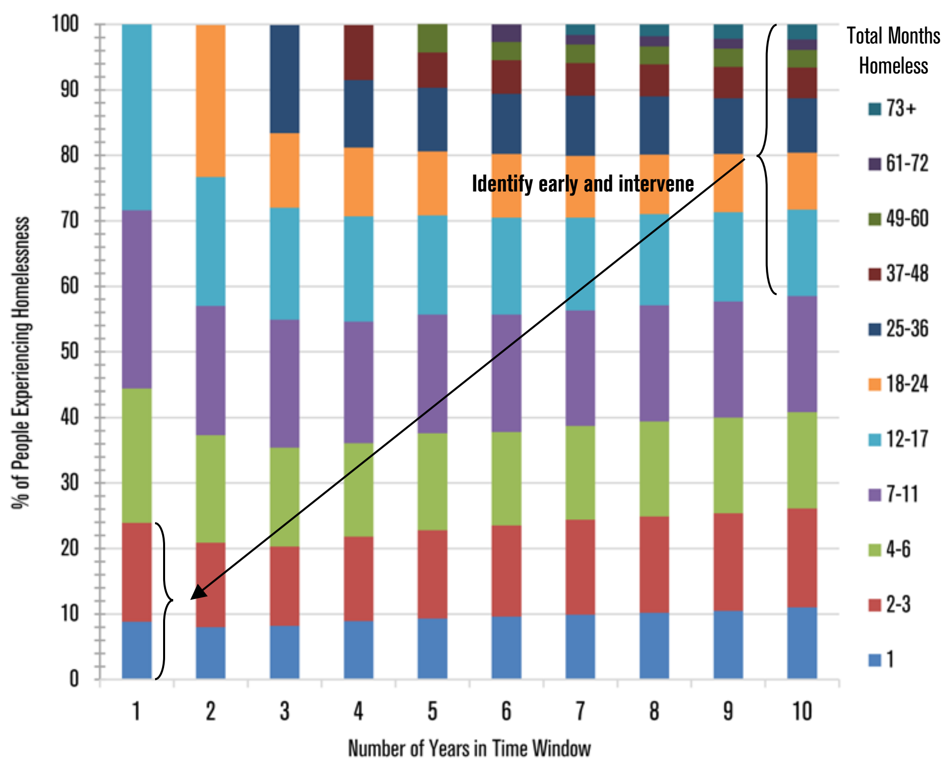
In addition to housing chronically homeless individuals, the most promising strategy for combating homelessness is to have tools for differentiating level of need among newly homeless individuals and to intervene early with intensive help for individuals who are likely to become persistently homeless.

A breakout of the different durations of homelessness for people who were homeless anytime within a 10-year time window is shown in *Figure 1*. This profile of time spent homeless is based on linked administrative records that provide fifteen years of history for over one-million residents of Los Angeles County who experienced homelessness. The source data is described in the text box on the following page and the *Methodology* chapter.

42% of people who become homeless over a two-year period are homeless for 12 or more months.

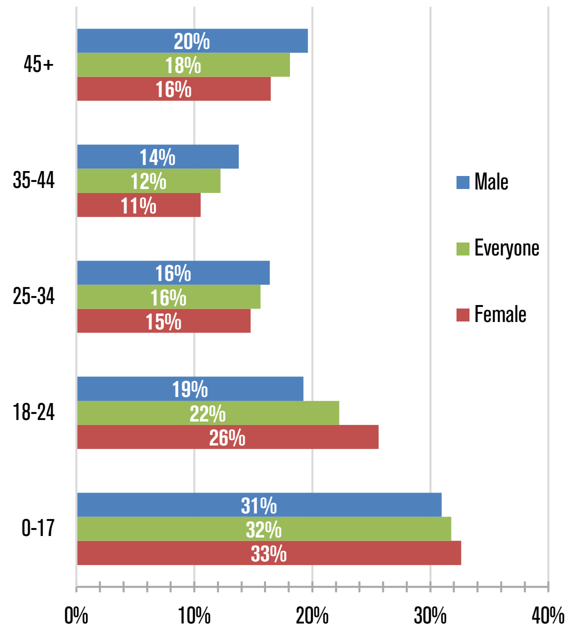
Using just a one-year time window, we see that only 28 percent of individuals who experienced homelessness were homeless for all of the year. However, this narrow window leaves out time spent homeless during

Figure 1: Total Months of Homelessness for Everyone who Experiences Homelessness during Intervals of 1 to 10 Years



the preceding year as well as recurrent episodes of homelessness in following years. When we expand the time window to two or more years, 42 percent of the total population that experiences homelessness is homeless for 12 or more months – they are persistently homeless.

Figure 2: Age when First Homeless



It is both beneficial for individuals who will go on to become persistently homeless and in the public interest to identify these high-risk individuals as soon as they become homeless and intervene immediately to support them in becoming stably housed before they are impacted by the problems that accompany protracted homelessness.

Early identification of high-risk individuals supports a form of progressive engagement in which more intensive interventions that otherwise would have been deferred until after individuals have been shown to be long-term homeless can be deployed immediately. Early intervention for high-risk individuals is important because the longer people remain homeless, the

Data Description

The administrative records used for this study include over one-million residents of Los Angeles County who were homeless sometime within a 15-year window. These individuals received some type of public benefits during this period: Medi-Cal, food stamps/SNAP, CalWORKs cash aid, or General Relief cash aid.

Individuals were counted as being homeless if they did not have a place of their own to sleep. This was based on using the address of an office of the Los Angeles County Department of Public Social Services as their mailing address. This indicated that they did not have a home address of their own.

The definition of homelessness used in this report includes individuals who are couch surfing, which is broader than HUD's criteria of sleeping in a place not meant for human habitation. Persistently homeless individuals were homeless more than once within three years. This group is not limited to individuals who also have disabilities, so it is broader than HUD's criteria for chronic homelessness.

The screening tools use all of the records that fit each of the two target populations and have benchmark dates for becoming unemployed or entering adulthood within a ten-year time window that provides three years of pre-benchmark historical information and three years of post-benchmark follow-up information.

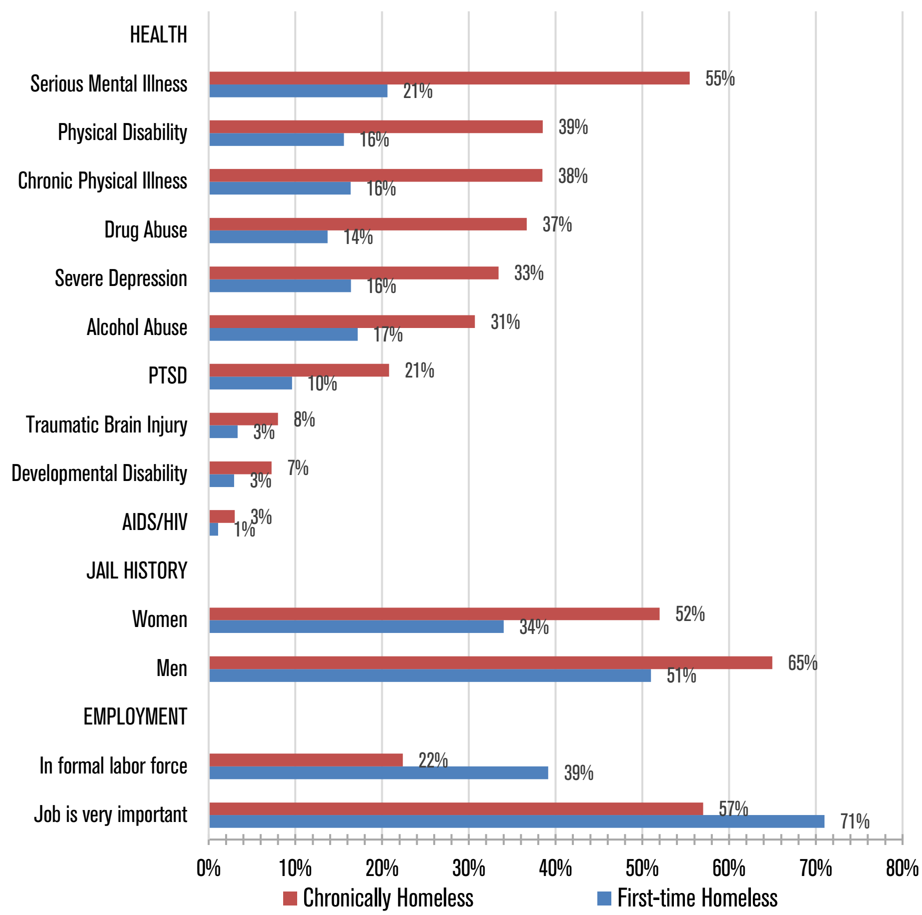
more social disconnection and legal, medical and behavioral health problems emerge and grow as increasingly formidable barriers to escaping homelessness.

There is a first day of homelessness for everyone who becomes homeless, however, on that first day it is difficult to differentiate those who will find rapid exits from those who will remain stuck in homelessness. This study presents two screening tools for quickly identifying and helping high-risk individuals, often before the first day of homelessness.

The first tool identifies workers who have lost their jobs and are likely to become persistently homeless in the next three years. It can be used at the time of unemployment for individuals who have never been homeless, are not currently homeless, or are currently homeless.

The second tool identifies young adults who are 18 to 24 years old and likely to become persistently homeless within the next three years. Similar to the first tool, this tool can also be used for individuals who have never been homeless, are not currently homeless, or are currently homeless.

Figure 3: Newly Homeless compared to Chronically Homeless



Source: Los Angeles Homeless Services Authority, 2016 and 2017 demographic surveys of unsheltered individuals. Respondents identified an average of two reasons, so total responses exceed 100 percent.

The screening tools address the needs of two specific adult groups within the overall population that experiences homelessness. Neither population includes children, who make up roughly one third of Los Angeles County resident who experienced homelessness, as shown in *Figure 2,* which breaks out everyone identified as being homeless within the 10-year window by age and gender.

The bi-modal age distribution of the homeless population, with concentrations of older and younger individuals, that has been reported in other studies (Culhane et al., 2013) can be seen in *Figure 2*.[1]

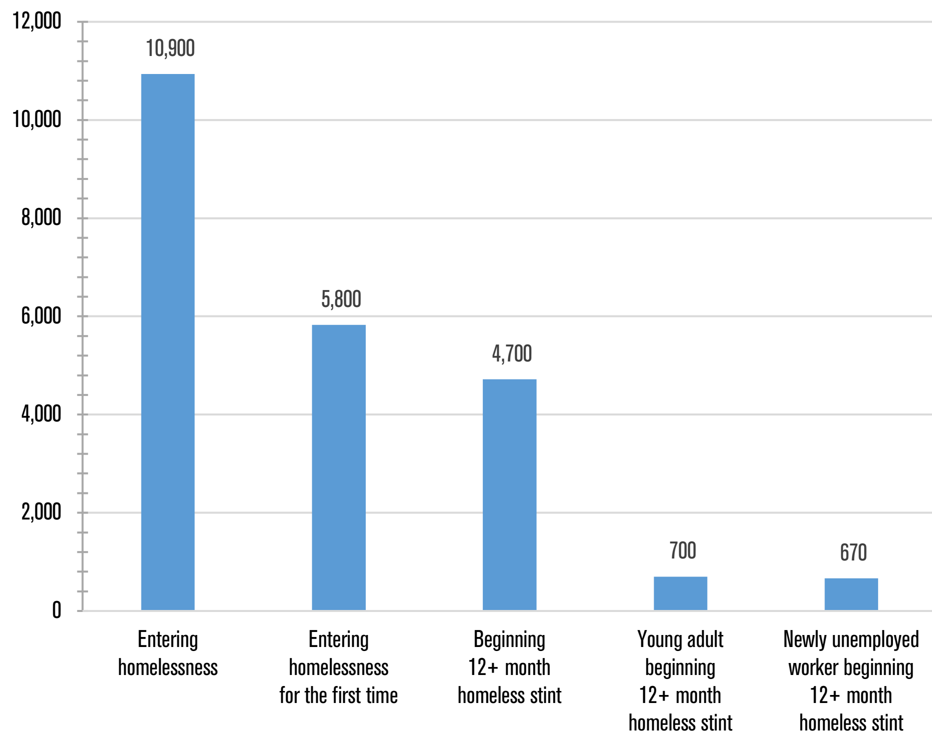
Females who experienced homelessness are more highly concentrated in the 18 to 24 age range than males (26 vs. 19 percent of each gender group), which is important for understanding the population addressed by the young adult screening tool.[2]

Extended homelessness is associated with extensive personal distress. Survey responses from Los Angeles' homeless count (*Figure 3)*, show that every reported health condition is two to three times more prevalent among individuals who are chronically homeless than among new entrants into homelessness. Incarceration histories increase, particularly among women, and there is less interest in developing skills and finding a job. Less intensive interventions are more feasible at the onset of homelessness *if* high-risk individuals can be identified early.

Among the one-million public benefits recipients in this study who experienced homelessness, an average of 10,900 people began a new

An average of 10,900 LA County residents began a new homeless stint each month, and 4,700 of them become persistently homeless.

Figure 4: Monthly Entrants into Homelessness among Public Benefits Recipients



homeless stint each month.[3] This included individuals who had previously been homeless and were beginning a new stint. The entrants into homelessness included an average of 5,800 individuals who were becoming homeless for the first time. Out of the total monthly entrants into homelessness, an average of 4,700 went on to have stints that lasted 12 or more months. This is shown in *Figure 4*.

The two screening tools presented in this study identify 29 percent of the individuals who became persistently homeless. Fifteen percent, or an average of 700 a month, were young adults who were 18 to 24 years old. Fourteen percent, or an average of 670 a month, were workers who had just lost their jobs.

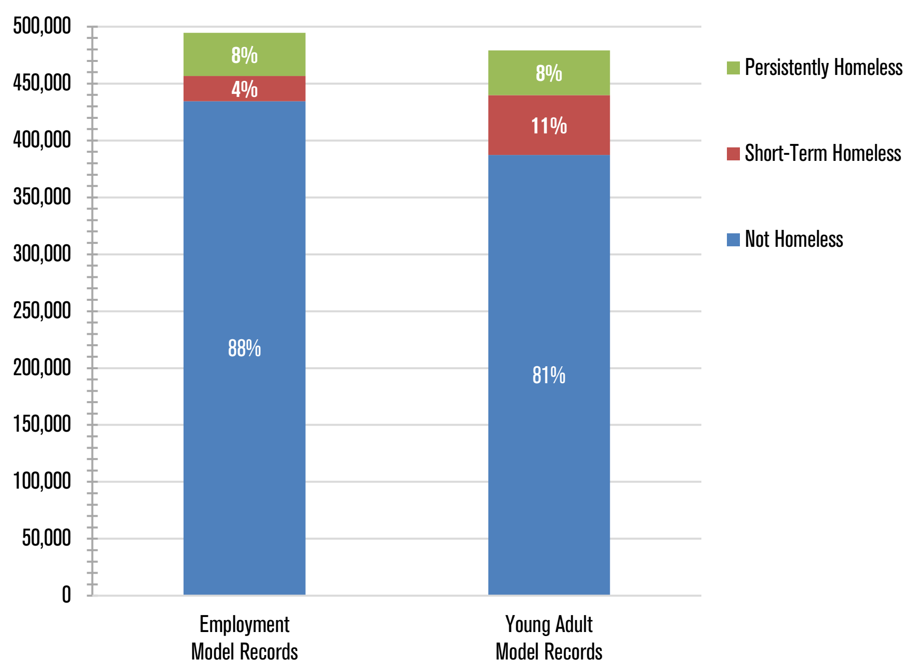
These two screening tools each provide rifle-shot targeting for identifying distinctive groups of people who become persistently homeless. An additional array of targeting tools is needed to identify other high-risk groups, as well as groups who do not become persistently homelessness but need specific types of short-term interventions.

Records of 920,575 people who were homeless were used to develop the youth and employment screening tools.

A total of 920,575 records were used to develop the youth and employment screening tools. Forty-six percent of the records were used just for the youth model, 48 percent just for the employment model, and 6 percent were used for both models. This report is one of the few large-scale, longitudinal studies of homelessness, utilizing linked administrative records from multiple public agencies serving poor residents (Metraux et al., 2018).[4]

The composition of each data set, broken out by duration of homelessness, is shown in *Figure 5*. In the case of the population of workers who became

Figure 5: Records Used to Develop Employment and Youth Screening Tools



unemployed, 12 percent experienced homelessness in the three years following unemployment, and 8 percent become persistently homeless. In the case of the young adults, 19 percent experienced homelessness within a three-year time window, and 8 percent became persistently homeless.

Each screening tool is designed to identify the 8 percent of individuals in each population who go on to become persistently homelessness. Persistent homelessness is defined as 12 consecutive months of homelessness, or two or more episodes of homelessness within three years.

The next chapter describes the attributes of workers who lost their jobs and became persistently homeless. The following chapter describes young adults who became persistently homeless, which is followed by a chapter discussing public costs, and finally by the chapter describing the statistical methods used to develop the two screening tools and the results from testing the reliability of the tools.

Overall, the first part of this report describes the attributes and needs of persistently homeless workers and young adults. The last part presents the multivariable analyses conducted to develop the predictive models.

Workers using old cars as sleeping quarters. Herald Examiner Collection, 1954. Courtesy of Los Angeles Public Library.

Workers Who Lose Their Jobs and Become Persistently Homeless

Demographics

The ethnic, gender and age distributions of the eight percent of workers who lost their jobs and became persistently homeless are shown in *Figure 6* and compared to the distributions for the other 92 percent of workers who also lost their jobs but did not become persistently homeless, including 88 percent who did not become homeless at all.

African Americans made up the largest share of persistently homeless workers (45 percent), followed by Latinos (36 percent), then European Americans (16 percent), and other ethnicities (10 percent).[5]

The majority of workers who lose their jobs and did not become persistently homeless were Latino

Sixty-two percent of persistently homeless workers were men and 38 percent were women. On the other hand, the majority of workers who lost their jobs and did not become persistently homeless were women.

The age distribution, both for workers who became persistently homeless and those who did not, was similar for workers who were 18 through 54 years of age, with a drop-off for older workers. A demographic breakout of workers who became persistently homeless after unemployment is shown in *Figure 7*.

African American workers were more than twice as likely as any other ethnic group to become persistently homeless after unemployment.

Figure 6: Distribution of Persistently Homeless Workers by Ethnicity, Gender and Age

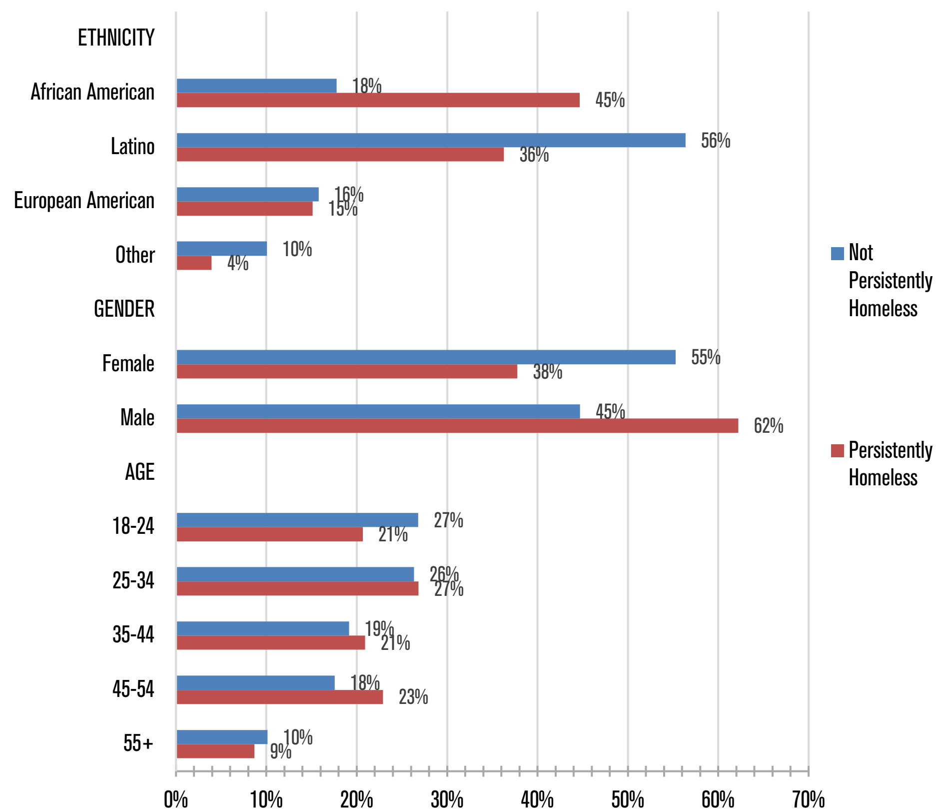
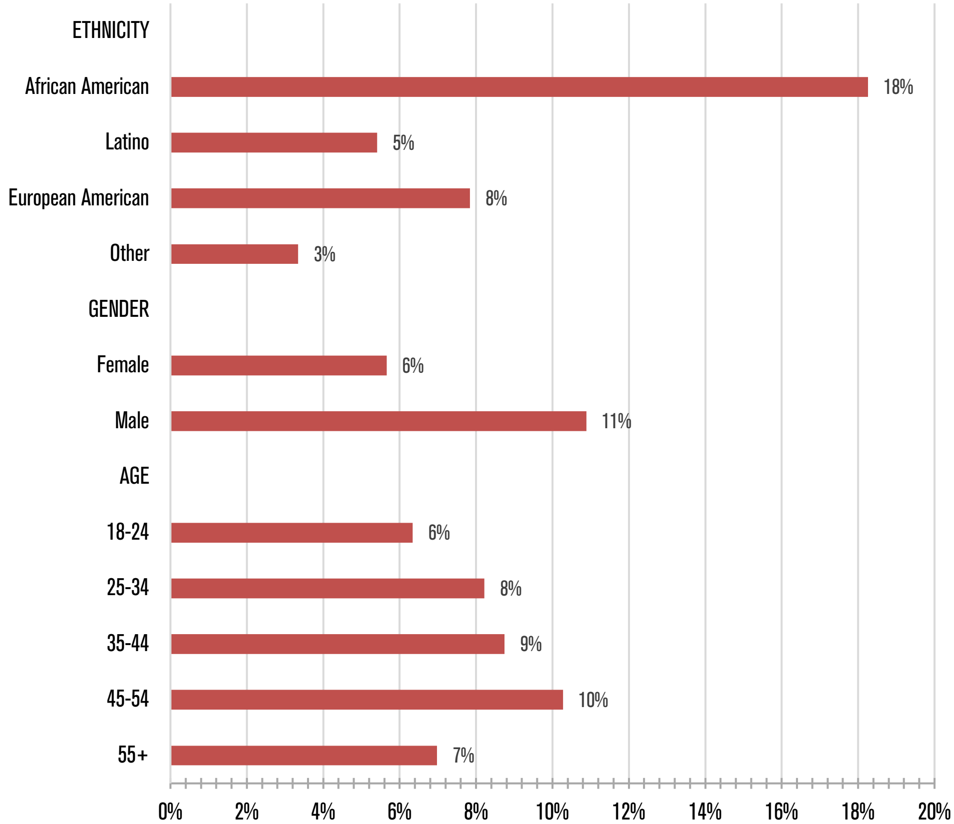


Figure 7: Rate of Persistent Homeless among Unemployed Workers by Attribute



African American workers were more than twice as likely as any other ethnic group to become persistently homeless after unemployment. This was the outcome for 18 percent of African American workers who lost their jobs.

Ethnic minorities in other developed countries experience greater risks of homelessness as well. An Australian study found that during periods of housing and job shortages, Indigenous Australians have significantly higher risks of entering homelessness. This mirrors the disproportionate numbers of Americans of African ancestry entering into homelessness and becoming persistently homeless (Johnson et al., 2018).[6]

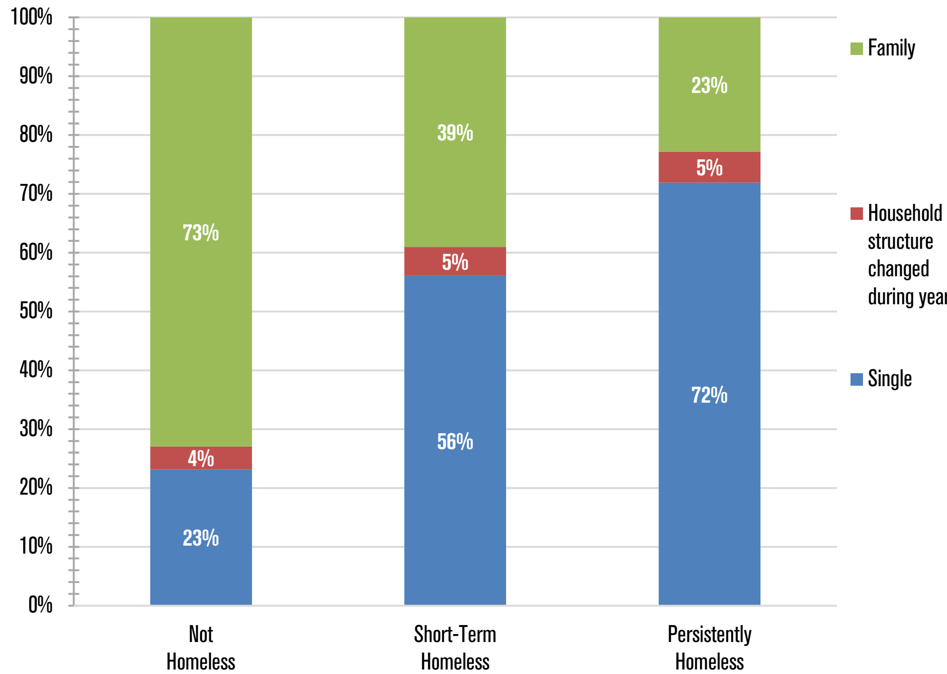
Men were almost twice as likely as women to become persistently homeless after unemployment – 11 versus 6 percent.

Ten percent of workers 45 to 54 years of age who lost their jobs became persistently homeless – the highest rate of any age group.

The household structure of unemployed workers who did not become homeless, experienced short stints of homelessness, or became persistently homeless is shown in *Figure 8*.

Nearly three-quarters (73 percent) of unemployed workers who did not become homeless were part of a family – they had an adult partner and/or children with them. These proportions were reversed for workers who

Figure 8: Homeless Outcomes and Household Structure in the 12 Months before Unemployment



became persistently homeless – nearly three-quarters (72 percent) were single. Support from another adult or a source of cash aid such as CalWORKs in paying the rent was strongly associated with being able to avoid homelessness when workers lost their jobs.

In summary, the risk of becoming persistently homeless after losing a job was particularly high for African Americans, was compounded for men and single individuals, and became progressively higher as individuals aged, until they were 55 or older.

Employment

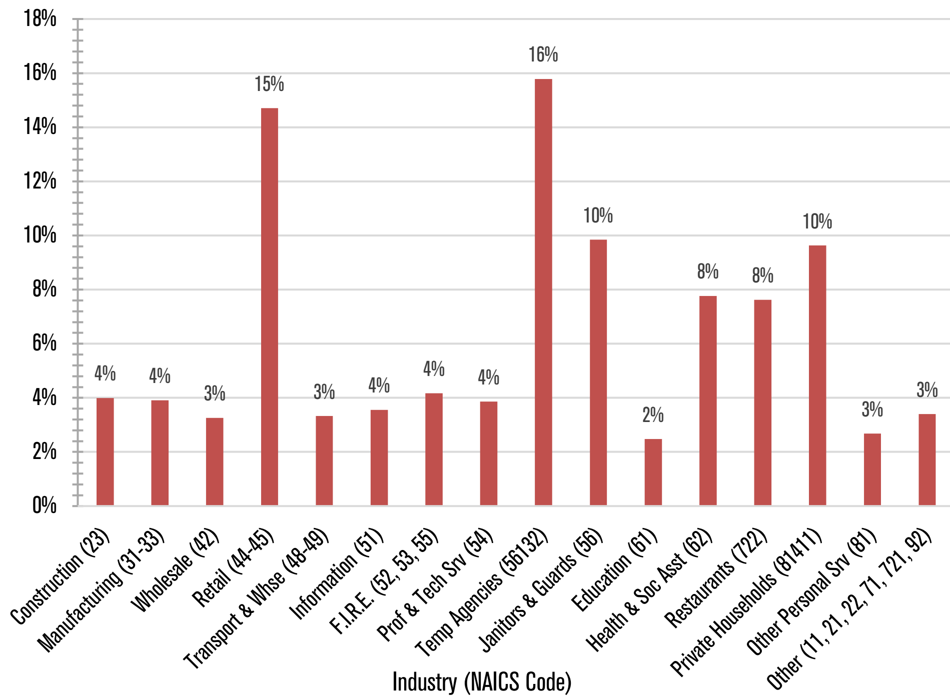
Persistent homelessness is associated with less consistent employment and lower earnings.

The industries in which workers who became persistently homeless lost their jobs tend to pay low wages and have high job turnover, as seen in *Figure 9*. Temporary employment agencies discharged the largest share of persistently homeless workers (16 percent), followed by retail stores (15 percent), janitors and security guards (10 percent), and private households (10 percent). These four groups of employers accounted for half of all workers who became unemployed and persistently homeless.

Persistent homelessness is associated with less consistent employment and lower earnings, as can be seen in *Figures 10* and *11*.

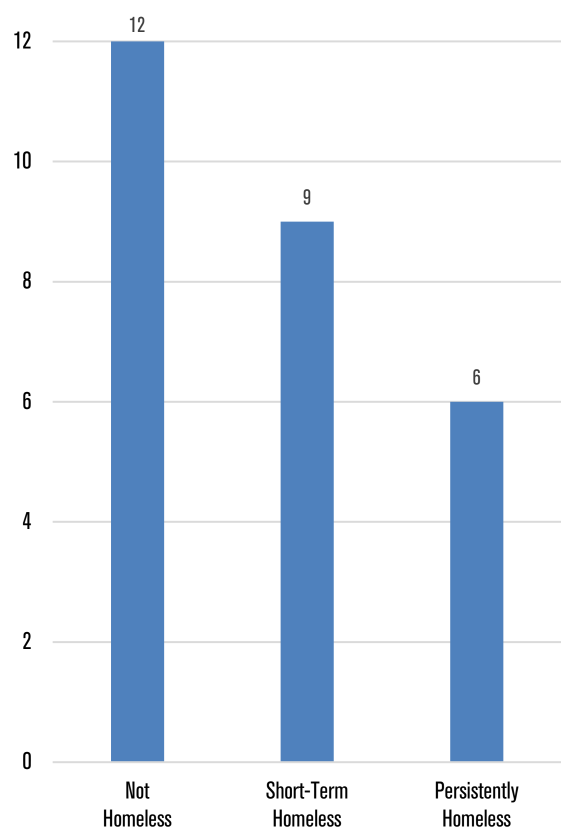
These employment outcomes are based on payroll records submitted by employers for work in the formal economy and do not include informal work such as recycling, panhandling, or day labor.

Figure 9: Industries in which Persistently Homeless Workers Lost Jobs



Other research using administrative data to track the employment and earnings of homeless workers has shown similar findings: job loss is a precipitating event leading to homelessness.[7] Our finding also is consistent with prior research on the effect of homelessness on employment, highlighting the vulnerability of single adults (Fargo et al., 2010).[8]

Figure 10: Median Months of Work in Past Year

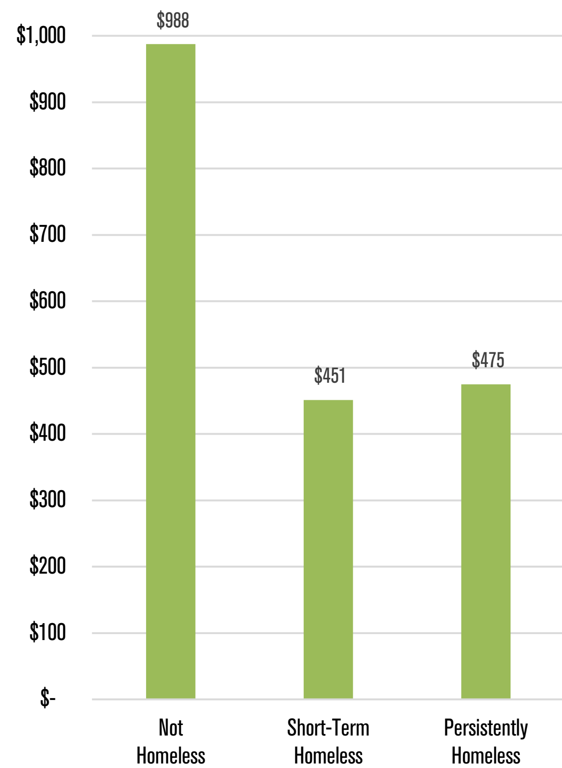


Workers who did not become homeless typically worked all 12 months in the year before they became unemployed. Workers with short stints of homelessness worked nine months, and workers who were persistently homeless worked only 6 months.

The impact of intermittent employment on earnings was compounded by lower wages or fewer hours of

The industries in which workers who became persistently homeless lost jobs tend to pay low wages and have high job turnover.

Figure 11: Median Monthly Earnings when Working in Past Year



work for persistently homeless workers, who typically earned only $475 a month when employed. In contrast, workers who did not become homeless typically had monthly earnings that were more than twice as high – $988 a month.

Workers who became persistently homeless had been unemployed more often than other workers, as shown in *Figure 12*.

All of the workers in the study population had at least one unemployment episode, which was the benchmark event for assessing whether they were subsequently homeless, and if so, for how long. Forty-six percent of persistently homeless workers had previous unemployment episodes in the past five years compared to just 26 percent of workers who did not become homeless and 28 percent of workers who had short stints of homelessness.

Figure 12: Number of Times Unemployed in Past 5 Years

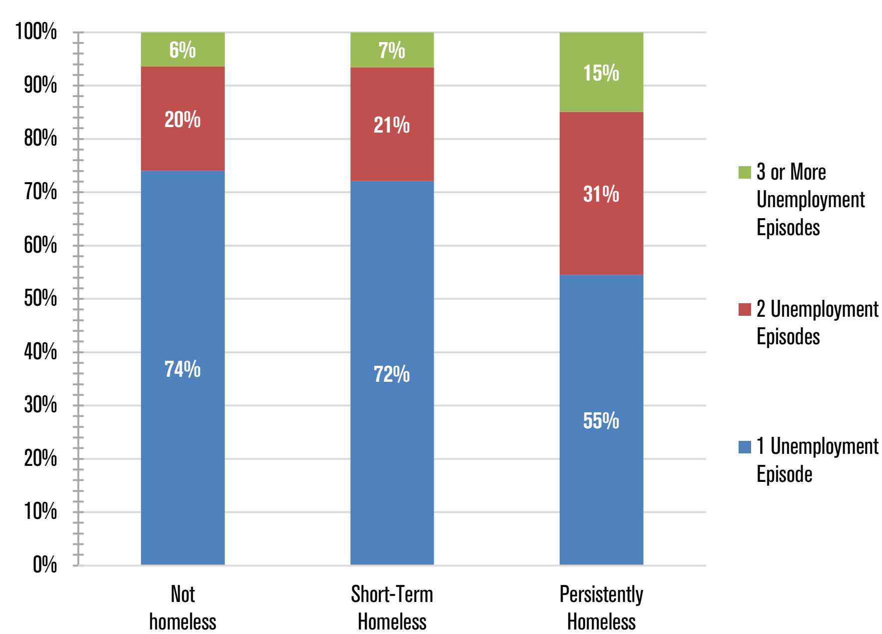
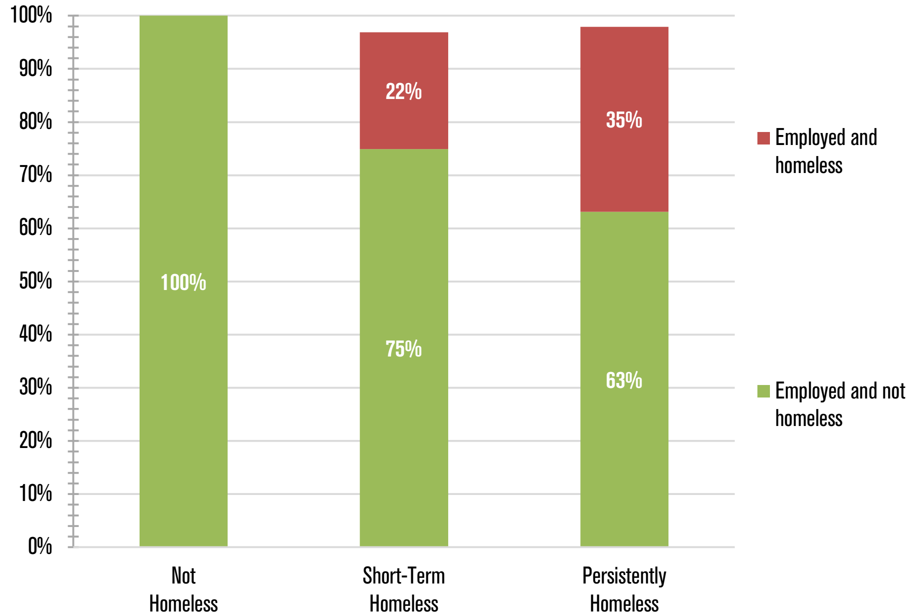


Figure 13: Employment and Homeless Status in Month before Losing Job



Workers who became persistently homeless were more likely to be homeless before they become unemployed, as shown in *Figure 13*. Over a third (35 percent) of workers who became persistently homeless were already homeless when they became unemployed, compared to a fifth of the workers with short homeless stints, and none of the workers who did not become homeless. This can be a self-reinforcing downward spiral – low earnings cause workers to lose housing, and the instability inherent in homelessness makes it harder to hold on to a job.

In summary, the industries in which workers who became persistently homeless lost jobs paid low wages and had high turnover. Persistent homelessness was associated with inconsistent employment and low earnings. Workers who became persistently homeless were more likely than other workers to have previously been unemployed. They also were more likely than other workers to already have been homeless when they lost their jobs, with the instability inherent in homelessness making it harder for them to hold on to their jobs.

Over a third of workers who became persistently homeless were already homeless when they became unemployed.

Barriers

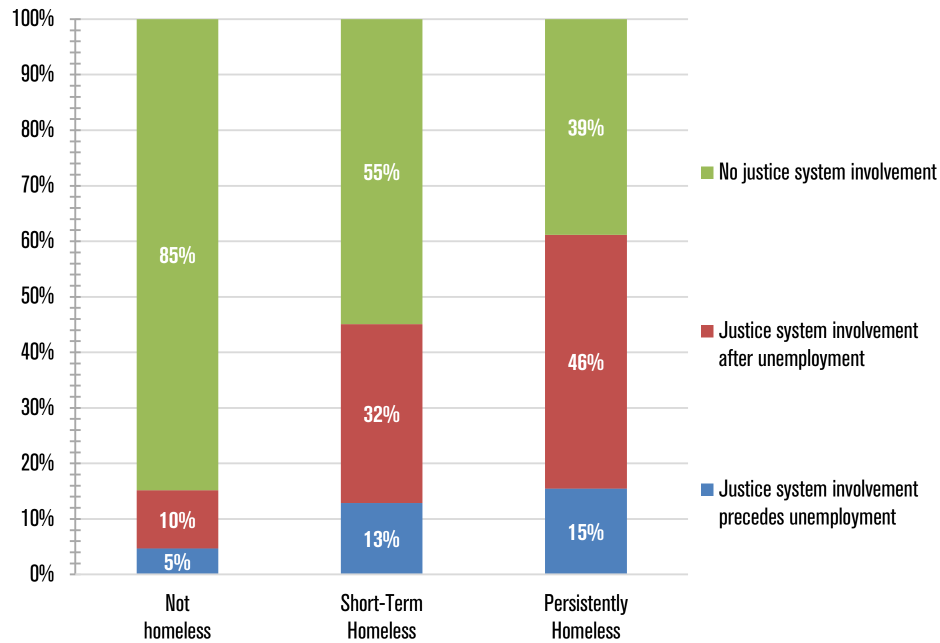
Criminal Justice History

Justice system involvement in the form of adult probation or jail stints often accompanied homelessness for workers who experienced even short episodes of homelessness after unemployment, as shown in *Figure 14*.

Most workers were not involved with the justice system prior to becoming unemployed. The rates of prior involvement ranged from five percent for workers who did not become homeless to 13 percent for workers who had short homeless stints and 15 percent for workers who became persistently

46% of workers who were persistently homeless became involved with the justice system.

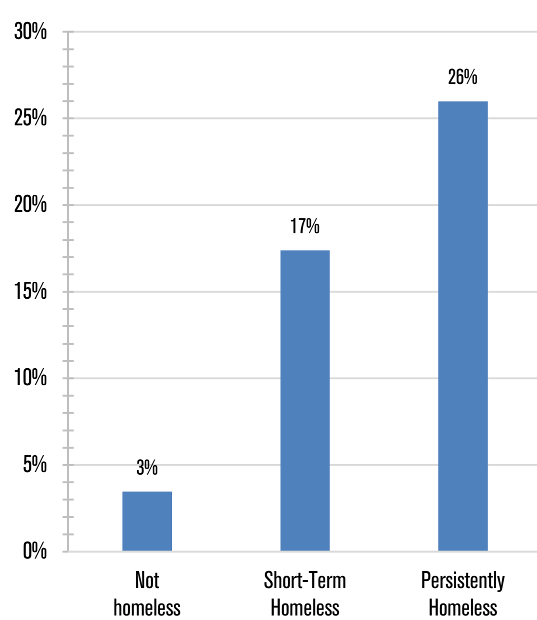
Figure 14: Homelessness and Justice System Involvement



homeless. The higher rate for workers who became homeless may be due to the fact that they were more likely to have had prior episodes of unemployment as well as to have been homeless when they lost their jobs.

After they became unemployed, 32 percent of workers with short homeless stints and 46 percent of workers who were persistently homeless became involved with the justice system. Justice system involvement is associated with unemployment and homelessness, and this involvement creates a barrier to future employment.

Figure 15: Medical Diagnosis of Substance Abuse



Substance Abuse

Substance abuse was diagnosed more frequent among workers who experienced homelessness than among other workers, as shown in *Figure 15*. Three percent of workers who did not become homeless were diagnosed with a substance abuse related health condition compared to 17 and 26 percent, respectively, of workers who were short-term and persistently homeless.

This chart under-reports actual rates of substance abuse because it is limited to individuals with medical diagnoses made in county health care facilities or who received substance abuse services. This includes diagnoses both before and after unemployment but leaves out workers whose problems were not severe enough to come to the attention of the health care system or who were cared for at a private health care facility. The medical diagnostic codes used to identify substance abuse are listed in *Appendix Table A-9*.

Substance abuse is a factor in the lives of many workers who experience homelessness. This issue should be addressed as part of the package of re-employment services for these workers.

Disabilities

One-tenth of workers (10 percent) had disabling conditions lasting three or more years while they were still employed, indicating that they were able to be gainfully employed despite having physical or mental limitations.[9] This includes 7.6 percent of workers with a physical limitation and 2.6 percent with a mental limitation. It's likely that more effective help in treating and managing disabilities would have helped some of these workers retain their jobs.

10% had disabling conditions lasting three or more years while they were still employed.

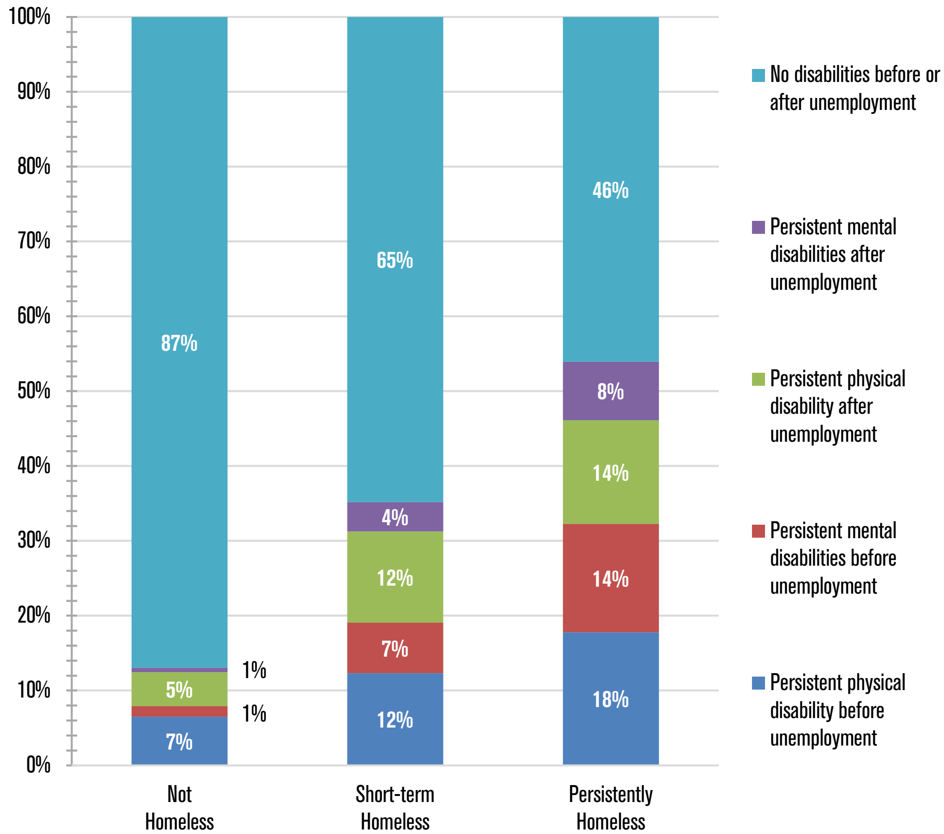
The prevalence of disabilities is strongly associated with experiences of homelessness and the duration of those experiences, as shown in *Figure 16*. Before they became unemployed, 8 percent of workers who did not become homeless, 19 percent of workers with short homeless stints, and 32 percent of workers who became persistently homeless had been identified as having a physical or mental disability.

After they because unemployed, there were proportional increases in the rates of disabilities, including an additional 6 percent of workers who did not become homeless, 16 percent of workers with short homeless stints, and 22 percent of workers who became persistently homeless.

These findings suggest three conclusions. First, the presence of a disability does not preclude employment. Second, rapid and effective help in becoming re-employed is likely to reduce the emergence of post-unemployment disabilities seen among workers who become persistently homeless. And, third, effective help in treating and managing both physical and mental disabilities will improve the prospects of persistently homeless workers for obtaining new jobs.

An earlier study in Alameda County, California (Zuvekas and Hill, 2001) explored whether homeless individuals could start and maintain income (both earned income and public assistance) over a 6-month period,

Figure 16: Presence, Timing and Type of Persistent Disabilities among Workers



depending on their homeless, health and disability status. All of these issues were barriers to employment, and correlated with lower employment levels.[10] These findings support the conclusion that workers who otherwise would become persistently homeless will benefit from support in treating and managing disabilities. Mental illness does not preclude employment. A survey of individuals with recent histories of homelessness who had a mental illness found that over two-thirds (69 percent) wanted to work (Poremski and Hwang, 2016).[11]

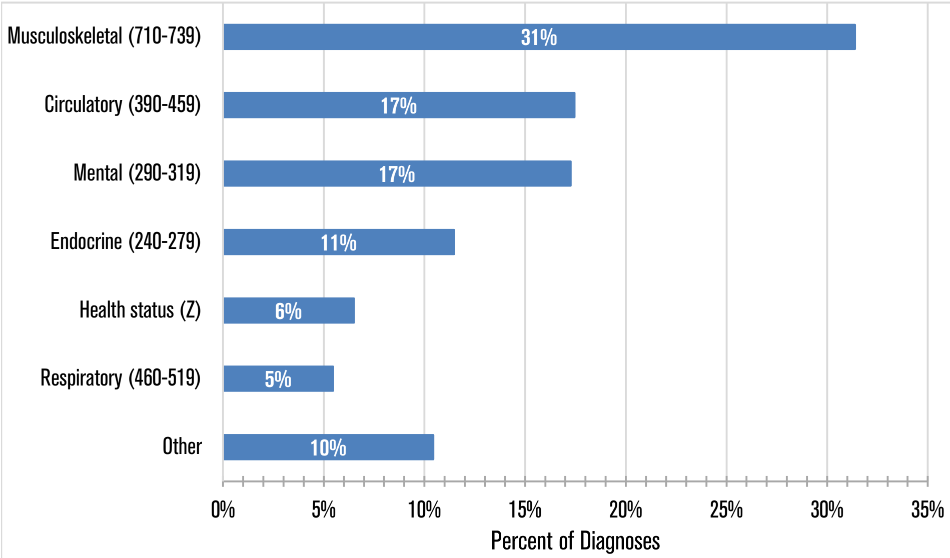
The most frequent disabilities were back, joint and arthritic conditions.

Medical diagnostic codes were available for two-fifths (42 percent) of the workers with disability flags in their public benefits records. Their health conditions are shown in *Figure 17*.

The most frequent problems were with the musculoskeletal system, accounting for almost one-third (31 percent) of disabilities. Over three-quarters (77 percent) of the problems in this category had to do with back, joint and arthritic conditions. Some of these workers need to be redirected to occupations that are less physically demanding.

The next most frequent category of problems were with the circulatory system. Hypertension accounted for three-quarters of these disabling conditions.

Figure 17: Medical Diagnoses for Persistently Homeless Workers with Disabilities



ICD-9-CM body system code range for diagnoses shown in parenthesis.

Mental disorders were the next most frequent category of problems, with episodic mood disorders accounting for 44 percent of these conditions.

Endocrine, nutritional, and metabolic diseases and immunity disorders were the fourth most frequent category of problems. Diabetes accounted for 85 percent of these conditions.

In summary, most workers were not involved with the justice system prior to becoming unemployed, however rates of prior involvement tripled for workers who became persistently homeless. Substance abuse is a frequent problem and becomes more frequent as individuals are homeless longer. Many workers have held jobs despite having disabling health conditions. These problems are much more frequent after individuals lose their jobs and become persistently homeless. Inadequate support in treating and managing disabling conditions is likely to have contributed to loss of jobs, and medical support in caring for these conditions should be included in re-employment services. Some workers will need to be redirected to occupations that place less stress on their backs and joints.

Conclusions

All low-wage workers face some level of risk that they will become persistently homeless if they lose their jobs, but this risk is disproportionately high for workers who are African American, male and single. It is important that screening to identify unemployed workers who are likely to become persistently homeless be carried out in a way that includes full representation of these groups with especially high-risks.

Some high-risk workers have barriers to employment resulting from substance abuse and involvement in the criminal justice system. This indicates that some need behavioral health services to overcome substance abuse problems as well as legal services to expunge or lessen their criminal justice records.

A quarter of high-risk workers are part of a family unit and a third are homeless before they lose their jobs. This indicates that some workers need affordable child care and many workers need affordable transitional housing.

Almost a third of workers who become persistently homeless have held down jobs despite having limiting physical or mental conditions. These disabilities become much more frequent during post-employment homelessness. These workers are likely to have better employment and job retention prospects if they receive health care support in treating and managing their conditions. Workers with back, joint and arthritic problems will benefit from looking for work in occupations that are less physically demanding than their previous jobs.

Workers who become persistently homeless often have histories of job turnover, under-employment and low earnings. This indicates that many high-risk workers need human capital investments in the form of education and training that will enable them to compete for better jobs. They may also need wage subsidies to encourage employers to give them an opportunity to demonstrate their capabilities.

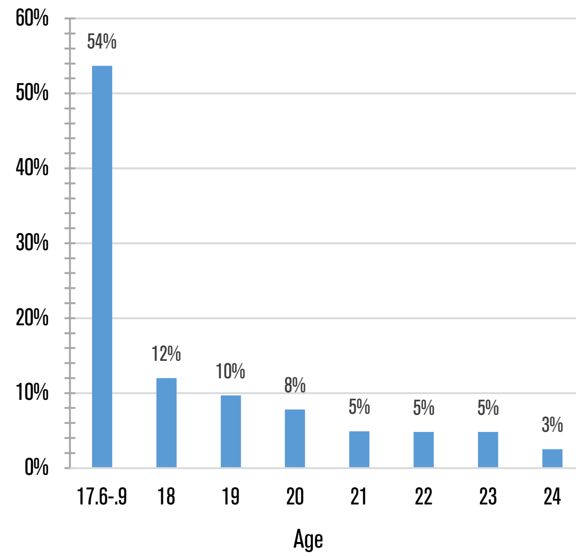
Youth sleeping in cardboard shantytown. By Javier Mendoza, Herald Examiner Collection, 1987. Courtesy of Los Angeles Public Library.

Young Adults Who Become Persistently Homeless

Demographics

The young adult screening tool is derived from information about individuals 18 to 24 years of age who received some form of public benefits – Medi-Cal, Food Stamps/SNAP, or cash aid.

Figure 18: Age at Benchmark Month that began 3-Year Study Window

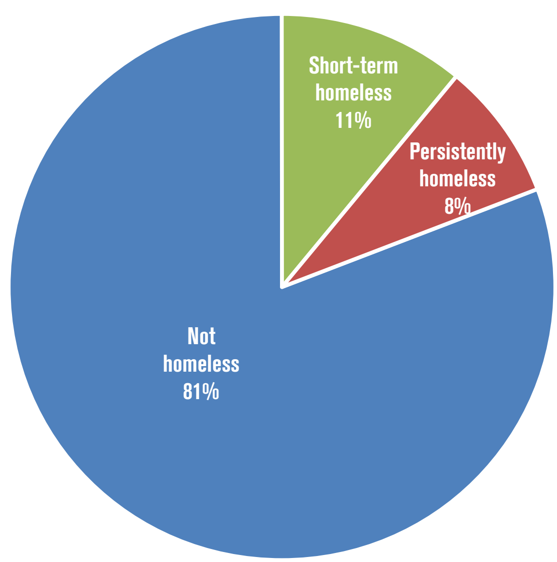


Outcomes for each youth were tracked throughout a three-year study window, beginning with their eighteenth year if they were receiving public benefits. The ages of youth were rounded to the nearest full year, which meant they were counted as being eighteen in the second half of their seventeenth year.

If youth were not receiving benefits when were eighteen, the study window started as soon as they began receiving benefits, up through 24 years of age. It was essential for youth to be receiving public benefits because their benefits records were the source of information about their homeless status.

Two thirds of youth entered the study window when they were eighteen, including the second half of their seventeen year and all of their eighteenth year, as shown in *Figure 18*.

Figure 19: Homeless Status of Young Adults in 3 Year Study Window



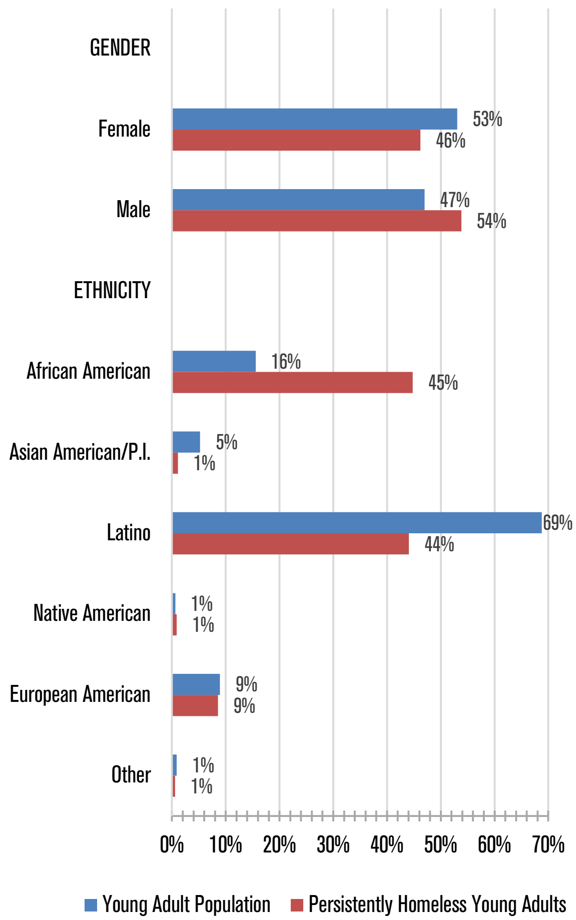
The study population includes a mix of youth who became 18 and were emancipated before AB 12 took effect in January 2012, extending eligibility for foster care services beyond age 18 to age 21, and those who were born afterwards and were eligible for this extended support.

Five percent of youth in the study population received foster care services. A fifth of these youth (19 percent) became 18 years of age after AB 12 took effect and were eligible for extended support. Four-fifths became 18 before that date and were emancipated into adulthood without the extended support provided by AB 12.

Eighty-one percent of all low-income young adults did not experience homelessness, as shown in *Figure 19*. Eleven percent had stints of homelessness that cumulatively lasted less than 12 months.

Eight percent were homeless for 12 or more months in a single episode, or experienced two or more episodes over three years. We describe these individuals as having been persistently homeless.

Figure 20: Distribution of Young Adult Population and Persistently Homeless Young Adults by Gender and Ethnicity



The young adult screening tool is designed to identify the eight percent of young adults who will become persistently homeless within three years.

8% of young adults receiving public benefits became persistently homeless.

The gender and ethnic distribution of the population of young adults receiving public benefits, as well as the subset of this population that became persistently homeless is shown in *Figure 20*.

Based on gender, a majority of the population was female, but a majority of those who become persistently homeless were male.

Based on ethnicity, a majority of the population was Latino, but a majority of those who were persistently homeless was African American. Other ethnicities accounted for 16 percent of the total population and 11 percent of those who became persistently homeless.

The percent of young adults in each demographic group who become persistently homeless is shown in *Figure 21*. What stands out is that 23

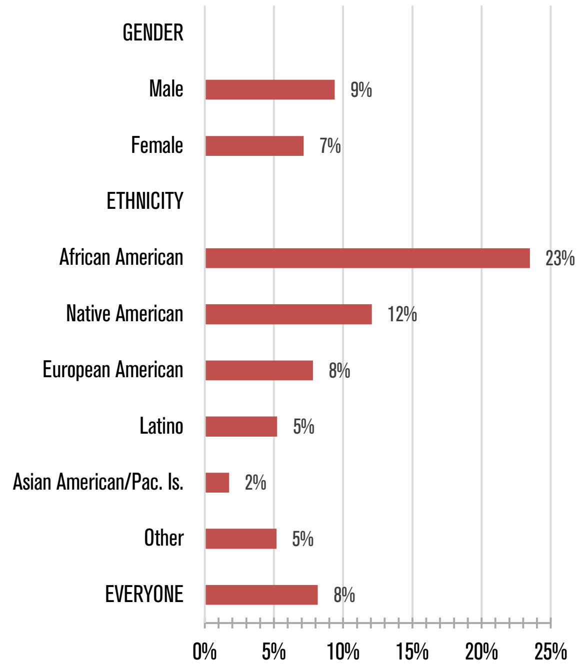
23% of African American youth receiving public benefits become persistently homeless.

percent of African American youth become persistently homeless – a rate roughly triple the average for young adults.

Family connections of young adults at the beginning and end of the study window are shown in *Figure 22*, with youth broken out by gender and homeless history.

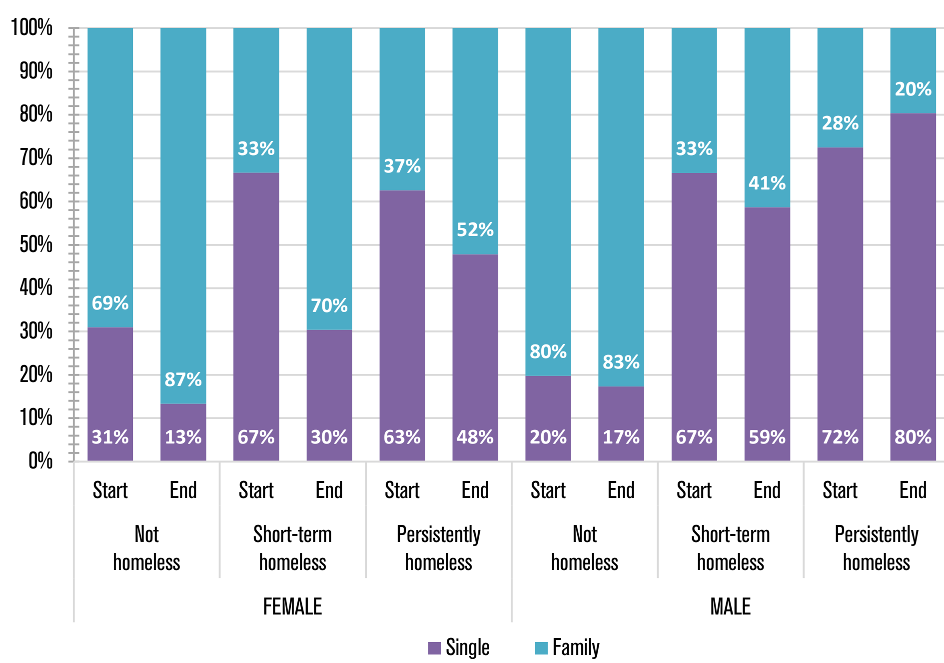
With one exception, more young adults were part of a family group at the end of the study window than at the beginning. The exception was persistently homeless males, who were the most solitary group when they entered the study window, and even more solitary three years later, with only one out of five in a household with another adult or child.

Figure 21: Rate of Persistent Homelessness among Young Adults in 3 Year Study Window by Gender and Ethnicity



Males with short homeless stints were less solitary. The share connected with a family increased from 33 to 41 percent from the start to the end of

Figure 22: Percent of Young Adults in a Family Unit by Age, Homeless Status, and Gender at Beginning and End of 3 Year Study Window



the 3-year study window. Over four-fifths of males who were not homeless were connected with a family.

Females who experienced homeless were similarly solitary at the start of the study window but by the end, females with short homeless stints were more frequently connected to a family than females who were persistently homeless (70 vs. 52 percent).

By the end of the study window, females who did not experience homelessness had the most frequent family connections of any group – 87 percent were part of a family unit.

At the start of the study window, more males were connected to families than females (70 vs. 63 percent). Family connections for both males and females increased by end of the window, but at the end, more females had family connections than males (82 vs. 73 percent)

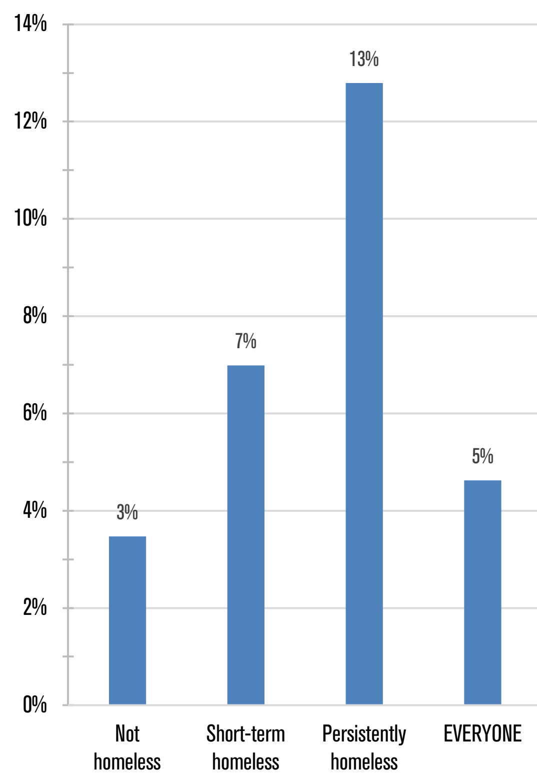
In summary, 8 percent of youth receiving public benefits experienced persistent homelessness during the three year study window, but the risk was far greater for African American youth, with 23 percent experiencing persistent homelessness. Males had a slightly higher rate of persistent homelessness than females (9 vs. 7 percent). Youth who were part of a family unit had lower rates of homelessness and the share of youth who were part of a family increased during the three year study window.

13% of persistently homeless young adults had been in the foster care system.

Foster Care

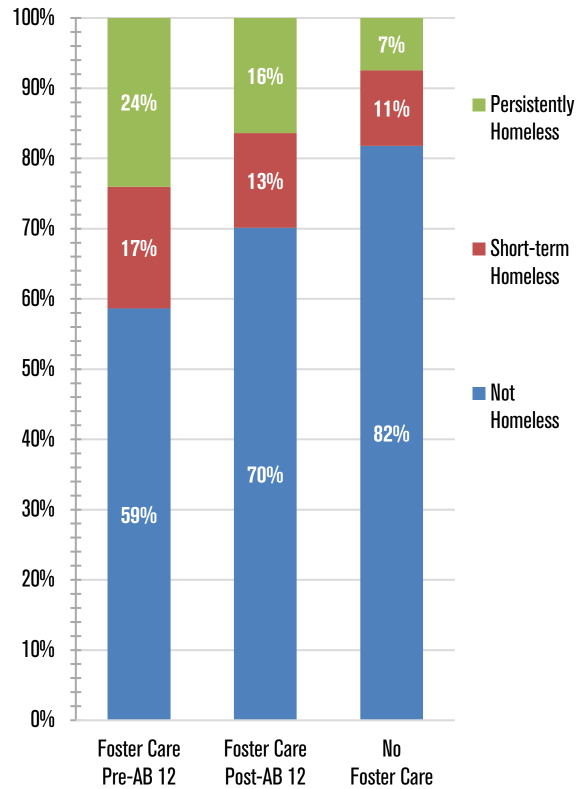
Only five percent of the low-income young adult population spent time in the foster care system, as shown in *Figure 23*, but a foster care history was associated with higher rates of homelessness. Three percent of young adults who were not homeless had a foster care history, 7 percent of those with short homeless stints had this history, and 13 percent of those who were persistently homeless had been in the foster care system.

Figure 23: Percent of Young Adults who Received Foster Care Services



Outcomes for youth from foster care have been found to be poor. However, previous research has also found that the longer youth received foster care support, the better their education and employment outcomes were. And income support and job preparation services were associated with achieving better employment outcomes (Barnow et al., 2015).[12]

Figure 24: Homeless Outcomes during 3-Year Study Window for Foster Care Youth Pre- vs. Post-AB 12



The enactment in January 2012 of Assembly Bill 12, "California Fostering Connections to Success Act," provided additional help for foster youth by extending foster care services from age 18 to age 21. Homeless outcomes for foster youth based on whether their eighteenth birthday came before or after AB 12 took effect are shown in *Figure 24*.

AB 12 reduced homelessness, but 16% of youth receiving extended benefits still became persistently homeless.

The rate of persistent homelessness among youth who were eligible for extended services under AB 12 was a third lower than for older youth who were not eligible and who were emancipated into independent adulthood at age 18.[13] Twenty-four percent of youth who were not eligible for AB 12 were persistently homeless compared to 16 percent of youth who were eligible for extended help.

These rates of persistent homelessness are much higher than the seven percent rate for youth who did not receive foster care services, but it is clear that the extended support for foster youth was valuable for preventing homelessness. The study window ended for most youth when they were 21 years of age. It is outside the scope of this study to identify the extent to which extended foster care support helped prevent homelessness after youth emancipated from foster care at age 21.

Positive impacts of extended foster care on reducing homelessness have previously been reported based on a survey of 616 21-year-old California foster youth (Courtney et al., 2018).[14] The study found that each year a youth participated in extended foster care decreased the odds of becoming homeless or couch-surfing by 28 percent, decreased the odds of experiencing an additional instance of homelessness by 32 percent and decreased the total number of days youth were homeless by 15 days.

A foster care history was associated with higher rates of homelessness. Forty percent of all youth in the study window with foster care histories experienced either short-term or persistent homelessness. However youth covered by AB 12, who remained eligible for foster care services until they were 21 years old, had better outcomes – 16 percent of these youth experienced persistent homelessness compared to 24 percent of older foster youth who emancipated into adulthood when they were 18 years old.

Homeless History

Youth who were homeless before reaching adulthood were 3 times more likely to become homeless as adults.

Homelessness puts youth at further risk of failing to continue education and prepare for employment, which in turn imperils their short- and long-term economic and housing stability (Milburn et al., 2009).[15]

Only 5 percent of young adults were homeless in the 6 years before they entered the study window, as shown in *Figure 25*. This includes 4 percent who were homeless for up to 12 months, 1 percent for 13 to 24 months, and only 0.3 percent for 25 or more months.

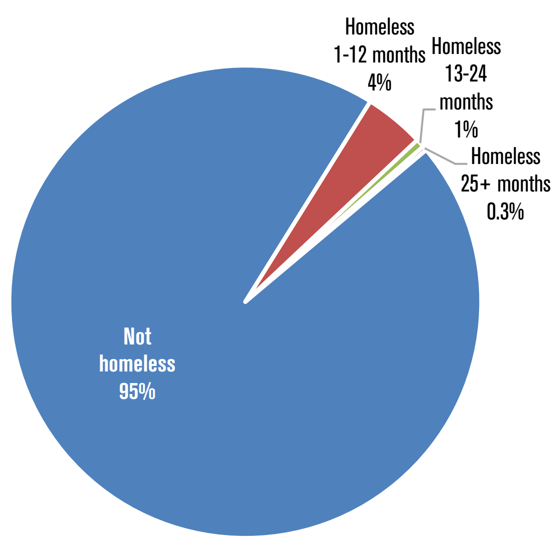
Young adults who experienced homelessness in the preceding six years were more than three times as likely to be homeless during the three year study window than those who had not previously been homeless (58 vs. 17 percent).

The duration of homelessness during the young adult study window was proportionate to the duration of childhood homelessness, as shown in *Figure 26*. Among youth who had been homeless up to 12 months, 56 percent were homeless after they entered the study window, including 26 percent who were persistently homeless.

Among youth who had been homeless 13 to 24 months before turning the study window, 65 percent were homeless during the window, including 37 percent who were persistently homeless.

Among youth who had been homeless more than two years before entering the study window, 80

Figure 25: Number of Months Homeless in the 6 Years <u>before</u> the Study Window

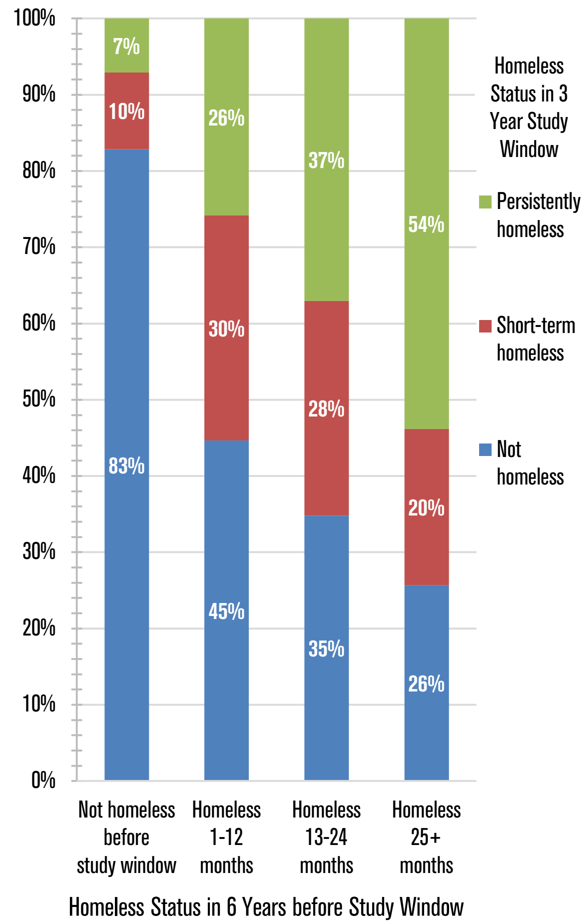


percent were homeless during the three-year window, including 54 percent who were persistently homeless.

Twenty-one percent of young adults experienced homelessness sometime from their 12th through 20th years of age, as shown in *Figure 27*. These homeless experiences include both short-term and persistent homelessness.

Seventy-nine percent were never homeless during the *nine-year* interval from childhood into young adulthood, compared to 81 percent shown in *Figure 19* who were not homeless in the *three-year* study window. Only 2 percent of youth had homeless experiences that were confined to their childhoods, and not repeated during the study window.

Figure 26: Homeless Outcome in 3 Year Study Window Based on Homeless Status in the Preceding 6 Years



Youth who were homeless when then entered adulthood were 10 times more likely to become persistently homeless.

Twelve percent of youth had homeless experiences that began when they entered study window of early adulthood.

Only seven percent of youth escaped homelessness until after entering the study window and then became homeless during the remainder of the window.

This demonstrates that the transition into adulthood is a high-risk interval for low-income youth.

Of the 12 percent of youth who were homeless when they entered the study window, 57 percent had short homeless stints and 43 percent became persistently homeless, as shown in *Figure 28*.

Of the youth who were *not* homeless when entering adulthood, only 9 percent experienced homelessness and of these, 4 percent were persistently homeless.

The likelihood of becoming persistently homeless was more than 10 times greater for youth who were homeless when they entered adulthood.

In summary, only five percent of young adults were homeless in the 6 years before they entered three-year study window of early adulthood, but they were more than three times as likely to be homeless during the study window as those who had not previously been homeless. The transition into adulthood is a high-risk interval for low-income youth. Sixty-three percent of the homeless experiences of young adults began as they entered the study window. The likelihood of becoming persistently homeless was more than ten times greater for youth who were homeless when they entered adulthood than for those who were not.

Employment History

Nearly half (46 percent) of young adults had some employment in the three-year study window, as shown in *Figure 29*.

Importantly, persistently homeless youth had the highest employment rate, with 56 percent having wage and salary earnings in the formal economy in the three-year study window.

Figure 27: Homeless Status in 3 Year Study Window

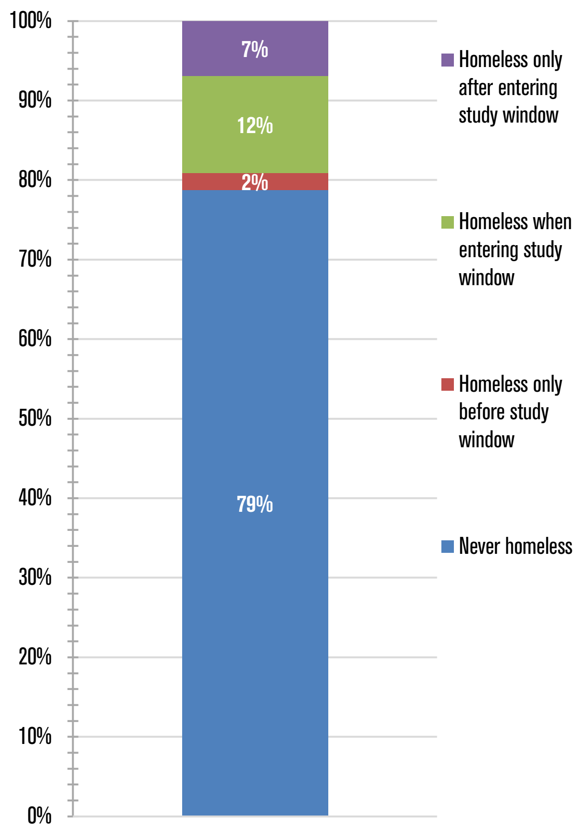
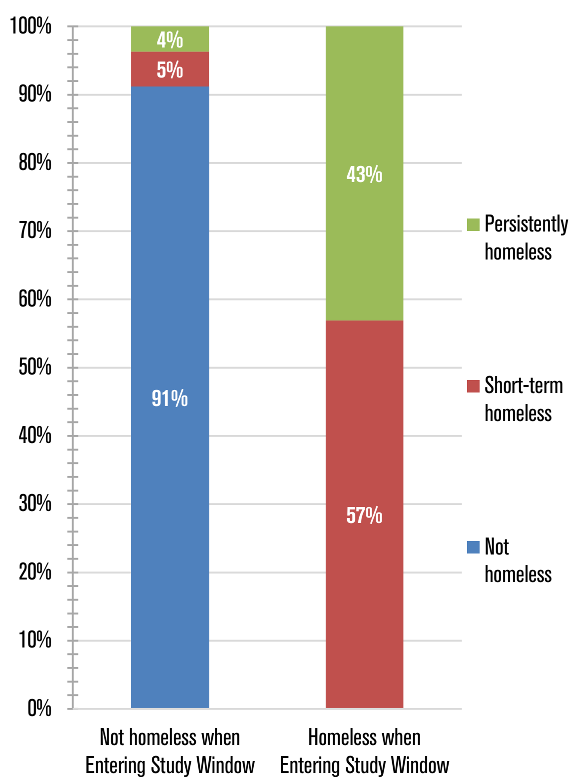


Figure 28: Three-Year Outcomes based on Homeless Status when Entering the Study Window



Persistently homeless youth had the highest employment rate but low earnings.

This demonstrates a strong drive among these youth to support themselves through work.

Other youth may have had stronger family or public aid support that made it less essential to support themselves through work. Only 46 percent of youth who were not homeless had earned income, with an even lower employment rate of 39 percent among youth with short homeless stints.

The typical young adult that had a job in the three year study window was employed during about one-third of those months, as shown by the median (50th percentile) outcome in *Figure 30*.

Persistently homeless young adults typically had only 10 months with earned income, compared to 12 months for youth who were not homeless or had short homeless stints. Possible explanations include less ability to compete in the labor market or difficulty holding on to jobs because of homeless living conditions.

Monthly employment rates in the three-year study window of young adulthood are shown in *Figure 31,* broken out by homeless status.

Figure 29: Employment by Homeless Status of Young Adults in the 3 Year Study Window

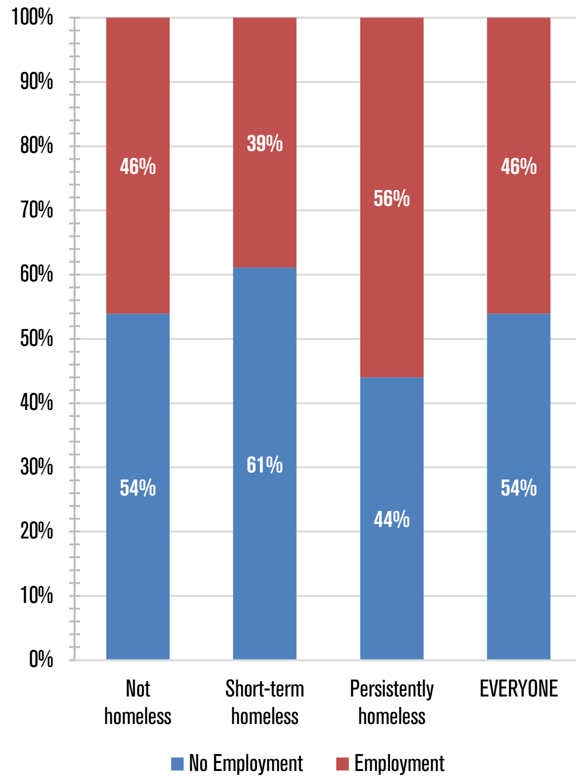


Figure 30: Number of Months Worked by Young Adults in the 3 Year Study Window

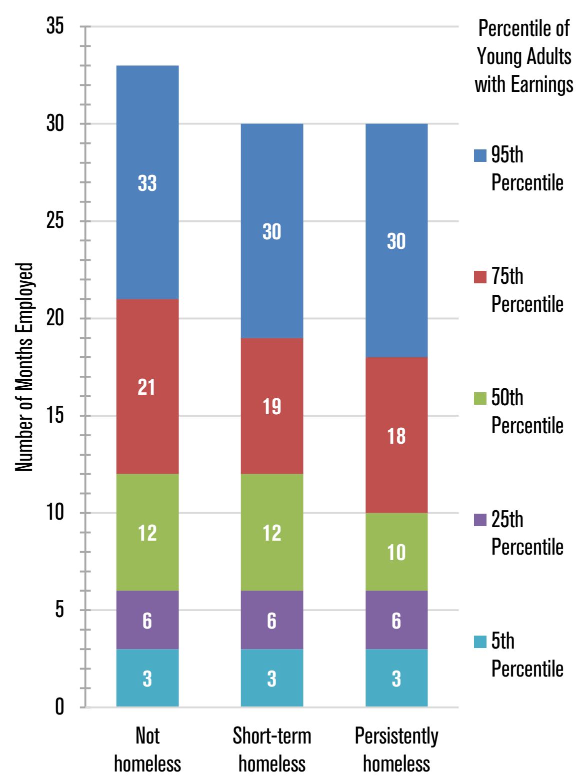
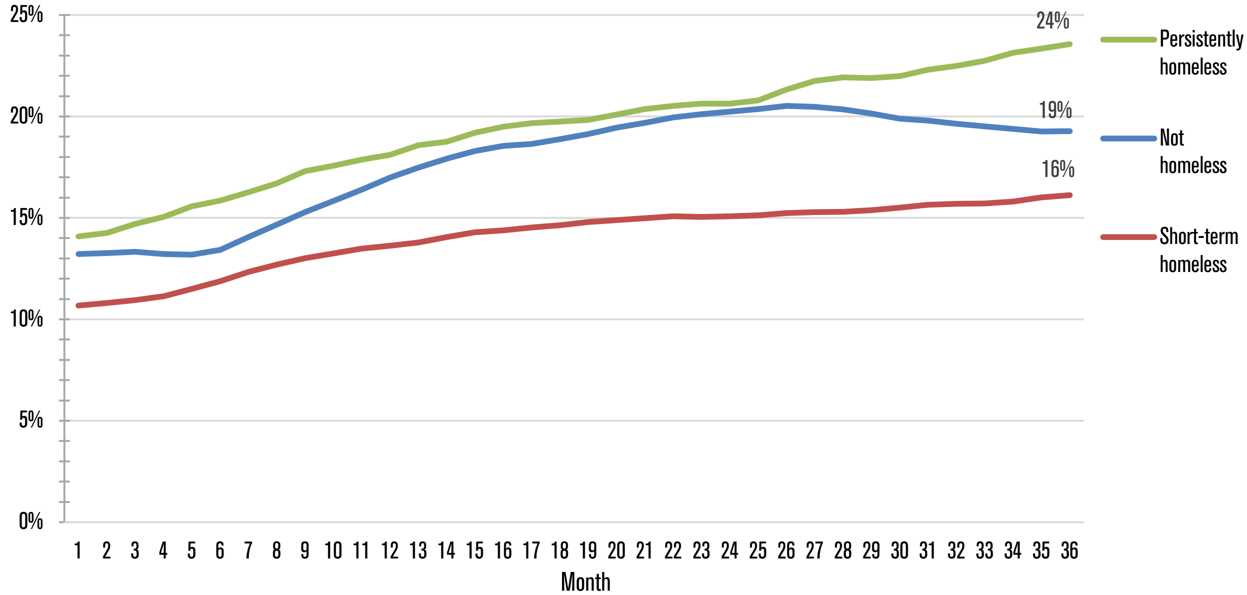


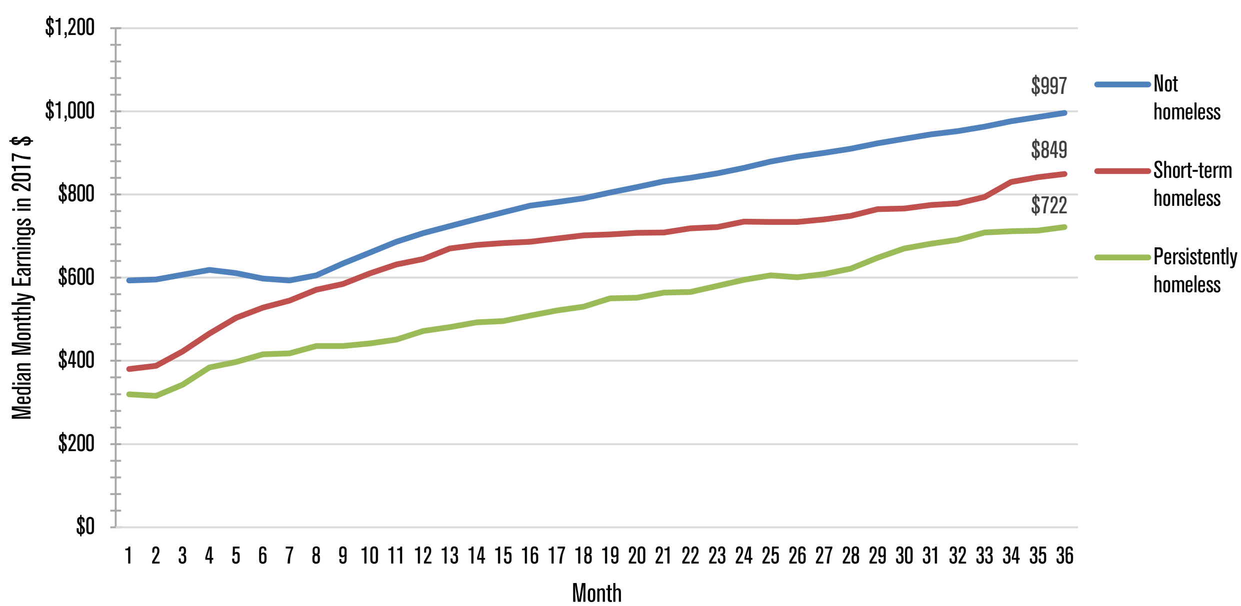
Figure 31: Monthly Employment Rate of Young Adults in the 3 Year Study Window



Employment rates increased for all youth cohorts over this three-year period, with the greatest increase and highest rate found among persistently homeless youth. However, having a job in any given month was the exception rather than the rule for all of these low-income youth. In the last month of the three-year study window, roughly a quarter of persistently homeless youth had a job, a fifth of youth who were not homeless, and a sixth of youth with short homeless stints.

The median monthly earnings of youth in the months when they were employed are shown in *Figure 32*. Earnings for each of the three cohorts

Figure 32: Median Monthly Earnings of Young Adults when Employed in the 3 Year Study Window



increased as they grew older, however their earnings were unlikely to be sufficient to pay for housing and living expenses in Los Angeles County. At the end of their twentieth year, youth who had not experienced homelessness had the highest earnings - $997 a month. This was followed by $849 a month for youth who had short homeless stints. Youth who were persistently homeless had the lowest earnings - $722 a month. Possible explanations for their low earnings include fewer hours worked or lower hourly wages.

All dollar values throughout this report are adjusted to 2017 dollars for the Los Angeles-Riverside-Orange County, California area.

The family connections of youth who were not homeless or had short homeless stints may have made it more feasible for them to remain housed and maintain more stable employment connections. It is important to improve employment opportunities and earnings for all low-income youth, particularly for youth who are trying to escape persistent homelessness by earning enough money to house themselves.

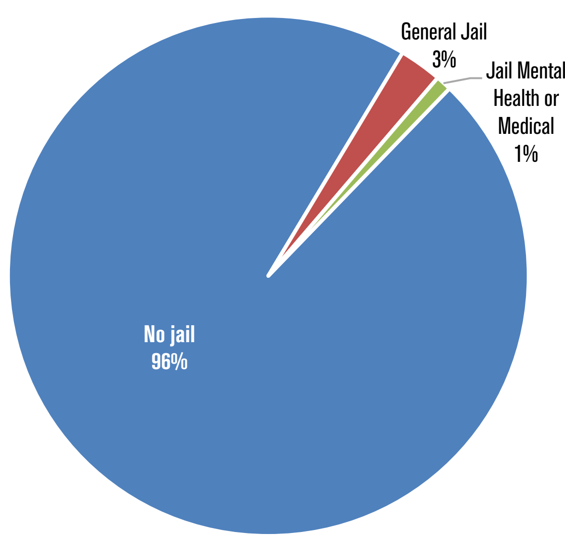
In summary, nearly half of young adults had some employment in the formal economy during the three-year study window. However, less than a fifth of youth had a job in any given month and median earnings were less than $1,000 a month when they were employed. These earnings are unlikely to be sufficient to pay for housing and living expenses. Persistently homeless youth had higher rates of employment in the formal economy than their peers who were not homeless or had short homeless stints, however, they typically were employed for fewer months and had lower earnings in the months when they were employed. Possible explanations include less ability to compete in the labor market or difficulty working regularly because of the instability inherent in homeless living conditions.

Jail

Half of youth who spent time in jail also spent time homeless.

Young adults rarely spent time in jail in the three-year study window, as shown in *Figure 33*. Only four percent were incarcerated, three percent in general jail facilities and one percent in jail mental health or medical facilities. These numbers do not represent the total history of justice system involvement because

Figure 33: Jail Time for Young Adults in the 3 Year Study Window



juvenile probation data was not available.

Persistently homeless young adults were incarcerated more frequently than their peers who were not homeless or who had short homeless stints, as shown in *Figure 34*.

Fourteen percent of persistently homeless young adults spent time in jail in the three-year study window. Ten percent were incarcerated in general jail facilities and four percent in jail mental health or medical facilities.

Incarceration rates dropped in half for short-term homeless – seven percent spent time in jail. And only three percent of youth who were not homeless spent time in jail.

The lens is reversed in *Figure 35*, which shows homeless outcomes based on jail status. Half of youth who spent time in jail also spent time homeless.

Among the one percent of youth who were incarcerated in jail mental health or medical facilities, 34 percent were persistently homeless and 21 percent had short homeless stints.

Figure 34: Jail Outcomes by Homeless Status of Young Adults in the 3 Year Study Window

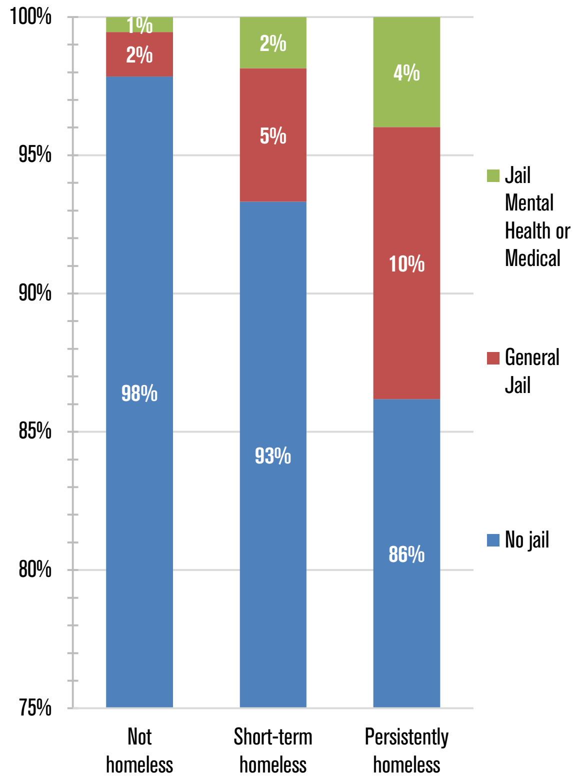
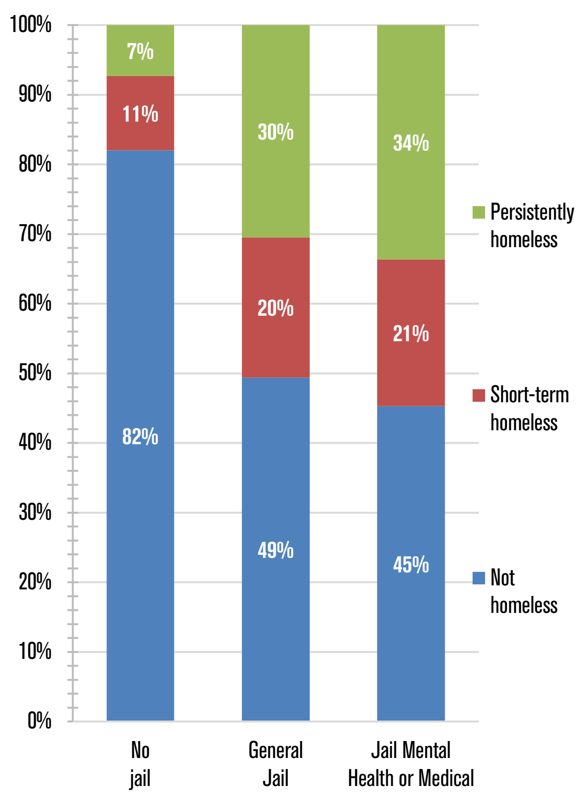


Figure 35: Homeless Outcomes by Jail Status of Young Adults in the 3 Year Study Window



African American youth were incarcerated more often than any other ethnic group.

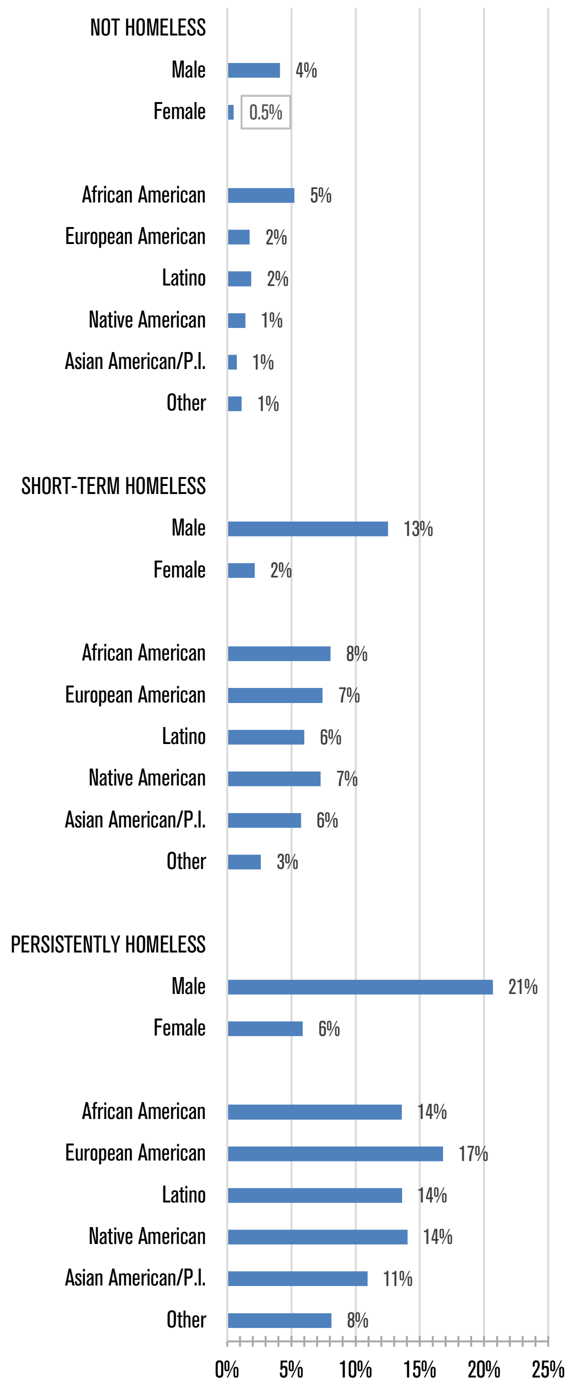
Homeless outcomes for the three percent of youth who spent time in general jail facilities were similar. Thirty percent were persistently homeless and 20 percent had short homeless stints.

There is a strong association between incarceration and homelessness.

Males were incarcerated seven times more often than females (7 vs. 1 percent), with rates rising from four percent for males who were not homeless, to 13 percent for short-term homeless, to 21 percent for persistently homeless males. These rates are shown in *Figure 36*.

African American youth were incarcerated more often than any other ethnic group. The overall rate was 8 percent for African Americans, 4 percent for both Native Americans and European Americans, 3 percent for Latinos, 1 percent for Asian Americans and Pacific Islanders, and 2 percent for Other Ethnicities.

Figure 36: Incarceration Rate for Young Adults in the 3 Year Study Window by Homeless Status, Gender and Ethnicity

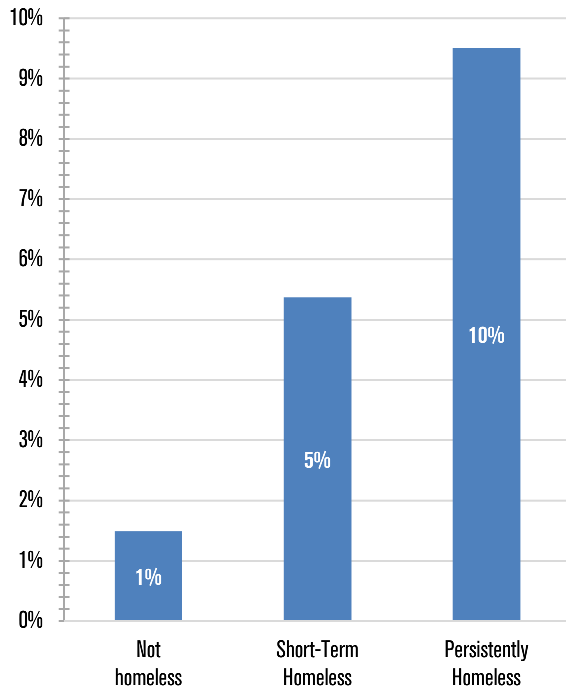


Incarceration rates increased for youth who had short episodes of homelessness, and increased still more for youth who were persistently homeless.

Seventeen percent of persistently homeless European Americans were incarcerated, followed by 14 percent of African Americans, Latinos and

Native Americans, 11 percent of Asian Americans and Pacific Islanders, and 8 percent of other ethnicities.

Figure 37: Medical Diagnosis of Substance Abuse based on Homeless Outcomes of Young Adults



Substance abuse problems increase the likelihood of justice system encounters as well as the difficulty of maintaining steady employment. The rate of medical diagnoses of substance abuse problems among youth with different homeless outcomes is shown in *Figure 37*.

The share of youth who were diagnosed with a substance abuse problem increased in proportion to the duration of homelessness. Only one percent of youth who were not homeless were diagnosed with a medical condition related to substance abuse, five percent of youth with short homeless stints, and 10 percent of youth who were persistently homeless. This distribution shows proportionate disparities based on homeless outcomes rather than actual rates of substance abuse, which are likely to be higher because many youth with substance abuse problems have not had this condition diagnosed within the county health care system.

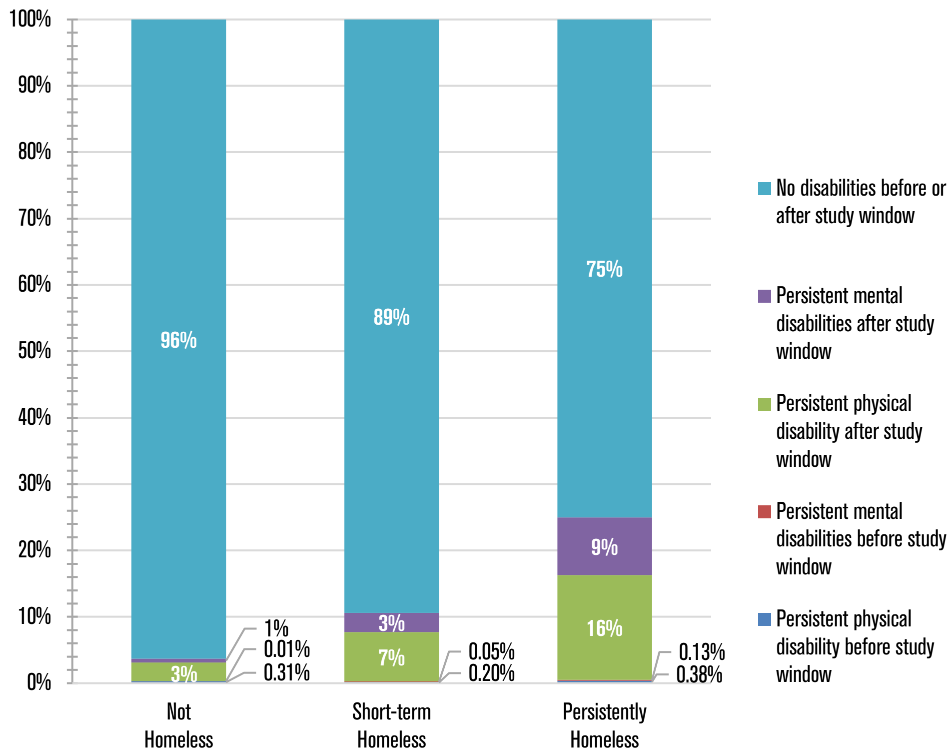
Substance abuse problems increase the likelihood of justice system encounters as well as the difficulty of holding a job.

In summary, only four percent of young adults spent time in jail in the three-year study window, but homeless outcomes were much worse for those who were incarcerated. Over half of youth who were incarcerated spent time homeless, including almost a third who were persistently homeless. Males were incarcerated seven times more often than females and African Americans were incarcerated twice as often as any other ethnic group. Substance abuse problems increase the likelihood of justice system encounters and are much more prevalent among youth who are persistently homeless.

Disabilities

National data reported by HUD shows that unaccompanied youth aged 18 to 24, have a high risk of health and mental health problems, as well as more frequent justice system encounters, as their time on the streets

Figure 38: Presence, Timing and Type of Disabilities among Young Adults



lengthens.[16] This is borne out by the rate of persistent disabilities among young adults with different homeless outcomes shown in *Figure 38*.

25% of persistently homeless youth were found to have disabilities.

Only a fraction of a percent of youth were identified as having disabilities before they entered the study window, but this changed as they progressed through young adulthood. During the three-year study window, 4 percent of young adults who were not homeless, 10 percent of young adults with short homeless stints, and 25 percent who were persistently homeless were found to have persistent disabilities.

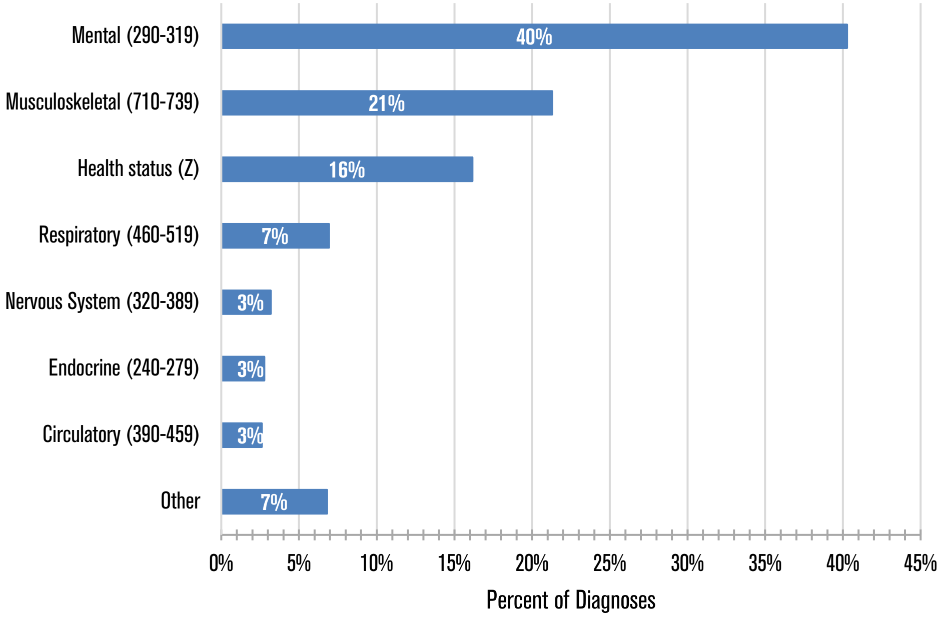
Effective early intervention for young adults who are on a path toward persistent homelessness can reduce the rapid emergence of long-term physical and mental disabilities that result from continued homelessness.

Medical diagnostic codes were available for one-fifth (19 percent) of the young adults with disability flags in their public benefits records. Their health conditions are shown in *Figure 39*.

The most frequent problems were with mental disorders, accounting for two-fifths of disabilities. Over a third (39 percent) of the problems in this category had to do with episodic mood disorders, a quarter (24 percent) with psychoses, and 17 percent with anxiety disorders.

The second most frequent problems were with the musculoskeletal system, accounting for over a fifth (21 percent) of disabilities. Four-fifths of the problems in this category had to do with joint and back conditions. Some

Figure 39: Medical Diagnoses for Persistently Homeless Young Adults with Disabilities



ICD-9-CM body system code range for diagnoses shown in parenthesis.

of these youth need to be directed to occupations that do not require heavy lifting.

The third most frequent category of problems were conditions that affect the health status of young adults and required health services. These conditions accounted for 16 percent of persistent disabilities among young adults. Orthopedic aftercare accounted for nearly all (99 percent) of these conditions.

Endocrine, nutritional, and metabolic diseases and immunity disorders were the fourth most frequent category of problems. Diabetes accounted for 85 percent of these conditions.

In summary, disabilities emerged rapidly among young adults who were homeless. A quarter of persistently homeless youth had persistent disabilities at the end of the three-year study window. The largest share of these disabilities were for mental conditions. Effective early intervention for young adults who are on a path toward persistent homelessness can reduce the rapid emergence of long-term physical and mental disabilities that result from continued homelessness.

Conclusions

Youth who become persistently homeless are far more likely to be solitary, disconnected from any family unit. Youth who experienced homelessness in six years the preceding adulthood were more than three times as likely to be homeless as young adults than those who had not previously been homeless. The risk of persistent homelessness was especially high for:

- African American youth
- Youth who had been in the foster care system
- Youth who were homeless as children
- Youth who were homeless when they enter adulthood
- Youth who had been incarcerated

It is important that screening to identify young adults who are likely to become persistently homeless be carried out in ways that effectively reach these groups with especially high-risks.

Substance abuse problems increase the likelihood of justice system encounters and are much more prevalent among youth who are persistently homeless. Many high-risk young adults need behavioral health services to overcome substance abuse problems and some need legal services to expunge or lessen their criminal justice records.

Only five percent of the young adult population spent time in the foster care system, but 13 percent of those who were persistently homeless had been in the foster care system.

The enactment of California Assembly Bill 12 in 2012 has improved outcomes for foster youth, but not eliminated the problem of homelessness. Youth who were eligible for foster care services under AB 12 had better outcomes – 16 percent of these youth experienced persistent homelessness compared to 24 percent of older foster youth who emancipated into adulthood when they were 18 years old, before the bill took effect.

Disabilities emerged rapidly among young adults who were homeless – a quarter of persistently homeless youth had persistent disabilities at the end of the three-year study window. The largest share of these disabilities were for mental conditions. Effective early intervention for young adults who are on a path toward persistent homelessness can reduce the rapid emergence of long-term physical and mental disabilities that result from continued homelessness.

Persistently homeless youth have higher employment rates but lower earnings than their peers who are not stuck in homelessness. This demonstrates a strong drive to earn enough money to pay for housing but little success in obtaining sustaining employment. Many high-risk young adults need human capital investments in the form of education and training that will enable them to compete for better jobs. They may also need wage subsidies to encourage employers to give them an opportunity to demonstrate their capabilities.

Trailer homes provided for returning veterans who did not have housing. Herald Examiner Collection, 1945. Courtesy of Los Angeles Public Library.

Public Costs

Cost Trajectories

The screening tools can help avert prolonged distress for vulnerable individuals. They can also help avoid ongoing high public costs for individuals who are persistently homeless. This chapter uses cost and service use data from the records of the two study populations to identify the local public costs for unemployed workers and young adults who are persistently homeless. *The cost factors for each public service and the sources of cost information are shown in Appendix Table A-1.*

Two broad trends shape public costs. First, within each population group, some individuals are stuck in homelessness, leading to more frequent use of public services. Most other individuals never experience homelessness or are able to quickly escape homelessness, leading to less frequent use of public services.

Public costs for homeless individuals increase as they age.

The second broad trend is that young people typically use fewer public services than older people because they are healthier and less entangled with the justice system, and because as they emancipate, they are disconnected from the social safety net for children. Health care and incarceration account for most public costs for homeless individuals, and these institutional connections become more frequent as people age.

The levels of service use among persistently homeless individuals within each of the two study populations over the three-year time window in

Figure 40: Percent of Persistently Homeless Using Services Anytime Over 3 Years: Young Adults and Unemployed Workers

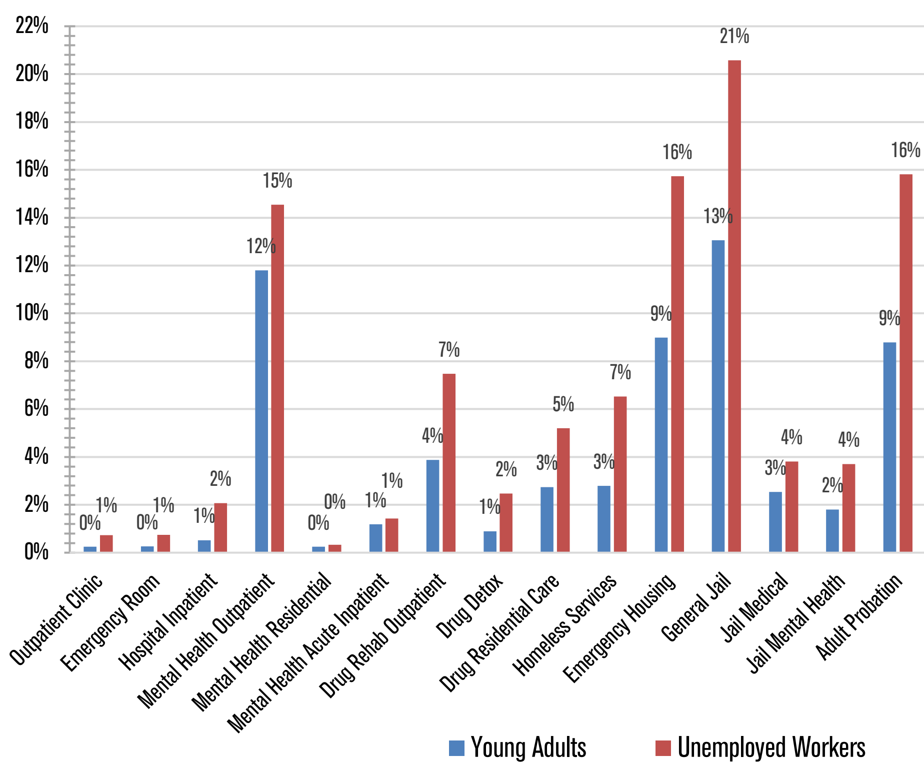
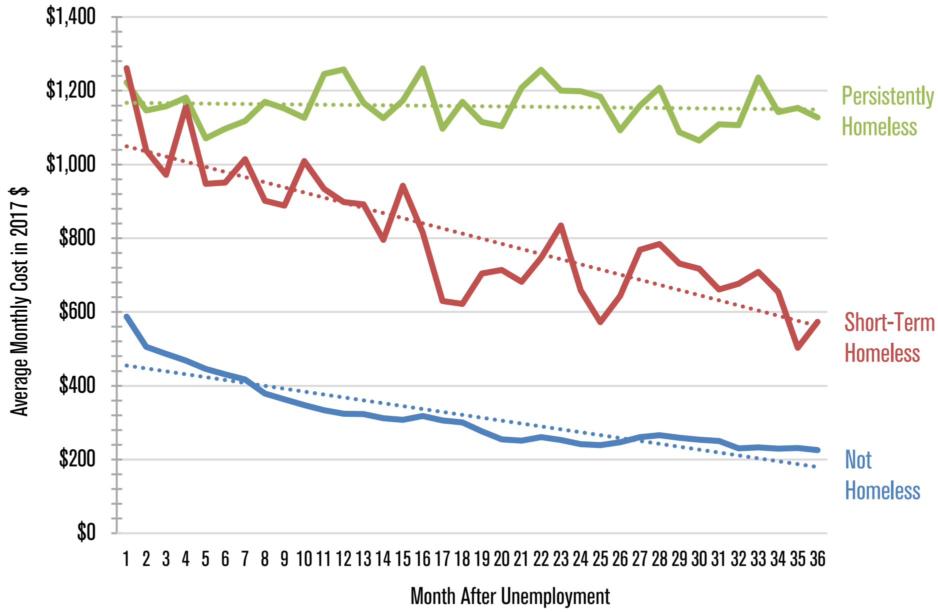


Figure 41: Trajectory of Monthly Local Public Costs for Workers who lose their Jobs



which outcomes were assessed are shown in *Figure 40*. The median age for young adults was 18.7 years at the beginning of the time window and for unemployed workers it was 36 years, so this is a comparison of the frequency with which younger versus somewhat older persistently homeless individuals used services.

Local public costs for persistently homeless workers do not decline - they stay high.

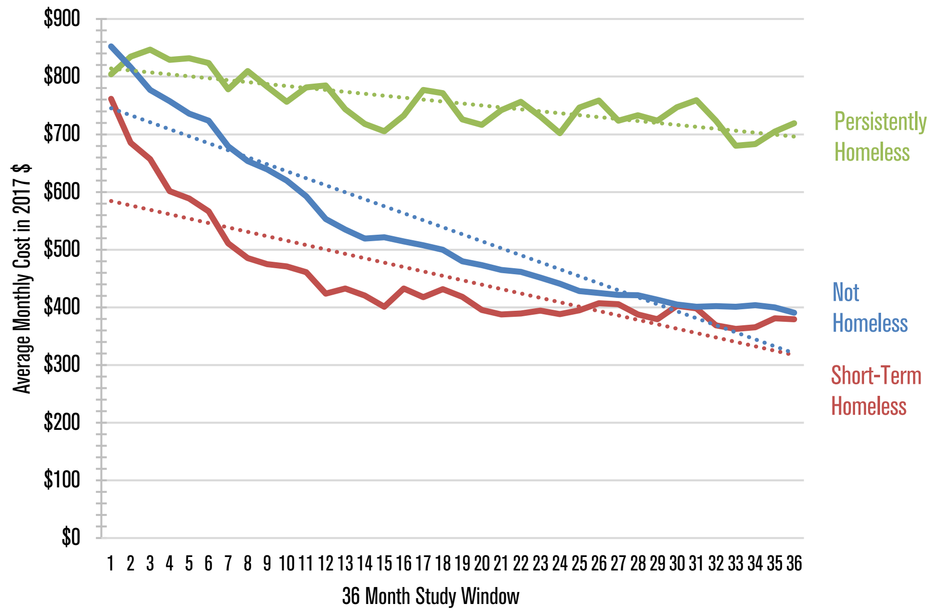
The most expensive services were used more frequently by older workers. For example: hospital inpatient care, which cost an average of $9,158 a day, was used four times more often by unemployed workers than by young adults; hospital emergency rooms, which cost $1,123 per visit, were used 2.9 times more often, and jail medical and mental health facilities, which cost an average of $1,200 a day, were used 1.8 times more often.

Unemployed Workers

Average monthly costs over the three years after workers lose their jobs are shown in *Figure 41*, broken out by homeless status. Two things stand out. First, local public costs for unemployed and persistently homeless workers are much higher than for other unemployed workers. At the end of three years, the monthly costs for persistently homeless workers were two times higher than the costs for short-term homeless and five time higher than the costs for workers who did not experience homelessness.

The second thing that stands out is that costs for these persistently homeless workers do not decline - they stay high. In contrast, costs for workers with short homeless stints are 55 percent lower at the end of three years than they were at the beginning. The decrease in monthly costs for workers

Figure 42: Monthly Local Public Costs for Young Adults during the 3 Year Study Window



who did not become homeless was even greater, dropping 62 percent by the end of three years.

Young Adults

Average monthly costs, excluding foster care, over the three-year study window for young adults are shown in *Figure 42*. Foster care is excluded because most of the five percent of youth in the study group who received foster care services were emancipate before the enactment of AB 12, which extended foster care services to age 21, and therefore were exiting the foster care system.

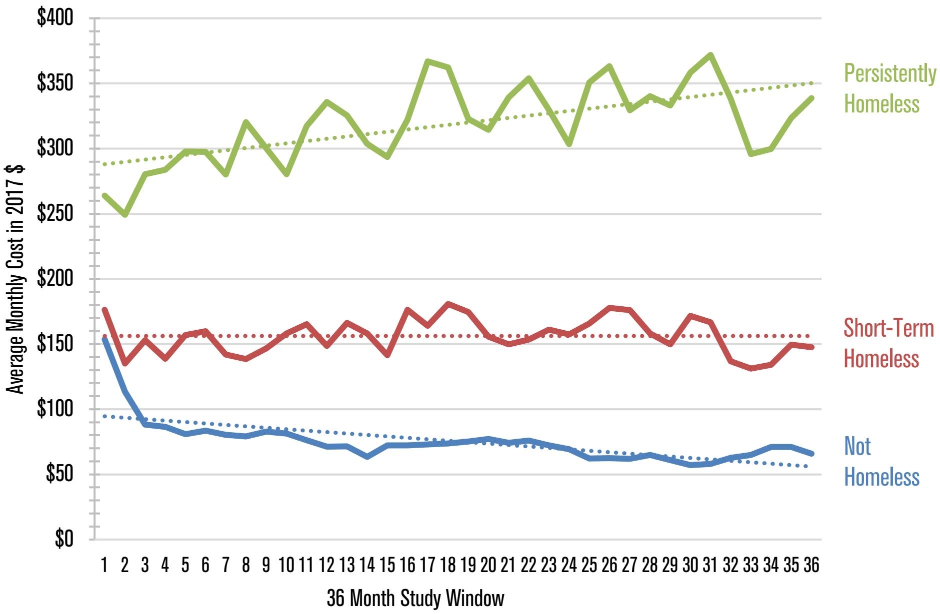
Persistently homeless youth had increasing costs for health and mental health care, substance abuse, homeless and justice system services.

Overall, public costs declined over the first three years of adulthood for these youth. However, the decline was least for youth who became persistently homeless (15 percent), versus 46 percent for youth with short homeless stings and 51 percent for youth who did not become homeless.

At the end of three years, the monthly costs for persistently homeless youth were two times higher than the costs for both short-term homeless and youth who do not experience homelessness.

When social service costs are removed, leaving health care, mental health, substance abuse, homeless, and justice system services, as shown in *Figure 43*, costs increased 21 percent for persistently homeless young adults while remaining constant for young adults with short homeless stints, and declining 42 percent for those who did not become homeless. The costs for these services over the three-year time window were comparatively low, but the upward trajectory for these high-cost services suggests a

Figure 43: Monthly Local Public Costs for Young Adults Excluding Social Services



likelihood of long-term high costs for persistently homeless youth who remain homeless for significant portions of their adult life.

In summary, public costs for persistently homeless individuals have upward cost trajectories that are likely to continue increasing as they age. In contrast, public costs for individuals with short homeless stints and even more so for individuals who do not become homeless, decrease over time. The cost difference between individuals who are persistently homeless and their peers who avoid this outcome, as well as the upward cost trajectories for persistently homeless individuals suggest that significant public costs can be avoided by intervening early to prevent persistent homelessness.

Local Public Costs after Three Years

Unemployed Workers

Local public costs for unemployed workers in the third year after they lost their jobs are shown in *Figure 44*. Three things stand out. First, annual costs were more than $10,000 higher for workers who became persistently homeless than for those who avoided homelessness.

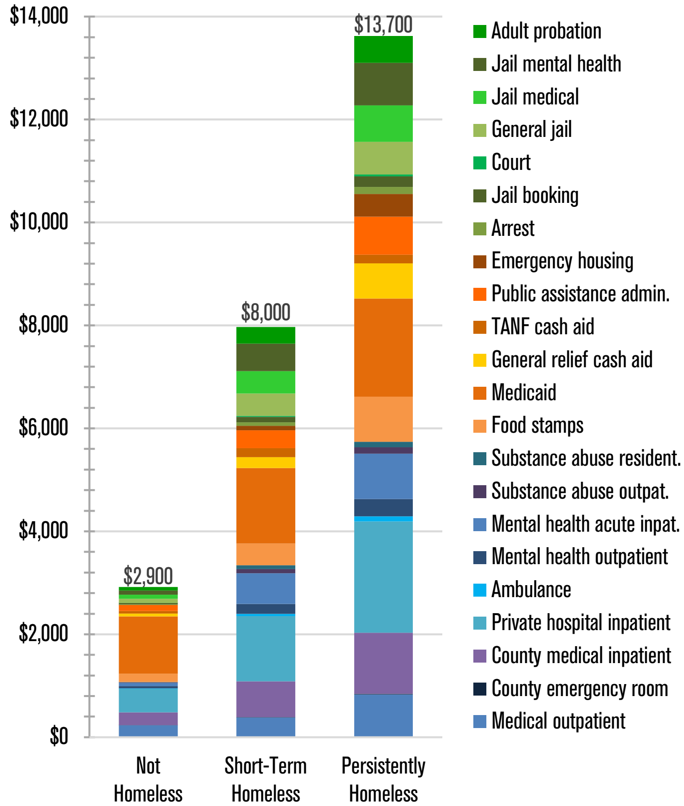
The second thing that stands out is that annual health care costs, shown by blue hues at the bottom of the columns in *Figure 44*, were $4,700, or five times, higher for workers who became persistently homeless than for workers who did not become homeless.

The third thing that stands out is that annual justice system costs, shown by green hues at the top of the columns in *Figure 44*, were $2,700, or nine

Annual costs were $10,000 higher for workers who became persistently homeless than for those who avoided homelessness.

times, higher for workers who became persistently homeless than for workers who did not become homeless.

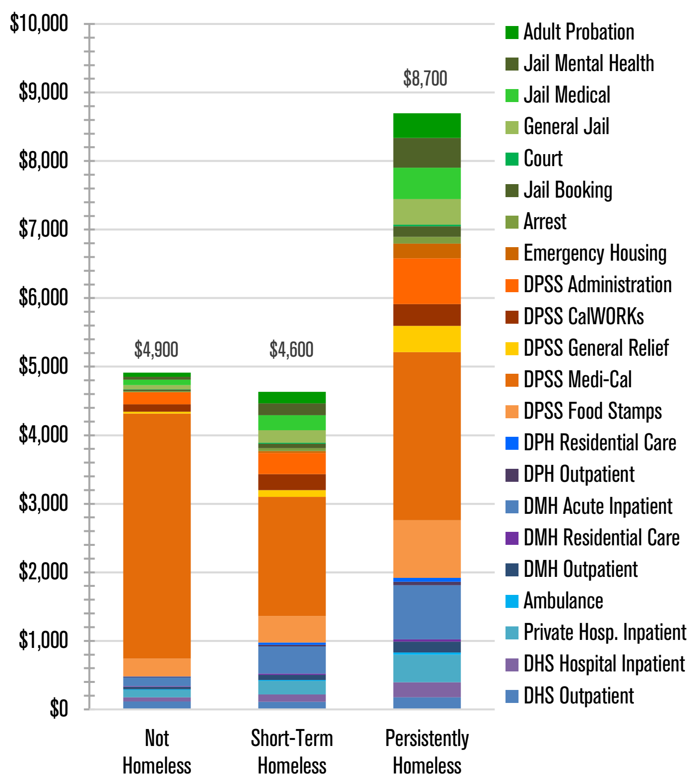
Figure 44: Total Annual Local Public Costs for Unemployed Workers in Year 3 after Unemployment by Homeless Outcome



Young Adults

Local public costs for young adults in the third year after they entered the study window are shown in *Figure 45*. Foster care costs were excluded because a majority of the youth had lost eligibility for that service.

Figure 45: Total Annual Local Public Costs for Young Adults in Year 3 of Study Window by Homeless Outcome



Note: Foster care costs not included.

Three things stand out. First, annual costs were similar for young adults who did not become homeless and those who had short homeless stints, because those who did not become homeless received more public assistance benefits than those with short homeless stints. However annual costs were more than $3,800 higher for youth who became persistently homeless than for youth who avoided homelessness.

The second thing that stands out is that annual health care costs, shown by blue hues at the bottom of the columns in *Figure 45*, were $1,400, or four times, higher for youth who became persistently homeless workers than for youth who did not become homeless.

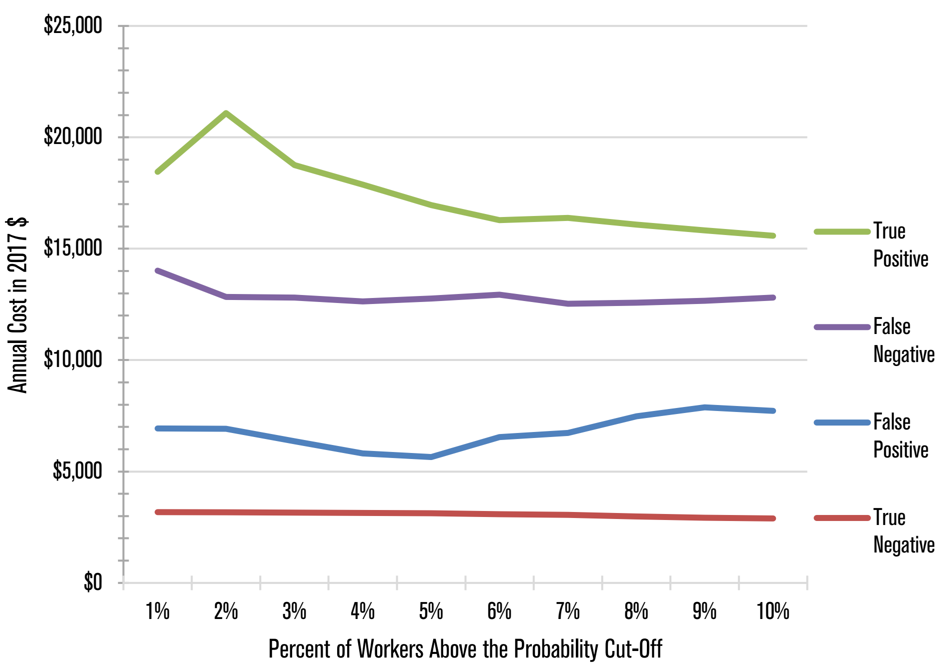
The third thing that stands out is that annual justice system costs, shown by green hues at the top of the columns in *Figure 45*, were $1,900, or seven times, higher for youth who became persistently homeless youth than for youth who did not become homeless.

Annual costs were more than $3,800 higher for youth who became persistently homeless than for youth who avoided homelessness.

Unemployed Workers in Year 3 Based On Results from Screening by Predictive Model

Each person who is screened by either of the two predictive models presented in this report will receive probability values between 0 and 1 for the likelihood that he or she will become persistently homeless. The screening tools are explained in detail in the following chapter, but the key point for understanding the impact on public costs from using the models is that the differing probabilities for each individuals can be used to rank and prioritize the entire screened population for access to a specific intervention.

Figure 46: Total Local Public Costs for Unemployed Workers in Year 3 based on Probability Cut-Off from Predictive Model Used to identify who is Eligible for Help



The models are most accurate at high probability levels. Lowering the probability cut-off point that is used to determine who will receive access to an intervention has the effect of including more high-need individuals but also of making slight, incremental increases in the share of people who do not become persistently homeless but are mistakenly included in the target population.

The dataset used to develop the model for identifying workers who become persistently homeless after losing their jobs included an average of 8,700 workers who lost jobs in the formal economy each month. Out of these monthly cohorts of job losers, 670 workers, or 7.7 percent of the total, went on to become persistently homeless.

The model is designed to predict long-term outcomes, so annual local public costs in the third year after unemployment are shown in *Figure 46*. There are four outcomes from the screening model. *True positives* are workers correctly identified as becoming persistently homeless. *False negatives* are workers who became persistently homeless but have a probability score below the cut-off level. *False positives* are workers who do not become persistently homeless but have a probability score above the cut-off level. And *true negatives* are workers who are correctly identified as not becoming persistently homeless.

Any cut-off point that includes less than 7.7 percent of the screened population above the cut-off point is automatically going to include some *false negatives* because the population prioritized for receiving help will be smaller than the population that becomes persistently homeless. At every cut-off level, the model will also produce some *false positives* because the probabilities are not completely accurate.

Table 1: Size and Local Public Cost of the Monthly Target Population Based on the Percent of the Screened Population above the Employment Model Cut-off Point

Percent of Screened Workers above the Cut-off Point for Services	Approximate Monthly Size of Target Population in Los Angeles County	Approximate Annual Public Cost for Each Worker in the Targeted Population
1 percent	90	$15,900
2 percent	170	$17,100
3 percent	260	$14,800
4 percent	350	$13,500
5 percent	440	$12,400
6 percent	520	$12,00
7 percent	610	$11,800

The potential monthly target population based on different cut-off points for the percent of job losers who are at greatest risk of becoming persistently homeless who can be served each month is shown below in *Table 1*. The share of the screened population that is shown being served ranges from one to seven percent, the number of people served each month ranges from 87 to 609, and the average annual local public cost for each person in the third year after unemployment, if there is no intervention to help them exit homelessness, ranges from $11,795 to $15,896 based on the mix of *true positives and false positives* in the population above the cut-off point.

In summary, persistent homelessness results in high local public costs that are not found among individuals who avoid homelessness or have short homeless stints. Annual costs were more than $10,000 higher for unemployed workers and $3,800 higher for young adults who became persistently homeless than for their counterparts who avoided homelessness. These costs increase over time as individuals become older. The target population identified by the employment model has ongoing high public costs. The cut-off point for determining which high-risk unemployed workers who have been prioritized through the model will receive services can be adjusted to match the capacity of programs that serve those workers.

Conclusions

Individuals who become persistently homeless use more public services and have far higher public costs than their peers who do not become homeless. These costs are ongoing and increase as individuals become older.

Health care costs were five times higher for persistently homeless workers and four times higher for persistently homeless youth than for their counterparts who did not become homeless.

Justice system costs were nine times higher for persistently homeless workers and seven times higher for persistently homeless youth than for their counterparts who did not become homeless.

Using predictive screening tools to identify high-risk individuals and intervene early before they become persistently homeless can help them avoid hardship and help the public avoid ongoing high costs from ongoing, intensive and increasing use of local services.

Works Progress Administration workers building La Brea Avenue, Los Angeles, 1936. Daily News Negative, courtesy of UCLA Islandora Repository.

Methodology

Introduction

This study presents two screening tools to predict persistent homelessness. The employment model works by predicting whether recently unemployed workers will experience persistent homelessness. The young adult model predicts whether youth who are entering adulthood while receiving public benefits will become persistently homeless. These tools make it possible to provide targeted interventions such as short-term subsidized employment or transitional youth services before these individuals enter into costly, protracted spells of homelessness.

The tools use administrative data to prioritize homeless adults with the highest risk of becoming persistently homeless. This approach requires some mechanism to accurately identify or predict which high -risk workers or young adults will become persistently homeless before there is substantial preventable personal harm and public costs, and before the crisis of being homeless has diminished their capacity to work and their identity as a member of society. The statistical predictive models presented in this report address that need.

No other study has developed models to predict persistent homelessness for low wage workers who lose their job or for youth who transition to adulthood while receiving public benefits.

There is a growing interest in using predictive models to combat homelessness by identifying high-risk homeless persons, such as the work done by Economic Roundtable in identifying individuals with high public costs (see Toros and Flaming, 2016, 2018).[17] Other studies have also identified predictors of homelessness and developed methods for providing more efficient homelessness prevention services (Bryne et al., 2015; Chan et al., 2017; Shinn et al., 2013).[18] However, to our knowledge, there is no other study that has developed models to predict persistent homelessness for either low wage workers after losing their job or youth who are transitioning into adulthood while receiving public assistance. Data sources, study populations, data preparation, variable selection, model development, results, and assessment are presented in the following sections.

Data and Populations

We used Los Angeles County administrative data for this study, as shown in *Figure 47*. All source data were de-identified. The main data source was a 10-year time window of records for public benefits recipients from the LEADER eligibility data system managed by the Department of Public Social Services (DPSS). This database provided the study population as well as information about demographics, aid, employment, and homelessness histories of individuals.

Homelessness histories were based on case addresses. If in any given month the case address was a DPSS office, homeless shelter or any other non-residential address, a person was assumed to be homeless in that month.

Figure 47: Modeling Time Line

The LEADER system also contains a homeless flag based on self-declared status filled in during an intake assessment. However, since the system does not turn off this flag at the end of a homelessness episode, and since there is a large overlap between the flag and homeless addresses at the beginning of an episode, clients' homeless status was determined based on their address. This practice is also followed by DPSS in their assessment of homelessness.

The second data source was the County's integrated Enterprise Linkages Project (ELP) database (see Bryne, et al., 2012),[19] which includes records of services provided to County clients by the departments of Children and Family Services (DCFS), Health Services (DHS), Mental Health (DMH), and Public Health (DPH). In addition, incarceration and adult probation histories of individuals were available. Finally, the Homeless Management Information System (HMIS) data was used to augment information about homelessness.

The study and tracking time windows for the training and validation data are shown by the bottom 4 arrows of *Figure 47*. In the employment model, we included all individuals who had at least one employment spell over 4 years. In the young adult model, we included all youth (18 to 24 years old) who were receiving public assistance benefits in the form of cash aid, Cal Fresh (food stamps) or Medi-Cal. The green arrow shows this window. The target population was comprised of individuals who became persistently homeless during the 36 months after becoming unemployed or their first month in aid as a young adult. This window is shown by the yellow arrow, data was tracked to identify persistently homeless individuals. The blue arrow shows the five-year staggered time window for employment and homelessness data for the study population. Finally, the gray arrow illustrates the two-year staggered time window for ELP data about service utilization from health, social service and justice system agencies. The structure of model variables is discussed later.

The size of the study and target populations for the *employment model* are shown in *Table 2*. Over four years, we identified 494,584 individuals who were employed at least once and became unemployed during this window. These persons had 673,139 employment spells since some of them were employed more than once. Almost a quarter (22.2 percent) of this population had been homeless at least once over 10 years. Over 72,000

Table 2: Study and Target Populations of the Employment Model

Population Category	Population Size	Percent	Employment Spells
Study Population	494,584	100%	673,139
Homeless at least once	109,769	22.2%	163,240
Homeless after Unemployment	72,594	14.7%	105,587
Target Population – Persistently Homeless	37,905	7.7%	58,166

individuals became homeless within three years following an unemployment incident, and almost 38,000 of them were identified as persistently homeless – 7.7 percent of the study population.

The sizes of the study and target populations for the *young adult model* are shown in *Table 3*. We identified 479,111 young adults receiving public assistance during a four-year window. Almost a quarter (24.8 percent) of this population had been homeless at least once over the 10 years of data used for this analysis. Over 106,000 individuals became homeless within three years following their entry into the time window of young adulthood, and over 39,000 of them were identified as persistently homeless – 8.2 percent of the study population.

Table 3: Study and Target Populations for the Young Adult Model

Population Category	Population Size	Percent
Study Population	479,111	100.0%
Homeless at least once	118,582	24.8%
Homeless after becoming Young Adult in Aid	106,456	22.2%
Target Population – Persistently Homeless	39,133	8.2%

In summary, the data used to develop the two models was drawn from a four-year rolling window for identifying benchmark dates when workers became unemployed or youth entered adulthood. Then, three-year outcomes for whether individuals with these benchmark events became persistently homeless were tracked in a five-year rolling follow-up window. Each of the two data sets used to develop the predictive models included nearly half a million people.

Data Preparation and Variable Selection

We integrated several data sources using multi-tiered fuzzy matching algorithms. All these datasets include information on factors that may have an effect on our outcomes of interest—becoming persistently homeless following becoming unemployed or first month in aid as a young adult. These include demographic variables (e.g., age, gender, ethnicity); clinical variables (e.g., ICD-9-CM medical diagnoses), and utilization variables for all types of services in the current and previous years (e.g., number or days of hospital stays, number of emergency room visits, number of mental health service encounters, days in jail, and number of incarcerations).

First, we generated a binary target or outcome variable for each model. In the employment model, it flags whether or not a person became persistently homeless after becoming unemployed. Persistent homelessness was defined as a person becoming homeless more than once or continuously for 12 or more months within three years after becoming unemployed. In the young adult model, the target variable flags whether or not a youth became persistently homeless after the first month of receiving assistance as a young adult. Persistent homelessness was defined the same way—a person who became homeless more than once or continuously for 12 or more months within three years after becoming a young adult while receiving public assistance.

The next step was to identify any potential variables that would have an effect on becoming persistent homeless for each model. Since each data source has many variables, this step required a laborious process to prepare all potential variables for the variable selection procedure.

We prepared the data by transforming variables to augment their predictive power. For example, continuous fields may be binned (such as the age category, which was modified into 3 groups—18 to 40, 41 to 57, and 58 or older. Binning is the process of reducing the number of levels of a predictor to a smaller number of bins (i.e. consolidations of levels to achieve parsimony and to find a relationship between the bins and the target rate (See Lund, 2016).[20] Some categorical variables were clustered such as ethnicity and diagnostic codes. A majority of the variables were transformed into binary (1 or 0) variables, for example, whether or not an individual had been hospitalized in the last year. These variables equal 1 if a condition exists (such as hospitalization) and 0 if the condition does not exist. All these binary variables were generated for the current and previous years. We generated many count variables that show the number of occurrences of a variable such as emergency room visits and days in probation. All count variables were also generated for the current and previous years. For homelessness and employment variables we used data going back five years. For other service utilization variables we included one or two years of history.

Data preparation was followed by the variable selection process, which is the method of selecting a particular set of predictors or independent variables for use in predictive models. The main objective of variable selection is to choose a reduced number of attributes that improve the accuracy of the prediction and to remove unneeded, irrelevant and redundant variables. The process also provides a better understanding of the model and generates simpler variables that can be computed more quickly (See Guyon and Elisseeff, 2003).[21] A parsimonious model is desirable because fewer variables reduce complexity, so a model becomes easier to understand and explain.

Predictive models can easily be beset by dimensionality and overfitting to minor or even random variables. Goodness-of-fit must be balanced against model complexity in order to avoid overfitting—that is, to avoid building models that explain the data at hand, but fail in out-of-sample predictions (Vandekerckhove, Matzke and Wagenmakers, 2015).[22] In the predictive analytics practice, applying first a method of automatic variable construction yields improved performance and a more compact set of variables. There are a number of commonly used methods that were applied in this study.

Filter methods assess the relationship between predictor variables and the target variable to compute the importance of variables. Various statistical methods such as correlation analysis or F-test can be used to measure the predictive power of single factors. Wrapper methods find the best combinations of variables to determine predictive power by applying different approaches such as forward, backward and stepwise selections that are explained below. Finally, embedded models such as Least Absolute Shrinkage and Selection Operator (LASSO) perform variable selection as part of the model construction process or, in other words, select variables as part of learning. Random forest is another embedded model that was applied in this study.

We used multiple variable selection methods in developing the models. Following data preparation, which generated over 350 potential predictors, we first selected relevant diagnostic codes and service factors. To do this we combined our experience in Los Angeles with statistical tests of association--applying chi-squared and t-tests to verify if any of these factors help separate persistently homeless persons from others. This step used the filter method of variable selection described above. After eliminating redundant and irrelevant factors we reduced the list of predictor variables from 350 to approximately 200. The list of these variables is shown in *Appendix Table A-2* for both models.

In the second iteration, we applied forward and backward selection and LASSO[23] methods to reduce our variable set further. The forward selection technique begins with only the intercept and then sequentially adds the variable that most improves the fit. The process terminates when no significant improvement can be obtained by adding any variable. In contrast, the backward elimination technique begins by calculating statistics for a model, including all of the independent variables. Then variables are deleted from the model one by one until all of the remaining variables are statistically significant at a specified level. At each step, the variable showing the smallest contribution to the model is deleted.[24]

Automated selection methods are often criticized for producing biased results (See Flom and Cassell, 2009).[25] However, advanced versions of these methods, particularly applying the Schwarz Bayesian information criterion (SBC) statistic or using a validation sample, generate accurate results (See Dziak et al., 2012, Lund et. al. 2017).[26] Using the SBC statistic as the selection and stopping criteria causes the predictor to be added that gives the lowest (best) SBC for the new model among all predictors currently available or removes predictors that produce the largest (worst) value of the SBC statistic. This stops at the step where adding or removing any variable increases the SBC statistic. SBC is a widely used penalized measure of fit for logistic regression models that favors the selection of a parsimonious model and avoids over fitting (see Judge et al., 1985).[27] The LASSO method applies a regularization process by penalizing variables, shrinking the coefficients of less important variables to zero. Only variables that have non-zero regression coefficients are selected while the values of the selected coefficients and penalty term minimize the prediction error (Fonti and Belitser 2017, Tibshirani 2011).[28] Using a combination of these three methods we reduced our predictor variable list from over 200 to 52 for the employment model and to 60 for the young adult model. The variables selected in the second round of the selection process are shown in *Appendix Table A-2* for both models.

We used the SAS high-performance procedure HPGENSELECT with the binomial distribution and logit link function to apply forward and backward selection with SBC and LASSO methods. (See Johnston and Rodriguez, 2014, SAS 2017).[29] The HPGENSELECT procedure is designed for predictive modeling. It provides variable selection methods for building models, and it supports standard distributions and link functions for generalized linear models.

In summary, over 350 potential predictors of persistent homelessness were developed and then narrowed down to 52 variables for unemployed workers and 60 variables for young adults for use in the next stage of model development.

Model Development

The models were developed to be transparent so that it is possible to explain how specific types of information are used to make predictions.

The variable selection process yielded over 50 potential variables to be trained in our predictive models. In the next step, we built several models to predict future persistent homelessness following an unemployment incidence or a youth's transition into adulthood while receiving public assistance. We used the high-performance SAS procedures HPLOGISTIC and HPFOREST to develop and assess predictive models (See Nord and Keeley, 2016, SAS 2017).[30]

Regression models are the mainstay of predictive analytics. The focus lies on establishing a mathematical equation as a model to represent the interactions between the several factors that are associated with an outcome, such as persistent homelessness. Based on the performance of predictive models in our past research and the need for interpretability and transparency, we adopted logistic regression models to predict future persistent homelessness. The results described in the next section represent the outcomes of these models. We also compare the performance of the logistic regression employment model to the random forest employment model, which is an increasingly common method used by scientists for prediction. This assessment is presented later.

One critical aspect of our model development methodology is avoiding the application of black-box analytics. Black-box refers to algorithmic predictive modeling techniques, particularly machine-learning techniques such as neural networks, k-nearest neighbors and support vector machine algorithms that do not explain their reasoning or explain it in a limited way. These algorithms are very useful for classification and prediction. However, they do not explain how given types of information are used to make predictions and are ill-suited for work where transparency is critical.

Our approach is consistent with conclusions reached in using predictive models to suggest medical diagnoses to clinicians. Diagnostic results from predictive models should not appear as black boxes, but rather should allow clinicians to explore the reasons for proposed diagnoses and provide feedback (Wang et al., 2018).[31] In this study, our focus is not only on prediction but also on interpretability and transparency (see Shumueli, 2010).[32]

To make the screening tools understandable and credible to the general public it is important to have reasonable explanations for how information is being used to make predictions. Since our predictive models are intended to be screening tools, we need to know which factors contribute to the final score that prioritizes workers or young adults for special assistance as well as the weights assigned to the variables that produced the score. Moreover, the predictive models require data elements from multiple public service domains ranging from hospitals to jails. Knowing the importance of input factors used in the models is critical for managing the logistics of data integration when the models are implemented. Consequently, we chose to develop logistic regression models that clearly explain the classification or decision process.

Logistic regression predicts the values of a discrete variable (persistently homeless or not) based on known values of multiple variables (see Allison, 2012).[33] In a nutshell, logistic regression models the probability of a binary outcome given various input variables. It transforms prediction probabilities with values ranging from 0 to 1 using the logistic function.

The performance of the logistic regression models is presented in the Validation and Assessment Section and is quite robust. Moreover, this performance is compared against the random forest model to assess if the predictive power is good in comparison. The results show that the logistic regression model produces comparable prediction accuracy without giving up transparency.

The screening tools are intended to be used by agencies that serve workers who have recently become unemployed or adolescence youth who receive public benefits. Ideally, the models can be implemented as system-based screening tools, generating risk scores from screening an entire integrated database and flagging high-risk individuals. While system-based implementation would be the most efficient mode, the tools can also be used to screen clients individually using a simple interface like Excel. Simple models with easy-to-populate variables are especially important if the models are used to screen people individually.

Keeping this requirement a priority, we performed a sensitivity analysis on several variables in the model that might not be available when doing person-by-person screening. Consequently, we dropped the number of variables from 52 to 32 for the employment model and from 60 to 20 for the young adult model, particularly eliminating all medical diagnostic variables that did not contribute much to model accuracy but would be difficult to enter into a manual tool. The results from these final models are presented in the next section.

In summary, multiple models were built and tested to predict future persistent homelessness following unemployment or a youth's transition into adulthood while receiving public assistance. The models were developed to be transparent so that it is possible to explain how specific types of information are used to make predictions. To make the models as simple and usable as possible, only 32 variables were used in the final employment model and 20 variables in the final young adult model.

Results

Employment Model

The frequency with which each variable used in the employment model is found among persistently homeless unemployed workers versus other unemployed workers is shown in *Appendix Table A-3*. The concordance index (C-statistic) was used to assess the predictive strength of the model. Significance of the estimated parameters (p-values) and odds ratios were evaluated as well. The odds-ratios for the likelihood of persistent homelessness based on the presence of each variable used in the employment model are presented in *Appendix Table A-4*.

The *Parameter Estimates* shown in *Table A-4* are the factors that drive the employment model.

As shown in *Appendix Table A-3*, persistently homeless workers included a much higher proportion of males, African Americans and single-individual households than the overall population that experienced unemployment. Their employment history was relatively shorter and average and maximum earnings were much lower than the rest of the population. The largest differences are observed in homelessness measures. While 42.5 percent of the persistently homeless group experienced homelessness during the year before they became unemployed, only 4 percent of others were homeless during that time. In the month preceding unemployment, almost 30 percent of the persistently homeless group was homeless in contrast to less than 2 percent of other unemployed workers.

Persistently homeless workers also showed higher rates of engagement with health and behavioral health services. There were large group differences for emergency medical service encounters (9 vs. 3 percent), outpatient medical clinic visits (14 vs. 8 percent), and outpatient mental health services (4.5 vs. 1.5 percent). The proportion who were disabled at the time of unemployment or had a disability history was also much higher among the persistently homeless group.

The rate of engagement in the criminal justice system was very high among persistently homeless workers compared to other workers who became unemployed. Over 20 percent were jailed during the last year compared to only 5 percent of other workers. Their average number of days in jail was more than 4 times greater than for the rest of the population—7.2 days vs. 1.6 days.

Finally, social services data showed that a very high portion of persistently homeless workers were receiving cash aid at the time of unemployment (42.3 percent), while 75 percent of other workers received only non-cash aid such as Medi-Cal or Food Stamps/SNAP.

Adjusted odds ratios are presented in *Appendix Table A-4*. All variables are statistically significant at the one percent level. The results reflect the differences we observe from the descriptive comparisons discussed above. Logistic regression models generate odds ratios that are used to assess the likelihood of a particular outcome (being a persistently homeless person in this study) if a certain factor (one of the model variables) is present. It is a relative measure showing how likely a person with a certain attribute (say, male) is to experience the outcome (persistently homeless) relative to another person without the attribute (female). In this way we capture the strength of relationship between the factor (gender) and the outcome. Adjusted odds ratios are generated after controlling for all other variables in the model, which means holding all other factors constant.

Odds ratios for binary variables (for example, jailed or not) are in general higher than the odds ratios for interval variables (for example, days in jail) and are interpreted differently. *Appendix Table A-4* shows whether a variable is binary, nominal or interval to assist the interpretation of odds ratios.

For example, the odds ratios show that workers who had been jailed in the past two years are 1.82 times more likely to be persistently homeless in the future than workers who had not been jailed. On the other hand, the odds ratio for each additional 10 days of jail is only .979, decreasing the likelihood (or odds) of being persistently homeless by 2 percent.

Adjusted odds ratios show that being younger than 58, African American, Alaskan American or American Indian and belonging to a single individual household significantly increased unemployed workers' odds of becoming persistently homeless. Being homeless in the past (particularly in the last year or the last month before becoming unemployed) yielded very strong odds ratios. In general, recent employment decreased the odds of becoming persistently homeless in the future, while having health or behavioral health issues increased the odds—except that medical outpatient services decreased the odds. Criminal justice involvement and not receiving any form of public assistance at the time of unemployment also increased the odds of becoming persistently homeless in the future.

We summarize the effects of all variables in *Appendix Table A-5*. It lists the effects estimated by the model and gives a plot of the LogWorth values for these effects. The LogWorth for each model effect is defined as -log10 (p-value). This transformation adjusts p-values to provide an appropriate scale for graphing. The table shows that the most statistically significant variables are household type, type of public benefits at time of unemployment, if homeless a month prior to unemployment time, if homeless last year, if arrested last year, ethnicity, marital status, age, disability, and amount of earnings in the last year.

Young Adult Model

The frequency with which each variable used in the young adult model is found among persistently homeless youth versus other youth is shown in *Appendix Table A-6*. The odds-ratios for the likelihood of persistent homelessness based on the presence of each variable used in the young adult model are presented in *Appendix Table A*-7.

The *Parameter Estimates* shown in *Table A*-7 are the factors that drive the young adult model.

As shown in *Appendix Table A-6*, persistently homeless youth include a much higher proportion of African Americans than the overall population of young adults receiving aid. The largest differences are observed in homelessness measures. While almost 60 percent of persistently homeless youth experienced homelessness at the first month in aid as a young adult, only 7 percent of other youth were homeless during that time. While 10 percent of persistently homeless youth experienced homelessness in the past year, only 1.4 percent of other youth experienced homelessness the past year.

Persistently homeless youth also had higher rates of engagement with health and behavioral health services. There were significant differences between persistently homeless youth and other youth in terms of using outpatient mental health services (3 vs. 1.5 percent), mental health services

(8 vs. 3 percent), and alcohol or substance abuse services (2 vs. 0.3 percent). The proportion who were disabled at the time of entry into adulthood while receiving public benefits was also much higher among persistently homeless youth (12 vs. 2 percent).

The rate of engagement in the criminal justice system was very high among persistently homeless youth compared to other youth. Over 10 percent were jailed during the past year compared to only 2.6 percent of other youth. Social services data showed that a higher portion of persistently homeless youth were receiving cash aid when they entered adulthood (30.6 percent) compared to 80 percent of the other youth who only received only non-cash aid such as Medi-Cal or Food Stamps/SNAP.

Finally, engagement with the foster care system was more frequent among persistently homeless youth. Thirteen percent of persistently homeless youth were in foster care while 96 percent of other youth were not.

Adjusted odds ratios are presented in *Appendix Table A*-7. The results reflect the differences we observe from descriptive comparisons. Odds ratios for binary variables (for example, jailed or not) are in general higher than the odds ratios for interval variables (for example, days in jail) and are interpreted differently. *Appendix Table A*-7 shows whether a variable is binary, nominal or interval to assist the interpretation of odds ratios.

Adjusted odds ratios show that being African American, Alaskan American or American Indian, or belonging to a single-person household significantly increased a youth's odds of becoming persistently homeless. Being homeless in the past, particularly in the first month of being a young adult receiving public assistance, yielded very high odds ratios. In general, recent employment, having behavioral health issues, being arrested in the past, receiving cash aid, and foster care placements also increased a youth's odds of becoming persistently homeless in the future.

We summarize the effects of all variables in *Appendix Table A-8*. It lists the effects estimated by the model and gives a plot of the LogWorth values for these effects. The table shows that the most statistically significant variables are being homeless at the time of entering adulthood, ethnicity, type of public benefits being received, disability status, foster care history, and arrest history in the past year.

Overall, the model predicts future persistent homelessness very well based on outcomes produced from the data set used to develop the model. However, in predictive analytics it is necessary to evaluate the out-of-sample prediction power as well, that is, prediction power for cases other than those used to develop the model. The next section presents the validation results.

In summary, the variables that have predictive power in the models also identify attributes associated with persistent homelessness. Both unemployed workers and *young adults who became persistently homeless included a much higher proportion of African Americans and higher rates of engagement with health and behavioral health services and the criminal justice system than the overall populations that they were part of. Persistently homeless workers also included a much higher proportion of males and single-individual households, and their employment histories were shorter and earnings lower than those of other workers. Persistently homeless youth were much more likely to have experienced homelessness during the first month in aid as a young adult, to have been in foster care, and to have entered adulthood with disabilities.*

Validation and Assessment

In predictive analysis, the biggest danger to having a model that produces generalizable results is overfitting the training data, which produces over-optimistic estimates of predictive accuracy. A good way to avoid this problem is to partition the data into a training and validation set. We then evaluate model performance not on the training set, that is, the data used to build the model, but rather on a holdout or validation sample that the model "did not see." We often observe strong predictive power based on in-sample performance if the model over-fits the data. In those cases the model only explains well the training data, and out-of-sample performance is very poor. Since a predictive model is intended to be applied to new data with unknown outcomes, validation is needed to assess a model's performance. Out-of-sample validation enables us to identify overfitting if the performance is significantly better with the training data set than with the validation data set.

The creation of a holdout sample can be achieved in several ways. The most commonly used method, which was adopted in this study, is a random partition of the sample into training and holdout sets. Having large data sets of nearly half a million individuals for both the employment and young adult models, we held out half of the data for validation so that we fit each model to half of the data and validated it on the other half.[34] Since the data often has multiple records produced at different times for the same individuals, we partitioned the data randomly by individuals so that the same individual did not appear in both the training and validation samples.

We next present statistics to measure model performance using the validation sample. Then we compare the performance of the logistic regression and random forest models.

Model Fit Statistics for the Employment Model

The employment model achieved a very strong C-statistic, .894, which is the probability that the predicted outcome is better than chance. The C-statistic is used to compare the goodness of fit of logistic regression models. Values for this measure range from 0.5 to 1. A value of 0.5 indicates that the model is no better than chance at making a prediction of membership in a group (in this case, the persistently homeless group). A value of 1 indicates that the model perfectly identifies those who are within a group and those who are not. Models are typically considered reasonable when the C-statistic is higher than 0.7 and strong when it exceeds 0.8 (Hosmer and Lemeshow 2000).[35]

Another widely used measure of model performance is the *Average Square Error or Brier Score,* which is the mean squared difference between the predicted probability and the actual outcome. The lower the *Brier score* is for a set of predictors, the better the classification performance of the model. (Zero is a perfect score.) The *Brier score* for the model is 0.057, which is also a very strong statistic. Moreover, the performance measures were almost identical for training and validation samples indicating that over-fitting was not a problem.

Predictive Performance for the Employment Model

In addition to model fit statistics, we used *sensitivity*, *specificity*, *positive predictive value* (*PPV*), a*ccuracy*, *area under the receiver operating characteristics* (*ROC*) *curve*, and *lift curve* to assess the out-of-sample model performance. All these values are presented for different percentiles of the validation sample for the employment model in terms of predicted risk—top 1 percent, 5 percent, 10 percent, and so on. *Table 4* presents sensitivity, specificity, PPV, and accuracy statistics for different cutoff points for the validation (out-of-sample) employment model cohort.[36]

The *percentile* identifies the percent of the screened population that is targeted for services.

The *probability cut-off* is the minimum score from the employment model that is required to be in the group that is targeted for services.

The *sensitivity statistic* measures the proportion of future persistently homeless persons correctly identified by the model with high scores (scores above the cutoff). It is also known as the true positive rate and reflects how well the model performs in identifying people who become persistently homeless in the future after becoming unemployed. The *specificity statistic* measures the proportion of not-persistently homeless persons correctly identified by the model with low scores (scores below the cutoff). If the level is too low, this is translated into to a high false positive rate (1-specificity) meaning a large number of not persistently homeless persons would be incorrectly identified as having a high risk of becoming persistently homeless.

Table 4: Predictive Performance of the Employment Model (Validation Results)

Percentile	Probability Cut-off	Sensitivity	1 - Specificity	Accuracy	PPV	Cumulative Population
1%	0.8430	9.2%	0.2%	91.8%	81.2%	3,338
2%	0.7295	17.3%	0.5%	92.2%	76.3%	6,674
3%	0.6225	24.7%	0.9%	92.5%	72.7%	10,003
4%	0.5280	30.8%	1.4%	92.6%	68.1%	13,334
5%	0.4420	36.3%	2.0%	92.6%	64.3%	16,661
10%	0.2145	54.9%	5.7%	90.9%	48.5%	33,358
15%	0.1435	66.9%	10.0%	88.0%	39.4%	50,112
20%	0.1035	75.5%	14.6%	84.5%	33.3%	66,768
25%	0.0770	81.9%	19.5%	80.6%	29.0%	83,424
30%	0.0605	86.4%	24.5%	76.5%	25.5%	99,853
35%	0.0475	89.9%	29.7%	72.1%	22.7%	116,665
40%	0.0375	92.5%	34.9%	67.6%	20.5%	133,164
50%	0.0270	95.8%	45.5%	58.2%	17.0%	166,373
75%	0.0135	99.0%	72.2%	34.1%	11.7%	248,443
100%	0	100.0%	100.0%	8.8%	8.8%	333,308

The *accuracy statistic* is the proportion of observations that are correctly classified. It measures the proportion of true positives and true negatives out of all persons.[37]

The *PPV statistic* estimates the accuracy of the model by measuring the proportion of true positives (correctly classified future persistently homeless persons) within the population predicted to become persistently homeless. In other words, it is the probability that persons with a high score truly became persistently homeless. If PPV equals 1 this means that the model identifies all persistently homeless persons correctly with no false positives. The higher the false positives, the lower the PPV.

The first column of *Table 4* shows the percentile of the population sorted by descending order of predicted probability of becoming persistently homeless, which is shown in the second column. Percentiles are computed based on the total population of the validation sample, 333,308. For example, the first row shows the results for the top 1 percent or the first percentile, and the sixth row shows the measures for the top 10 percent or the first decile.

If the top 3 percent of persons at risk of becoming persistently homeless are considered, we see that the model identifies approximately 10,000 individuals who are predicted to become persistent homeless in the future. We know that out of 333,000 employment spells in the validation sample approximately 29,500 of them became persistent homeless in the future (8.8 percent). The probability threshold is 62.2 percent. Twenty-five percent sensitivity shows that the model captured 25 percent of the 29,500 when targeting the top 3 percent, which is quite impressive. The 1-specificity value is only 1 percent, verifying that the model correctly identifies 99 percent of those who do not become persistently homeless in the future.

The PPV value of 72.7 percent and accuracy value of 92.5 percent for the top 3 percent are also very high. The model achieves a PPV result of almost 73 percent, meaning that out of 10,000 persons that the model predicted to be persistently homeless, almost three quarters are true positives and the remaining portion is false positives. PPV is an important measure for assessing the effectiveness of the model. At higher probability thresholds PPV increases, but at the cost of lower sensitivity values. At lower probability thresholds, sensitivity increases, but at the cost of lower PPV values or higher numbers of false positives.

Table 5: Prediction Performance showing Predicted Homeless Populations

Percentile	Probability	Cumulative Population	True Positives	False Positives	Homeless False Positives	PPV	Adjusted PPV
0	NA	0	0	0	0	0%	0%
1	0.8430	3,338	2,711	627	273	81.2%	89.39%
2	0.7295	6,674	5,093	1,581	677	76.3%	86.45%
3	0.6225	10,003	7,269	2,734	1,112	72.7%	83.78%
4	0.5280	13,334	9,087	4,247	1,604	68.1%	80.18%
5	0.4420	16,661	10,710	5,951	2,150	64.3%	77.19%
10	0.2145	33,358	16,173	17,185	5,019	48.5%	63.53%
15	0.1435	50,112	19,729	30,383	7,970	39.4%	55.27%
20	0.1035	66,768	22,264	44,504	10,297	33.3%	48.77%
25	0.0770	83,424	24,153	59,271	12,104	29.0%	43.46%
30	0.0605	99,853	25,472	74,381	13,601	25.5%	39.13%
35	0.0475	116,665	26,510	90,155	14,911	22.7%	35.50%
40	0.0375	133,164	27,266	105,898	16,021	20.5%	32.51%
50	0.0270	166,373	28,227	138,146	17,570	17.0%	27.53%
75	0.0135	248,443	29,168	219,275	19,454	11.7%	19.57%
100	0	333,308	29,473	303,835	20,217	8.8%	14.91%

If we consider 10,000 *randomly* chosen persons, the PPV value would be only 8.8 percent, which is the ratio of true positives to the population size. This means that a random selection without using any knowledge or model would yield only 8.8 percent true positives. The remaining 91.3 percent would be false positives. When we compare this number to the model PPV for 10,000 persons (73 percent), we get a ratio of 8.2 which shows that the model is performing more than 8 times better than random selection at the 3 percent threshold. This measure is known as the *lift* of a model and shows the effectiveness ratio between the results obtained with and without the predictive model.

The trade-off to be weighed when using the model is between, on the one hand, using lower probability thresholds in order to identify as many persistently homeless individuals as possible while accepting a substantial number of not persistently homeless individuals as part of the mix, and, on the other hand, using higher probability thresholds to identify a smaller population in which a higher proportion of individuals will be persistently homeless. The model is highly accurate in distinguishing persistently homeless individuals from others. However, it is still necessary to calibrate the probability cut-off level that will be used to determine who within the screened population will be offered the intervention.

Table 5 provides insights for making this decision by tabulating the number of false and true positives at different probability levels. In many interventions the selection of the threshold is based on the capacity and funding of the program. Hence, for example, if the goal of the program is to serve 10,000 persons, then the appropriate threshold would be .6225. If the goal is 50,000 then the threshold would be .1435.

Table 5 shows that the model performs very well for the top 5 percent and fairly well for the top decile (10 percent). For the top 5 percent, 60 percent of the targeted 16,673 persons were true positives. The PPV value drops to below 50 percent at the 10th percentile.

Table 5 also includes a column labeled "Homeless False Positives," which represents false positives that fell short of becoming persistently homeless in the next 3 years but were observed to be homeless starting from 6 months following the time of unemployment. They had one homeless episode of less than 12 months. When we add these numbers to true positives, PPV values shown as "Adjusted PPV," values increase significantly. At the top 3 percent threshold, approximately 84 percent of persons were observed to be homeless after the unemployment incidence. At the top decile threshold PPV increases from 48.5 percent to 63.5 percent. Hence the data shows that at high thresholds the model identifies future persistently homeless individuals accurately. Furthermore, a significant proportion of false positives also become homeless with a risk of becoming persistently homeless after 3 years.

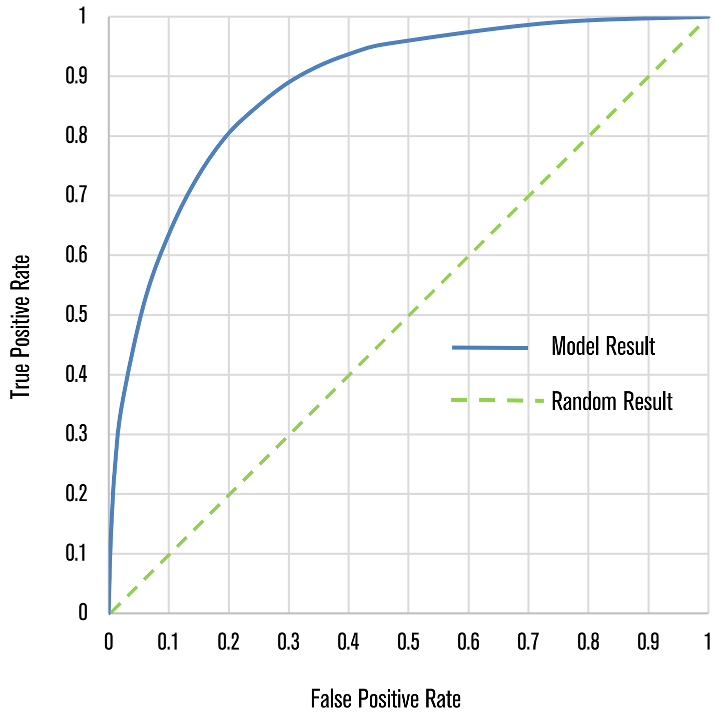
The models predict future persistent homelessness very accurately.

ROC and Lift Curves for the Employment Model

Another way of assessing the predictive power of a logistic regression model is the area under the ROC curve, which shows the trade-off between true positives (sensitivity) and false positives (1-specificity) at all possible thresholds. (See Gonen, 2007 for ROC analysis for predictive models.)[38] The ROC curve for the employment model is shown in *Figure 48*.

Figure 48: Prediction Results for Unemployed Workers becoming Persistently Homeless in the Next 3 Years

ROC Curve: Area Under the Curve = 0.89



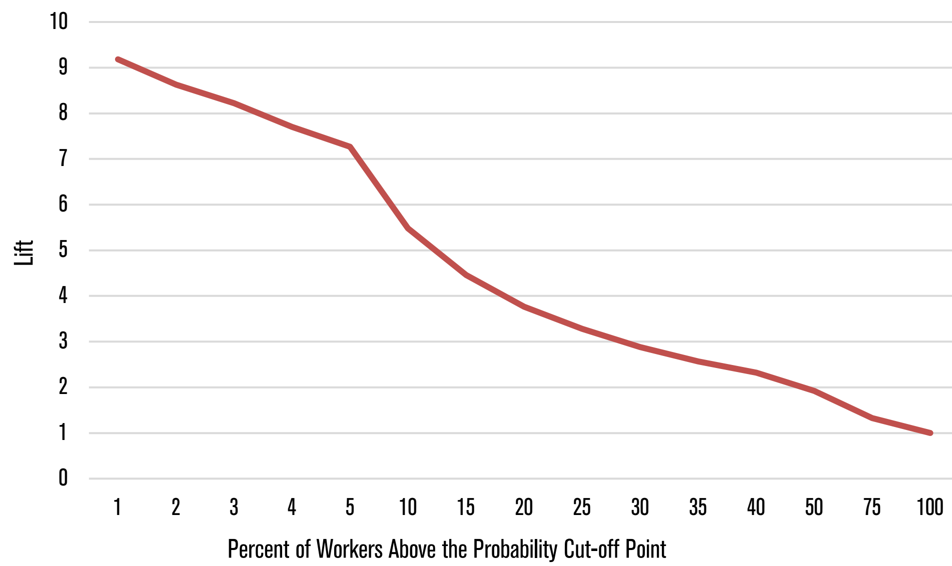
The closer the curve follows the vertical axis and then the top border, the more accurate the model. Conversely, the closer the curve comes to the 45-degree diagonal, the less accurate the model is. The area under the curve (AUC) measures the accuracy of the model where 1 represents a perfect model and 0.5 (same as the diagonal line) shows a useless model.

The employment model generated a very high AUC of 0.892 for the validation sample, indicating an 89.2 percent probability that a randomly selected unemployed person who becomes persistently homeless in the future will receive a higher model score than a randomly selected homeless person who does not become persistently homeless. In the predictive analytics literature, models with AUC exceeding 0.8 are thought to have good predictive power while AUC values below 0.7 indicate poor model performance.

The ROC curve illustrates the trade-off between increasing true positives —finding as many homeless persons as possible who will be persistently homeless in the future—and false positives—decreasing potential program effectiveness by including homeless persons who will not be persistently homeless in the future. It can be used to help select a cutoff value with the ideal balance between these two considerations.

The *lift curve* provides a similar picture. The x axis on the bottom of the graph represents the expected number of true positives we would predict if we did not have a model but simply selected cases at random. It provides a benchmark against which we can see the performance of the model.

Figure 49: Lift Chart for Employment Predictive Model



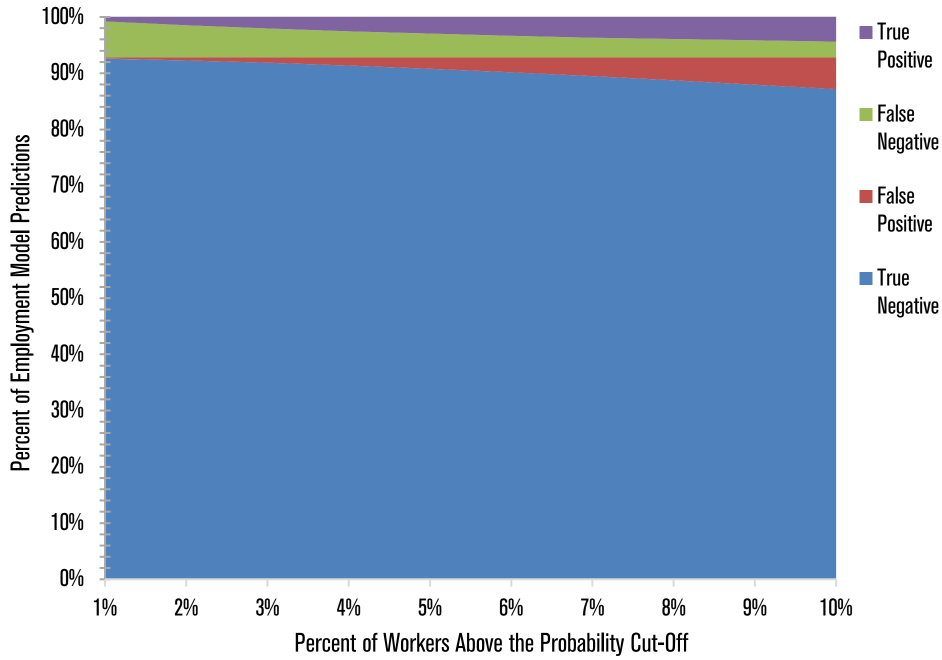
A good model will give us a high lift when we act on only a few cases, i.e., those with the highest probability scores. As we include more cases with lower scores, the lift will decrease.

The lift curve of the employment model for all thresholds is presented in *Figure 49*. The lift is quite high for cases with a high probability of being in the persistently homeless group. For example, for the top one percent, the model generates a lift of 9.2. This means that the model identifies 9.2 times more future persistently homeless workers (true positives) than random selection. This is presented against the baseline lift of 1. At slightly lower thresholds, such as the top three percent, lift drops to 8.2 because to classify more true positives we have to accept a larger share of false positives. For the top 5 percent the lift is 7.3, and for the top ten percent the lift is 5.5.

The overall prediction results from the employment model are shown in *Figure 50*, based on the percent of screened workers who are above the cut-off level for services (bottom axis). The largest task the model performs is correctly identifying workers who do not become persistently homeless – *true negatives*. These correctly excluded cases make up roughly 90 percent of employment model predictions. The remainder of the model's work is to differentiate outcomes for the tenth of workers whose futures are less clear.

The most important task the model performs is correctly identifying workers who do become persistently homeless – *true positives*. The higher cut-off level for probability scores used to target workers for services, the more accurate these predictions are. However, since eight percent of unemployed workers are known are known to become persistently homeless, cut-off levels that include fewer than eight percent of workers necessarily produce *false negatives* – workers who become persistently homeless but are not targeted for services.

Figure 50: Predictive Results from Employment Model by Probability Cut-Off Level



As the probability cut-off level drops to capture a larger share of workers who become persistently homeless, the share of *false negatives* decreases, but the share of incorrectly targeted workers, or *false positives*, increases.

The ratio of workers who become persistently homeless and are correctly targeted for services (*true positives*) versus workers from the same cohort who are incorrectly excluded from services (*false negatives*) is equal when six and a half percent of screened job losers are above the probability cut-off – a cut-off value of 0.32.

Model Comparison for the Employment Model

Finally, we present a model comparison between the logistic regression model we developed and the random forest model, which is a very powerful algorithm and increasingly the "standard tool" used for prediction. The drawback is that it is not transparent – the variables it uses are not explained. Our intention is to determine whether the predictive power of the logistic regression model is good enough relative to the random forest algorithm to justify our model selection, which prioritized transparency over accuracy. A recent study presented a large-scale benchmark experiment for comparing the performance of logistic regression and random forest in binary classification. Random forest performed better than logistic regression based on the level of accuracy measured in approximately 69 percent of the datasets (Couronne, Probst and Boulesteix, 2017).[39]

Random forest is an "ensemble learning" technique consisting of the aggregation of a large number of decision trees, which results in reduced variance compared to a single decision tree. It combines predictions from many classification or regression trees to construct more accurate predictions using bootstrap methods (see Breiman, 2001).[40]

Our comparison is made in several ways. *Table 6* shows the goodness-of-fit statistics of both models for the training and validation samples. For both samples, random forest yields slightly better fit statistics—in the validation sample the AUC is .012 higher and the misclassification rate is .003 lower. Training and validation results are almost identical for both samples, verifying the absence of overfitting.

Table 6: Measures of Fit for Employment Models

Statistic	Logistic Regression		Random Forest	
	Training	Validation	Training	Validation
Area Under the Curve (AUC)	.892	.894	.898	.916
Misclassification Rate	.072	.074	.070	.071

These comparisons verify that even though the random forest model performs slightly better, the improvement does not warrant its selection over the logistic regression model due to the loss in interpretability. Our logistic regression model performs very accurately, is transparent, and with only 32 variables is simple enough to be used in a manual screening tool.

Model Fit Statistics and Predictive Performance for the Young Adult Model

The young adult model achieved a very strong C-statistic, .88, which is the probability that the predicted outcome is better than chance. The *Brier score* for the model is 0.05, which is also a very strong statistic. Moreover, the performance measures were almost identical for training and validation samples indicating that over-fitting was not a problem.

In addition to model fit statistics, we used *sensitivity*, *specificity*, *positive predictive value* (*PPV*), *accuracy*, *area under the receiver operating characteristics* (*ROC*) *curve*, and *lift curve* to assess the out-of-sample model performance. All these values are presented for different percentiles of young adults in terms of predicted risk in *Table* 7, similar to what was shown earlier for the employment model in *Table 4*. *Table* 7 presents *sensitivity, specificity, PPV,* and *accuracy* statistics for different cutoff points for the validation (out-of-sample) cohort. These statistics were explained earlier for *Table 4*.

The first column of *Table* 7 shows the percentile of young adults sorted by descending order of predicted probability of becoming persistently homeless, shown in the second column. Percentiles are computed based on the total population of the validation sample, 239,555. For example, the first row shows the results for the top 1 percent or the first percentile and the sixth row shows the measures for the top 15 percent or the first decile.

If the top 3 percent of persons at risk of becoming persistently homeless are considered, we see that the model identifies approximately 7,300 individuals who are predicted to become persistent homeless in the future. We know that out of 333,000 employment spells in the validation sample, approximately 19,600 of them became persistent homeless in the future (8.2 percent). The probability threshold is 59.7 percent. Twenty-four percent sensitivity shows that the model captured 23.8 percent of the 19,600 when targeting the top 3 percent, which is quite impressive. The 1-specificity value is only 1.2 percent verifying that the model correctly identifies 99 percent of those who do not become persistently homeless in the future.

Table 7: Predictive Performance of the Young Adult Model (Validation Results)

Percentile	Probability	Sensitivity	1 - Specificity	Accuracy	PPV	Cumulative Population
1	0.7595	8.80%	0.30%	92.30%	72.32%	2,392
2	0.6505	16.40%	0.70%	92.50%	67.10%	4,796
3	0.5970	23.80%	1.20%	92.70%	63.87%	7,305
5	0.4705	35.10%	2.30%	92.60%	57.53%	11,977
11	0.2280	60.10%	7.00%	90.30%	43.36%	27,176
15	0.1240	69.00%	10.20%	88.10%	37.70%	35,917
20	0.0800	76.40%	14.80%	84.50%	31.57%	47,452
30	0.0430	86.50%	24.80%	76.10%	23.73%	71,451
41	0.0275	90.80%	36.40%	65.80%	18.19%	97,947
55	0.0245	94.80%	52.00%	51.90%	13.99%	132,882
91	0.0165	99.20%	90.60%	16.80%	8.89%	218,685
100	0.0035	100.00%	100.00%	8.20%	8.19%	239,555

The PPV value of 63.9 percent and accuracy value of 92.7 percent for the top 3 percent are also very high. The model achieves a PPV result of almost 64 percent, meaning that out of 7,300 persons that the model predicted to be persistently homeless, almost two thirds are true positives and the remaining one-third are false positives.

If we consider 7,300 *randomly* chosen persons, the PPV value would be only 8.2 percent, which is the ratio of true positives to the population size. This means that a random selection without using any knowledge or model would yield only 8.2 percent true positives. The remaining 91.8 percent would be false positives. When we compare this number to the model PPV for 7,300 persons (64 percent), we get a ratio of 8 which shows that the model is performing 8 times better than random selection at the 3 percent threshold. This measure is known as the *lift* of a model.

Similar to the employment model, this model is also highly accurate in distinguishing persistently homeless individuals from others. However, it is still necessary to calibrate the probability cut-off level that will be used to determine who within the targeted population will be offered the intervention.

Insights for making this decision are shown in *Table 8*, which provides the number of false and true positives at different probability levels. In many interventions the selection of the threshold is based on the capacity and funding of the program. Hence, for example, if the program has the capacity to serve 12,000 persons, then the appropriate threshold would be .4705. If the target is around 5,000 then the threshold would be .6505.

The numbers show that the model performs very well for the top 5 percent and fairly well for the top decile (10 percent). For the top 5 percent, 57.5 percent of the targeted 11,977 persons were true positives. The PPV value drops to below 50 percent at the 10th percentile.

Table 8: Prediction Performance showing Predicted Homeless Populations

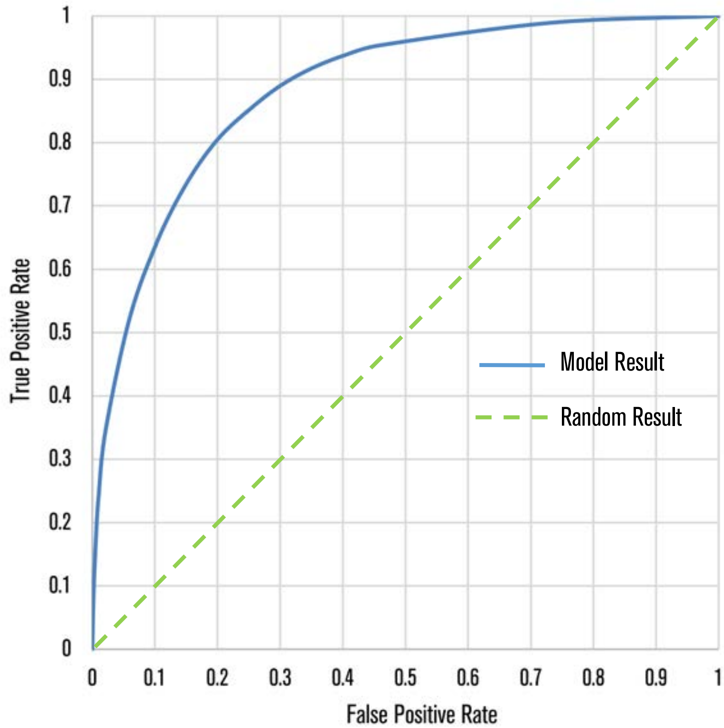
Percentile	Probability	Cumulative Population	True Positives	False Positives	Homeless False Positives	PPV	Adjusted PPV
0	NA	0	0	0	0	0%	0%
1	0.7595	2,392	1,730	662	338	72.32%	86.5%
2	0.6505	4,796	3,218	1,578	711	67.10%	81.9%
3	0.5970	7,305	4,666	2,639	1,160	63.87%	79.8%
5	0.4705	11,977	6,890	5,087	1,798	57.53%	72.5%
11	0.2280	27,176	11,784	15,392	4,422	43.36%	59.6%
15	0.1240	35,917	13,539	22,378	6,366	37.70%	55.4%
20	0.0800	47,452	14,980	32,472	8,829	31.57%	50.2%
30	0.0430	71,451	16,955	54,496	12,814	23.73%	41.7%
41	0.0275	97,947	17,812	80,135	15,300	18.19%	33.8%
55	0.0245	132,882	18,589	114,293	17,879	13.99%	27.4%
91	0.0165	218,685	19,451	199,234	21,010	8.89%	18.5%
100	0.0035	239,555	19,608	219,947	21,492	8.19%	17.2%

Table 8 also includes a column labeled "Homeless False Positives", which represents false positives that fell short of becoming persistently homeless in the next 3 years but were observed to be homeless starting from the sixth month after entering adulthood while in aid. They had one homeless episode lasting less than a year in the three years following their entry into

adulthood. When we add these numbers to true positives, PPV values shown as "Adjusted PPV" values increase significantly. At the top 3 percent threshold, approximately 80 percent of persons were observed to be homeless after their first month as young adults in aid. At the 11th percentile threshold PPV increases from 43 percent to 60 percent. Hence, the data shows that at high thresholds the model identifies future persistently homeless individuals accurately. Furthermore, a significant proportion of false positives also become homeless with a risk of becoming persistently homeless after 3 years.

Figure 51: Prediction Results for Young Adults becoming Persistently Homeless in the Next 3 Years

ROC Curve: Area Under the Curve = 0.88

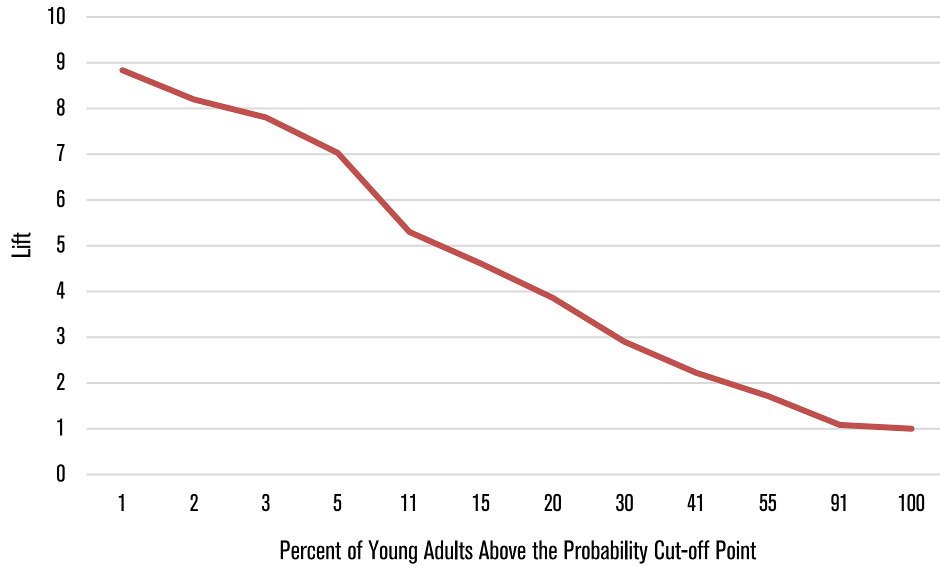


ROC Curve and Lift Curves for the Young Adult Model

The ROC curve for the young adult model is shown in *Figure 51*. Our model generated a very high AUC of 0.88 for the validation sample, indicating an 88 percent probability that a randomly selected unemployed person who becomes persistently homeless in the future will receive a higher model score than a randomly selected homeless person who does not become persistently homeless in the future.

The lift curve of the young adult model for all thresholds is presented in *Figure 52*. The lift provided by the model is presented against the baseline lift of 1, which represents random results. It is quite high for cases with a high probability of being in the persistently homeless group. For example, for the top 5 percent, the model generates a lift of 6.5. This means that the model identifies 6.5 times more future persistently homeless persons (true positives) than random selection.

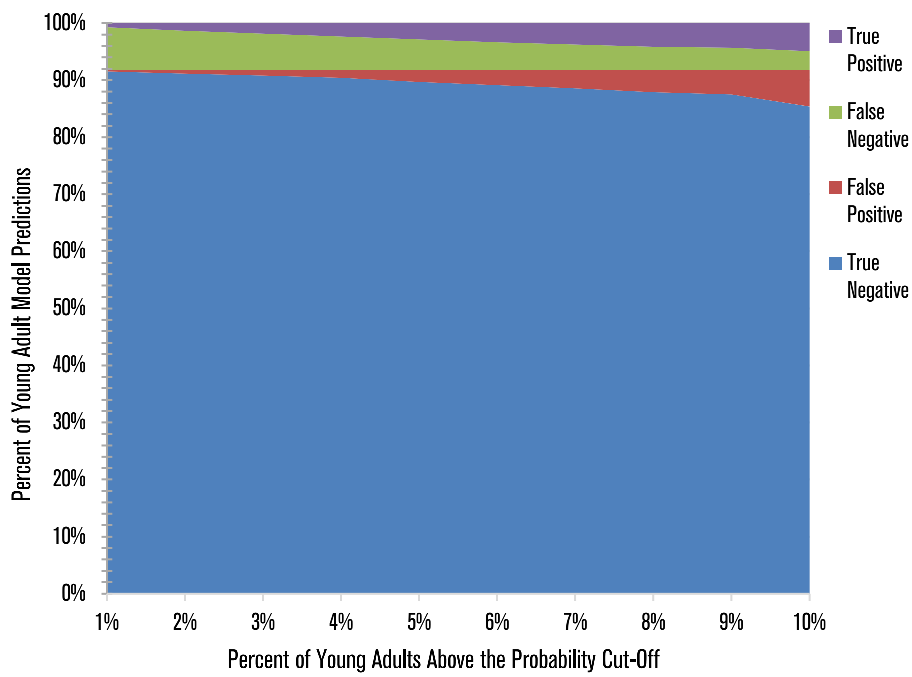
Figure 52: Lift Chart for Young Adult Predictive Model



At slightly lower thresholds, such as the top 10 percent, lift drops to 5.3 because to classify more true positives we have to accept a larger share of false positives.

The overall prediction results from the employment model are shown in *Figure 53*, based on the percent of screened young adults who are above the cut-off level for services (bottom axis). The model correctly identifies roughly 90 percent of young adults who do not become persistently homeless – *true negatives*. The remaining one-tenth of cases include young adults whose futures are less clear.

Figure 53: Predictive Results from Young Adult Model by Probability Cut-Off Level



The higher cut-off level for probability scores used to target young adults for services, the more accurate these predictions are. However, since eight percent of young adults are known are known to become persistently homeless, cut-off levels that include fewer than eight percent of young adults necessarily produce *false negatives* – youth who become persistently homeless but are not targeted for services.

As the probability cut-off level drops to capture a larger share of youth who become persistently homeless, the share of *false negatives* decreases, but the share of incorrectly targeted workers, or *false positives*, increases.

The ratio of young adults who become persistently homeless and are correctly targeted for services (*true positives*) versus young adults from the same cohort who are incorrectly excluded from services (*false negatives*) is equal when seven percent of screened young adults are above the probability cut-off – a cut-off value of 0.32.

The performance of the two predictive models was evaluated using data sets different from those used to develop the models. This was done to be sure that the models performed well for the overall populations they will be used to screen and that their accuracy was not limited to the specific data sets used to develop the models. A variety of statistical tests all demonstrated that both models perform very well and are highly accurate.

Conclusions

Both predictive models are very accurate and particularly strong when using high probability cutoff levels, generating small numbers of false positives and high numbers of true positives. A key strength of the models is that the accuracy of predictions was validated using three years of post-prediction data. Another key strength is that the models are transparent and identify distinctive attributes of high-cost individuals. The results confirm that local public costs for targeted individuals are likely to be high and to increase over time.

In the absence of broadly representative, local longitudinal data that is linked across service providers and that that can be used to develop tools comparable to those presented in this report, it is reasonable to use these screening tools in metropolitan areas throughout the United States. The study population used to develop these tools includes everyone who was homeless during fifteen years, a total of over one million people, in the most populous county in the United States. The large and broadly representative study population used to develop these tools can reasonable be assumed to share many of the attributes and face many of the same obstacles as their counterparts in other urban centers.

The screening tools can be reconfigured to use locally available data and still retain a high level of accuracy, provided that key attributes of individuals are addressed. This includes demographic characteristics, homeless and employment histories, and use of services provided by the health, behavioral health, social service, and justice systems.

The tools are particularly useful for prioritizing unemployed workers and young adults for services because each individual who is screened is given a probability of becoming persistently homeless. These probabilities can be used to rank everyone who is screened for access to services. Prioritizing individuals for access to early, comprehensive interventions is important because the resources that are most effective for preventing homelessness, including subsidized housing and employment, are scarce in relation to the demand for those resources.

The purpose of the models is to target individuals for additional help, so there are no adverse consequences to individuals if they are incorrectly targeted.

The optimal probability cutoff level for individuals who will be targeted for services is not simply an empirical decision. One important factor is program capacity for helping unemployed workers obtain new jobs and for helping young adults make a successful transition into adulthood. Another factor is the extent to which costs avoided by averting persistent homelessness will be relied upon to fund delivery of services.

Both models are system-based tools. Depending on the model, they require information about healthcare, justice system involvement, foster care, employment, homeless history, and demographics that is available only from those institutional systems. Cooperation of public agencies is necessary to protect the privacy of personal information while providing the data required for the tools.

Because of the level of effort required to obtain and integrate the necessary data, the most efficient use of the tool is for regular, ongoing system-wide screening of linked records rather than screening clients individually. By predicting how likely each person in the entire identified population of homeless resident is to become persistently homeless, it is possible to prioritize individuals for access to the scarce supply of services.

Because the tools do not correctly identify all high-risk individuals, the screening process should include an option to override the probability score based on the judgment of service providers. Allowing overrides permits service providers to adapt to changing populations and conditions and to be responsive to unique circumstances.

The descriptive information in this report and the factors used in the predictive models provide extensive information about the characteristics and needs of individuals who become persistently homeless. This information identifies needs that should be addressed but it does not define the program models for addressing those needs. Programs models should be structured using evidence-based findings about best practices for helping unemployed workers obtain sustaining employment and helping high-risk young adults make a successful transition to adulthood.

The strong validation results for these models show that it is possible to develop many other predictive models that will target other homeless groups for specific types of interventions. Each model is likely to provide rifle-shot targeting because discrete population groups with distinctive attributes are needed to produce accurate predictive results. An updated typology of homelessness that breaks out distinct homeless trajectories will be valuable for mapping the full range of groups that should be targeted for interventions that will minimize the harm, cost and duration of homelessness.

Social Security Board Records Office, Baltimore, Maryland, 1937. Courtesy of PICRYL.

Appendix Tables

Table A-1: Cost Factors for Local Public Services in 2017 Dollars

Service	Cost	Source
Outpatient visit to Los Angeles County Department of Health Services (DHS) outpatient clinic	$823	Los Angeles County
Emergency room visit to DHS hospital	$1,213	Los Angeles County
1 inpatient day at DHS hospital	$9,158.35 Average	OSHPD records for Los Angeles County DHS hospitals in 2014, adjusted to 2017 dollars. This study used varying average costs based on 3-digit ICD-9 diagnosis.
Ratio of total private hospital inpatient cost to total Los Angeles County DHS inpatient cost	1.82	OSHPD records for Los Angeles County general hospitals in 2014.
1 emergency medical transportation trip to hospital	$553	Cost data from Santa Clara County adjusted to 2017 dollars. The ratios of trips to hospital encounters are: 0.2327 trips per emergency room visit; 0.1711 trips per inpatient admission; and .0618 trips per psychiatric encounter.
Outpatient visit to Los Angeles County Department of Mental Health (DMH)	$217.44 Average	Los Angeles County. Service cost varies by provider.
1 day of residential care by DMH	$115	Los Angeles County
1 day of acute inpatient care by DMH	$600	Los Angeles County
Outpatient visit to Los Angeles County Department of Public Health (DPH) substance abuse treatment program	$55	Los Angeles County
1 day DPH substance abuse residential program	$115	Los Angeles County
1 day DPH substance abuse detox program	$375	Los Angeles County
1 month of Los Angeles County Department of Public Social Services (DPSS) food stamp/SNAP/Cal Fresh benefits	$139.66	Food Stamp Program Participation and Benefit Issuance Report (DFA 256)
1 month of DPSS Medi-Cal benefits	$644.45	Los Angeles County 2017-18 Budget, Analysis of the Medi-Cal Budget (includes local administrative cost)
1 month of DPSS General Relief assistance	$202.85	California Department of Social Services GR 237 report
1 month of DPSS CalWORKs assistance per person in caseload	$225.48	California Department of Social Services CalWORKs Annual Summary
DPSS administrative cost	30% of budget, 0.431 ratio to benefits	Los Angeles County 2018-19 Recommended Budget Volume I (Medi-Cal benefits excluded)
1 month of foster care services	$2,139	California Department of Social Services
1 service encounter funded by the Los Angeles Homeless Services Authority (LAHSA)	$40	Los Angeles County
1 night of emergency or temporary housing funded by LAHSA	$40	Los Angeles County
Arrest by police	$405	Los Angeles County Sheriff Department hourly patrol rate of $135 x 3 hours
Jail booking by Los Angeles County Sheriff Department	$662.37	Los Angeles County
Court cost	$110	Cost data from Santa Clara County adjusted to 2017 dollars – average cost excluding jury trials.
1 day of incarceration in general jail facility	$99.42	Los Angeles County
1 day of incarceration in jail medical facility	$1,309.17	Los Angeles County
1 day of incarceration in jail mental health facility	$1,309.17	Los Angeles County
1 month of adult probation	$555	Los Angeles County

Table A-2: List of Potential Factors to Predict Persistent Homeless Persons

Model Variables	Employment Model		Young Adult Model	
Demographic and Household Factors	**2nd Round**	**Final Model**	**2nd Round**	**Final Model**
Age	✓	✓		
Gender	✓	✓		
Ethnicity	✓	✓	✓	✓
Marital Status	✓	✓		
Household Type	✓	✓		
Homelessness Factors (-1, -2, -5 years)				
If homeless	✓	✓	✓	✓
Number of Months of Homelessness	✓	✓	✓	
If homeless Month before Unemployment	✓	✓		
If homeless as an young adult in aid			✓	✓
Employment and Earnings (-1, -2, -5 years)				
Number of Months Employed	✓	✓	✓	
If Employed	✓	✓	✓	✓
Duration of the most recent Employment	✓	✓	✓	
The ranking of the employment spell	✓	✓		
Average Earnings	✓	✓	✓	
Maximum Earnings	✓	✓	✓	
The year became unemployed	✓			
Health Diagnoses (-1, -2 years)				
001-139 Infections and Parasitic				
042 HIV Disease				
140-239 Neoplasms	✓			
240-279 Endocrine and Metabolic and Immune	✓			
250 Diabetes				
280-289 Blood and Blood Organs				
290-319 Mental Health Disorders			✓	
291 Alcohol-induced Mental Illness				
292 Drug-induced Mental Illness			✓	
295 Schizophrenic Disorders				
296 Episodic Mood Disorders			✓	
298 Other Nonorganic Psychoses	✓			
311 Depressive Disorders				
309 Adjustment Reaction				
303-5 Alcohol and drug dependence				
320-389 Nervous System	✓			
390-459 Circulatory System	✓			
402-429 Heart Disease				
451-459 Vein and lymphatic Disease				
460-519 Respiratory System			✓	
470-478 Other Disease-Upper Respiratory Tract				

Model Variables	Employment Model		Young Adult Model	
Demographic and Household Factors	**2nd Round**	**Final Model**	**2nd Round**	**Final Model**
480-488 Pneumonia and Influenza				
490-496 Chronic Pulmonary Disease				
520-579 Digestive System			✓	
569-73 and 76-78 and 85-94 and 96 liver, pancreas, intestines and kidneys				
580-629 Genitourinary System	✓			
590-599, 614-616 Urinary Disease				
680-709 Skin and Subcutaneous				
710-739 Musculoskeletal System				
780-799 Ill-defined Conditions			✓	
799 Other ill-defined and unknown causes of morbidity and mortality				
800-999 Injury and Poisoning	✓		✓	
V01-V89 Factors Influencing Health				
Health and Behavioral Health Services (-1, -2 years)				
If any Emergency Medical Service (EMS) encounters	✓	✓	✓	
Number of EMS encounters				
If any hospital inpatient admissions				
Number of hospital inpatient admissions			✓	
If any outpatient hospital/clinic visits	✓	✓	✓	
Number of outpatient hospital/clinic visits	✓	✓		
If any Private Public Partnership (PPP) clinic visits			✓	
Number of PPP clinic visits				
Days of hospital inpatient stays				
Any service received from Health Services	✓		✓	
If disabled at the time of unemployment/adolescence youth	✓	✓	✓	✓
If any outpatient visit with mental health targeted services	✓	✓	✓	✓
Number of outpatient visits with mental health targeted services				
If any Mental Health admission for Acute Care or Residential Treatment				
Number of Mental Health admissions for Acute Care or Residential Treatment				
Days of Mental Health services for Acute Care or Residential Treatment				
Any service received from Mental Health Services			✓	✓
If any Mental Health outpatient admission				
If any Alcohol and Drug Abuse (ADA) outpatient visits	✓	✓	✓	
Number of ADA outpatient visits			✓	
If any ADA residential services				
Months of ADA residential services			✓	✓
If any detox treatments	✓	✓		
Number of detox treatments				
If any narcotic treatment				
Number of narcotic treatments				
Days of detox treatments				

Model Variables	Employment Model		Young Adult Model	
Demographic and Household Factors	**2nd Round**	**Final Model**	**2nd Round**	**Final Model**
Days of ADA residential services	✓			
Days of narcotic treatments	✓			
Any service received from Public Health Services	✓	✓	✓	✓
Criminal Justice (-1, -2 years)				
If in probation	✓	✓		
Days of probation				
Frequency of probation times				
If arrested	✓	✓	✓	✓
Number of arrests				
Number of days in jail	✓	✓	✓	
If housed in medical or mental health facilities	✓			
Number of days in medical or mental health facilities			✓	
Social services				
If aided at the time of Unemployment	✓	✓		
If cash aided at the time of Adolescence Youth in aid			✓	✓
Disability history while on aid	✓	✓	✓	
If needed mental health services while on aid				
Number of months with disability while on aid				
Foster Care				
If history of foster care			✓	✓

Table A-3: Averages of Model Variables for Persistent Homeless Persons and the Rest of the Population (Other) for the Employment Model

Variable	Persistently Homeless	Other
Demographics and Household		
Age 18-40 (Percent)	62%	65.5%
Age 41-57 (Percent)	34.3%	28.4%
Age 58+ (Percent)	3.7%	6%
Male (Percent)	62.3%	45.2%
Female (percent)	37.7%	54.8%
African American (Percent)	44.5%	18.2%
Alaskan American and American Indian (Percent)	1.4%	0.4%
Hispanic (Percent)	37.1%	56.3%
Other Ethnicity (Percent)	2.4%	10.0%
European American (Percent)	14.6%	15.1%
Single individual household at time of unemployment (percent)	78%	30%
Family households at the time of unemployment (percent)		
Married individuals (percent)	4%	25%
Other than single and married individuals (percent)	12.7%	8.3%
Single individuals	83.3%	66.7%
Employment and Earnings	74%	68%
Employed one to two years earlier (Percent)	59.4%	71.5%
Employed three to five years earlier (Percent)	62.3%	67.5%
Months employed last year (Median)	9	12
Months employed in three to five years earlier (Mean)	10	13
Duration of the most recent employment (Median Months)	7	13
If the first unemployment (percent)	74%	68%
Average earnings last year (median)	$469	$1,060
Maximum earnings last year (median)	$701	$1,465
Homelessness		
Homeless last year (Percent)	42.5%	4%
Homeless one to two years earlier (Percent)	29%	3.5%
Homeless three to five years earlier (Percent)	34%	6.2%
Months of homelessness three to five years earlier (Mean)	3.55	.05
Homeless month before unemployment started (Percent)	29.5%	1.7%
Health and Behavioral Health		
Emergency Medical Service encounter this year (Percent)	8.9%	3.4%
Number of Outpatient Admissions to Hospitals last year (mean)	.5	.3
Outpatient Admissions to Hospitals last year (Percent)	14%	8%
Disability History (Percent)	15.6%	3.8%
Disabled at the time of Unemployment (percent)	27%	6.2%
Mental Health Outpatient Service encounter last year (Percent)	4.5%	1.5%
Number of alcohol and substance abuse services last year (mean)	.08	.01
Detox services (percent)	1.1%	.1%
Criminal Justice		

Variable	Persistently Homeless	Other
Number of days in jail last year (mean)	7.2	1.6
Jailed in last year (percent)	21.7%	4.8%
Jailed in one to two years earlier (percent)	18.3%	4.2%
In probation last year (percent)	9.4%	2.2%
Social Services		
Cash aid at the time of unemployment (percent)	42.3%	13%
Not aided at the time of unemployment	15.6%	12%
Non-cash aid at the time of unemployment (percent)	42.1%	75%

Table A-4: Logistic Regression Adjusted Odds Ratios, Parameter Estimates and Types of Predictor Variables for the Employment Model

Variable	Variable Type	Parameter Estimate	Odds Ratio
Intercept		-3.144	
Demographics and Household			
Age 18-57	Nominal	.5896	1.8
Age 58+ (Reference Group)	Nominal	0	1
Male	Binary	.1531	1.17
Female (Reference Group)	Binary	0	1
African American	Nominal	.5298	1.7
Alaskan American and American Indian	Nominal	.7198	2.05
Hispanic	Nominal	.1422	1.15
Other Ethnicity	Nominal	-.5671	.57
European American (Reference Group)	Nominal	0	1
Single individual household at time of unemployment	Binary	1.0027	2.73
Family households at the time of unemployment (Reference Group)	Binary	0	1
Married individuals	Nominal	-.9007	.41
Other than single and married individuals	Nominal	.0923	1.1
Single individuals (Reference Group)	Nominal	0	1
Employment			
Employed one to two years earlier	Binary	.0553	1.06
Employed three to five years earlier	Binary	.2406	1.27
Months employed last year	Interval	-.0191	.981
Months employed in three to five years earlier	Interval	-.0089	.991
Duration of the most recent employment	Interval	-.0062	.994
Each additional unemployment spell	Ordinal	.0428	1.04
Average earnings last year (unit=$100)	Interval	-.0461	.954
Maximum earnings last year (unit=$100)	Interval	.0224	1.02
Homelessness			
Homeless last year	Binary	1.1987	3.32
Homeless one to two years earlier	Binary	.6638	1.94
Homeless three to five years earlier	Binary	.7159	2.05
Months of homelessness three to five years earlier	Interval	.0122	1.012
Homeless month before unemployment started	Binary	1.0489	2.85
Health and Behavioral Health			
Emergency medical service encounter this year	Binary	.23197	1.26
Number of outpatient admissions to medical clinic last year	Interval	-.03396	.967
Outpatient admission to medical clinic last year	Binary	-.0859	.918
No disability history	Binary	.7266	2.07
Disabled at the time of unemployment	Binary	.6902	1.99
Mental health outpatient service encounter last year	Binary	.2092	1.23
Number of alcohol and substance abuse services last year	Interval	.0569	1.06
Detox services last year	Binary	.6326	1.88
Alcohol or substance abuse services 1 to 2 years earlier	Binary	.1163	1.12

Variable	Variable Type	Parameter Estimate	Odds Ratio
Criminal Justice			
Number of days in jail last year (unit=10)	Interval	-.0192	.981
Jailed in last year	Binary	.5714	1.77
Jailed in one to two years earlier	Binary	.2053	1.23
In probation last year	Binary	.0803	1.08
Social Services			
Cash aid at the time of unemployment	Nominal	.3124	1.37
Not aided at the time of unemployment	Nominal	.9371	2.55
Non-cash aid at the time of unemployment (Reference Group)	Nominal	0	1

Table A-5: Effect Summary Report for the Employment Model

Source	LogWorth	Effect Summary	PValue
Household status	670.178		0.00000
Type of aid	399.822		0.00000
Homeless last month	314.383		0.00000
Homeless last year	270.976		0.00000
Ethnicity	254.442		0.00000
Marital status	186.747		0.00000
Homeless 3-5 yrs. earlier	154.241		0.00000
Disabled	143.608		0.00000
Jailed last year	116.555		0.00000
Avg. earnings last year	101.950		0.00000
Disability history	77.411		0.00000
Age group	59.583		0.00000
Max earnings last year	54.999		0.00000
Number of unemp. spells	42.840		0.00000
Homeless 1-2 years earlier	39.281		0.00000
Jailed 1-2 years earlier	29.666		0.00000
Months employed 5 years	23.378		0.00000
Gender	20.662		0.00000
Employed 3-5 yrs. earlier	20.156		0.00000
Emergency med. last year	14.584		0.00000
Outpatient last year	14.416		0.00000
Jail days last year	12.497		0.00000
Months employed last yr.	11.681		0.00000
Employed 1-2 yrs. earlier	10.841		0.00000
Mths. hmls. 3-5 yrs. earlier	10.285		0.00000
On probation last year	5.953		0.00000
Detox services last year	4.411		0.00004
MH outpat. last year	2.539		0.00289
Alc. or SA srvs. last year	1.704		0.01977

Table A-6: Averages of Model Variables for Persistent Homeless Persons and the Rest of the Population (Other) for the Young Adult Model

Variable *Note: Time of adolescence refers to the first month in aid as a young adult*	Persistent Homeless	Other
Demographics		
African American	44.8%	13%
Alaskan American and American Indian	.75%	.2%
Hispanic	44.1%	71%
Other Ethnicity	1.8%	6.8%
European American (Reference Group)	8.5%	9%
Employment		
Employed before the time of adolescence	40.5%	29%
Employed three to five years earlier	28.6%	18.9%
Homelessness		
Homeless last year	10.5%	1.4%
Homeless one to two years earlier	6.4%	1%
Homeless three to five years earlier	8.2%	2%
Homeless at the time of adolescence	59.9%	7%
Health and Behavioral Health		
Disabled at the time of adolescence	11.7%	2.1%
Mental Health Outpatient Service encounter last year for the first time	3.1%	1.5%
Mental Health Outpatient Service encounter more than 2 years earlier	2.9%	1.5%
If any mental health service encounter last year	7.8%	3.2%
If any mental health service encounter one to two years earlier	6.8%	2.8%
If received alcohol and substance abuse services all 3 past years	1.8%	.3%
Months of residential alcohol and substance abuse services last year	.4	.1
Criminal Justice		
Jailed in last year	10.5%	2.6%
Social Services		
Cash aid at the time of adolescence	30.6%	20.1
Non-cash aid at the time of adolescence (Reference Group)	69.4%	79.9
Foster Care		
If history of foster care	11.2%	3.5%

Table A-7: Logistic Regression Adjusted Odds Ratios, Parameter Estimates and Types of Predictor Variables for the Young Adult Model

Variable *Note: Time of adolescence refers to the first month in aid as a young adult*	Variable Type	Parameter Estimate	Odds Ratio
Intercept		-4.114	
Demographics			
African American	Nominal	1.2573	3.5
Alaskan American and American Indian	Nominal	1.076	2.93
Hispanic	Nominal	.0532	1.06
Other Ethnicity	Nominal	-.76135	.47
European American (Reference Group)	Nominal	0	1
Employment			
Employed before the time of adolescence	Binary	.4359	1.57
Employed three to five years earlier	Binary	.07697	1.09
Homelessness			
Homeless last year	Binary	.3399	1.4
Homeless one to two years earlier	Binary	.4411	1.55
Homeless three to five years earlier	Binary	.6457	1.92
Homeless at the time of adolescence	Binary	2.8416	17.1
Health and Behavioral Health			
Disabled at the time of adolescence	Binary	.7247	2.05
Mental Health Outpatient Service encounter last year for the first time	Binary	.2292	1.26
Mental Health Outpatient Service encounter more than 2 years earlier	Binary	.608	1.85
If any mental health service encounter last year	Binary	.4658	1.59
If any mental health service encounter one to two years earlier	Binary	.5535	1.75
If received alcohol and substance abuse services all 3 past years	Binary	.7762	2.11
Months of residential alcohol and substance abuse services last year	Interval	-.0186	.98
Criminal Justice			
Jailed in last year	Binary	.6724	1.95
Social Services			
Cash aid at the time of adolescence	Nominal	.4116	1.52
Non-cash aid at the time of adolescence (Reference Group)	Nominal	0	1
Foster Care			
If history of foster care	Nominal	.8821	2.416

Table A-8: Effect Summary Report for the Young Adult Model

Source	LogWorth	Effect Summary	PValue
Homeless at the time of adolescence	887.493		0.00000
Ethnicity	234.362		0.00000
Cash aid	96.346		0.00000
Disabled	82.872		0.00000
History of foster care	63.264		0.00000
Arrested last year	60.715		0.00000
Employed before	50.828		0.00000
Homeless 3-5 years earlier	46.944		0.00000
MH services 1-2 years earlier	25.156		0.00000
MH outpatient enc. last year	22.464		0.00000
Homeless 1-2 years earlier	18.835		0.00000
SA res. serv. duration last year	15.458		0.00000
Homeless last year	9.918		0.00000
SA services last 3 years	7.608		0.00000
MH services last year	5.079		0.00001
First time MH services last year	3.328		0.00047
Employed 3-5 years earlier	0.812		0.03401

Table A-9: IDC-9-CM Medical Diagnostic Codes Used to Identify Substance Abuse

Diagnostic Category	ICD-9-CM Code	Description
Alcohol withdrawal delirium		
	291.1	Alcohol-induced persisting amnestic disorder
	291.2	Alcohol-induced persisting dementia
	291.3	Alcohol-induced psychotic disorder with hallucinations
	291.4	Idiosyncratic alcohol intoxication
	291.5	Alcohol-induced psychotic disorder with delusions
	291.8	Other specified alcohol-induced mental disorders
	291.81	Alcohol withdrawal
	291.82	Alcohol-induced sleep disorders
	291.89	Other alcohol-induced disorders
	291.9	Unspecified alcohol-induced mental disorders
	303.00–303.03	Acute alcohol intoxication
	303.90–303.93	Other and unspecified alcohol dependence
	305.00–305.03	Alcohol abuse
	357.5	Alcoholic polyneuropathy
	425.5	Alcoholic cardiomyopathy
	535.30, 535.31	Alcoholic gastritis
	571	Alcoholic fatty liver
	571.1	Acute alcoholic hepatitis
	571.2	Alcoholic cirrhosis of liver
	571.3	Alcoholic liver damage, unspecified
	E860.0	Alcoholic beverages poisoning
Amphetamines		
	304.40–304.43	Amphetamines dependence
	305.70–305.73	Nondependent amphetamine abuse
Cannabis		
	304.30–304.33	Cannabis dependence
	305.20–305.23	Nondependent cannabis abuse
Cocaine		
	304.20–304.23	Cocaine dependence
	305.60–305.63	Nondependent cocaine abuse
	968.5	Poisoning by cocaine
	E938.5	Cocaine, adverse effects
Drug-induced mental disorders		
	292	Drug withdrawal
	292.11	Drug-induced psychotic disorder with delusions
	292.12	Drug-induced psychotic disorder with hallucinations
	292.2	Pathological drug intoxication
	292.81	Drug-induced delirium
	292.82	Drug-induced persistent dementia
	292.83	Drug-induced persistent amnestic disorder
	292.84	Drug-induced mood disorder
	292.85	Drug-induced sleep disorders

Diagnostic Category	ICD-9-CM Code	Description
	292.89	Other drug-induced mental disorder
	292.9	Unspecified drug-induced mental disorder
Hallucinogens		
	304.50–304.53	Hallucinogen dependence
	305.30–305.33	Nondependent hallucinogen abuse
	969.6	Poisoning by hallucinogens (psychodysleptics)
	E854.1	Accidental poisoning by hallucinogens (psychodysleptics)
	E939.6	Hallucinogens, adverse effects
Opioids		
	304.00–304.03	Opioid type dependence
	304.70–304.73	Combinations of opioids with any other
	305.50–305.53	Nondependent opioid abuse
	965	Poisoning by opium
	965.01	Poisoning by heroin
	965.02	Poisoning by methadone
	965.09	Poisoning by other opiates and related narcotics
	E850.0	Heroin poisoning
	E935.0	Heroin, adverse effects
Sedatives, hypnotics, anxiolytics, tranquilizers, barbiturates		
	304.10–304.13	Sedatives, hypnotics, or anxiolytic dependence
	305.40–305.43	Nondependent sedative, hypnotic, or anxiolytic abuse
Other		
	304.60–304.63	Other, specified drug dependence
	304.80–304.83	Combinations excluding opioids
	304.90–304.93	Unspecified drug dependence
	305.90–305.93	Other, mixed or unspecified drug abuse
	648.30–648.34	Drug dependence complicating pregnancy, childbirth, or the puerperium
	V654.2	Counseling, substance use

End Notes

[1] It should be noted that the population shown in *Figure 2* differs somewhat from those in the following studies in that it shows the age distribution when individuals were first homeless rather than the age distribution of the point-in-time population. Culhane, D., S. Metraux, T. Byrne, M. Stino, J. Bainbridge. 2013. *The Age Structure of Contemporary Homelessness: Evidence and Implications For Public Policy*, Analyses of Social Issues and Public Policy 13(1) December 2013. See also: Culhane, D., S. Metraux, T. Byrne. 2013. *Aging Trends in Homeless Populations*. Contexts 12(2): pp. 66-68 (May 2013). This research studied 22 years of New York City shelter records, combined with the 2010, 2000 and 1990 U.S. Census of Population and Housing.
https://journals.sagepub.com/doi/pdf/10.1177/1536504213487702

[2] The administrative records used for this study include only binary gender categories, so it is not possible to identify individuals using categories other than female and male.

[3] A report released by the Economic Roundtable in 2015 (*All Alone: Antecedents of Chronic Homelessness*, p.12, https://economicrt.org/publication/all-alone/) put the average monthly number of entrants into homelessness at 13,300. This report puts the number lower, at 10,900, for two reasons. First, extensive effort was made to de-duplicate records, which reduced number of people identified as having experienced homelessness. Second, a more conservative indictor of homelessness was used – whether a public benefits recipient used the address of an office of the Los Angeles County Department of Public Social Services as their address, rather than whether they had a homeless flag in their record. The address criteria was preferable for this report because it is more reliable for determining the duration of homelessness.

[4] Another such study is: Metraux, S., J. Fargo, N. Eng and D. Culhane. 2018. "*Employment and Earnings Trajectories During Two Decades Among Adults in New York City Homeless Shelters*" *Cityscape*, Vol. 20, No. 2, The Housing-Health Connection (2018), pp. 173-202.

[5] These ethnic categories roll up more detailed categories as follows. African American includes individuals identified as African American or Black. Asian American / Pacific Islander includes individuals identified as Cambodian, Chinese, Filipino, Guamanian, Hawaiian, Japanese, Korean, Laotian, Samoan, Vietnamese, and Other Asian. European American includes individuals identified as White (not of Hispanic origin). Latino includes individuals identified as Hispanic. Native American includes individuals identified as American Indian or Alaska Native. Other includes individuals identified as Other, Unknown or Unspecified.

[6] Johnson, G., Scutella, R., Tseng, Y., Wood G. 2018. "*How do housing and labour markets affect individual homelessness?*" Housing Studies, November 2018.
https://www.researchgate.net/publication/328797374_How_do_housing_and_labour_markets_affect_individual_homelessness

[7] Metraux, S., J. Fargo, N. Eng and D. Culhane. 2018. "*Employment and Earnings Trajectories During Two Decades Among Adults in New York City Homeless Shelters*" Cityscape, Vol. 20, No. 2, The Housing-Health Connection (2018), pp. 173-202.

[8] Fargo, J., N. Eng, S. Metraux, D. Culhane. 2010. *Trends in earnings and employment before and after the first instance of homelessness: A multi-cohort analysis* (conference paper). 138th APHA Annual Meeting and Exposition, November 2010.

[9] At some point within the 14-year time window provided by public benefits records, one-fifth (22 percent) of workers had flags in their records indicating a disability. The prevalence of these flags increased with age. One-fifth (21 percent) of these disability flags were removed within three years. This indicates that at some point in their public benefits histories, 17 percent of workers were identified as having persistent disabilities. Among workers with disability flags, 23 percent also had a NSA (needs special assistance) flag indicating a mental disability.

[10] Zuvekas, S and Hil, S. *Income and employment among homeless people: the role of mental health, health and substance abuse*. The Journal of Mental Health Policy and Economics: 30 April 2001 https://doi.org/10.1002/mhp.94

[11] Poremski, Daniel and Hwang, Stephen. (2016). "Willingness of Housing First Participants to Consider Supported-Employment Services." *Psychiatric Services* (Washington, D.C.). 67. appips201500140. DOI: 10.1176/appi.ps.201500140.

[12] Barnow, B. S., Buck, A., O'Brien, K., Pecora, P., Ellis, M. L., and Steiner, E. (2015). "Effective services for improving education and employment outcomes for children and alumni of foster care service: Correlates and educational and employment outcomes." *Child and Family Social Work*, *20*(2), 159–170.

[13] Four-fifths (81 percent) of the foster youth in the study population emancipated into adulthood at 18 years of age, before AB 12 took effect, and one-fifth (19 percent) were eligible for extended support until they became 21 years of age. Consequently, foster care data in this report that does not break out the pre- and post-AB 12 cohorts is skewed toward youth who did not receive extended foster care services.

[14] Courtney, M. E., Okpych, N. J., & Park, S. (2018). *Report from CalYOUTH: Findings on the relationship between extended foster care and youth's outcomes at age 21*. Chicago, IL: Chapin Hall at the University of Chicago., p. 12.

[15] Milburn, N, E. Rice, M. Rotheram-Borus, S. Mallett, D. Rosenthal, P. Batterham, S. May, A. Witkin, and N. Duan. 2009. "Adolescents Exiting Homelessness Over Two Years: The Risk Amplification and Abatement Model" *Journal of Research on Adolescence*, December 1; 19(4): 762–785.

[16] U.S. Department of Housing and Urban Development. 2017. *HUD 2017 Continuum of Care homeless assistance programs homeless populations and subpopulations*. Accessed on November 27, 2018, at https://www.hudexchange.info/resource/reportmanagement/published/CoC_PopSub_NatlTerrDC_2017.pdf

[17] Toros, H. and Flaming, D. (2016). Identifying and Housing High-Cost Homeless Residents. Technical report. Economic Roundtable. Retrieved from https://economicrt.org/publication/ silicon-valley-triage-tool.

Toros, H. and Flaming, D. (2018). "Prioritizing Homeless Assistance Using Predictive Algorithms: An Evidence-Based Approach". Cityscape: A Journal of Policy Development and Research 20 (1), pp: 117–146.

[18] Byrne, T.; Treglia, D.; Culhane, D. P.; Kuhn, J. and Kane, V. (2015). "Predictors of Homelessness Among Families and Single Adults After Exit From Homelessness Prevention and Rapid Re-Housing Programs: Evidence From the Department of Veterans Affairs Supportive Services for Veteran Families Program" *Housing Policy Debate*, DOI:10.1080/10511482.2015.1060249.

Chan, H.; Rice, E.; Vayanos, P.; Tambe, M. and Morton, M. (2017). "Evidemce from the Past: AI Decision Aids to Improve Housing systems for Homeless Youth" *Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI) 2017 Fall Symposium*

Series. *Retrieved from http://www-bcf.usc.edu/~vayanou/papers/2017/AAAIsymposium2017-evidence-past-ai-accepted.pdf*.

Shinn, M.; Greer, A. L.; Bainbridge, J.; Kwon, J. and Zuiderveen (2013). "Efficient Targeting of Homelessness Prevention Services for Families" *American Journal of Public Health* 103 (S2): pp. S324-S330.

[19] Byrne, T.; Metraux, S.; Moreno, M.; Culhane, P.; Toros, H. and Stevens, M. (2012) "Los Angeles County's Enterprise Linkages Project: An Example of the Use of Integrated Data Systems in Making Data-Driven Policy and Program Decisions" California Journal of Public Policy. 4 (2): pp. 95-112. DOI 10.1515/cjpp-2012-0005.

[20] Lund, B. (2016). Finding and Evaluating Multiple Candidate Models for Logistic Regression, *Proceedings of the SAS Global Forum 2016 Conference*, Paper 7860-2016.

[21] Guyon, I. and Elisseeff, A. (2003). "An Introduction to Variable and Feature Selection" Journal of Machine Learning Research 3: pp. 1157-1182.

[22] Vandekerckhove, J.; Matzke, D. and Wagenmakers, E-J. (2015). "Model Comparison and the Principle of Parsimony" The Oxford Handbook of Computational and Mathematical Psychology. DOI: 10.1093/oxfordhb/9780199957996.013.14.

[23] LASSO (least absolute shrinkage and selection operator) is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the statistical model it produces.

[24] In the traditional implementation of forward and backward selection, the statistic used to gauge improvement in fit is an F statistic that reflects an effect's contribution to the model if it is included. At each step, the effect that yields the most significant F statistic is added or the predictor producing the least significant F statistic is dropped

[25] Flom, P.L. and Cassell, D.L. (2009). "Stopping stepwise: Why stepwise and similar selection methods are bad, and what you should use" Retrieved from https://www.lexjansen.com/pnwsug/2008/DavidCassell-StoppingStepwise.pdf.

[26] Dziak, J, Coffman, D., Lanza, S., Li, R. (2012). Sensitivity and Specificity of Information Criteria, The Methodology Center, Pennsylvania State University, Technical Report Series #12-119. Retrieved from https://methodology.psu.edu/media/techreports/12-119.pdf.

Lund, B. (2017). SAS® Macros for Binning Predictors with a Binary Target, *Proceedings of the SAS Global Forum 2017 Conference*, Paper 969-2017.

[27] Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H., and Lee, T.-C. (1985). The Theory and Practice of Econometrics. 2nd ed. New York: John Wiley & Sons.

[28] Fonti, V. and Belitser, E. (2017) "Feature Selection Using LASSO" Research Paper in Business Analytics. VU Amsterdam. Retrieved from https://beta.vu.nl/nl/Images/werkstuk-fonti_tcm235-836234.pdf.

Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73(3):273–282.

[29] Johnson, G. and Rodriguez, R. N. (2014). "Introducing the HPGENSELECT Procedure: Model Selection for Generalized Linear Models and More" *Retrieved from https://support.sas.com/rnd/app/stat/papers/2014/hpgenselect2014.pd*.

SAS (2017). SAS/STAT 14.3 User's Guide: High-Performance Procedures. SAS Institute Inc., Cary, NC, USA.

[30] Nord, C. and Keeley, J (2016). An Introduction to the HPFOREST Procedure and its Options", *Midwest SAS Users Group Conference*, Paper AA20.

SAS (2017). SAS/STAT 14.3 User's Guide: High-Performance Procedures. SAS Institute Inc., Cary, NC, USA.

[31] Wang F, Casalino LP, Khullar D. "Deep Learning in Medicine—Promise, Progress, and Challenges." JAMA Internal Medicine. Published online December 17, 2018. doi:10.1001/jamainternmed.2018.7117

[32] Shmueli, G. (2010) "To Explain or to Predict?" Statistical Science 25 (3): pp. 289-310. DOI: 10.1214/10-STS330.

[33] Allison, P. (2012). *Logistic regression Using SAS: Theory and Application, Second Edition*. Cary, NC: SAS Institute.

[34] When data is scarce, cross-validation or resampling methods such as bootstrapping and k-fold validation are preferred, but since they are computationally intensive and do not produce different results with large data sets, we used the hold-out sample approach.

[35] Hosmer, D.W. and Lemeshow, S. (2000). *Applied Logistic Regression* (2nd ed.). New York: John Wiley and Sons.

[36] Sensitivity = True Positive /(True Positive + False Negative); Specificity = True Negative /(False Positive + True Negative); PPV = True Positive /(True Positive + False Positive); Accuracy = (True Positive + True Negative) /All

[37] Misclassification rate is often used as a performance measure, which is 1-correct classification rate.

[38] Gonen, M. (2007). Analyzing Receiver Operating Characteristics with SAS. Cary, NC: SAS Press Series.

[39] Couronne, R.; Probst, P. and Boulesteix, A-L. (2017), "Random Forest versus Logistic Regression: a Large Scale Benchmark Experiment" Technical Report Number 205, Dept. of Medical Informatics, Biometry and Epidemiology, LMU Munich. The mean difference was approximately .03 for accuracy and .043 for the area under the curve (AUC).

[40] Breiman (2001) "Random forests". *Machine. Learning,* **45:** pp. 5–32. http://doi.org/10.1023/A:1010933404324.