Pre-Analysis Plan

Youth Homelessness Prevention: A Randomized Control Trial

James Sullivan — **University of Notre Dame**
David Phillips — **University of Notre Dame**

I. Introduction

Historically, policies addressing homelessness have focused on providing housing services to those who are already homeless. These interventions have been criticized, however, for treating the symptoms of homelessness rather than their cause (NCH, 2006)[1]. More recently there has been a "fundamental redirection in the nation's homelessness assistance policies" (Culhane et al., 2010)[2] as policy makers have increasingly focused on homelessness prevention efforts.

At the local level, many of the largest counties in the US, including Santa Clara[3] and Los Angeles[4], have passed ballot measures over the past 5 years providing specific support for preventing and mitigating homelessness. In King County, WA, specifically, voters passed a levy, which included $19 million to develop a homelessness prevention program aimed at youth and families. This project will be conducted in the context of the Youth and Family Homelessness Prevention Initiative (YFHPI) in King County, Washington, where lawmakers are interested in assessing the effectiveness of the program's case management component.

Because the past policy focus has primarily been on treatment, there is little evidence about which prevention programs work best and for whom. Our research project attempts to address this gap in the research by comparing the efficacy and cost effectiveness of two prevention strategies within the sample population. We will use a lottery to measure the effectiveness of (1) a combined program of progressive case management and flexible financial assistance relative to (2) only flexible financial assistance. In practice, average financial assistance is similar between the two groups. Thus, the primary difference between the services provided to the two groups is the level of case management. Our research design considers people assigned to the combined program (case management plus access to flexible funds) the treatment group, and the people assigned to flexible funds only the control group.

The results of this study will also be informative to policymakers and service providers in other communities that are interested in the most effective means of homelessness prevention. Enrollment began in May of 2018 and will continue until we have enrolled 600 participants. Researchers had not received data on outcomes at the time of filing of this pre-analysis plan; only baseline characteristics, case management hours, and financial assistance amounts. Given the potentially wide-ranging effect of homelessness prevention, we identified five outcome domains: a) Housing Stability, b) Health, c) Crime, d) Public Benefits & Employment, and e) Child Welfare.

---

[1] National Coalition for the Homeless (NCH). 2006. "McKinney-Vento Act."
http://www.nationalhomeless.org/publications/facts/McKinney.pdf
[2] Culhane, Dennis P., Stephen Metraux, and Thomas Byrne. 2010. "A Prevention-Centered Approach to Homelessness Assistance: A Paradigm Shift?" Housing Policy Debate, 21(2): 295-315.
[3]https://www.sccgov.org/sites/osh/HousingandCommunityDevelopment/AffordableHousingBond/Pages/home.aspx
[4] https://data.lacounty.gov/stories/s/yg4r-nckc

II. Key Data Sources

A. King County Homeless Management Information Systems

The Homeless Management Information System collects client-level data from all publicly contracted homeless service providers in King County. From this information system, we can observe date-specific service records to track outcomes for both treatment and control group participants. Among other variables, HMIS data will allow us to observe vulnerability assessment scores and entry into emergency shelters or transitional housing. HMIS will act as the primary database for the purposes of merging County records. HMIS ID, name, SSN, date of birth, and other characteristics will be used to link records together.

King County already has access to these records and we have signed a data sharing agreement to facilitate the use of these records for the evaluation.

B. Infutor (Consumer Reference Database)

We will also measure housing stability using comprehensive consumer reference data on household addresses in the U.S. These data allow us to determine how frequently study participants move or have an official address—information that is difficult to obtain through surveys or administrative data records. Infutor is a consumer reference data company that aggregates records from various sources (e.g. utility bills) to construct address histories for most people in the United States. Similar address history data has been used, for example, to longitudinally follow residents affected by rent control in San Francisco (Diamond et al. 2018). In the context of housing stability, this outcome is particularly useful. Compared to shelter entry, the existence of a formal address in such data is a more comprehensive measure of homelessness as it can provide information on unsheltered homelessness and those depending on others for housing.

LEO has signed a Data Licensing Agreement with Infutor that allows for the data to be matched with study records for evaluation purposes.

C. King County Department of Community and Human Services

Prospective clients participate in an in-person interview with a caseworker. The caseworker conducts the informed consent process with interested and eligible clients. During the interview, the caseworker asks for demographic information, age, family composition, gender, ethnicity, employment, education level, language used for consent, agency, etc. This information is recorded in the King County Department of Community and Human Services (DCHS) Clarity Prevention Database, an information management system that will also allow us to track group assignment and services received through YFHPI (duration, frequency, and topic of interaction).

D. Washington State Department of Social and Health Services

We have identified a clear process for linking to additional outcomes through the Research and Data Analysis section of the Washington State Department of Social and Health Services (DSHS). They have internal data on many means-tested programs: TANF, SNAP (Employment Securities Department), subsidized healthcare (Health Care Authority), etc. Child protection and foster care data (Children's Administration) are also internal to DSHS. For their own purposes, they have linked these data to other datasets: arrests (State Patrol and

Juvenile Rehabilitation Administration), inpatient and emergency department visits (Health Care Authority), and earnings (UI records).

The health outcomes are particularly important given the well-documented correlation between homelessness and health and the high health care costs of the homeless. Within 30 days of hospital discharge, 70% of hospital visits by homeless patients result in an inpatient readmission, observation status stay, or emergency department visit (Doran et al., 2013). While previous studies have examined self-reported health (Gubits, et al. 2016), ours would be the first RCT studying effects on outcomes tracked by administrative data.

As an extension to the study focused on housing stability and health outcomes, we also plan to examine the effect of YFHPI on crime. If we are able to procure the data, we plan to track bookings, court appearances and incarceration rates for study participants. Analysis of these outcomes may significantly affect the cost-benefit balance for homelessness prevention services.

We have started the process of seeking approvals from individual agencies to access this data for this project and submitting the project for review by the Washington State IRB. We anticipate that we will be able to procure access to the requested data.[5]

III. Contrast in Services Received between Treatment and Control Groups

As an intermediate outcome we plan to measure take-up for both components of the intervention: case management and emergency financial assistance. We expect the treatment group to receive more case-management services, and we expect that, on average, clients in either the treatment or control group will receive similar amounts of flexible funds.

A. Financial Assistance receipt

1. Primary measure of financial take-up: YFHPI financial assistance
    - a) Continuous measure of dollars received from YFHPI flexible funds
    - b) Constructed using HMIS data
2. Alternative measure of financial receipt: all financial assistance
    - a) Continuous measure of dollars received from financial assistance, both YFHPI and other programs tracked by HMIS
    - b) Constructed using HMIS data

B. Case Management receipt

1. Primary measure of case management receipt: time spent with YFHPI agency case manager
    - a) Continuous measure of hours spent with case manager after intake appointment
    - b) Constructed using HMIS data

[5] The University of Notre Dame Institutional Review Board already approved this research protocol. We are now in the process to get IRB approval from the Washington State IRB.

IV. Hypotheses – Analysis by Outcome Domains

We will analyze outcomes by domain. Within each domain, we indicate a primary measure, which will be the focus of our analysis. We also list alternative measures of interest, which we may investigate to further understand the main effects on the primary outcome measure. We will split outcomes into short-term versus long-term time ranges. We will measure short-term outcomes from the start date to 3 months later. We will measure long-term outcomes from the start date to 24 months later.

A. Indicators for Housing Stability:

1. **Primary Measure for Short-Term Housing Stability: Used homelessness services (excluding Homelessness Prevention)**
    a) Measuring homelessness as implied by the use of homelessness services
        (1) Dummy for whether or not participant received any of the following services: Emergency Shelter, Street Outreach, Permanent Supportive Housing, Rapid Re-Housing, Transitional Housing, Diversion, or Coordinated Entry
        (2) Constructed using HMIS data
    b) Hypothesis: expect treatment group to be less likely to be designated as homeless as measured by use of services outside of Homelessness Prevention.

2. **Alternative Measure: Whether a participant enters a homeless shelter**
    a) Measuring shelter admittance rate
        (1) Dummy for whether or not participant entered shelter
        (2) Constructed using HMIS data
    b) Hypothesis: expect treatment group to have a lower shelter entry rate

3. **Alternative Measure: Number of days spent in a shelter**
    a) Measuring days spent in shelter
        (1) Continuous
        (2) Constructed using HMIS data
    b) Hypothesis: expect treatment group to have lower average number of days spent in a shelter

4. **Alternative Measure: Returns to YFHPI**
    a) Measuring the rate at which a client might return to YHFPI
        (1) Dummy for whether or not client received any emergency financial assistance after exiting the program
        (2) Constructed using HMIS data
    b) Hypothesis: expect treatment group to be less likely to return for financial assistance.

5. **Alternative Measure: Formal address change**
    a) Indicator for whether a participant has changed addresses since random assignment

(1) Dummy for whether most recent address (defined by address with latest verification or start date) is different from the address at the time of random assignment; coded as zero if no formal address after random assignment.

(2) Constructed using consumer reference database Infutor.

c) Hypothesis: expect treatment group to be less likely to move

6. Alternative Measure: Any formal address

a) Measuring whether the participant has a formal address

(1) Dummy for having any address with a last verification or start date after random assignment

(2) Constructed using consumer reference database Infutor

b) Hypothesis: expect treatment group to be more likely to have an address

7. Alternative Measure: Whether a participant is evicted

a) Measuring eviction rate

(1) Dummy for whether or not participant was evicted

(2) Constructed using eviction court records[6]

b) Hypothesis: expect treatment group to have a lower eviction rate

8. Alternative Measure: Neighborhood characteristics for most recent address

a) Measuring the characteristics of the neighborhood at the participant's neighborhood based on the most recent address. Characteristics include neighborhood rent, median income, fraction white, fraction with a bachelor's degree or higher, and crime incidents per square mile.

(1) Continuous

(2) Using the consumer reference database Infutor we can identify the most recent address observed

(3) Only observed for sub-sample ever matching to an Infutor record.

(4) Using the data from the American Community Survey (2019, 5-year estimates) we can calculate all characteristics except crime for the ZIP code of the address. Crime incidents come from the Seattle police.

b) Hypothesis: expect treatment group participant to live in communities with higher median rents. Other characteristics expect to move in the same direction as their correlation with median rent.

B. Indicators for Health

1. Primary Measure for Health: expected overall cost of healthcare

a) Includes emergency department, inpatient visits, and outpatient visits. Cost weights constructed by RDA.

b) The outcome will be measured in dollars as a weighted sum, weighting the number of each type of visit by the cost of such visits.

c) Constructed using the Health Authority data

d) Hypothesis: expect treatment group to have lower health spending

2. **Alternative Measure for Health: indicator for any emergency department visit**
   a) Measuring any ED stay
      (1) Dummy for whether or not participant was admitted to the emergency room (ER inpatient and/or ER outpatient).
      (2) Constructed using the Health Authority data
   b) Hypothesis: expect treatment group to be less likely to have ED visit

3. **Alternative Measure for Health: indicator for any hospital inpatient visit**
   a) Measuring any hospital inpatient stay
      (1) Dummy for whether or not participant was admitted to the hospital for inpatient
      (2) Constructed using the Health Authority data
   b) Hypothesis: expect treatment group to be less likely to have inpatient stay

4. **Alternative Measure for Health: indicator for any outpatient visit**
   a) Measuring any hospital outpatient stay
      (1) Dummy for whether or not participant was recorded an outpatient, non-ED visit
      (2) Constructed using the Health Authority data
   b) Hypothesis: expect treatment group to be less likely to have outpatient stay

5. **Alternative Measure for Health: indicator for Substance Use Disorder treatment** need
   a) Measuring any SUD treatment need
      (1) Dummy for whether or not participant needed SUD treatment
      (2) Aggregated by RDA based on arrest, prescription, diagnosis, and SUD treatment service information.
   b) Hypothesis: expect treatment group to be less likely to need SUD treatment

6. **Alternative Measure for Health: indicator for any mental health treatment need**
   a) Measuring any mental health treatment need.
      (1) Dummy for whether or not participant needed mental health treatment.
      (1) Aggregated by RDA based on prescription, diagnosis, and mental health treatment service information
   b) Hypothesis: expect treatment group to be less likely to need mental health treatment

7. **Alternative Measure for Health: indicators by diagnosis type and measure**
   a) We will split the above measures by diagnosis and type of treatment

C. Indicators for Crime

1. **Primary Measure for Crime: Any arrest**

a) Measuring police arrests
   (1) Dummy for whether or not ANY member of household was ever arrested
   (2) Constructed using data from Washington State Patrol and Juvenile Rehabilitation Administration
b) Hypothesis: expect treatment group to be less likely to be arrested

2. Alternative measure: Juvenile Incarceration

a) Indicator for whether or not youth in household was incarcerated in Juvenile Rehabilitation Facility
   (1) Dummy for whether or not ANY member of household was incarcerated in a juvenile rehabilitation facility
   (2) Constructed using data from the Juvenile Rehabilitation Administration
b) Hypothesis: expect treatment group to be less likely to be incarcerated

3. Alternative Measure: Number of arrests

a) Measuring police arrests
   (1) Continuous; count of arrests
   (2) Constructed using Washington State Patrol data
b) Hypothesis: expect treatment group to have fewer arrests

4. Primary Measure for Crime: Any charge

a) Measuring court charges
   (1) Dummy for whether or not ANY member of household was ever charged with a crime
   (2) Constructed using data from Washington State Patrol and Juvenile Rehabilitation Administration
b) Hypothesis: expect treatment group to be less likely to be charged

5. Alternative Measure: Types of charges

a) Measuring charges
   (1) Dummy; any charge in category
   (2) Separate outcomes by type of crime. Type can include severity (felony, misdemeanor, gross misdemeanor) and category (assault, theft, sex, dv, custody, alcohol/drug, trespassing, driving, license, weapons, supervision, murder, failure to comply, and other).
   (3) Constructed using Washington State Patrol data
b) Hypothesis: expect treatment group to have fewer arrests of various types

D. Indicators for Public Benefits & Employment

1. Primary Measure for Benefits: receiving any public assistance

a) Indicator for whether or not participant receives any public assistance

(1) Public assistance includes any ESA service, TANF/SFS, SNAP, HEN
(2) Constructed using the Economic Service Administration database
(3) Dummy: 1 if yes: 0 if no

b) Hypothesis: expect treatment group to be less likely to receive public assistance

2. Primary Measure for Employment: hours worked per quarter

a) Measuring hours employed
(1) Continuous measure for hours worked per quarter
(2) Constructed using data from the Employment Security Department

b) Hypothesis: expect treatment group to work more hours

3. Alternative Measure: earnings per quarter

a) Measuring income from employment
(1) Continuous measure for dollars per quarter of earnings
(2) Constructed using data from the Employment Security Department

b) Hypothesis: expect treatment group to have greater earnings

4. Alternative Measure: receipt of different types of assistance

a) Indicator for whether or not participant receives a particular type of assistance
(1) Measured separately for TANF/SFA, SNAP, HEN
(2) Constructed using the Economic Service Administration database
(3) Dummy: 1 if yes: 0 if no

b) Hypothesis: expect treatment group to be less likely to receive public assistance of various types

E. Indicators for Child Welfare – Child Protection

1. Primary measure: Any contact with Children's Administration

a) Indicator for whether any child in the household received any service or interacted with Washington Children's Administration
(1) Dummy: 1 if yes: 0 if no
(2) Constructed using data from the Children's Administration

b) Hypothesis: expect treatment group to be less likely to interact with child protective services

2. Alternative Measure: Indication of abuse

a) Indicator of whether Children's Administration has record that the child was ever maltreated, physically abused, neglected, or sexually abused. This will be analyzed collectively and individually.
(1) Dummy: 1 if yes: 0 if no
(2) Constructed using data from the Children's Administration

b) Hypothesis: expect treatment group to be less likely to experience abuse

3. **Alternative Measure: Foster Care Status**

   a) Indicator of whether Children's Administration had placed the child out of the home
      (1) Dummy: 1 if yes: 0 if no
      (2) Constructed using data from the Children's Administration

   b) Hypothesis: expect treatment group to be less likely to be placed in foster care

4. **Alternative Measure: Behavioral Services**

   a) Indicator of whether youth received behavioral rehabilitation services
      (1) Dummy: 1 if yes: 0 if no
      (2) Constructed using data from the Children's Administration

   b) Hypothesis: expect treatment group to be less likely to need behavioral rehabilitation services

V. Sub-Group Analysis

We are interested in determining whether the intervention matters more for certain populations relative to others. This is particularly relevant because of the resource-intensive nature of the intervention; given its expense it may be important to target the intervention towards the population most likely to benefit from it.

A. Sub-Group Analysis: Risk score
   1. Eligible range is 15 to 28.
   2. Linear interaction of score with treatment effects.

B. Sub-Group Analysis 2: Race/Ethnicity
   1. Black = Self-identified as Black or African American; Other = everyone else
   2. White = Self-identified; Other = everyone else

C. Sub-Group Analysis 3: Gender – Males vs. Females
   - Female = Self-identified as female; Male= Self-identified as male

D. Sub-Group Analysis 4: Age
   - Above vs. below the median age at baseline

E. Sub-Group Analysis 5: Disability
   - Any Household Member has a Disability= Self-identified as member with disability

F. Sub-Group Analysis 6: Predicted outcomes (halves)

For each outcome we construct, we will predict the outcome in the control group using a regression of the baseline value of all primary outcomes, risk score dummies, gender, race dummies, ethnicity dummies, month of interview dummies, and months between baseline and follow-up dummies. We will then predict the outcome for all observations using observed characteristics and the coefficients from that regression. We will finally split the sample into halves based on that index and test for heterogeneous effects in the two groups. To avoid endogenous stratification we compute these statistics with a repeated split sample procedure as in Abadie, Chingos, and West (2018, RESTAT).

VI. Data Analysis

A. Estimates

For continuous outcomes, we will estimate treatment effects by OLS using the following regression:

$$Y_i = \alpha_0 + T_i\beta_0 + X_i\gamma_0 + \epsilon_i$$

Y_i is the outcome. T_i is an intent-to-treat dummy indicating the random assignment of person i. In the case of non-compliance, T_i takes on the value of the original random assignment. The vector X_i includes a set of person-level characteristics collected at baseline, and ϵ_i is an error term. The coefficient on the treatment dummy β_0 will give us the difference in means between the treatment and comparison groups, the estimated impact of the program. The full estimation sample will include about 300 individuals in the treatment group and 300 individuals in the control group.

For dichotomous outcomes, we will use a linear probability model:

$$Y_i = \alpha_1 + T_i\beta_1 + X_i\gamma_1 + \epsilon_i$$

where Y_i is the dummy variable.

B. Covariates

We plan to include the following list of covariates in our regressions:

1. Value of dependent variable at baseline, if available
2. Age
3. Age squared
4. Gender – 1= female; 0 otherwise
5. Household size
6. Set of mutually exclusive variables for race/ethnicity
7. Set of mutually exclusive variables for imminent risk of homelessness (e.g. self-reported, vacate notice, recent major trauma or event)
8. Dummy for history of homelessness
9. Number of household children
10. Indicator for whether household has a member with a disability
11. Set of mutually exclusive dummies for month of random assignment

C. Standard Errors

We will use simple heteroskedasticity-robust standard errors.

D. Multiple Hypothesis Testing

We have limited our primary outcomes to a causal chain with one primary outcome per domain, making multiple hypothesis testing less of a concern. Nonetheless, we will report classic p-values. This provides the reader with full information that they can use to make multiple hypothesis testing corrections if they desire.