

**Pre-Analysis Plan**  
**Testing the Impact of Subsidized Public Transit for Unemployed Persons**

**David Phillips**  
University of Notre Dame

**Matthew Freedman**  
University of California Irvine

**May 2023**

**I. Introduction**

There is substantial evidence to show that transportation poses a barrier for low-income individuals searching for jobs, especially for those residing in neighborhoods further away from job opportunities (Phillips, 2014). If an individual can't pay for transportation to job interviews or training, job search intensity wanes and poverty persists.

Several researchers have studied the impact of free public transportation on people's travel habits, finding that immediate transit use and mobility increase significantly. For example, one field experiment that provided free public transportation to a large group of employed individuals found a 12% increase in overall travel and a 23% increase in travel during off-peak hours (Bull et al., 2021). Another large-scale experiment that provided free transit to individuals with low income documented a doubling of transit use (Brough et al., 2022). These results show high potential for an uptake in transit usage and could suggest an even larger jump for a low-income population. However, a follow-up paper showed limited effects of the transit subsidy on hours worked (Brough et al., 2023). Null results for employment may have occurred because that study sample was too detached from the labor market or because free transit affects employment outcomes beyond hours worked and wage rates. The Wilson Sheehan Lab for Economic Opportunities (LEO) has partnered with the Seattle Department of Transportation (SDOT) and Uplift Northwest (Uplift), a temporary staffing agency serving a population of overwhelmingly low-income and/or homeless people, to bridge this gap by pairing free transit to people enrolled in Uplift's job matching programs and studying its impact on employment and self-sufficiency outcomes using Uplift's detailed data on participants' interactions with the labor market.

LEO will be conducting a randomized controlled trial (RCT) to evaluate this program. Clients of Uplift who consent to participation will be randomly assigned to receive either a one-time \$10 transit card or a fully subsidized transit card. Outcomes of interest include transit usage and labor market outcomes. We will also measure benefits utilization (such as SNAP and TANF) and housing stability.

**II. Evaluation Design**

*A. Research Questions*

The research question we seek to answer in this study is whether there is an impact of receiving a fully subsidized ORCA card on employment and well-being.

*B. Eligibility*

All temporary employees at Uplift will be eligible for the study. For new temporary employees at Uplift, there will be a representative from SDOT there to introduce the study and conduct enrollment. Existing Uplift employees will also be free to approach the SDOT representative to enroll in the study. Clients who wish to participate will read and sign a consent form and be randomized on the spot.

*C. Randomization*

We expect a sample size of 832 over two-years of enrollment. SDOT does not have sufficient funding to provide fully-subsidized transit passes to more than half of a study population this large. Randomization is therefore both the most equitable way and the most rigorous way—with respect to causal evidence creation—to

distribute fully-subsidized transit passes. The process of randomization will be automatic, happening entirely within the intake survey, which will be administered by a SDOT representative at Uplift. Consenting individuals will be randomly assigned independently with a 50% chance to be selected for the treatment group and a 50% chance to be selected for the control group.

#### *D. Intervention*

The intervention is simply receiving a fully subsidized one-year transit pass. Control group participants will receive a one-time \$10 transit card.

#### *E. Power Calculations*

About 832 consenting individuals will participate after two years of enrollment with 50% in each group (based on historical enrollment numbers and 80% consent rate). Assuming only 70% of participants take up the subsidized transportation and an  $R^2$  on the controls of 0.10, the study is powered to detect a 155 hour change in time worked annually, or a change of about 3 hours worked per week. This is within Uplift Northwest and SDoT's goal of increasing hours worked to 15 hours per week.

### **III. Key Data Sources**

The following section summarizes the planned primary data sources for this project. Notably, securing access to these data sources is partially complete. Given this, any outcomes for which we do not already have data secured may ultimately be excluded if there are barriers to gathering the needed information.

#### *A. Uplift Northwest*

We plan to use data about clients of Uplift Northwest participating in the study, including baseline identifying information, as well as information about the type of job opportunities, hours worked, and income. In particular, Uplift NW employs participants directly, so it has access to data including time sheets and characteristics of jobs worked (location, job title, industry, wage rate, hours worked, shift worked, and other job benefits). These data will allow us to measure the effect of free transit on employment decisions and outcomes. LEO and Uplift Northwest have executed a DSA allowing for the regular transfer of this data.

#### *B. King County Metro*

We plan to collect data from King County Metro to measure the effect of free transit on transit use. These data will allow us to see for each participant how often and what modes of public transit was used, as well as whether the participant is enrolled in any other transit subsidy programs. LEO and King County Metro are amending an existing DSA to allow the research team for this study to access the necessary data.

#### *C. King County Homeless Management Information System (HMIS)*

We plan to use data on homeless shelter stays from HMIS to measure the effect of free transit on housing stability. LEO is working to create a DSA that will allow us to access HMIS data for this study.

#### *D. Infutor Data Solutions*

We also plan to use data from Infutor to measure the effect of free transit on housing stability. Infutor holds data on the address histories of individuals in the United States. With this data, we will be able to measure how frequently study participants have an address change, allowing us to measure housing stability. LEO has an existing relationship with Infutor which will allow us to collect this data on study participants.

*E. Washington State Department of Social and Health Services (DSHS)*

We plan to use data from several Washington state agencies housed in DSHS. These data will include employment, public benefits (such as SNAP and TANF), contact with the criminal justice system, and healthcare use. This will allow us to assess the effects of transit subsidies on a broad set of outcomes.

#### **IV. Hypotheses: Analysis by Outcome Domains**

There are multiple secondary domains, which are final outcomes that might be directly affected by the subsidy. These secondary outcomes will be measured both for the time-period when the transit subsidy is active, as well as after the transit subsidy has expired.

*A. First Stage: Travel habits and use of public transportation*

1. Primary measure: Number of boardings using study transit card
  - a) Description of measure: Measures the number of boardings on the study cards
    - (1) Continuous measure of the number of boardings during peak and off-peak transit times per day
    - (2) Constructed using King County Metro boardings data
  - b) Hypothesis: Expect treatment group to board transit more frequently as compared to the control group
2. Alternative measure: Number of trips using study transit card
  - a) Description of measure: Measures the number of trips, defined as continuous transit within two hours, on the distributed cards
    - (1) Continuous measure of the number of trips during peak and off-peak transit times per day
    - (2) Constructed using King County Metro boardings data
  - b) Hypothesis: Expect treatment group to have more transit trips as compared to the control group
3. Alternative measure: Use of study card after subsidy expires
  - a) Description of measure: Measures whether study participants load value to their ORCA LIFT card after the subsidy expires
    - (1) Continuous variable for the amount of value that is loaded to the study card
    - (2) Constructed using King County Metro sales data
  - b) Hypothesis: Expect control group to load more value to study cards than the treatment group during the duration of the transit subsidy
  - c) Hypothesis: Expect treatment group to load more value to study cards than the control group after the transit subsidy expires
4. Alternative measure: Subsets of boardings by mode, line, time of day and location
  - a) Description of measure: Measures number of boardings by mode, timing, and location date/time of ORCA card taps, as well as geographical information including latitude/longitude of taps, routes used, and stops visited.
    - (1) Continuous variable counter number of boardings for subsets of boardings

(2) Constructed using King County Metro boardings data

B. *Primary Domain: Employment*

1. Primary measure (Uplift data): Hours worked per week at jobs offered by Uplift Northwest
  - a) Description of measure:
    - (1) Continuous variable for hours worked per week
  - b) Hypothesis: Expect treatment group to work more hours per week as compared to the control group
  
2. Alternative measure (UI data): Hours worked per quarter at all UI-covered jobs
  - a) Description of measure:
    - (1) Continuous variable for hours worked per quarter
  - b) Hypothesis: Expect treatment group to work more hours per quarter as compared to the control group
  
3. Alternative measure (Uplift data): Income earned from jobs offered by Uplift Northwest
  - a) Description of measure: Income from jobs offered by Uplift Northwest
    - (1) Continuous variable for income earned
  - b) Hypothesis: Expect treatment group to earn greater income through employment with Uplift Northwest as compared to the control group
  
4. Alternative measure (UI data): Total earnings
  - a) Description of measure:
    - (1) Continuous variable for total earnings
  - b) Hypothesis: Expect treatment group to have higher total earnings as compared to the control group
  
5. Alternative measure (Uplift data): Job characteristics
  - a) Description of measure:
    - (1) Continuous variables for distance of jobs from place of residence, any employment, wage rate, number of shifts worked, hours per shift, time of day worked, presence of non-wage benefits, average wages of job title
  - b) Hypothesis: Expect treatment group to work more jobs that are further from their place of residence and of higher quality, compared to the control group
  
6. Alternative measure (UI data): Job characteristics
  - a) Description of measure:
    - (1) Measures of any work, wage rate, industry (particularly temporary vs permanent)
  - b) Hypothesis: Expect treatment group to have a higher quality employment
  
7. Alternative measure (Uplift data): Job performance
  - a) Description of measure: Indicators for negative reports from employers, including incidents and no-show
    - (1) Internal Uplift records of having any incident or having a no-show for a job

- b) Hypothesis: Expect treatment group to be less likely to have negative job incidents and no-shows

C. *Secondary Domain: Well-Being*

1. Primary measure (HMIS data): Homeless Program Use
  - a) Description of measure: Indicator for enrolling in any program in HMIS that is limited to people who are already homeless
  - b) Hypothesis: Expect treatment group spend less time in homeless shelters as compared to the control group
2. Alternative measure (Uplift data): Professional development opportunities utilized
  - a) Description of measure: Indicator for any participation in Uplift Northwest's assortment of job skills trainings and opportunities to earn technical certifications
  - b) Hypothesis: Expect treatment group to utilize more professional development opportunities as compared with control group
3. Alternative measure (Uplift data): "points" accrued
  - a) Description of measure:
    - (1) Continuous variable for points accrued for development and training
  - b) Hypothesis: Expect treatment group to earn more points as compared with the control group
4. Alternative measure (public benefits): Receiving any public assistance
  - a) Description of measure: Indicator for whether participant receives any public assistance
    - (1) Public assistance includes any ESA service, TANF/SFS, SNAP, HEN
    - (2) Constructed using the Economic Service Administration database (RDA data)
    - (3) Dummy: 1 if yes; 0 if no
  - b) Hypothesis: Ambiguous; transit may make it easier to enroll in public assistance programs, but could also reduce eligibility by increasing income
5. Alternative measure (public benefits): Receipt of different types of assistance
  - a) Description of measure: Indicator for whether participant receives a particular type of assistance
    - (1) Measured separately for TANF/SFA, SNAP, HEN
    - (2) Constructed using the Economic Service Administration database (RDA data)
    - (3) Dummy: 1 if yes; 0 if no
  - b) Hypothesis: Ambiguous; transit may make it easier to enroll in public assistance programs, but could also reduce eligibility by increasing income
6. Alternative measure (health): Expected overall cost of healthcare
  - a) Description of measure: Includes emergency department, inpatient visits, and outpatient visits. Cost weights constructed by RDA.

- (1) 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.
    - (2) Constructed using the Health Authority data
  - b) Hypothesis: Ambiguous; treatment group may have better physical health but may use subsidized transit to access more healthcare services
  
7. Alternative measure (health): Indicator for any emergency department visit
  - a) Description of measure: Measures any ED stay
    - (1) Constructed using the Health Authority data
    - (2) Dummy for whether an individual was admitted to the emergency room (ER inpatient and/or ER outpatient).
  - b) Hypothesis: Ambiguous; treatment group may have better physical health but may use subsidized transit to access more healthcare services
  
8. Alternative measure (health): Indicator for any hospital inpatient visit
  - a) Description of measure: Measuring any hospital inpatient stay
    - (1) Dummy for whether patient was admitted to the hospital for inpatient
    - (2) Constructed using the Health Authority Data (RDA data)
  - b) Hypothesis: Ambiguous; treatment group may have better physical health but may use subsidized transit to access more healthcare services
  
9. Alternative measure (health): Indicator for any outpatient visit
  - a) Description of measure: Measuring any hospital outpatient stay
    - (1) Dummy for whether participant was recorded as outpatient, non-ED visit
    - (2) Constructed using Health Authority Data (RDA data)
  - b) Hypothesis: Ambiguous; treatment group may have better physical health but may use subsidized transit to access more healthcare services
  
10. Alternative measure (criminal justice involvement): Arrest within one year
  - a) Description of measure: Measures whether an individual is arrested within one year of enrolling in the study
    - (1) Dummy for whether an individual is arrested
    - (2) Categorical variable describing the charge for which the individual was re-arrested
    - (3) Constructed using Research and Data Analysis (RDA) data
  - b) Hypothesis: Expect treatment group to have fewer arrests as compared to those in the control group
  
11. Alternative measure (Infutor data): Whether individual moves
  - a) Description of measure: Measures whether individual moves
    - (1) Dummy for whether individual who matches to the Infutor database moves
    - (2) Constructed using address information from Infutor consumer reference data
    - (3) This analysis is limited to the sample that matches to address history data at baseline.

- b) Hypothesis: Expect treatment group to be more likely to move as compared to the control group

12. Alternative measure (Infutor data): Move location

- a) Description of measure: Measures neighborhood characteristics of housing moves including whether out of county, proximity to transit, and accessibility of jobs by transit
  - (1) Dummy for whether individual who matches to the Infutor database moves to location with selected characteristic
  - (2) Constructed using address information from Infutor consumer reference data
  - (3) This analysis is limited to the sample that matches to address history data at baseline.
- b) Hypothesis: Expect treatment group to be less likely to move out of county

**V. Balance Checks**

*A. Demographics*

- 1. Race
- 2. Ethnicity
- 3. Gender
- 4. Age
- 5. Baseline transit use

*B. Values of outcome measures in the quarter prior to random assignment*

- 1. Employment
  - a) Hours worked (Uplift)
  - b) Hours worked (All)
  - c) Total earnings
  - d) Location of jobs worked
  - e) Measures of job quality
- 2. Travel habits and use of public transportation (survey questions from study enrollment form)
  - a) Number of days per week used public transit
  - b) Modes of transportation used in last two weeks
  - c) Mode of transportation to Uplift the day of study intake
- 3. Benefits utilization
  - a) Receiving any public assistance
  - b) Expected overall cost of healthcare
- 4. Criminal justice outcomes
  - a) Arrests
- 5. Housing stability
  - a) Homeless program use

**VI. Subgroup Analysis**

We are interested in determining whether the intervention is more effective for certain populations relative to others.

#### *A. Causal forest*

Our primary method for analyzing heterogeneity will be the causal forest method of Wager and Athey (2018). For covariates, we will include lagged values of all outcomes from the balanced panel and all variables listed in the section above on balance checks. Results of this causal forest analysis will inform which simple subgroup analyses we report.

#### *B. Payment method at baseline*

Identifies whether at baseline the person paid for transit with an ORCA card, cash, or some other method/nonpayment. The latter option is particularly important for understanding if the presence of fare evasion moderates treatment effects.

### **VII. Data Analysis**

#### *A. Estimates*

We will estimate intent-to-treat (ITT) treatment effects by OLS using the following regression:

$$Y_{it} = \alpha_0 + T_i\beta_0 + X_i\gamma_0 + \epsilon_{it}$$

We will analyze the data using a balanced panel in time relative to random assignment where  $i$  denotes the person and  $t$  denotes time relative to random assignment. We will analyze data at a weekly time frequency or, if not available weekly, at the finest unit of time available in the data.  $Y_{it}$  is the outcome.  $T_i$  is an intent-to-treat dummy indicating random assignment. 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  $\epsilon_{it}$  is an error term. The coefficient on the treatment dummy  $\beta_0$  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 416 individuals in the treatment group and 416 individuals in the control group.

#### *B. Covariates*

We will report all analyses without covariate adjustment and with covariate adjustment. We plan to include the following list of covariates in our regression:

1. Value of dependent variable at baseline, if applicable
2. Set of mutually exclusive dummies for month of enrollment

#### *C. Standard Errors*

Standard errors will be clustered at the individual-level, as treatment assignment is randomized on the individual-level.

#### *D. Multiple Hypothesis Testing*

The research team has limited their primary outcomes to those described above, which each fall under distinct domains. Classic p-values will be reported for all outcomes, which will provide a reader with full information that they can use to make multiple hypothesis testing corrections if they desire.

### **VIII. References**



- Brough, Rebecca, Freedman, Matthew, Phillips, David C., 2022. Experimental Evidence on the Effects of Means-Tested Public Transportation Subsidies on Travel Behavior. *Regional Science and Urban Economics*, vol. 96, p. 103803. <https://doi.org/10.1016/j.regsciurbeco.2022.103803>
- Brough, Rebecca, Freedman, Matthew, Phillips, David C., 2023. Eliminating Fares to Expand Opportunities: Experimental Evidence on the Impacts of Free Public Transportation on Economic Disparities (Working Paper). [https://conference.nber.org/conf\\_papers/f178633/f178633.pdf](https://conference.nber.org/conf_papers/f178633/f178633.pdf)
- Bull, Owen, Muñoz, Juan C., Silva, Hugo E., 2021. The Impact of Fare-Free Public Transport on Travel Behavior: Evidence from a Randomized Controlled Trial. *Regional Science and Urban Economics*, vol. 86, p. 103616. <https://doi.org/10.1016/j.regsciurbeco.2020.103616>
- Phillips, David C., 2014. Getting to Work: Experimental Evidence on Job Search and Transportation Costs. *Labour Economics*, vol. 29, pp. 72–82. <https://doi.org/10.1016/j.labeco.2014.07.005>
- Wager, S. and Athey, S., 2018. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523), pp.1228-1242.