eAppendix for: Helped into Harm: Mediation of a housing voucher intervention on mental health and substance use in boys

1 Baseline covariates

- Adolescent characteristics: site (Boston, Chicago, LA, NYC), age, race/ethnicity (categorized as black, latino/Hispanic, white, other), number of family members (categorized as 2, 3, or 4+), someone from school asked to discuss problems the child had with schoolwork or behavior during the 2 years prior to baseline, child enrolled in special class for gifted and talented students.
- Adult household head characteristics included: high school graduate, marital status (never vs ever married), whether had been a teen parent, work status, receipt of AFDC/TANF, whether any family member has a disability.
- Neighborhood characteristics: felt neighborhood streets were unsafe at night; very dissatisfied with neighborhood; poverty level of neighborhood.
- Reported reasons for participating in MTO: to have access to better schools.
- Moving-related characteristics: moved more then 3 times during the 5 years prior to baseline, previous application for Section 8 voucher.

2 Additional variables included in imputation

- Baseline variables:
  - Child-specific: child had health problems that limited activity, health problems that necessitated special medicine or equipment, child had behavioral problems or emotional problems in class, child was suspended or expelled in the past year
  - Household-specific: household owned a car, household received SSI, household member was victimized in the past 6 months, head of household: chatted with a neighbor at least once per week, very likely to tell on a neighborhood kid in trouble, had no family living in the neighborhood, had no friends living in the neighborhood, very sure about finding an apartment, and had primary or secondary reason for moving being because of drugs or crime

- Interim variables: household head work status, working at least 35 hours per week, and warmth of interaction between parent and child during home observation
- Final variables: youth behavioral problem index

3 Estimator

Our estimation approach is described in **?**. Briefly, we use a targeted minimum loss-based estimator of $\theta = E(Y_{a'}, G_{a^\star})$. Targeted minimum loss-based estimation is a type of estimation approach based on solving the efficient influence function, given below (from **?**) for the binary Z of moving with the voucher we use here.

The efficient influence function for θ in the nonparametric model M is equal to $D_P(o) - \theta$, where

$$
\begin{aligned}
D_P(o) = & \frac{1\{a = a'\}}{\mathsf{g}(a' \mid w)} \mathsf{c}(a', z, m, w)\{y - \mathsf{b}(a', z, m, w)\} && (1)\\
& + \frac{1\{a = a'\}}{\mathsf{g}(a' \mid w)} \{\mathsf{u}(1, a', w) - \mathsf{u}(0, a', w)\} \left\{z - \mathsf{q}(1 \mid a', w)\right\} && (2)\\
& + \frac{1\{a = a^\star\}}{\mathsf{g}(a^\star \mid w)} \left\{ \sum_{z=0}^{1} \mathsf{b}(a', z, m, w)\mathsf{q}(z \mid a', w) - \mathsf{v}(a^\star, w) \right\}, && (3)\\
& + \mathsf{v}(a^\star, w), \text{ where} && (4)
\end{aligned}
$$

$$
\mathsf{c}(a, z, m, w) = \frac{\mathsf{g}(a \mid w)}{\mathsf{g}(a^\star \mid w)} \frac{\mathsf{q}(z \mid a, w)}{\mathsf{r}(z \mid a, m, w)} \frac{\mathsf{h}(a^\star \mid m, w)}{\mathsf{h}(a \mid m, w)} \tag{5}
$$

$$
\mathsf{u}(z, a, w) = \mathsf{E}\left\{ \mathsf{b}(A, Z, M, W)\mathsf{c}(A, Z, M, W), \middle| Z = z, A = a, W = w \right\}, \tag{6}
$$

$$
\mathsf{v}(a, w) = \mathsf{E}\left\{ \int_{\mathcal{Z}} \mathsf{b}(a', z, M, W)\mathsf{q}(z \mid a', W)\mathrm{d}\nu(z) \middle| A = a, W = w \right\}. \tag{7}
$$

Consequently, we fit nuisance parameters: $\mathsf{g}(a \mid w)$, $\mathsf{q}(z \mid a, w)$, $\mathsf{r}(z \mid a, m, w)$, $\mathsf{c}(a, z, m, w)$, and $\mathsf{b}(a, z, m, w)$. We then use these fits in fitting 2 additional nuisance parameters: $\mathsf{u}(z, a, w)$ and $\mathsf{v}(a, w)$. We use the Superlearner ensemble approach in model fitting,**?** weighting fits from the following algorithms: generalized linear models (including with interaction terms), bayesian generalized linear models, generalized additive models, and regression splines (MARS). We plug these fits into the TMLE algorithm, given in **?**.

This estimation approach is robust under the following conditions:

1. $\mathsf{v}_1 = \mathsf{v}$ and either $(\mathsf{q}_1, \mathsf{h}_1, \mathsf{r}_1) = (\mathsf{q}, \mathsf{h}, \mathsf{r})$ or $(\mathsf{b}_1, \mathsf{q}_1) = (\mathsf{b}, \mathsf{q})$, or
2. $\mathsf{g}_1 = \mathsf{g}$ and either $(\mathsf{q}_1, \mathsf{h}_1, \mathsf{r}_1) = (\mathsf{q}, \mathsf{h}, \mathsf{r})$ or $(\mathsf{b}_1, \mathsf{q}_1) = (\mathsf{b}, \mathsf{q})$.

Thus, $\mathsf{q}(z \mid a, w)$ always needs to be correctly specified.

eTable 1: Effect of randomized voucher receipt on self-reported aspects of school and social environments (measured at the final timepoint) among boys in the Moving to Opportunity, 1994-2010. Survey weighted and combined across 10 imputed datasets. All results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-001.

Outcome	Estimate	95% confidence interval
Hangs out at a friend's house $\geq$once per week	-0.06	(-0.11, -0.01)
Has one or more close friends	-0.03	(-0.05, -0.00)
Sees baseline neighborhood friends a few times a week	-0.07	(-0.12, -0.03)
Suspended or expelled from school in the past 2 years	0.05	(0.01, 0.10)