Pre-Registration for "The Effects of Moving to Opportunity on Youth Political Behavior" Using the "AsPredicted" Template on the OSF

1. **Have any data been collected for this study already?**
   - **Yes, at least some data have been collected for this study already**
   - **No, no data have been collected for this study yet**

   Yes. All data for this study have already been collected, but the research team does not have access to the data at the time of pre-registration. The research team will be licensing data about the Department of Housing and Urban Development's Moving to Opportunity (MTO) program from the National Bureau for Economics Research. The MTO data will be matched at the individual level with voter registration records. The matching process has not yet begun at the time of pre-registration.

2. **What's the main question being asked or hypothesis being tested in this study?**

   This paper uses the Moving to Opportunity (MTO) housing experiment to evaluate the causal impact on political participation of being assigned to — and for some, complying and living in — less-distressed neighborhoods. Our outcome of interest is registration and voter turnout. We examine outcomes for participants who were adults, teens, and children at the time of random assignment.

   We hypothesize that when children are assigned to a low-poverty voucher condition, their political participation will be higher later in life than those who are assigned to a no-voucher condition. Those assigned to a standard voucher condition (unrestricted) will exhibit weaker effects. We expect those assigned as adults or teens to be much less positively affected, likely exhibiting null or negative effects. This is consistent with studies of childhood anti-poverty interventions on adult voting (Holbein 2017), and with prior studies exploring the effect of MTO on other outcomes, such as income, that are immediate predictors of voting (Chetty, Hendren, and Katz 2016). Furthermore, we expect that longer exposure to better neighborhoods as a child increases voting rates as an adult. The MTO data include administrative data on individual program assignment and utilization, which are the key predictors. If the compliance variables are available, we expect that individuals who complied with the treatment more fully and for longer periods of time will have larger gains in voting rates.

3. **Describe the key dependent variable(s) specifying how they will be measured.**

   Our outcome of interest is registration and voter turnout. We examine outcomes for adults, teens, and children at the time of random assignment.

   Registration is measured as the proportion of MTO subjects in each random assignment group who correctly matched to the voter file. This estimand is unbiased if we assume that our ability to match is equal across treatment groups and attrition via incapacitation is equal across groups. However, it is possible that a subject who was registered to vote at some point after random assignment was subsequently removed from the voter file (death, name change, imprisonment, subsequent moves, etc.), but this cannot be observed. To that end, we define the primary estimand as the percent registered among all possible voters in each group and thus incorporate attrition into this estimand. However, with this approach, bias can still result from differences in ability to match the voter file with MTO data, but this cannot be observed.

To measure turnout, we examine general and primary elections from 2000 through 2018. In examining turnout, rather than looking at a single election, we average turnout over a number of elections in order to minimize the influence of idiosyncratic variation. Because we know the date when the subject registered to vote, subject age, and the date of assignment in the MTO experiment, we can measure participation before 2000 as a pre-treatment covariate (if applicable and available), post-MTO participation, and post-registration date participation.

Voter turnout can be examined as proportion voting among all subjects or only among those registered. Both approaches are subject to potential bias from differential non-matching across treatment and control. We will evaluate turnout both ways, for robustness. We plan to test for balance in pre-treatment covariates among missing and non-missing subjects, and if missingness is balanced between treatment and control, this will provide some assurance that missingness will not cause bias. However, such balance does not account for unobserved covariates that may be related to both missingness and potential outcomes; thus false negative and false positive are both possible even after accounting for covariate balance.

4. How many and which conditions will participants be assigned to? (optional)

Households were assigned to one of three groups: 1) receive a traditional Section 8 housing voucher (traditional voucher condition); 2) receive a housing voucher that could only be used in a Census tract with less than 10% poverty (low-poverty voucher condition); or 3) receive no voucher (with the ability to remain in public housing) (control condition).

5. Specify exactly which analyses you will conduct to examine the main question/hypothesis.

All treatment effects can be measured by comparing the control to the treatment groups using randomization inference that clusters the subjects by MTO site and family. We will also include pre-treatment covariates available from L2 and the MTO baseline survey, including pre-treatment turnout (if applicable and available), race, and gender. We will also conduct our analysis with and without the weights to adjust for uneven probabilities of randomization (Orr et al. 2003). If compliance variables are available, we will measure a complier average causal effect (CACE); if not, we will measure an intent to treat effect (ITT).

Each analysis will be conducted separately by age at the time of random assignment. As noted above, a person who was younger at the time of random assignment, especially in years prior to adulthood, is likely to be more affected by the treatment. As such, we will test for different effects by age in three ways: 1) conducting subgroup analysis, with the sample divided between people over 18, 13 – 18, and under 13 at random assignment; and 2) interacting these age subgroups by treatment.

6. Any secondary analyses?

We will test the moderating effects of gender, race/ethnicity (for racial subgroups with sufficient power), duration of exposure/moving frequency (if available), and city at random assignment (if sufficiently powered).

Additionally, we will conduct mediation analysis, using nonparametric procedures (Imai et al. 2010) to estimate the average causal mediation effect and average direct effect of each respective mediator. For example, prior research suggests the MTO intervention led to gains in education and later earnings for children whose families received vouchers to live in low-poverty neighborhoods. These are plausible mechanisms for higher rates of political participation later in life. Mediation analysis will allow us to explore the degree to which any gains in political participation among children assigned to the low-poverty vouchers are driven by higher educational attainment and improved earnings. We also test other mediators as explained below.

Education will be coded as a set of indicator variables: high school completion (or GED), college attendance, and if sufficiently powered, college completion. We lack full data on college attendance. We estimate that we will have a college attendance measure for 76% of the adult cohort based on their responses to the interim and final household evaluations. We estimate having a college attendance measure for 73% of the teen cohort from the interim household survey, given the proportion of the teen cohort that was college age by 2002 and the interim survey response rate. Depending on data availability and the portion of the children's cohort over the age of 18 interviewed in the final youth evaluation, we estimate having a college attendance measure for 23-38% of the children's cohort from the final youth evaluation, final household, and the interim household surveys. We will have a measure of college completion for the 71% of adults who completed the final household evaluation in 2008. We will not have a measure of college completion for the teen or children's cohort. Finally, we will have a pre-treatment measure of whether adult household members were high school graduates or had a GED.

Income will be coded as raw annual earnings. Income data will come from two possible sources: income derived by L2 from commercial data (c. 2017) and income from the MTO data. The latter includes a measure of individual income from the respondent's main job, recorded by the MTO final evaluation surveys for the household and youth (2008). If available, we will supplement the adult measure from the household survey with an MTO measure of income from *all* of their employers in the previous year, measured in raw earnings or income categories depending on whether the respondent supplied the raw dollar amount. If available, we will also use a measure of total household income from the MTO final evaluation (measured in raw earnings). From these available measures, we estimate having a 2008 MTO measure of income for approximately 71% of the adult cohort and 16-23% of the children's cohort, assuming that all members of the children's cohort who were 18 or older as of December 31, 2007 have an applicable measure of income. None of the teen cohort will have the 2008 MTO income measure. If available, we will also calculate a pre-treatment measure of income for the head of household from their earnings per hour, their hours worked per week, and the number of months worked in the preceding year.

AFDC use: Available pre-treatment for all heads of households. It will also be used as a post-treatment mediator, if available, for the heads of households in the final household evaluation and the subset of the children's cohort that was administered the final youth evaluation.

Marital status: Available pre-treatment for all adults in the household, available for non-head members of the household from the interim household evaluation, and available for non-head members of the household from the final household evaluation. If available, we will also include a measure of martial status for the head of household from the final household evaluation. Assuming that this measure is applicable for those 18 and over, the final household evaluation will provide a measure of marital status for 71% of the teen cohort and 38% of the children's cohort, given the survey response rate.

Parental status: Only indirectly and imprecisely available, and only for the subset of the children's cohort that was administered the final youth evaluation. It can be derived from a question asking who the respondent lives with, since one of the answer choices is their own children. If available, we will also use an item on the final household evaluation measuring the number of biological children each household member has, as well as a direct measure of parental status from the final youth evaluation.

Time in jail or prison: Available for household members *but not head of household* from the final household evaluation in 2008. If available, we will also measure for how long and the last time household members were in jail or prison, both appearing on the final household evaluation. Members of the teen cohort would be 23-32 in this wave, and members of the children's cohort would be 10-26. Assuming that these measures are applicable for all those 18 and over, the final household evaluation will provide measures of time in jail or prison for 71% of the teen cohort and 38% of the children's cohort, given the survey response rate.

Index of social connection: Only non-neighborhood specific measures are available post-treatment, though neighborhood-specific measures are available pre-treatment. For the heads of household, we measure how often they had friends over, how often they visited friends, and how often they go to church. For the subset of the children's cohort participating in the final youth evaluation, we measure how they met their friends, how often they visit friends they had at the time of random assignment, and how often they attend youth religious activities. For the heads of household, we also have pre-treatment measures of how often they talk to neighbors, how many friends and family members live in their neighborhood, and their likelihood of telling a neighbor if the neighbor's child had gotten into trouble.

Index of crime exposure: For the heads of household, we measure in 2008 whether they have recently seen drug use or sales, the safety of streets, and whether abandoned buildings, graffiti, and trash are a problem. For the subset of the children's cohort participating in the final youth evaluation, we measure whether they have recently seen drug use or sales and the safety of the streets. For the heads of household, we also have pre-treatment measures of street safety and, if available, whether they have been victims of crime in the recent past.

Index of perceived racial discrimination: For the heads of household, we measure in 2008 whether they have faced discrimination in the past 6 months in their own neighborhood and with police. For the subset of the children's cohort participating in the final youth evaluation, we measure whether they have faced discrimination from the police in the past 6 months.

Dissatisfaction with neighborhood: For the subset of the children's cohort participating in the final youth evaluation and for the heads of household, we measure in 2008 satisfaction with their neighborhood. For the heads of household, we also have a pre-treatment measure of satisfaction with their neighborhood.

Under-served by police: Available for the heads of household from the 2008 final household evaluation.

Moving frequency and housing stability: If available, we will use the following measures from the 2002 interim household survey to construct measures of moving frequency and housing stability: length of time the head of household has been in their current housing space; whether the head of household was without housing in the past year; the amount of time it has been since the head of household last rented or owned their own unit; the length of time that the head of household has lived in their current neighborhood; and the number of times the head of household moved since the year of random assignment.

In addition to the poverty level mentioned above as a compliance variable, we will test as mediators variables that measure the neighborhood's distress and isolation: a set of variables measuring economic distress (percent unemployed; percent TANF-receiving; percent female-headed household; aggregate income per working-age adult; percent under 18); level of racial segregation; and tract-level crime rates. These will be measured for the focal neighborhood, which is the first neighborhood after treatment, and if available, subsequent neighborhoods.

We will apply the FDR correction to the set of mediation tests to adjust for multiple comparisons, as needed.

We can also check our results for sensitivity to false positive and negative matches by restricting our data by various levels of the match probability. The matching procedure (see below) will return a posterior probability that two entries are a match, as well as a false discovery rate and false negative rate at different probability cutoffs for the fuzzy matching algorithm. These cutoffs correspond to lower bounds for the posterior probability at which a match will be accepted as true. Thus, we will restrict our dataset to more or less restrictive match thresholds (0.85, 0.9, 0.95, 0.99) and test whether our primary results substantively differ with match criteria.

7. **How many observations will be collected or what will determine the sample size? No need to justify decision, but be precise about exactly how the number will be determined.**

   The MTO initiative randomized approximately 4,600 poor families. The approximate numbers of individual MTO participants are: 4,616 adults, 2,331 teens (13 – 18 at random assignment), and 8,945 children (under age 13 at random assignment).

8. **Anything else you would like to pre-register? (e.g., data exclusions, variables collected for exploratory purposes, unusual analyses planned?)**
   The matching procedure is as follows:

   To perform the match, we use the fuzzy-matching algorithm (see the *fastLink* package in R; Enamorado, Fifield, and Imai 2018), which returns multiple matches with a probabilistic estimate for each pairwise match of being correct. We will exact match on gender, such that all female MTO participants will be matched to individuals in the voter file who are listed as "female" or "unknown," and all male MTO participants will be matched to individuals in the voter file who are listed as "male" or "unknown". To reduce the number of pairwise comparisons, we will then derive clusters of individuals in the voter file and the MTO dataset with maximally similar first names (see Enamorado, Fifield, and Imai 2018). We then derive partial matches based on first name, last name, middle name, suffix, and birthdate. The probabilistic estimate returned by the algorithm will allows us to restrict our final dataset to different thresholds to check the sensitivity of our results to potential false positives.