Impact Evaluation of the NPower Tech Fundamentals Program: Pre-Analysis Plan

Tania Barham, University of Colorado Boulder
Brian Cadena, University of Colorado Boulder
Patrick S. Turner, University of Notre Dame, Lab for Economic Opportunities (LEO)

March 22, 2021
Note: We registered this trial and posted our pre-analysis plan after randomization began in November 2020, but prior to the receipt of any application, program, or outcome data.

Introduction

This plan outlines the hypotheses and empirical approach used to measure the impact of the NPower Tech Fundamentals program. This pre-analysis plan has been created prior to any data collection.

Overview of the Study

We will implement a randomized controlled trial to evaluate the impact of the NPower Tech Fundamentals program. NPower Inc. works with young adults and veterans and their spouses to create career pathways into the information technology industry through training and job placement. The population served by NPower is economically disadvantaged and composed of racial and ethnic groups that are underrepresented in technology careers.

At the beginning of the program, NPower students participate in a tuition-free, 16-week Tech Fundamentals training course. Students participate in instructor-led training and are given the opportunity to work toward industry-recognized certification. Throughout the program, they are exposed to mentors from the IT industry, complete employment readiness workshops, and receive on-going support from a social support manager who helps them identify and overcome barriers to program completion. Graduates of the training course are placed in a 7-week paid internship with a local company to further develop hard and soft skills. Upon completion, program alumni receive job placement assistance to help them match to an entry level IT position with top employers in the area.

Experimental Design
LEO will conduct a randomized controlled trial evaluation of the NPower Tech Fundamentals program in Harlem, NY and Dallas, TX. The research team will work with NPower to introduce a lottery into the current recruitment process. At each location during the year, NPower enrolls two cohorts of program participants that roughly begin in March and September. Recruitment for the next cohort begins as soon as the last cohort has launched.

The research study relies on experimental variation in program access to identify causal estimates of program impacts. The study uses nonexperimental variation in program access, generated by in-person interviews with program staff, to explore the degree to which impacts would scale to a broader applicant pool.

Applicants apply to the Tech Fundamentals online through an application. Within that application, potential study participants will be introduced to the research study and asked to provide informed consent. Those who consent will receive a gift card of nominal value to thank them for their participation. The NPower recruitment manager then reviews the application to ensure the applicant meets the specific criteria for program eligibility. Eligible applicants are invited to the program office for interviews with the recruitment manager, social support manager, and program director to assess their suitability. They recruit to fill 45–50 spots in each cohort, enrolling two cohorts at a site each year.

We will implement stratified randomization after individuals complete the online application. The recruitment manager will review the application for eligibility and will determine whether the applicant should be invited to interview with program staff. The recruitment manager will enter eligible and suitable applicants into a lottery to determine treatment status. Randomization will be stratified based on site (Harlem and Dallas), gender (male vs. female/other), and educational attainment (HS diploma/GED vs. some college/Associates vs. currently enrolled/Bachelors). We will generate a series of random treatment assignments to ensure equal treatment probability (two-thirds) across strata, and assign an applicant the next treatment status on the list at the time they enter the lottery.[1] Those who are assigned to the treatment group will be invited to interview with program staff. Based on their performance in the interviews, treatment group members will either be (i) invited to enroll in the program, (ii) placed on a wait list, or (iii) informed they were not accepted into the program. Those who are assigned to the control group will not be invited to continue the process. The assigned treatment status thus represents an intent-to-treat indicator among NPower applicants eligible for interview.

As part of the initial application, we will collect baseline demographic and socioeconomic data, as well as identifiers needed to link to administrative records, for all consenting applicants. This group includes all applicants who enter the lottery, as well as applicants who are screened out by NPower prior to the lottery.

The proposed project has two main aims. The first is to better understand the program's current selection process and to describe the type of applicant who is admitted into the final step of the program's application process. The second is to directly measure the causal impact of the program on individual outcomes. Using administrative data, such as earnings data, we will compare the outcomes of three groups of applicants: those who apply but do not enter the lottery, those who are assigned to the control group, and those who are assigned to the treatment group. Because the assignment of lottery participants to the treatment group and control group is random, comparing these two groups yields the causal impact of the program.

Proposed Timeline

NPower operates two cohorts of their program each year, and will recruit study participants over 4 cohorts at two locations. The current timeline of the study is scheduled as follows:

[1] This approach is similar to the one used by Barham, Cadena, and Turner (2019) in an evaluation of ReHire Colorado, a transitional jobs program. To construct the list, we will draw short sequences of 0s and 1s in equal proportion and randomly vary the length of the sequence. Thus, over short sets of applicants within a strata, the treatment probability will remain roughly 66 percent. NPower program staff who administer the recruitment process will not have access to the prepopulated treatment assignments, nor the algorithm used to construct the randomization lists.

November 2020: Begin enrolling study participants

August 2025: Observe six quarters of earnings data for all 6 cohorts, pull initial education data

While we aim to recruit for this timeline, we will allow for some flexibility in implementation as our partner copes with challenges presented by COVID-19 and virtual enrollment and program delivery in the near future. See below for more details surrounding contingency plans.

Sample Selection and Treatment Assignment

The study sample will follow current NPower applicant requirements. In Harlem, the program works specifically with young adults. Applicants must be 18–25 years old, have at least a high school diploma or equivalent credential, and have an income under 200 percent of the local median income. In Dallas, the program targets military veterans and their spouses. Applicants must be at least 21 years of age and either be a currently serving or honorably discharged veteran or the spouse of a military veteran. At both locations, applicants must be legally authorized to work in the U.S. without requiring employer sponsorship.

We aim for an experimental sample size of 1,500 study participants divided 2:1 between treatment and control with an expected take-up rate (share of treatment group who enroll in the program) of 54 percent.[2] We would reach our target sample by recruiting roughly 126 study participants at 2 locations (Harlem and Dallas) over 3 years with 2 cohorts per year. In a given recruitment cycle, NPower receives roughly 300 applications. Thus, our total study sample will include an additional 1,000 applicants with which to explore program selection.

Data Sources

NPower Intake Application

Socioeconomic characteristics (age, education, parent's education, family income, gender, marital status, number of children); personal identifiers (name, date of birth, address, SSN, phone number);

NPower Recruitment Data

NPower will share information about each applicant's performance throughout the interviews, including when they dropped out of the recruitment process.

NPower Program Data

Information on attendance, participation, and program completion.

Earnings Data from Unemployment Insurance Records (UI) and US Treasury Tax Data

[2] We believe this take-up rate is conservative. The estimate is based on conversations with program leadership and their understanding of historical acceptance rates. To fill a class of 45 students, we project enrolling roughly 126 applicants into the lottery, with 84 being placed into the treatment group. Of those, 45 will pass the interviews and enroll in the program, 5 will pass the interviews and be placed on the waitlist, and 34 will not be accepted into the program. If the waitlist members enroll in subsequent cohorts, or more applicants are accepted, the take-up rate could be higher.

We will attempt to access administrative earnings data from two potential sources: quarterly state unemployment insurance records; individual-level tax returns. Participants will share personal identifiable information to facilitate data linkage. To the extent that outliers may impact the analysis we will use three methods to examine their impact. We will top-code earnings at the 99th percentile, we will trim earnings at the 99th percentile, and estimate the impact of the program on income rank (with no top-coding or trimming of the data).

Quarterly UI Data: Using this data we will construct measures of quarterly earnings, quarterly employment, and if possible, number of employers and industry of employment. To accommodate the rolling admissions, quarterly outcomes will be measured relative to their application quarter (based on start date of the cohort). For example, Quarter 0 for individuals who apply for a cohort that begins July 22, 2021 would be calendar quarter 2021Q3.

Treasury Data: Using this data we will construct measures of annual earned income. This data has the benefit of more fully capturing income earned across multiple sources (including contract work not observed in UI data). To accommodate rolling admissions, annual outcomes will be measured relative to their application year (based on start date of the cohort). For example, Year 0 for individuals who apply for a cohort that begins July 22, 2020 would be calendar year 2020.

Secondary Outcomes

We will endeavor to link study participants to additional administrative data sets to measure secondary outcomes. If possible, these will include post-secondary education data from the National Student Clearinghouse, SNAP and TANF receipt data, Medicaid participation and claims, credit bureau records, address histories from Infutor data solutions, and LinkedIn utilization and network data.

Hypotheses

Selection Characteristics: NPower tries to recruit and select program participants who will be most successful in their program using a number of written and technical assessments.

Data: Pre-application employment and earnings; NPower application

Primary Characteristics: Age; marital status; felony conviction status; educational attainment; father's education; mother's education; veteran status; pre-application values of primary and secondary outcomes; outcome of NPower interviews (i.e., NPower's subjective scoring of program readiness)

Short- and Medium-Run Earnings and Education Outcomes: Participation in the NPower Tech Fundamentals Program may decrease earnings in the short-run as participants engage in the 23-week training program. In the medium-run (two quarters after application and beyond), we anticipate that the program will increase earnings as individuals move from low-wage employment to higher-wage employment. While earnings likely improve, the effect on post-secondary educational attainment is ambiguous. The program may substitute for education in the short-run, but participants may pursue further educational attainment after the program to the extent that their careers require further credentials or degrees for advancement.

Data: Post-application quarterly employment and earnings; indicator for any post-secondary enrollment; indicator for completion of Associate's degree; indicator for completion of Bachelor's degree

Primary Outcome: Average earnings in the second year after application. In the quarterly UI data, this would represent earnings in quarters 4–7. In the tax data, this would represent Year 2. Because earnings data tend to be noisy, we will also construct an indicator of whether the individual earned more than $30,000 in the year (full-time work at $15/hour).

Estimation Methodology

Selection
We will use Ordinary Least Squares regression to compare the pre-application characteristics of those accepted to the Tech Fundamentals to those who were not accepted to the Tech Fundamentals program. The goal of the analysis is to describe the extent to which the NPower screening process selects applicants of differing characteristics.

Treatment effects
We will use Ordinary Least Squares regression to estimate intent-to-treat treatment effects. We will regress the outcome of interest on the treatment group indicator and present the coefficients on treatment that are estimated from two different specifications:

(1) A model containing only stratification fixed effects (Gender-Education-City-Cohort)
(2) A model with controls selected by the post-double selection LASSO procedure of Belloni, Chernozhukov, and Hansen (2014). The set of potential control variables will include all available baseline controls.

For continuous outcomes like earnings, we will also test for equality of distributions using a Pearson's chi-squared test and/or Kolmogorov-Smirnov test, and use quantile regressions.

We will report heteroskedasticity-robust standard errors, as well as randomization-based p-values. Specifically, we will simulate 10,000 placebo treatment assignments by reseeding the randomization tool that created the initial treatment assignments. We will report the share of t-statistics from re-estimating the model with the placebo treatment statuses that are larger in absolute value than the t-statistic estimated from the actual treatment assignment (Young 2019).

Heterogeneous Effects
We will explore subgroup heterogeneity across six groups defined by three baseline characteristics: gender (male & female/other), education (HS grad/GED & some college), and baseline employment (employed & not employed). We will also look at effects of the program separately by the target population at each site: veterans and young adults.

We will also characterize heterogeneity using the machine learning procedure of Chernozhukov et al. (2018) or related methods.

Multiple Hypothesis Testing

Our primary outcomes of interest are defined above. However, our domains of interest—i.e., "labor market outcomes", "education outcomes", "public health insurance utilization", etc.—can be measured by multiple variables. When reporting estimates of individual outcomes, we will report standard p-values and also p-values adjusted to account for multiple outcomes (Westfall and Young 1993). We will also summarize the impact on a domain by following the methodology employed by Finkelstein et al. (2012).

To correct for concerns of multiple comparisons with subgroup analysis, we will follow the two approaches used in Chetty et al. (2016).

(1) We will conduct parametric joint F tests of the null hypothesis that the effect on earnings is zero for both subgroups when dividing the sample by gender (male vs. female/other), education (HS grad/GED vs. some college), and baseline employment (employed vs. not employed). We will re-estimate the treatment effect within each subgroup and test whether the effects for each subgroup are jointly zero.
(2) We will conduct a permutation test that corrects for the possibility of finding a significant effect across each of the 8 pre-specified subgroups (male, female/other, HS grad/GED, some college, employed, not employed, veterans, young adults). Using 10,000 placebo treatment assignments, we will re-estimate the treatment effects within each subgroup. Using the reported p-value from the subgroup, we will report the share of placebo samples where at least one of the six subgroups had a p-value below the corresponding actual p-value. This approach adjusts the p-value to account for the likelihood one would, through random chance, find an effect of at least the statistical significance we find across any of the subgroups.

Missing Data

If a control variable is missing for a significant fraction of the population, we will omit the control from our regressions. For control variables that are missing from a small share of respondents, we will impute the value of the control at the average. As a robustness check, we will include a dummy variable that indicates whether the individual had an imputed value for that control.

Our approach to measuring outcomes relies on linking participants to administrative data. Thus, we will only be able to include study participants in the analysis for whom we can observe the PII needed to link to the dataset (e.g., SSN). If a study participant is missing a needed identifier, we will exclude that individual from the analysis. In the case of earnings, if a person has the appropriate identifier (e.g., a valid SSN) but does not return records in the data (e.g., a quarter of earnings) we will impute a value of zero for earnings and employment for that quarter. To the extent possible, we will attempt to determine whether we have the right person if an individual's PII returns no records of earnings across all periods.

Contingency Plans

Our research design and power calculations relied on assumptions surrounding consent rates and program take-up among the treatment group. Our target sample of 1,300 individuals in the experimental sample assumes most applicants will consent to be a part of the study, and roughly 54 percent of the treatment group will enroll in the program. We will monitor consent rates, the quality of the SSN data, and take-up rates. If the research team determines that low consent, prevalence of missing SSNs, or low take-up rates will meaningfully impact power, we will extend study recruitment beyond the currently planned four cohorts and/or two sites. If consent rates are low, we will work with NPower to address the issue, potentially increasing the gift card incentive amount or structuring a different incentive (e.g., $100 gift card awarded randomly to one-tenth of consenters.)

COVID-19

The current public health crisis (COVID-19) may impact program and evaluative operations. At this time, we anticipate two potential concerns related to COVID and address how we will adapt the study in response.

(1) *Program Model:* The pandemic may substantially impact program delivery, such that the traditional in-person environment is shifted to a hybrid or virtual program model. To the extent that the program delivered during the early cohorts of the evaluation is substantially different from the traditional program model, we will work with NPower to extend study recruitment to maintain statistical power needed to estimate the impact of the traditional program model or the on-line model if they don't go back to the traditional model. We will take two approaches in analysis:
    (a) Report impacts with/without COVID cohorts
    (b) Include a Treatment x COVID dummy interaction in the analysis to control for differences in impacts across COVID cohorts
    (c) Include type of delivery and Treatment x Delivery Type dummies in the analysis to examine if the effect differs by delivery method.

    In addition, we may recruit additional cohorts to improve power among non-COVID cohorts.
(2) *Recruitment*: COVID-19 may impact NPower's ability to meet the recruitment targets needed to achieve a 66 percent randomization rate. In the event that NPower is unable to recruit a full cohort under the 66 percent randomization protocol, we will provide a randomly selected set of control group members for NPower to interview for the program.
(3) *Implementation Delays*: Recruitment issues during COVID-19 could also delay the implementation of the randomization protocol. If randomization is postponed, we would still aim to recruit at least 6 cohorts of study participants at both sites, which would delay the time until analysis could be completed.

References

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen, 2014, "Inference on Treatment Effects after Selection among High-Dimensional Controls," *The Review of Economic Studies*, 81(2):608–50.

Chetty, Raj, Nathan Hendren, and Lawrence F. Katz, "The effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment," *American Economic Review*, 106(4):855–902.

Chernozhukov, Victor, Mert Demirer, Esther Duflo, and Ivan Fernandez-Val. Generic machine learning inference on heterogenous treatment effects in randomized experiments. *NBER Working Paper No. w24678*. National Bureau of Economic Research, 2018.

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, Katherine Baicker, Oregon Health Study Group, 2012, "The Oregon Health Insurance Experiment: Evidence from the First Year," *The Quarterly Journal of Economics*, 127(3):1057–1106.

Young, Alwyn, 2019, "Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results," *The Quarterly Journal of Economics*, 134(2):557–98.