Documentation of the MTO Public Use Datasets

for the *New England Journal of Medicine (NEJM)* Article:

"Neighborhoods, Obesity, and Diabetes – A Randomized Social Experiment"

by

Jens Ludwig, Lisa Sanbonmatsu, Lisa A. Gennetian, Emma Adam,
Greg J. Duncan, Lawrence F. Katz, Ronald C. Kessler, Jeffrey R. Kling,
Stacy Tessler Lindau, Robert C. Whitaker, and Thomas W. McDade

New England Journal of Medicine, Vol. 365 no. 16 pp. 1509-1519

Published October 20, 2011

Documentation last updated:

December 9, 2013

Data were provided by the U.S. Department of Housing and Urban Development. The contents of this document do not necessarily reflect the views or policies of the U.S. Department of Housing and Urban Development or the U.S. Government.

This documentation was prepared with the assistance of Matt Sciandra and Ray Yun Gou, research analysts at the National Bureau of Economic Research.

Contents

Also included are the following tables:

DOCUMENTATION OF

MTO PUBLIC USE DATASETS

RELATED TO *NEW ENGLAND JOURNAL OF MEDICINE (NEJM)* ARTICLE:

"Neighborhoods, Obesity, and Diabetes – A Randomized Social Experiment"

***New England Journal of Medicine*, Vol. 365 no. 16 pp. 1509-1519**

1. Overview

The public use files (PUFs) are designed for a rough replication of the Moving to Opportunity (MTO) analysis in Tables 1, 2 and 3 of the article "Neighborhoods, Obesity, and Diabetes – A Randomized Social Experiment" published in the *New England Journal of Medicine* (*NEJM*) on October 20, 2011 (available at www.nejm.org/doi/pdf/10.1056/NEJMsa1103216). The article shows that moving from a high-poverty to lower-poverty neighborhood was associated with modest but potentially important reductions in the prevalence of extreme obesity and diabetes.

The two PUFs available at www.nber.org/mtopuf and that will also be made available at the Inter-university Consortium for Political and Social Research's (ICPSR) Replication Datasets collection (www.icpsr.umich.edu/icpsrweb/ICPSR/index.jsp) are:

- **mto_nejm_puf_cells_20131025.dta** – To preserve confidentiality, the data have been collapsed to 81 cells (ranging in size from 14 to 143 respondents). The cells are constructed using data for all 3,273 adults interviewed as part of the MTO long-term survey. Note that the *NEJM* sample (n = 3,186) was slightly smaller because it limited the analyses to female adults for whom we had either obesity or diabetes information. The 81 cells are homogenous on MTO site and treatment group status to allow for the estimation of impacts. When possible, the cells have also been divided by treatment compliance status and race. The file contains the mean, standard deviation, and sum of weights for all outcomes and mediators. There are 200 variables on the file (see Appendix A for a complete list of variables).

- **mto_nejm_puf_pseudo_20131025.dta** – The cell-level data have been expanded to a pseudo-individual level dataset (n=3,273) that for each outcome mimics that outcome's mean value, standard deviation, and approximate number of observations within a cell. However, the outcome values for individual records within a cell do not reflect the actual values on the original individual level file. The file has 75 variables (see Appendix B for a complete list).

The U.S. Department of Housing and Urban Development (HUD) provided the MTO data and must be acknowledged in any paper using these PUFs. The contents of this document do not necessarily reflect the views or policies of HUD or the U.S. Government.

MTO is a randomized housing experiment administered by HUD that gave low-income families living in high-poverty areas in five cities the chance to move to lower-poverty areas. Families were randomized

into one of three treatment groups: (1) a low-poverty voucher group who received a housing voucher that could only be used in a low-poverty (< 10%) census tract, (2) a traditional voucher group who received an unrestricted hosing voucher, or (3) a control group who did not receive a voucher but remained eligible for any government assistance to which they otherwise would have been entitled.

The PUFs contain information on:

- long-term obesity and diabetes outcomes
- housing mobility, neighborhood safety, health care access, and social networks;
- census-tract characteristics including neighborhood poverty and minority concentration, share of single female-headed families, and share of college-educated residents;
- demographic information such as age, gender, race/ethnicity, employment status, and education level;
- other baseline indicators such as income, neighborhood safety and satisfaction, and reasons for wanting to move via MTO; and
- program information such as treatment group, randomization site, and treatment compliance.

Note that the *NEJM* sample was slightly smaller than the sample presented here because the *NEJM* analyses excluded adult males (who comprise about 2% of the MTO interviewed adults) and excluded individuals with missing information on the two main outcomes.

Note that a few variables included on the PUFs have been modified to ensure data confidentiality (see variables with "rad_" in their names). Before the data were aggregated, some demographic information was replaced with group averages. Census tract characteristics have also been rounded.

The PUFs allow one to estimate impact estimates for the voucher groups jointly or separately and to estimate impacts controlling for baseline site. It is not possible to include controls for the full set of baseline covariates that were used in the *NEJM* article because of the aggregation of the data. Tables 3-7 in this memo compare estimates using individual-level data versus the pseudo-individual level data. These tables illustrate that, despite these limitations, the estimates using the PUFs produce means, point estimates, and standard errors that are quite similar to those in the original *NEJM* article. To more precisely estimate impacts and use a more complete set of covariates, the MTO individual-level data are needed. A more complete restricted access dataset will be archived later with ICPSR.

2. Background on the MTO Experiment

The MTO demonstration was authorized by the U.S. Congress in section 152 of the Housing and Community Development Act of 1992. HUD launched MTO to test whether offering housing vouchers to families living in public housing projects in high-poverty neighborhoods of large inner cities could improve their lives and the lives of their children by allowing them to move to lower-poverty neighborhoods.

From 1994 to 1998, the MTO demonstration enrolled 4,604 low-income households in Baltimore, Boston, Chicago, Los Angeles, and New York. Eligibility for MTO was limited to households with children in

public or other government-subsidized, project-based housing in selected high-poverty areas. Enrolled families were assigned at random to one of three groups:

1. The **low-poverty voucher (LPV) group** (also called the experimental group) received Section 8 rental assistance certificates or vouchers that they could use only in census tracts with 1990 poverty rates below 10 percent. The families received mobility counseling and help in leasing a new unit. One year after relocating, families could use their voucher to move again if they wished, without any special constraints on location.

2. The **traditional voucher (TRV) group** (also called the Section 8 group) received regular Section 8 certificates or vouchers that they could use anywhere; these families received no special mobility counseling.

3. The **control group** received no certificates or vouchers through MTO, but continued to be eligible for project-based housing assistance and whatever other social programs and services to which they would otherwise be entitled.

Forty-eight percent of families in the LPV group and 63% of families in the TRV group "complied" with the treatment by moving using a housing voucher obtained through MTO. For more details on the motivation for and structure of the MTO experiment, please see the Chapter 1 of the MTO Final Impacts Evaluation report (Sanbonmatsu et al., 2011).

Because it was implemented as an experiment, MTO overcomes some of the empirical challenges of identifying neighborhood effects on people's life outcomes that have limited previous research. An MTO-type experiment enables us to determine whether moving to a lower-poverty neighborhood itself, rather than some other characteristic of the individuals or families that might be related to both their propensity to move and their behavioral outcomes, directly *caused* improvements in health, economic security, or some other outcome of interest. Because of random assignment, the control group's experience shows, on average, what would have happened to the families in the treatment groups had they not been offered a voucher through MTO.

Researchers have collected survey data on MTO participants at different points in time:

- Baseline (1994-1998): At the time families applied for the program, the household head filled out a survey with information about the household and basic information about each household member.
- Short-Term Site by Site Findings - Preliminary studies conducted a few years into the program by research teams at each site
- Canvasses (1997 and 2000): Families were canvassed and asked a limited set of questions.
- Interim evaluation (2002): Abt Associates, along with the National Bureau of Economic Research (NBER), conducted an evaluation of the program 4 to 7 years on average after random assignment (interviews were completed with 3,519 adults). The interim survey asked questions about: housing, neighborhood, employment and education, income and public assistance, outlook and social networks, physical and mental health, and household composition. (Links to the main findings from the interim evaluation are available at www.mtoresearch.org/interim.htm. Orr, et al.

(2003) and Kling, Liebman, and Katz (2007) are the most comprehensive sources for the interim findings.)

- Final impacts evaluation (2008-2010): 10-15 year follow-up with families. This wave of data collection is the basis for the *NEJM* article.

In addition, researchers have also conducted qualitative interviews with families (see www.mtoresearch.org/qualitative.htm).

3. Sample, Sources, and Measures

3.1 Sample

The PUF datasets consist of data on the 3,273 adults interviewed as part of the long-term MTO evaluation. Researchers at NBER conducted the evaluation and the Institute for Social Research (ISR) at the University of Michigan (through a subcontract) interviewed MTO adults and youth. ISR completed the interviews between June 2008 and April 2010. One adult was selected for interview from each LPV and control group household and, for budgetary reasons, one adult was selected from a random two-thirds subsample of the TRV group households. The overall effective response rate (ERR) for the adult survey was 89.6%, and the ERRs by MTO treatment group were similar: 90.8% for the low-poverty voucher group, 86.6% for the traditional voucher group, and 90.0% for the control group. The long-term adult survey instrument is available at www.mtoresearch.org/instruments/final_hhold.pdf.

3.2 Data Sources

The data analyzed for the *NEJM* article and included in the PUFs come from physical measurements, dried blood spot assays, survey data, and also census data linked to participants' residential address histories.

Physical Measurements: During the long-term survey, interviewers measured height and weight using a modified version of the protocols from the University of Michigan Health and Retirement Study (HRS; ISR 2008). For height, respondents removed their shoes and stood with their heels and shoulders against a wall. Height was marked on the wall with the use of a rafter angle square and measured to the nearest 0.6 cm (0.25 in.) with a metal tape measure. For weight, respondents also removed their shoes as well as heavy outer clothing and items from their pockets. Weight was measured to the nearest 0.23 kg (0.5 lb) with a digital electronic floor scale (Health o meter [Pelstar], model 800KL), which had a maximum capacity of 180 kg (397 lbs). If height or weight could not be measured or if subsequent quality checks revealed an unusual value after the interview, respondent self-reports were obtained.

Respondents were provided with a form indicating their height and weight measurements and were encouraged to share this record with their doctor on their next visit. More detailed protocols for the height and weight measurements can be found at www.mtoresearch.org/instruments.html.

Dried Blood Spot Assays: During the long-term survey, interviewers collected a small sample of blood via finger prick following HRS protocols. An autoretractable lancet finger stick was used to collect up to five drops of whole-blood capillary samples on specimen-collection paper (Whatman no. 903) . To avoid risks

to participants, respondents with a history of a bleeding disorder or who were taking medication that could affect coagulation were excluded. Respondents were given information about values that are considered within normal limits by the American Health Association: total cholesterol < 240 mL, high-density lipoprotein (HDL) cholesterol > 40 mL, glycosylated hemoglobin (HbA1c) < 7 mg/dL, and C-reactive protein (CRP) < 3 mg/L. Interviewers allowed the blood spots to dry for 15 minutes and then placed them in a foiled envelope with a desiccant packet. Interviewers mailed the samples once per week to a central location for proper freezer storage at –20°C to –30°C. Because the interviews for Section 8 adults started about eight months after experimental and control interviews, more of the Section 8 assays were done in the later batches by the laboratory. Samples were assayed at a laboratory with Clinical Laboratory Improvement Amendments certification (FlexSite Diagnostics) with the use of a Roche COBAS Integra immunochemical analyzer that was validated for use with dried blood spots and certified by the National Glycohemoglobin Standardization Program. A more detailed protocol for the dried blood spot collection can be found at www.mtoresearch.org/instruments.html.

Survey Data: The adult survey asked respondents a variety of questions about their economic circumstances, physical and mental health, neighborhoods, housing, social networks, and other topics. The adult survey was designed to take about 75 minutes, with an additional 20 minutes allotted for physical measurements and dried blood spot collection.

Census Data Linked to Address Histories: To help us understand the neighborhood conditions in which MTO families were living during the course of the program, we reconstructed each family's residential history from random assignment onward. Our strategy was to assemble a best guess of the family's residential history from administrative records and previous canvasses and surveys of MTO families and then ask MTO adults to confirm or correct their full history. After constructing the residential histories, we geocoded all addresses to 1990 and 2000 Census tracts and linked the tracts to data from those two decennial censuses as well as to the 2005-09 American Community Surveys 5-year averages data. The tract characteristics were then linearly interpolated and extrapolated to capture neighborhood characteristics at baseline, 1 year after random assignment, 5 years after random assignment, and 10 years after random assignment, and to calculate the tract characteristics of all the addresses that families lived at from the time of random assignment through May 2008, weighting each tract by the duration of time the family lived in that tract. Table 2 in the *NEJM* article includes measures of tract share poor (the fraction of tract residents living below the poverty level), tract share minority (the fraction of tract residents who are members of racial or ethnic minority groups), tract share households with single female-headed families (the fraction of tract households that include single female-headed families), and tract share college graduates (the fraction of tract residents ages 25 and older whose highest level of educational attainment is a college degree).

3.3 Measures

Obesity Measures: Height and weight were used to calculate body mass index (BMI), equal to weight in kilograms divided by height in meters squared. Moderate obesity was defined as BMI $\geq$ 30, severe obesity as BMI $\geq$ 35, and very severe obesity as BMI $\geq$ 40.

Diabetes Measure: Glycated hemoglobin (HbA1c) was assayed from the dried blood spots collected during the long-term survey. A single HbA1c measurement provides an integrated assessment of a person's average blood glucose levels over the preceding several months; fasting is not required before a

sample is obtained (Saudek et al., 2008). HbA1c values of 6.5% or higher were coded as indicating diabetes (American Diabetes Association, 2010).

4. Construction of the Cell-Level PUF

The cell-level PUF includes the data on the 3,273 adults interviewed as part of the MTO long-term evaluation. Individual-level data have been aggregated into 81 cells. Cells are homogenous by randomization site and treatment group and largely homogenous by the magnitude of the analysis weight. In addition, cells were further divided by treatment compliance status and race (non-Hispanic African-Americans) when cells were sufficiently large. Code used to collapse the data can be found in Appendix C. The table below lists some of the key variables on the file, and Appendix A contains a complete list of variables on the file.

Table 1. Key Variables on the Cell-Level PUF Dataset

Description	**Variables**
Treatment Group Categories	**ra_group** 1 = Low-poverty voucher (LPV) group (also called the "Experimental" group) 2 = Traditional voucher (TRV) group (also called the "Section 8" group) 3 = Control group
Treatment Group Dummy Variables	**ra_grp_exp** – flag for the LPV (or experimental) group **ra_grp_s8** – flag for the TRV (or Section 8) group **ra_grp_control** – flag for the control group
Compliance Status	**f_svy_cmove** –flag indicating that the family moved using an MTO housing voucher or certificate (LPV or TRV) 1 = core mover (complier) 0 = not a core mover
Site Categories	**ra_site** – the MTO site at which the family enrolled: 1 = Baltimore 2 = Boston 3 = Chicago 4 = Los Angeles 5 = New York City
Site Dummy Variables	**x_f_site_balt** – Baltimore site flag **x_f_site_bos** – Boston site flag **x_f_site_chi** – Chicago site flag **x_f_site_la** – Los Angeles site flag (New York is the omitted category in the regression models, and New York cells can be identified via the **sgx_ra_site_3g_all_nyc** flag)
Cell Information	**cell_id_ad** – cell identification number, ranging from 1 to 81 **cell_numobs** – number of individual observations collapsed into the cell **mn_f_wt_totsvy** – average analysis weight for the cell
Outcome Mean†	**mn_**[*original outcome name*] – weighted mean of the outcome for the observations comprising the cell
Outcome Standard Deviation†	**sd_**[*original outcome name*] – weighted standard deviation of the outcome for the observations comprising the cell
Outcome Sum of Weights†	**wt_**[*original outcome name*] – sum of the weights for observations in the cell with valid data for the specific outcome (e.g., weights can vary slightly from outcome to outcome)

† The portion of the variable name following the "mn_", "sd_", or "wt_" prefix uses the convention of f_c9010t_ for census tract characteristics, f_ph_/f_db_ for physical health, f_nb_ for neighborhood outcomes, f_spl_ for residential mobility, f_sn_ for social networks, and x_f_/x_rad_/cov_ for baseline covariates.

5. Construction of the Expanded Pseudo-Individual PUF

To make it easier to calculate standard errors for impact estimates, we expanded the aggregated cell-level data to mimic an individual-level file. We expanded each cell by the number of observations that had been aggregated into the cell (e.g., if the cell had originally had 20 records, we add an additional 19 duplicate records for the cell). This yields a total of 3,273 observations on the expanded file and mimics the N's, weight, and weighted means of the original file. To calculate the standard errors of the impacts correctly, each outcome needs to have the same standard deviation within a cell as it did in the original file. To mimic the standard deviation of the original observations comprising the cell, we artificially set one record within each cell to a value above the mean (mean + V2), one record to a value below the mean (mean – V2), and set the rest of the observations to the outcome's mean (or to missing if there were missing data for some observations).[2] These new values are assigned to variables starting with the prefix: ps_*. In addition to the ps_* variables, we have created pseudo *dichotomous* variables that start with the prefix psbi_* and are restricted to values of 0, 1, or missing. These dichotomous measures allow for the rough replication of our logit analyses. These pseudo dichotomous variables have been constructed to approximate the number of non-missing observations and the weighted mean for the outcome for each cell. The "pseudo-individual level" file roughly mimics the original data for the purposes of running impact estimates by site and treatment group. However, the pseudo-individual data CANNOT be used to examine cross-variable correlations, to run regressions other than basic impact estimates, or to run impact estimates that control for baseline characteristics other than site. This is because the values of the ps_* and psbi_* variables have been artificially set and while they capture the mean and the standard deviation for the variable across the cell, they DO NOT reflect the actual values in the original dataset. Within a cell, any relationship between variables is strictly an artifact of the way the data were expanded. The ps_* and psbi_* variables can be used as the dependent variable in an impact estimate; however, they should not be used on the right-hand side as a covariate. As the first record within each cell has been set to a high value for that outcome, outcomes will appear artificially correlated with each other. Appendix C shows the code used to expand the cell-data to create the pseudo individual-level data.

Below is a description of some of the key variables on the expanded pseudo-individual file. For a complete list of variables see Appendix B.

[2] V2 = square root of [(variance of population * number of observations in the cell)/2]

Table 2. Key Variables on the Expanded Pseudo-Individual PUF Dataset

Description	**Variables**
Treatment Group Categories	**ra_group** 1 = Low-poverty voucher (LPV) group (also called the "Experimental" group) 2 = Traditional voucher (TRV) group (also called the "Section 8" group) 3 = Control group
Treatment Group Dummy Variables	**ra_grp_exp** – flag for the LPV (or experimental) group **ra_grp_s8** – flag for the TRV (or Section 8) group **ra_grp_control** – flag for the control group
Compliance Status	**f_svy_cmove** –flag indicating that the family moved using an MTO housing voucher or certificate (LPV or TRV) 1 = core mover (complier) 0 = not a core mover
Site Categories	**ra_site** – the MTO site at which the family enrolled: 1 = Baltimore 2 = Boston 3 = Chicago 4 = Los Angeles 5 = New York City
Site Dummy Variables	**x_f_site_balt** – Baltimore site flag **x_f_site_bos** – Boston site flag **x_f_site_chi** – Chicago site flag **x_f_site_la** – Los Angeles site flag (New York is the omitted category in the regression models, and New York observations can be identified via the **sgx_ra_site_3g_all_nyc** flag)
Pseudo-Individual Level Measures for Outcomes, Mediators, and Select Baseline Characteristics†	**ps_**[*original outcome name*] – These are synthetic variables that for each cell mimic the original data in terms of the number of observations, weighted mean of the data, and standard deviation of the data. HOWEVER, these variables were constructed using the cell-level PUF data and are NOT actual individual-level data.
Pseudo-Individual Level Binary Measures for Dummy Variable Outcomes and Mediators†	**psbi_**[*original outcome name*] – These are synthetic dummy variables (with values of 0,1, or missing) that approximate the original data in terms of the number of observations and the weighted mean. HOWEVER, these variables were constructed using the cell-level PUF data and are NOT actual individual-level data. These variables include the obesity (psbi_f_ph_bmi_obese_srm_ad, psbi_f_ph_bmi_obese2_srm_ad, and psbi_f_ph_bmi_obese3_srm_ad) and diabetes measures (psbi_f_db_hba1c_diab_final).

† See note to Table 1.

6. Replicating Results Using the PUFs

6.1 Estimation of Control and Treatment Means (NEJM Table 1)

Table 1 in the *NEJM* article presents mean baseline characteristics for the control group and the MTO treatment (voucher) groups. These group means and tests of the differences can be roughly replicated using the expanded pseudo-individual file. See Table 3 of this document for a comparison of the results published in the *NEJM* article and the values generated using the expanded pseudo-individual level PUF.

In Stata, the weighted mean of a baseline characteristic (such as "never married") can be calculated by using the average weight variable (*mn_f_wt_totsvy*) and restricting to either the control group (*ra_grp_control* = 1) or one of the treatment groups (*ra_grp_exp* = 1 or *ra_grp_s8* = 1):

```
summarize ps_x_f_ad_nevmarr [aw=mn_f_wt_totsvy] if ra_grp_control==1

summarize ps_x_f_ad_nevmarr [aw=mn_f_wt_totsvy] if ra_grp_exp==1

summarize ps_x_f_ad_nevmarr [aw=mn_f_wt_totsvy] if ra_grp_s8==1
```

To test the significance of the difference in means, we use a weighted regression of the treatment group dummy (either the LPV group or the TRV group) on the baseline characteristic. We use the t-statistic on the treatment group coefficient to calculate the p-value for a two-tailed t-test of two samples with equal variance. In Stata (using the LPV group for this example):

```
regress ps_x_f_ad_nevmarr ra_grp_exp [pw=mn_f_wt_totsvy] if ra_grp_s8==0

scalar sc_diff_pv = ttail(e(df_r), abs(_b[ra_grp_exp] / _se[ra_grp_exp])) * 2
```

where the Section 8 group is excluded (*ra_grp_s8* == 0), *e(df_r)* is the regression degrees of freedom, *_b[ra_grp_exp]* is the coefficient on the treatment dummy variable, and *_se[ra_grp_exp]* is the standard error of the treatment variable.

See Appendix D for the Stata program that generates the means using the expanded pseudo-individual PUF.

Table 3. Comparison of Baseline Characteristics Using Individual-Level Data vs. Expanded Pseudo-Individual Data: Means by Treatment Group (NEJM Table 1)*

Characteristic	Low-Poverty Voucher		Traditional Voucher		Control	
	Individual-Level Data	Pseudo-Individual Data	Individual-Level Data	Pseudo-Individual Data	Individual-Level Data	Pseudo-Individual Data
	(N=1425)	(N=1456)	(N=657)	(N=678)	(N=1104)	(N=1139)
	number (percent)					
Age†						
≤ 35 yr	196 (14.6)	199 (14.5)	94 (13.5)	96 (14.5)	163 (14.7)	164 (14.3)
36-40 yr	310 (21.5)	312 (21.3)	156 (23.9)	161 (21.3)	253 (23.3)	255 (22.7)
41-45 yr	347 (23.5)	355 (23.6)	143 (21.7)	148 (23.6)	257 (23.2)	265 (23.5)
46-50 yr	273 (18.6)	276 (18.4)	124 (20.5)	129 (18.4)	194 (17.1)	203 (17.4)
>50 yr	299 (21.7)	313 (22.2)	140 (20.4)	144 (22.2)	237 (21.7)	252 (22.1)
Race or ethnic group‡						
Black						
Any ethnicity	973 (65.0)		393 (63.9)		706 (66.1)	
Non-Hispanic		975 (63.2)		392 (63.2)		706 (63.5)
Hispanic	404 (31.5)	410 (31.1)	235 (33.0)	249 (31.1)	346 (30.3)	360 (30.3)
Never married	874 (62.6)	922 (62.3)	395 (63.5)	413 (62.3)	692 (64.3)	727 (63.7)
Age <18 yr at birth of first child	347 (25.1)	376 (24.9)	163 (28.0)	179 (24.9)	265 (25.0)	282 (24.6)
Employed	368 (27.1)	391 (27.1)	176 (26.0)	190 (27.1)	258 (23.9)	277 (24.5)
Enrolled in school	216 (16.0)	232 (16.1)	113 (17.7)	122 (16.1)	172 (16.9)	183 (16.7)
Received high school diploma	565 (38.3)	575 (38.1)	233 (34.3)	239 (38.1)	407 (35.9)	419 (36.1)
Received certificate of General Educational Development (GED)	235 (16.2)	236 (15.9)	124 (18.7)	126 (15.9)	204 (19.9)	211 (19.9)
Receiving Supplemental Security Income§	221 (15.9)	242 (16.1)	107 (17.1)	114 (16.1)	171 (16.3)	185 (16.4)

* Numbers are raw, unweighted data. Percentages were calculated with the use of sample weights to account for changes in random-assignment ratios across randomized groups and for subsample interviews. Percentages include imputed values. The individual-level data sample is the sample used for the *NEJM* article analysis: women for whom valid data on body mass index (BMI) or glycated hemoglobin (HbA1c) level were available in the long-term follow-up study (N=3,186). The pseudo-individual sample is all adults interviewed as part of the long-term follow-up study regardless of gender and whether or not they have valid BMI or HbA1c data (N=3,273).

† The age listed was that calculated as of December 31, 2007, just before the long-term follow-up began in June 2008.

‡ The individual-level data black race measure includes blacks of any ethnicity, i.e. including Hispanics, whereas the pseudo-individual data measure includes only non-Hispanic blacks (and those with missing ethnicity). An Hispanic person could be a member of any race.

§ Supplemental Security Income is a federal assistance program for aged, blind, and disabled people.

6.2 Estimation of Intention-to-Treat Effects (NEJM Tables 2 and 3)

The impacts presented in Tables 2 and 3 of the *NEJM* article are intention-to-treat (ITT) effects or the impacts of being offered an MTO voucher as part of the LPV or TRV groups. The impacts are calculated separately for each treatment group, i.e. different models for the LPV-Control and TRV-Control impacts. Using the expanded pseudo-individual file, the ITT effects on continuous dependent variables are estimated using a linear regression while the effects on dichotomous variables are calculated using a logistic regression and are presented as average marginal effects. Both regressions use the outcome or mediator as the dependent variable with the key independent variable being the dummy variable indicating assignment to either one of the two voucher groups (*ra_grp_exp* = 1 or *ra_grp_s8* = 1). The regressions also control for randomization site (using the four *x_f_site_** indicators, with New York City as the omitted category)[3], apply a probability weight (*mn_f_wt_totsvy*), and generate Huber-White standard errors.

In Stata, the command line to estimate the impact on adult (moderate) obesity of being offered an MTO LPV housing voucher, using the logistic regression, is as follows:

```
logit psbi_f_ph_bmi_obese_srm_ad ra_grp_exp x_f_site_balt x_f_site_bos
   x_f_site_chi x_f_site_la [pw=mn_f_wt_totsvy] if ra_grp_s8==0
```

where the Section 8 group is excluded (*ra_grp_s8* == 0) and the ITT impact of the voucher will be the coefficient on the variable *ra_grp_exp* or *_b[ra_grp_exp]*.

Using linear regression, the command line to estimate the impact on neighborhood poverty of being offered an MTO TRV housing voucher is as follows:

```
regress ps_f_c9010t_perpov_dw ra_grp_s8 x_f_site_balt x_f_site_bos
   x_f_site_chi x_f_site_la [pw=mn_f_wt_totsvy] if ra_grp_exp==0
```

See Appendix E for the Stata program that generates the ITT estimates using the expanded pseudo-individual PUF data.

Tables 4-7 below compare three sets of impact estimates:

1. using the *individual-level data* (female adults only as in the *NEJM* article) and controlling for *all covariates* (these results match those presented in the *NEJM* article);
2. using the *individual-level data* (female and male adults) and controlling *only for randomization site*; and
3. using the *expanded pseudo-individual data* (female and male adults) from the PUF and controlling *only for randomization site.* The PUF data have been collapsed into cells by site and group and then expanded and thus do not allow one to control for additional baseline characteristics in the regression models.

[3] As mentioned above, the results in the *NEJM* article cannot be replicated exactly because the additional independent variables used for the article could not be made available on the PUF.

Table 4. Comparisons of Neighborhood and Other Mediator ITT Estimates Using Individual-Level Data vs. Expanded Pseudo-Individual Data for the Low-Poverty Voucher Group (NEJM Table 2)*

	Control mean			Low-poverty housing voucher vs. control group								
				Intention to Treat Estimate (95% CI) †			P Value			N		
	Females Only	Females and Males		Females Only, Full Covariates	Females and Males, Site Covariates		Females Only, Full Covariates	Females and Males, Site Covariates		Females Only	Females and Males	
Variable	Indiv.-Level Data	Indiv.-Level Data	Pseudo-Indiv. Data	Individual-Level Data	Individual-Level Data	Pseudo-Individual Data	Indiv.-Level Data	Indiv.-Level Data	Pseudo-Indiv. Data	Indiv.-Level Data	Indiv.-Level Data	Pseudo-Indiv. Data
Mean number of moves‡	2.1	2.2	2.2	0.57 (0.42 to 0.71)	0.54 (0.39 to 0.68)	0.54 (0.39 to 0.68)	<.01	<.01	<.01	2404	2595	2595
Poverty rate in census tract (%)§												
Baseline	53.1	53.1	53.1	-0.37 (-1.23 to 0.50)	-0.49 (-1.41 to 0.42)	-0.50 (-1.39 to 0.40)	.41	.29	.28	2404	2555	2555
At 1 yr	50.0	49.9	49.9	-17.06 (-18.57 to -15.56)	-17.00 (-18.50 to -15.51)	-17.01 (-18.47 to -15.54)	<.01	<.01	<.01	2404	2552	2552
At 5 yr	39.9	39.9	39.9	-9.78 (-11.25 to -8.31)	-9.99 (-11.46 to -8.52)	-9.96 (-11.38 to -8.53)	<.01	<.01	<.01	2404	2544	2544
At 10 yr	33.0	33.0	32.9	-4.86 (-6.23 to -3.48)	-4.60 (-5.97 to -3.23)	-4.57 (-5.93 to -3.21)	<.01	<.01	<.01	2404	2540	2540
Mean census-tract characteristics (%) ¶												
Poor ‖	39.6	39.6	39.6	-9.14 (-10.26 to -8.02)	-9.04 (-10.15 to -7.92)	-9.03 (-10.13 to -7.94)	<.01	<.01	<.01	2404	2592	2592
Minorities	88.0	88.0	88.0	-6.23 (-7.58 to -4.89)	-6.05 (-7.44 to -4.67)	-6.05 (-7.43 to -4.68)	<.01	<.01	<.01	2404	2592	2592
Household headed by a woman	54.3	54.1	54.2	-7.95 (-9.08 to -6.82)	-7.84 (-8.96 to -6.72)	-7.84 (-8.95 to -6.72)	<.01	<.01	<.01	2404	2592	2592
College graduate	16.1	16.1	16.1	4.49 (3.68 to 5.30)	4.28 (3.49 to 5.07)	4.27 (3.50 to 5.05)	<.01	<.01	<.01	2404	2592	2592

Table 4. (continued)

Variable	Control mean			Low-poverty housing voucher vs. control group								
				Intention to Treat Estimate (95% CI) †			P Value			N		
	Females Only	Females and Males		Females Only, Full Covariates	Females and Males, Site Covariates		Females Only, Full Covariates	Females and Males, Site Covariates		Females Only	Females and Males	
	Indiv.-Level Data	Indiv.-Level Data	Pseudo-Indiv. Data	Individual-Level Data	Individual-Level Data	Pseudo-Individual Data	Indiv.-Level Data	Indiv.-Level Data	Pseudo-Indiv. Data	Indiv.-Level Data	Indiv.-Level Data	Pseudo-Indiv. Data
Respondents reporting collective efficacy (%) ** ††	58.9	58.9	58.9	8.20 (4.20 to 12.21)	7.64 (3.56 to 11.73)	7.60 (3.63 to 11.58)	<.01	<.01	<.01	2516	2581	2581
Respondents reporting feeling safe or very safe on streets near home during the day (%) ††	80.7	80.4	80.5	3.70 (0.52 to 6.87)	3.70 (0.46 to 6.93)	3.72 (0.58 to 6.86)	.02	.03	.02	2522	2587	2587
Respondents reporting having at least one friend who graduated from college (%) ††	53.4	53.2	53.1	6.90 (2.74 to 11.06)	6.87 (2.68 to 11.07)	6.88 (2.78 to 10.98)	<.01	<.01	<.01	2478	2543	2541
Respondents reporting access to local health care services, excluding emergency room (%) ††	93.4	93.5	93.4	-1.36 (-3.49 to 0.77)	-1.37 (-3.50 to 0.77)	-1.36 (-3.53 to 0.82)	.21	.21	.22	2526	2590	2590

Table notes appear below Table 5.

Table 5. Comparisons of Neighborhood and Other Mediator ITTs Using Individual-Level Data vs. Expanded Pseudo-Individual Data for the Traditional Voucher Group (NEJM Table 2)*

	Control mean			Traditional housing voucher vs. control group								
				Intention to Treat Estimate (95% CI) †			P Value			N		
	Females Only	Females and Males		Females Only, Full Covariates	Females and Males, Site Covariates		Females Only, Full Covariates	Females and Males, Site Covariates		Females Only	Females and Males	
Variable	Indiv.-Level Data	Indiv.-Level Data	Pseudo-Indiv. Data	Individual-Level Data	Individual-Level Data	Pseudo-Individual Data	Indiv.-Level Data	Indiv.-Level Data	Pseudo-Indiv. Data	Indiv.-Level Data	Indiv.-Level Data	Pseudo-Indiv. Data
Mean number of moves‡	2.1	2.2	2.2	0.58 (0.38 to 0.79)	0.60 (0.41 to 0.78)	0.60 (0.41 to 0.78)	<.01	<.01	<.01	1673	1817	1817
Poverty rate in census tract (%)§												
Baseline	53.1	53.1	53.1	-0.37 (-1.55 to 0.81)	-0.46 (-1.54 to 0.61)	-0.47 (-1.51 to 0.57)	.54	.40	.38	1673	1797	1797
At 1 yr	50.0	49.9	49.9	-13.50 (-15.33 to -11.67)	-13.47 (-15.12 to -11.82)	-13.44 (-14.99 to -11.90)	<.01	<.01	<.01	1673	1793	1793
At 5 yr	39.9	39.9	39.9	-6.26 (-8.41 to -4.11)	-6.90 (-8.61 to -5.20)	-6.88 (-8.48 to -5.27)	<.01	<.01	<.01	1673	1785	1786
At 10 yr	33.0	33.0	32.9	-2.87 (-4.80 to -0.95)	-3.74 (-5.34 to -2.13)	-3.74 (-5.29 to -2.18)	<.01	<.01	<.01	1673	1771	1771
Mean census-tract characteristics (%) ¶												
Poor ‖	39.6	39.6	39.6	-6.07 (-7.53 to -4.61)	-6.69 (-7.92 to -5.46)	-6.69 (-7.87 to -5.52)	<.01	<.01	<.01	1673	1817	1817
Minorities	88.0	88.0	88.0	-0.99 (-2.88 to 0.90)	-1.90 (-3.54 to -0.25)	-1.90 (-3.43 to -0.36)	.30	.02	.02	1673	1817	1817
Household headed by a woman	54.3	54.1	54.2	-5.03 (-6.55 to -3.51)	-5.79 (-7.08 to -4.50)	-5.79 (-7.03 to -4.55)	<.01	<.01	<.01	1673	1817	1817
College graduate	16.1	16.1	16.1	1.41 (0.29 to 2.52)	2.01 (1.14 to 2.88)	2.02 (1.16 to 2.87)	.01	<.01	<.01	1673	1817	1817

Table 5. (continued)

Variable	Control mean			Low-poverty housing voucher vs. control group								
				Intention to Treat Estimate (95% CI) †			P Value			N		
	Females Only	Females and Males		Females Only, Full Covariates	Females and Males, Site Covariates		Females Only, Full Covariates	Females and Males, Site Covariates		Females Only	Females and Males	
	Indiv.-Level Data	Indiv.-Level Data	Pseudo-Indiv. Data	Individual-Level Data	Individual-Level Data	Pseudo-Individual Data	Indiv.-Level Data	Indiv.-Level Data	Pseudo-Indiv. Data	Indiv.-Level Data	Indiv.-Level Data	Pseudo-Indiv. Data
Respondents reporting collective efficacy (%) ** ††	58.9	58.9	58.9	0.80 (-5.16 to 6.76)	3.74 (-1.44 to 8.93)	3.73 (-1.33 to 8.78)	.79	.16	.15	1752	1807	1806
Respondents reporting feeling safe or very safe on streets near home during the day (%) ††	80.7	80.4	80.5	5.00 (0.50 to 9.50)	4.49 (0.44 to 8.53)	4.23 (0.34 to 8.11)	.03	.03	.03	1756	1812	1811
Respondents reporting having at least one friend who graduated from college (%) ††	53.4	53.2	53.1	-2.11 (-8.33 to 4.11)	-0.25 (-5.62 to 5.12)	-0.03 (-5.23 to 5.17)	.51	.93	.99	1723	1778	1777
Respondents reporting access to local health care services, excluding emergency room (%) ††	93.4	93.5	93.4	0.64 (-2.11 to 3.40)	1.61 (-0.68 to 3.91)	1.71 (-0.61 to 4.03)	.65	.17	.15	1755	1810	1809

See table notes on next page.

Notes to Tables 4 and 5

* The analysis sample for the Females Only/Individual-Level Data columns matches that used for the *NEJM* article: women with a valid BMI or HbA1c measurement. The sample for the Females and Males/Individual-Level Data columns and the Pseudo-Individual Data columns includes all adults (females and males) interviewed as part of the long-term survey, regardless of whether or not they have valid BMI or HbA1c measurement. For the Females Only/Individual-Level Data columns, analyses of number of moves and census-tract characteristics were further limited to participants with valid addresses at baseline and years 1, 5, and 10. As above, that limitation does not apply to the Females and Males/Individual-Level Data columns and the Pseudo-Individual Data columns. The intention-to-treat estimates come from a regression that compares average outcomes across randomly assigned groups, with statistical control for baseline characteristics, which may differ slightly from the difference in raw group means presented here.

† Intention-to-treat estimates compare the average of the outcomes for everyone assigned to the intervention group with the average of the outcomes for controls. The Full Covariates columns included adjustment for the set of baseline covariates shown in Table 1 of the NEJM article (and Table 3 of this document) and indicators for survey-sample release (families were randomly selected with regard to the time at which they would first be contacted about participation in the long-term follow-up study), site, and random-assignment periods. The Site Covariates columns included adjustment only for site. The effects on continuous dependent variables were calculated with the use of linear regression; the effects on dichotomous variables were calculated with the use of logistic regression and are presented as average marginal effects.

‡ The total number of moves is the number from the time of randomization (1994 through 1998) to the beginning of long-term follow-up (May 2008).

§ Census-tract characteristics were recorded as of the time when a family lived in the tract and were interpolated with the use of 1990 and 2000 decennial census data and data from the American Community Survey, 2005 to 2009.

¶ Average duration-weighted census-tract characteristics give more weight to tracts in which families spent relatively more time during the study period.

|| The term “poor” is defined as having an annual income below the federal government’s poverty threshold.

** Collective efficacy is defined as the likelihood that adults will take action in response to youth spraying graffiti on local buildings. See Sampson, Raudenbush, and Earls (1997) for more details on collective efficacy.

†† In the NEJM article, these measures were also presented using data from the MTO interim evaluation, but the PUF do not include the interim data and those rows are therefore not replicated here.

Table 6. Comparisons of ITT Effects on Body Mass Index and Glycated Hemoglobin Outcomes Using Individual-Level Data vs. Expanded Pseudo-Individual Data for the Low-Poverty Voucher Group (NEJM Table 3)*

	Control mean			Low-poverty housing voucher vs. control group								
				Intention to Treat Estimate (95% CI) †			P Value			N		
	Females Only	Females and Males		Females Only, Full Covariates	Females and Males, Site Covariates		Females Only, Full Covariates	Females and Males, Site Covariates		Females Only	Females and Males	
Variable	Indiv.-Level Data	Indiv.-Level Data	Pseudo-Indiv. Data	Individual-Level Data	Individual-Level Data	Pseudo-Individual Data	Indiv.-Level Data	Indiv.-Level Data	Pseudo-Indiv. Data	Indiv.-Level Data	Indiv.-Level Data	Pseudo-Indiv. Data
BMI‡												
≥ 30	58.6	58.4	58.6	-1.19 (-5.41 to 3.02)	-1.17 (-5.41 to 3.08)	-1.30 (-5.46 to 2.86)	.58	.59	.54	2508	2550	2550
≥ 35	35.5	35.1	35.0	-4.61 (-8.54 to -0.69)	-4.21 (-8.15 to -0.27)	-4.27 (-8.15 to -0.40)	.02	.04	.03	2508	2550	2550
≥ 40	17.7	17.5	17.7	-3.38 (-6.39 to -0.36)	-3.22 (-6.23 to -0.22)	-3.35 (-6.35 to -0.35)	.03	.04	.03	2508	2550	2550
Glycated hemoglobin§												
≥ 6.5%	20.0	20.4	20.4	-4.31 (-7.82 to -0.80)	-4.20 (-7.83 to -0.58)	-4.19 (-7.75 to -0.63)	.02	.02	.02	2092	2130	2127

* The analysis sample for the Females Only/Individual-Level Data columns matches that used for the *NEJM* article: women with a valid BMI measurement (for the BMI analysis) or a valid glycated hemoglobin measurement (for the glycated hemoglobin analysis) in the long-term follow-up data collection. The sample for the Females and Males/Individual-Level Data columns and the Pseudo-Individual Data columns includes all adults (females and males) in the long-term follow-up data collection.

† Intention-to-treat estimates compare the average outcomes for all participants assigned to an intervention group with the average outcomes for controls. The Full Covariates columns included adjustment for the set of baseline covariates shown in Table 1 of the *NEJM* article (and Table 3 of this document) and indicators for survey-sample release (families were randomly selected with regard to the time at which they would first be contacted about participation in the long-term follow-up study), site, and random-assignment periods. The Site Covariates columns included adjustment only for site. The effects are calculated with the use of logistic regression and are presented as average marginal effects.

‡ BMI (the weight in kilograms divided by the square of the height in meters) was calculated from measured height and weight for most adults as part of the long-term follow-up data collection.

§ Glycated hemoglobin (HbA1c) was assayed from dried blood spots collected as part of the long-term follow-up data collection.

Table 7. Comparisons of ITT Effects on Body Mass Index and Glycated Hemoglobin Outcomes Using Individual-Level Data vs. Expanded Pseudo-Individual Data for the Traditional Voucher Group (NEJM Table 3)*

	Control mean			Traditional housing voucher vs. control group								
				Intention to Treat Estimate (95% CI) †			P Value			N		
	Females Only	Females and Males		Females Only, Full Covariates	Females and Males, Site Covariates		Females Only, Full Covariates	Females and Males, Site Covariates		Females Only	Females and Males	
Variable	Indiv.-Level Data	Indiv.-Level Data	Pseudo-Indiv. Data	Individual-Level Data	Individual-Level Data	Pseudo-Individual Data	Indiv.-Level Data	Indiv.-Level Data	Pseudo-Indiv. Data	Indiv.-Level Data	Indiv.-Level Data	Pseudo-Indiv. Data
BMI‡												
≥ 30	58.6	58.4	58.6	-0.14 (-6.27 to 5.98)	-1.00 (-6.24 to 4.24)	-1.13 (-6.22 to 3.96)	.96	.71	.67	1747	1788	1788
≥ 35	35.5	35.1	35.0	-5.34 (-11.02 to 0.34)	-5.12 (-10.05 to -0.20)	-5.09 (-9.86 to -0.33)	.07	.04	.04	1747	1788	1788
≥ 40	17.7	17.5	17.7	-3.58 (-7.95 to 0.80)	-2.66 (-6.48 to 1.17)	-2.49 (-6.28 to 1.30)	.11	.17	.20	1747	1788	1788
Glycated hemoglobin§												
≥ 6.5%	20.0	20.4	20.4	-0.08 (-5.18 to 5.02)	0.35 (-4.29 to 4.99)	0.20 (-4.25 to 4.65)	.98	.88	.93	1516	1554	1551

* The analysis sample for the Females Only/Individual-Level Data columns matches that used for the *NEJM* article: women with a valid BMI measurement (for the BMI analysis) or a valid glycated hemoglobin measurement (for the glycated hemoglobin analysis) in the long-term follow-up data collection. The sample for the Females and Males/Individual-Level Data columns and the Pseudo-Individual Data columns includes all adults (females and males) in the long-term follow-up data collection.

† Intention-to-treat estimates compare the average outcomes for all participants assigned to an intervention group with the average outcomes for controls. The Full Covariates columns included adjustment for the set of baseline covariates shown in Table 1 of the *NEJM* article (and Table 3 of this document) and indicators for survey-sample release (families were randomly selected with regard to the time at which they would first be contacted about participation in the long-term follow-up study), site, and random-assignment periods. The Site Covariates columns included adjustment only for site. The effects are calculated with the use of logistic regression and are presented as average marginal effects.

‡ BMI (the weight in kilograms divided by the square of the height in meters) was calculated from measured height and weight for most adults as part of the long-term follow-up data collection.

§ Glycated hemoglobin (HbA1c) was assayed from dried blood spots collected as part of the long-term follow-up data collection.

7. Summary

The MTO *NEJM* PUFs are designed to allow the user to roughly replicate the results in the *NEJM* article "Neighborhoods, Obesity, and Diabetes – A Randomized Social Experiment." The PUFs contain information for MTO adults on diabetes, obesity, and other outcomes and mediators. The data have been aggregated by site and by treatment group assignment to allow for the replication of impact estimates. Aggregation protects the confidentiality of the data. Under no circumstances should the user attempt to identify any individual in the datasets. Users who wish to explore the individual-level data that allow for more types of analyses and controlling for a more complete set of baseline covariates in estimating impacts can apply for access to the restricted access dataset when it becomes available via ICPSR.

References

American Diabetes Association. 2010. "Standards of medical care in diabetes – 2010." *Diabetes Care*. 33(1): S11-61.

Chobanian, Aram V., George L. Bakris, Henry R. Black, William C. Cushman, Lee A. Green, Joseph L. Izzo, Jr., Daniel W. Jones, et al., and the National High Blood Pressure Education Program Coordinating Committee. 2003. Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Hypertension*. 42(6): 1206-52.

Institute for Social Research. 2008. *Health and retirement study: Physical measures and biomarkers 2008*. Ann Arbor: University of Michigan. Available at http://hrsonline.isr.umich.edu/modules/meta/2008/core/qnaire/online/2008PhysicalMeasuresBiomarkers.pdf.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75(1): 83-119.

National Institutes of Health, National Heart, Lung and Blood Institute, and Obesity Education Initiative. 2004. *Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure*. Washington, DC: National Institutes of Health.

Orr, Larry, Judith D. Feins, Robin Jacob, Erik Beecroft, Lisa Sanbonmatsu, Lawrence F. Katz, Jeffery B. Liebman, and Jeffrey R. Kling. 2003. *Moving to Opportunity for Fair Housing Demonstration Program: Interim Impacts Evaluation*. U.S. Department of Housing and Urban Development, Office of Policy Development and Research, Washington, DC; available at www.huduser.org/Publications/pdf/MTOFullReport.pdf.

Sampson, Robert J., Stephen W. Raudenbush, and Felton Earls. 1997. "Neighborhoods and Violent Crime: a Multilevel Study of Collective Efficacy." *Science* 277(5328): 918-24.

Sanbonmatsu, Lisa, Jens Ludwig, Lawrence F. Katz, Lisa A. Gennetian, Greg J. Duncan, Ronald C. Kessler, Emma Adam, Thomas W. McDade, and Stacy Tessler Lindau. 2011. *Moving to Opportunity for Fair Housing Demonstration Program: Final Impacts Evaluation*. U.S. Department of Housing and Urban Development, Office of Policy Development and Research, Washington, DC; available at www.huduser.org/portal/publications/pubasst/MTOFHD.html.

Saudek, Christopher D., William H. Herman, David B. Sacks, Richard M. Bergenstal, David Edelman, and Mayer B. Davidson. 2008. "A new look at screening and diagnosing diabetes mellitus." *Journal of Clinical Endocrinology and Metabolism* 93(7): 2447-53.

Appendix A – Variable List for the Cell-Level PUF File
(mto_nejm_puf_cells_20131025.dta)

Note that this appendix is a variable list only. For further details on the variables, please see section 4 above (specifically Table 1), which explains the variable prefixes, as well as Appendix F, which includes detailed documentation of the outcome and mediator measures.

obs: 81
vars: 200

Variable Name	Type	Format	Variable Label
cell_id_ad	float	%9.0g	cell_id_ad - number id assigned to cell
ra_group	float	%34.0g	ra_group - Randomization grp (1=Low-PovVouch/Exp, 2=TradVouch/Sec8, 3=Control)
ra_grp_exp	float	%43.0g	ra_grp_exp - Low-Poverty Voucher (Experimental) grp (1=LPV)
ra_grp_s8	float	%43.0g	ra_grp_s8 - Traditional Voucher (Section 8) grp (1=TRV)
ra_grp_control	float	%31.0g	ra_grp_control - Control grp flag (1=control)
ra_site	float	%15.0g	ra_site - Site (1=Baltimore/2=Boston/3=Chicago/4=Los Angeles/5=New York City)
x_f_site_balt	float	%42.0g	x_f_site_balt - AD in Baltimore Site (1=Baltimore, 0=Not Baltimore)
x_f_site_bos	float	%45.0g	x_f_site_bos - AD in Boston Site (1=Boston, 0=Not Boston)
x_f_site_chi	float	%44.0g	x_f_site_chi - AD in Chicago Site (1=Chicago, 0=Not Chicago)
x_f_site_la	float	%40.0g	x_f_site_la - AD in LA Site (1=Los Angeles, 0 = Not Los Angeles)
sgx_rasite_3g_all_nyc	float	%43.0g	sgx_rasite_3g_all_nyc - New York site, all grps (1=New York City)
mn_f_wt_totsvy	float	%9.0g	mn_f_wt_totsvy - average of f_wt_totsvy for obs in cell
cell_numobs	long	%10.0g	cell_numobs - # of observations contributing to cell
cell_sumwgt	double	%10.0g	cell_sumwgt - sum of weights for the cell
mn_f_svy_cmove	float	%9.0g	mean of f_svy_cmove - Core Mover/Treatment Compliance Flag (1=Moved)
mn_x_rad_ad_ethrace_black_nh	float	%9.0g	mean of x_rad_ad_ethrace_black_nh - AD Black Non-Hispanic (1=Black)
mn_x_rad_ad_ethrace_hisp	float	%9.0g	mean of x_rad_ad_ethrace_hisp - AD Hispanic, any race (1=Hispanic)
mn_f_ph_bmi_obese_srm_ad	float	%9.0g	mean of f_ph_bmi_obese_srm_ad - AD obese (measured/self-reports): BMI>=30
mn_f_ph_bmi_obese2_srm_ad	float	%9.0g	mean of f_ph_bmi_obese2_srm_ad - AD obese 2 (measured/self-reports): BMI>=35
mn_f_ph_bmi_obese3_srm_ad	float	%9.0g	mean of f_ph_bmi_obese3_srm_ad - AD obese 3 (measured/self-reports): BMI>=40
mn_f_db_hba1c_diab_final	float	%9.0g	mean of f_db_hba1c_diab_final - AD diabetes (blood test): HbA1c>=6.5%
mn_f_spl_moves_n	float	%9.0g	mean of f_spl_moves_n - # of moves according to spell file addresses (0-13)
mn_rad_c9010t_perpov_yr1	float	%9.0g	mean of rad_c9010t_perpov_yr1 - Tract poverty 1 yr post-randomization
mn_rad_c9010t_perpov_yr5	float	%9.0g	mean of rad_c9010t_perpov_yr5 - Tract poverty 5 yrs post-randomization
mn_rad_c9010t_perpov_bl	float	%9.0g	mean of rad_c9010t_perpov_bl - Tract poverty at baseline address
mn_rad_c9010t_perpov_yr10	float	%9.0g	mean of rad_c9010t_perpov_yr10 - Tract poverty 10 yrs post-randomization
mn_f_c9010t_perpov_dw	float	%9.0g	mean of f_c9010t_perpov_dw - Duration-wgtd tract poverty
mn_f_c9010t_pminorty_dw	float	%9.0g	mean of f_c9010t_pminorty_dw - Duration-wgtd tract share minority
mn_f_c9010t_pfsfem_dw	float	%9.0g	mean of f_c9010t_pfsfem_dw - Duration-wgtd share single female-headed fam
mn_f_c9010t_pcolldeg_dw	float	%9.0g	mean of f_c9010t_pcolldeg_dw - Duration-wgtd share college grad
mn_f_sn_monit_graffiti_ad	float	%9.0g	mean of f_sn_monit_graffiti_ad - AD Likely/very likely to report graffiti
mn_f_nb_safe_safday_ad	float	%9.0g	mean of f_nb_safe_safday_ad - AD felt safe/very safe in nbhd during day
mn_f_sn_net_anyfrndgrad_ad	float	%9.0g	mean of f_sn_net_anyfrndgrad_ad - AD 1+ friends who graduated college (flag)
mn_f_ph_care_place2go_noner_ad	float	%9.0g	mean of f_ph_care_place2go_noner_ad - Place to go for routine care
mn_x_rad_ad_male	float	%9.0g	mean of x_rad_ad_male - AD male (1=male/0=female/0<x<1=imputed)
mn_x_rad_ad_le_35	float	%9.0g	mean of x_rad_ad_le_35 - AD age 35 or younger as of 12/31/07 (1=age <= 35)
mn_x_rad_ad_36_40	float	%9.0g	mean of x_rad_ad_36_40 - AD age 36 to 40 as of 12/31/07 (1=age 36-40)
mn_x_rad_ad_41_45	float	%9.0g	mean of x_rad_ad_41_45 - AD age 41 to 45 as of 12/31/07 (1=age 41-45)
mn_x_rad_ad_46_50	float	%9.0g	mean of x_rad_ad_46_50 - AD age 46 to 50 as of 12/31/07 (1=age 46-50)
mn_x_f_ad_nevmarr	float	%9.0g	mean of x_f_ad_nevmarr - At baseline, AD never been married (1=never married)

Appendix A, continued

Variable Name	Type	Format	Variable Label
mn_x_f_ad_parentu18	float	%9.0g	mean of x_f_ad_parentu18 - At baseline, AD < age 18 at birth of 1st child
mn_x_f_ad_working	float	%9.0g	mean of x_f_ad_working - At baseline, AD working for pay (1=working for pay)
mn_x_f_ad_edinsch	float	%9.0g	mean of x_f_ad_edinsch - At baseline, AD enrolled in school (1=in school)
mn_x_f_ad_edgradhs	float	%9.0g	mean of x_f_ad_edgradhs - At baseline, AD completed high school (flag)
mn_x_f_ad_edged	float	%9.0g	mean of x_f_ad_edged - At baseline, AD had a GED (1=had ged)
mn_x_f_ad_edgradhs_miss	float	%9.0g	mean of x_f_ad_edgradhs_miss - Missing flag for baseline GED/H.S. diploma
mn_x_f_hood_5y	float	%9.0g	mean of x_f_hood_5y - At baseline, hhhead living in nbhd. 5+ yrs (flag)
mn_x_f_hood_chat	float	%9.0g	mean of x_f_hood_chat - At baseline, hhhead chatted w/ neighbor>=1x/wk (flag)
mn_x_f_hood_nbrkid	float	%9.0g	mean of x_f_hood_nbrkid - At baseline, hhhead very likely tell on nbhd kid
mn_x_f_hood_nofamily	float	%9.0g	mean of x_f_hood_nofamily - At baseline, hhhead has no family living in nbhd
mn_x_f_hood_nofriend	float	%9.0g	mean of x_f_hood_nofriend - At baseline, hhhead has no friends living in nbhd
mn_x_f_hood_unsafenit	float	%9.0g	mean of x_f_hood_unsafenit - At baseline, nbhd. streets very unsafe at night
mn_x_f_hood_verydissat	float	%9.0g	mean of x_f_hood_verydissat - At baseline, hhhead very dissatisfied with nbhd
mn_x_f_hh_car	float	%9.0g	mean of x_f_hh_car - At baseline, hhld owned a car (1=owned a car)
mn_x_f_hh_disabl	float	%9.0g	mean of x_f_hh_disabl - At baseline, a hhld member had a disability (flag)
mn_x_f_hh_noteens	float	%9.0g	mean of x_f_hh_noteens - At baseline, no teens (ages 13-17) in hhld (flag)
mn_x_f_hh_afdc	float	%9.0g	mean of x_f_hh_afdc - At baseline, hhld receiving AFDC/TANF (1=receive welf)
mn_x_f_hh_victim	float	%9.0g	mean of x_f_hh_victim - At baseline, hhld member victimized past 6 mos (flag)
mn_x_f_hh_size2	float	%9.0g	mean of x_f_hh_size2 - At baseline hhld size is 2 or smaller (1=size is <=2)
mn_x_f_hh_size3	float	%9.0g	mean of x_f_hh_size3 - At baseline hhld size is 3 (1=size is 3)
mn_x_f_hh_size4	float	%9.0g	mean of x_f_hh_size4 - At baseline hhld size is 4 (1=size is 4)
mn_x_f_hous_fndapt	float	%9.0g	mean of x_f_hous_fndapt - At baseline, hhhead very sure of finding apt (flag)
mn_x_f_hous_mov3tm	float	%9.0g	mean of x_f_hous_mov3tm - At baseline, hhhead had moved >3x in 5 yrs (1=flag)
mn_x_f_hous_sec8bef	float	%9.0g	mean of x_f_hous_sec8bef - At baseline, hhhead applied for Section 8 before
mn_x_f_hous_movdrgs	float	%9.0g	mean of x_f_hous_movdrgs - At baseline 1st/2nd reason want to move=drug/crime
mn_x_f_hous_movschl	float	%9.0g	mean of x_f_hous_movschl - At baseline 1st/2nd reason want to move: schools
mn_x_f_release1	float	%9.0g	mean of x_f_release1 - Release 1 Sample AD for Final Survey (1=release 1)
mn_cov_hous_movjob	float	%9.0g	mean of cov_hous_movjob - At baseline 1st/2nd reason want to move: find job
mn_cov_hous_movapt	float	%9.0g	mean of cov_hous_movapt - At baseline 1st/2nd reason want to move: better apt
mn_cov_hh_femhoh	float	%9.0g	mean of cov_hh_femhoh - At baseline, head of household was female
mn_cov_hh_fdstmp	float	%9.0g	mean of cov_hh_fdstmp - At baseline, AD receiving food stamp
mn_cov_hh_wic	float	%9.0g	mean of cov_hh_wic - At baseline, AD receiving WIC
mn_cov_hh_medicd	float	%9.0g	mean of cov_hh_medicd - At baseline, AD receiving Medicaid
mn_cov_hood_store	float	%9.0g	mean of cov_hood_store - Baseline addr 30+ mins away from grocery store
mn_cov_hood_doctor	float	%9.0g	mean of cov_hood_doctor - Baseline addr 30+ mins away from doctor
mn_cov_hh_ssi	float	%9.0g	mean of cov_hh_ssi - At baseline, AD receiving SSI
sd_f_svy_cmove	float	%9.0g	std dev of f_svy_cmove - Core Mover/Treatment Compliance Flag (1=Moved)
sd_x_rad_ad_ethrace_black_nh	float	%9.0g	std dev of x_rad_ad_ethrace_black_nh - AD Black Non-Hispanic (1=Black)
sd_x_rad_ad_ethrace_hisp	float	%9.0g	std dev of x_rad_ad_ethrace_hisp - AD Hispanic, any race (1=Hispanic)
sd_f_ph_bmi_obese_srm_ad	float	%9.0g	std dev of f_ph_bmi_obese_srm_ad - AD obese (measured/self-reports): BMI>=30
sd_f_ph_bmi_obese2_srm_ad	float	%9.0g	std dev of f_ph_bmi_obese2_srm_ad - AD obese 2 (measured/self-reports): BMI>=35
sd_f_ph_bmi_obese3_srm_ad	float	%9.0g	std dev of f_ph_bmi_obese3_srm_ad - AD obese 3 (measured/self-reports): BMI>=40
sd_f_db_hba1c_diab_final	float	%9.0g	std dev of f_db_hba1c_diab_final - AD diabetes (blood test): HbA1c>=6.5%
sd_f_spl_moves_n	float	%9.0g	std dev of f_spl_moves_n - # of moves according to spell file addresses (0-13)
sd_rad_c9010t_perpov_yr1	float	%9.0g	std dev of rad_c9010t_perpov_yr1 - Tract poverty 1 yr post-randomization
sd_rad_c9010t_perpov_yr5	float	%9.0g	std dev of rad_c9010t_perpov_yr5 - Tract poverty 5 yrs post-randomization
sd_rad_c9010t_perpov_bl	float	%9.0g	std dev of rad_c9010t_perpov_bl - Tract poverty at baseline address
sd_rad_c9010t_perpov_yr10	float	%9.0g	std dev of rad_c9010t_perpov_yr10 - Tract poverty 10 yrs post-randomization
sd_f_c9010t_perpov_dw	float	%9.0g	std dev of f_c9010t_perpov_dw - Duration-wgtd tract poverty

Variable Name	Type	Format	Variable Label
sd_f_c9010t_pminorty_dw	float	%9.0g	std dev of f_c9010t_pminorty_dw - Duration-wgtd tract share minority
sd_f_c9010t_pfsfem_dw	float	%9.0g	std dev of f_c9010t_pfsfem_dw - Duration-wgtd share single female-headed fam
sd_f_c9010t_pcolldeg_dw	float	%9.0g	std dev of f_c9010t_pcolldeg_dw - Duration-wgtd share college grad
sd_f_sn_monit_graffiti_ad	float	%9.0g	std dev of f_sn_monit_graffiti_ad - AD Likely/very likely to report graffiti
sd_f_nb_safe_safday_ad	float	%9.0g	std dev of f_nb_safe_safday_ad - AD felt safe/very safe in nbhd during day
sd_f_sn_net_anyfrndgrad_ad	float	%9.0g	std dev of f_sn_net_anyfrndgrad_ad - AD 1+ friends who graduated college (flag)
sd_f_ph_care_place2go_noner_ad	float	%9.0g	std dev of f_ph_care_place2go_noner_ad - Place to go for routine care
sd_x_rad_ad_male	float	%9.0g	std dev of x_rad_ad_male - AD male (1=male/0=female/0<x<1=imputed)
sd_x_rad_ad_le_35	float	%9.0g	std dev of x_rad_ad_le_35 - AD age 35 or younger as of 12/31/07 (1=age <= 35)
sd_x_rad_ad_36_40	float	%9.0g	std dev of x_rad_ad_36_40 - AD age 36 to 40 as of 12/31/07 (1=age 36-40)
sd_x_rad_ad_41_45	float	%9.0g	std dev of x_rad_ad_41_45 - AD age 41 to 45 as of 12/31/07 (1=age 41-45)
sd_x_rad_ad_46_50	float	%9.0g	std dev of x_rad_ad_46_50 - AD age 46 to 50 as of 12/31/07 (1=age 46-50)
sd_x_f_ad_nevmarr	float	%9.0g	std dev of x_f_ad_nevmarr - At baseline, AD never been married (1=never married)
sd_x_f_ad_parentu18	float	%9.0g	std dev of x_f_ad_parentu18 - At baseline, AD < age 18 at birth of 1st child
sd_x_f_ad_working	float	%9.0g	std dev of x_f_ad_working - At baseline, AD working for pay (1=working for pay)
sd_x_f_ad_edinsch	float	%9.0g	std dev of x_f_ad_edinsch - At baseline, AD enrolled in school (1=in school)
sd_x_f_ad_edgradhs	float	%9.0g	std dev of x_f_ad_edgradhs - At baseline, AD completed high school (flag)
sd_x_f_ad_edged	float	%9.0g	std dev of x_f_ad_edged - At baseline, AD had a GED (1=had ged)
sd_x_f_ad_edgradhs_miss	float	%9.0g	std dev of x_f_ad_edgradhs_miss - Missing flag for baseline GED/H.S. diploma
sd_x_f_hood_5y	float	%9.0g	std dev of x_f_hood_5y - At baseline, hhhead living in nbhd. 5+ yrs (flag)
sd_x_f_hood_chat	float	%9.0g	std dev of x_f_hood_chat - At baseline, hhhead chatted w/ neighbor>=1x/wk (flag)
sd_x_f_hood_nbrkid	float	%9.0g	std dev of x_f_hood_nbrkid - At baseline, hhhead very likely tell on nbhd kid
sd_x_f_hood_nofamily	float	%9.0g	std dev of x_f_hood_nofamily - At baseline, hhhead has no family living in nbhd
sd_x_f_hood_nofriend	float	%9.0g	std dev of x_f_hood_nofriend - At baseline, hhhead has no friends living in nbhd
sd_x_f_hood_unsafenit	float	%9.0g	std dev of x_f_hood_unsafenit - At baseline, nbhd. streets very unsafe at night
sd_x_f_hood_verydissat	float	%9.0g	std dev of x_f_hood_verydissat - At baseline, hhhead very dissatisfied with nbhd
sd_x_f_hh_car	float	%9.0g	std dev of x_f_hh_car - At baseline, hhld owned a car (1=owned a car)
sd_x_f_hh_disabl	float	%9.0g	std dev of x_f_hh_disabl - At baseline, a hhld member had a disability (flag)
sd_x_f_hh_noteens	float	%9.0g	std dev of x_f_hh_noteens - At baseline, no teens (ages 13-17) in hhld (flag)
sd_x_f_hh_afdc	float	%9.0g	std dev of x_f_hh_afdc - At baseline, hhld receiving AFDC/TANF (1=receive welf)
sd_x_f_hh_victim	float	%9.0g	std dev of x_f_hh_victim - At baseline, hhld member victimized past 6 mos (flag)
sd_x_f_hh_size2	float	%9.0g	std dev of x_f_hh_size2 - At baseline hhld size is 2 or smaller (1=size is <=2)
sd_x_f_hh_size3	float	%9.0g	std dev of x_f_hh_size3 - At baseline hhld size is 3 (1=size is 3)
sd_x_f_hh_size4	float	%9.0g	std dev of x_f_hh_size4 - At baseline hhld size is 4 (1=size is 4)
sd_x_f_hous_fndapt	float	%9.0g	std dev of x_f_hous_fndapt - At baseline, hhhead very sure of finding apt (flag)
sd_x_f_hous_mov3tm	float	%9.0g	std dev of x_f_hous_mov3tm - At baseline, hhhead had moved >3x in 5 yrs (1=flag)
sd_x_f_hous_sec8bef	float	%9.0g	std dev of x_f_hous_sec8bef - At baseline, hhhead applied for Section 8 before
sd_x_f_hous_movdrgs	float	%9.0g	std dev of x_f_hous_movdrgs - At baseline 1st/2nd reason want to move=drug/crime
sd_x_f_hous_movschl	float	%9.0g	std dev of x_f_hous_movschl - At baseline 1st/2nd reason want to move: schools
sd_x_f_release1	float	%9.0g	std dev of x_f_release1 - Release 1 Sample AD for Final Survey (1=release 1)
sd_cov_hous_movjob	float	%9.0g	std dev of cov_hous_movjob - At baseline 1st/2nd reason want to move: find job
sd_cov_hous_movapt	float	%9.0g	std dev of cov_hous_movapt - At baseline 1st/2nd reason want to move: better apt
sd_cov_hh_femhoh	float	%9.0g	std dev of cov_hh_femhoh - At baseline, head of household was female
sd_cov_hh_fdstmp	float	%9.0g	std dev of cov_hh_fdstmp - At baseline, AD receiving food stamp
sd_cov_hh_wic	float	%9.0g	std dev of cov_hh_wic - At baseline, AD receiving WIC
sd_cov_hh_medicd	float	%9.0g	std dev of cov_hh_medicd - At baseline, AD receiving Medicaid
sd_cov_hood_store	float	%9.0g	std dev of cov_hood_store - Baseline addr 30+ mins away from grocery store
sd_cov_hood_doctor	float	%9.0g	std dev of cov_hood_doctor - Baseline addr 30+ mins away from doctor
sd_cov_hh_ssi	float	%9.0g	std dev of cov_hh_ssi - At baseline, AD receiving SSI
wt_f_svy_cmove	double	%9.0g	sum of wts f_svy_cmove - Core Mover/Treatment Compliance Flag (1=Moved)

Appendix A, continued

Variable Name	Type	Format	Variable Label
wt_x_rad_ad_ethrace_black_nh	double	%9.0g	sum of wts x_rad_ad_ethrace_black_nh - AD Black Non-Hispanic (1=Black)
wt_x_rad_ad_ethrace_hisp	double	%9.0g	sum of wts x_rad_ad_ethrace_hisp - AD Hispanic, any race (1=Hispanic)
wt_f_ph_bmi_obese_srm_ad	double	%9.0g	sum of wts f_ph_bmi_obese_srm_ad - AD obese (measured/self-reports): BMI>=30
wt_f_ph_bmi_obese2_srm_ad	double	%9.0g	sum of wts f_ph_bmi_obese2_srm_ad - AD obese 2 (measured/self-reports): BMI>=35
wt_f_ph_bmi_obese3_srm_ad	double	%9.0g	sum of wts f_ph_bmi_obese3_srm_ad - AD obese 3 (measured/self-reports): BMI>=40
wt_f_db_hba1c_diab_final	double	%9.0g	sum of wts f_db_hba1c_diab_final - AD diabetes (blood test): HbA1c>=6.5%
wt_f_spl_moves_n	double	%9.0g	sum of wts f_spl_moves_n - # of moves according to spell file addresses (0-13)
wt_rad_c9010t_perpov_yr1	double	%9.0g	sum of wts rad_c9010t_perpov_yr1 - Tract poverty 1 yr post-randomization
wt_rad_c9010t_perpov_yr5	double	%9.0g	sum of wts rad_c9010t_perpov_yr5 - Tract poverty 5 yrs post-randomization
wt_rad_c9010t_perpov_bl	double	%9.0g	sum of wts rad_c9010t_perpov_bl - Tract poverty at baseline address
wt_rad_c9010t_perpov_yr10	double	%9.0g	sum of wts rad_c9010t_perpov_yr10 - Tract poverty 10 yrs post-randomization
wt_f_c9010t_perpov_dw	double	%9.0g	sum of wts f_c9010t_perpov_dw - Duration-wgtd tract poverty
wt_f_c9010t_pminorty_dw	double	%9.0g	sum of wts f_c9010t_pminorty_dw - Duration-wgtd tract share minority
wt_f_c9010t_pfsfem_dw	double	%9.0g	sum of wts f_c9010t_pfsfem_dw - Duration-wgtd share single female-headed fam
wt_f_c9010t_pcolldeg_dw	double	%9.0g	sum of wts f_c9010t_pcolldeg_dw - Duration-wgtd share college grad
wt_f_sn_monit_graffiti_ad	double	%9.0g	sum of wts f_sn_monit_graffiti_ad - AD Likely/very likely to report graffiti
wt_f_nb_safe_safday_ad	double	%9.0g	sum of wts f_nb_safe_safday_ad - AD felt safe/very safe in nbhd during day
wt_f_sn_net_anyfrndgrad_ad	double	%9.0g	sum of wts f_sn_net_anyfrndgrad_ad - AD 1+ friends who graduated college (flag)
wt_f_ph_care_place2go_noner_ad	double	%9.0g	sum of wts f_ph_care_place2go_noner_ad - Place to go for routine care
wt_x_rad_ad_male	double	%9.0g	sum of wts x_rad_ad_male - AD male (1=male/0=female/0<x<1=imputed)
wt_x_rad_ad_le_35	double	%9.0g	sum of wts x_rad_ad_le_35 - AD age 35 or younger as of 12/31/07 (1=age <= 35)
wt_x_rad_ad_36_40	double	%9.0g	sum of wts x_rad_ad_36_40 - AD age 36 to 40 as of 12/31/07 (1=age 36-40)
wt_x_rad_ad_41_45	double	%9.0g	sum of wts x_rad_ad_41_45 - AD age 41 to 45 as of 12/31/07 (1=age 41-45)
wt_x_rad_ad_46_50	double	%9.0g	sum of wts x_rad_ad_46_50 - AD age 46 to 50 as of 12/31/07 (1=age 46-50)
wt_x_f_ad_nevmarr	double	%9.0g	sum of wts x_f_ad_nevmarr - At baseline, AD never been married (1=never married)
wt_x_f_ad_parentu18	double	%9.0g	sum of wts x_f_ad_parentu18 - At baseline, AD < age 18 at birth of 1st child
wt_x_f_ad_working	double	%9.0g	sum of wts x_f_ad_working - At baseline, AD working for pay (1=working for pay)
wt_x_f_ad_edinsch	double	%9.0g	sum of wts x_f_ad_edinsch - At baseline, AD enrolled in school (1=in school)
wt_x_f_ad_edgradhs	double	%9.0g	sum of wts x_f_ad_edgradhs - At baseline, AD completed high school (flag)
wt_x_f_ad_edged	double	%9.0g	sum of wts x_f_ad_edged - At baseline, AD had a GED (1=had ged)
wt_x_f_ad_edgradhs_miss	double	%9.0g	sum of wts x_f_ad_edgradhs_miss - Missing flag for baseline GED/H.S. diploma
wt_x_f_hood_5y	double	%9.0g	sum of wts x_f_hood_5y - At baseline, hhhead living in nbhd. 5+ yrs (flag)
wt_x_f_hood_chat	double	%9.0g	sum of wts x_f_hood_chat - At baseline, hhhead chatted w/ neighbor>=1x/wk (flag)
wt_x_f_hood_nbrkid	double	%9.0g	sum of wts x_f_hood_nbrkid - At baseline, hhhead very likely tell on nbhd kid
wt_x_f_hood_nofamily	double	%9.0g	sum of wts x_f_hood_nofamily - At baseline, hhhead has no family living in nbhd
wt_x_f_hood_nofriend	double	%9.0g	sum of wts x_f_hood_nofriend - At baseline, hhhead has no friends living in nbhd
wt_x_f_hood_unsafenit	double	%9.0g	sum of wts x_f_hood_unsafenit - At baseline, nbhd. streets very unsafe at night
wt_x_f_hood_verydissat	double	%9.0g	sum of wts x_f_hood_verydissat - At baseline, hhhead very dissatisfied with nbhd
wt_x_f_hh_car	double	%9.0g	sum of wts x_f_hh_car - At baseline, hhld owned a car (1=owned a car)
wt_x_f_hh_disabl	double	%9.0g	sum of wts x_f_hh_disabl - At baseline, a hhld member had a disability (flag)
wt_x_f_hh_noteens	double	%9.0g	sum of wts x_f_hh_noteens - At baseline, no teens (ages 13-17) in hhld (flag)
wt_x_f_hh_afdc	double	%9.0g	sum of wts x_f_hh_afdc - At baseline, hhld receiving AFDC/TANF (1=receive welf)
wt_x_f_hh_victim	double	%9.0g	sum of wts x_f_hh_victim - At baseline, hhld member victimized past 6 mos (flag)
wt_x_f_hh_size2	double	%9.0g	sum of wts x_f_hh_size2 - At baseline hhld size is 2 or smaller (1=size is <=2)
wt_x_f_hh_size3	double	%9.0g	sum of wts x_f_hh_size3 - At baseline hhld size is 3 (1=size is 3)
wt_x_f_hh_size4	double	%9.0g	sum of wts x_f_hh_size4 - At baseline hhld size is 4 (1=size is 4)
wt_x_f_hous_fndapt	double	%9.0g	sum of wts x_f_hous_fndapt - At baseline, hhhead very sure of finding apt (flag)
wt_x_f_hous_mov3tm	double	%9.0g	sum of wts x_f_hous_mov3tm - At baseline, hhhead had moved >3x in 5 yrs (1=flag)
wt_x_f_hous_sec8bef	double	%9.0g	sum of wts x_f_hous_sec8bef - At baseline, hhhead applied for Section 8 before

Appendix A, continued

Variable Name	Type	Format	Variable Label
wt_x_f_hous_movdrgs	double	%9.0g	sum of wts x_f_hous_movdrgs - At baseline 1st/2nd reason want to move=drug/crime
wt_x_f_hous_movschl	double	%9.0g	sum of wts x_f_hous_movschl - At baseline 1st/2nd reason want to move: schools
wt_x_f_release1	double	%9.0g	sum of wts x_f_release1 - Release 1 Sample AD for Final Survey (1=release 1)
wt_cov_hous_movjob	double	%9.0g	sum of wts cov_hous_movjob - At baseline 1st/2nd reason want to move: find job
wt_cov_hous_movapt	double	%9.0g	sum of wts cov_hous_movapt - At baseline 1st/2nd reason want to move: better apt
wt_cov_hh_femhoh	double	%9.0g	sum of wts cov_hh_femhoh - At baseline, head of household was female
wt_cov_hh_fdstmp	double	%9.0g	sum of wts cov_hh_fdstmp - At baseline, AD receiving food stamp
wt_cov_hh_wic	double	%9.0g	sum of wts cov_hh_wic - At baseline, AD receiving WIC
wt_cov_hh_medicd	double	%9.0g	sum of wts cov_hh_medicd - At baseline, AD receiving Medicaid
wt_cov_hood_store	double	%9.0g	sum of wts cov_hood_store - Baseline addr 30+ mins away from grocery store
wt_cov_hood_doctor	double	%9.0g	sum of wts cov_hood_doctor - Baseline addr 30+ mins away from doctor
wt_cov_hh_ssi	double	%9.0g	sum of wts cov_hh_ssi - At baseline, AD receiving SSI

Appendix B – Variable List for the Pseudo-Individual PUF File
(mto_nejm_puf_pseudo_20131025.dta)

Note that this appendix is a variable list only. For further details on the variables, please see section 5 above (specifically Table 2), which explains the variable prefixes, as well as Appendix F, which includes detailed documentation of the outcome and mediator measures.

obs: 3,273
vars: 75

Variable Name	Type	Format	Value Label	Variable Label
cell_id_ad	float	%9.0g		cell_id_ad - number id assigned to cell
ra_group	float	%34.0g	GROUP	ra_group - Randomization grp (1=Low-PovVouch/Exp, 2=TradVouch/Sec8, 3=Control)
ra_grp_exp	float	%43.0g	LPV	ra_grp_exp - Low-Poverty Voucher (Experimental) grp (1=LPV)
ra_grp_s8	float	%43.0g	TRV	ra_grp_s8 - Traditional Voucher (Section 8) grp (1=TRV)
ra_grp_control	float	%31.0g	CONTROL	ra_grp_control - Control grp flag (1=control)
ra_site	float	%15.0g	SITE	ra_site - Site (1=Baltimore/2=Boston/3=Chicago/4=Los Angeles/5=New York City)
x_f_site_balt	float	%42.0g	BALT	x_f_site_balt - AD in Baltimore Site (1=Baltimore, 0=Not Baltimore)
x_f_site_bos	float	%45.0g	BOS	x_f_site_bos - AD in Boston Site (1=Boston, 0=Not Boston)
x_f_site_chi	float	%44.0g	CHI	x_f_site_chi - AD in Chicago Site (1=Chicago, 0=Not Chicago)
x_f_site_la	float	%40.0g	LA	x_f_site_la - AD in LA Site (1=Los Angeles, 0 = Not Los Angeles)
sgx_rasite_3g_all_nyc	float	%43.0g	NYC	sgx_rasite_3g_all_nyc - New York site, all grps (1=New York City)
mn_f_wt_totsvy	float	%9.0g		mn_f_wt_totsvy - average of f_wt_totsvy for obs in cell
tmp_order_incell	float	%9.0g		tmp_order_incell - order of pseudo records within cell
ps_f_svy_cmove	float	%9.0g		pseudo obs f_svy_cmove - Core Mover/Treatment Compliance Flag (1=Moved)
ps_x_rad_ad_ethrace_black_nh	float	%9.0g		pseudo obs x_rad_ad_ethrace_black_nh - AD Black Non-Hispanic (1=Black)
ps_x_rad_ad_ethrace_hisp	float	%9.0g		pseudo obs x_rad_ad_ethrace_hisp - AD Hispanic, any race (1=Hispanic)
psbi_f_ph_bmi_obese_srm_ad	float	%11.0g	OBESE	pseudo dum f_ph_bmi_obese_srm_ad - AD obese (measured/self-reports): BMI>=30
psbi_f_ph_bmi_obese2_srm_ad	float	%12.0g	OBESE2	pseudo dum f_ph_bmi_obese2_srm_ad - AD obese 2 (measured/self-reports): BMI>=35
psbi_f_ph_bmi_obese3_srm_ad	float	%12.0g	OBESE3	pseudo dum f_ph_bmi_obese3_srm_ad - AD obese 3 (measured/self-reports): BMI>=40
psbi_f_db_hba1c_diab_final	float	%25.0g	HBA1C	pseudo dum f_db_hba1c_diab_final - AD diabetes (blood test): HbA1c>=6.5%
ps_f_spl_moves_n	float	%9.0g		pseudo obs f_spl_moves_n - # of moves according to spell file addresses (0-13)
ps_rad_c9010t_perpov_yr1	float	%9.0g		pseudo obs rad_c9010t_perpov_yr1 - Tract poverty 1 yr post-randomization
ps_rad_c9010t_perpov_yr5	float	%9.0g		pseudo obs rad_c9010t_perpov_yr5 - Tract poverty 5 yrs post-randomization
ps_rad_c9010t_perpov_bl	float	%9.0g		pseudo obs rad_c9010t_perpov_bl - Tract poverty at baseline address
ps_rad_c9010t_perpov_yr10	float	%9.0g		pseudo obs rad_c9010t_perpov_yr10 - Tract poverty 10 yrs post-randomization
ps_f_c9010t_perpov_dw	float	%9.0g		pseudo obs f_c9010t_perpov_dw - Duration-wgtd tract poverty

Variable Name	Type	Format	Value Label	Variable Label
ps_f_c9010t_pminorty_dw	float	%9.0g		pseudo obs f_c9010t_pminorty_dw - Duration-wgtd tract share minority
ps_f_c9010t_pfsfem_dw	float	%9.0g		pseudo obs f_c9010t_pfsfem_dw - Duration-wgtd share single female-headed fam
ps_f_c9010t_pcolldeg_dw	float	%9.0g		pseudo obs f_c9010t_pcolldeg_dw - Duration-wgtd share college grad
psbi_f_sn_monit_graffiti_ad	float	%31.0g	GRAF	pseudo dum f_sn_monit_graffiti_ad - AD Likely/very likely to report graffiti
psbi_f_nb_safe_safday_ad	float	%35.0g	SAFED	pseudo dum f_nb_safe_safday_ad - AD felt safe/very safe in nbhd during day
psbi_f_sn_net_anyfrndgrad_ad	float	%30.0g	FND_COL	pseudo dum f_sn_net_anyfrndgrad_ad - AD 1+ friends who graduated college (flag)
psbi_f_ph_care_place2go_noner_ad	float	%25.0g	PLACE_CARE	pseudo dum f_ph_care_place2go_noner_ad - Place to go for routine care
ps_x_rad_ad_male	float	%9.0g		pseudo obs x_rad_ad_male - AD male (1=male/0=female/0<x<1=imputed)
ps_x_rad_ad_le_35	float	%9.0g		pseudo obs x_rad_ad_le_35 - AD age 35 or younger as of 12/31/07 (1=age <= 35)
ps_x_rad_ad_36_40	float	%9.0g		pseudo obs x_rad_ad_36_40 - AD age 36 to 40 as of 12/31/07 (1=age 36-40)
ps_x_rad_ad_41_45	float	%9.0g		pseudo obs x_rad_ad_41_45 - AD age 41 to 45 as of 12/31/07 (1=age 41-45)
ps_x_rad_ad_46_50	float	%9.0g		pseudo obs x_rad_ad_46_50 - AD age 46 to 50 as of 12/31/07 (1=age 46-50)
ps_x_f_ad_nevmarr	float	%9.0g		pseudo obs x_f_ad_nevmarr - At baseline, AD never been married (1=never married)
ps_x_f_ad_parentu18	float	%9.0g		pseudo obs x_f_ad_parentu18 - At baseline, AD < age 18 at birth of 1st child
ps_x_f_ad_working	float	%9.0g		pseudo obs x_f_ad_working - At baseline, AD working for pay (1=working for pay)
ps_x_f_ad_edinsch	float	%9.0g		pseudo obs x_f_ad_edinsch - At baseline, AD enrolled in school (1=in school)
ps_x_f_ad_edgradhs	float	%9.0g		pseudo obs x_f_ad_edgradhs - At baseline, AD completed high school (flag)
ps_x_f_ad_edgradhs_miss	float	%9.0g		pseudo obs x_f_ad_edgradhs_miss - Missing flag for baseline GED/H.S. diploma
ps_x_f_ad_edged	float	%9.0g		pseudo obs x_f_ad_edged - At baseline, AD had a GED (1=had ged)
ps_x_f_hood_5y	float	%9.0g		pseudo obs x_f_hood_5y - At baseline, hhhead living in nbhd. 5+ yrs (flag)
ps_x_f_hood_chat	float	%9.0g		pseudo obs x_f_hood_chat - At baseline, hhhead chatted w/ neighbor>=1x/wk (flag)
ps_x_f_hood_nbrkid	float	%9.0g		pseudo obs x_f_hood_nbrkid - At baseline, hhhead very likely tell on nbhd kid
ps_x_f_hood_nofamily	float	%9.0g		pseudo obs x_f_hood_nofamily - At baseline, hhhead has no family living in nbhd
ps_x_f_hood_nofriend	float	%9.0g		pseudo obs x_f_hood_nofriend - At baseline, hhhead has no friends living in nbhd
ps_x_f_hood_unsafenit	float	%9.0g		pseudo obs x_f_hood_unsafenit - At baseline, nbhd. streets very unsafe at night
ps_x_f_hood_verydissat	float	%9.0g		pseudo obs x_f_hood_verydissat - At baseline, hhhead very dissatisfied with nbhd
ps_x_f_hh_car	float	%9.0g		pseudo obs x_f_hh_car - At baseline, hhld owned a car (1=owned a car)
ps_x_f_hh_disabl	float	%9.0g		pseudo obs x_f_hh_disabl - At baseline, a hhld member had a disability (flag)
ps_x_f_hh_noteens	float	%9.0g		pseudo obs x_f_hh_noteens - At baseline, no teens (ages 13-17) in hhld (flag)
ps_x_f_hh_afdc	float	%9.0g		pseudo obs x_f_hh_afdc - At baseline, hhld receiving AFDC/TANF (1=receive welf)

Appendix B, continued

Variable Name	Type	Format	Value Label	Variable Label
ps_x_f_hh_victim	float	%9.0g		pseudo obs x_f_hh_victim - At baseline, hhld member victimized past 6 mos (flag)
ps_x_f_hh_size2	float	%9.0g		pseudo obs x_f_hh_size2 - At baseline hhld size is 2 or smaller (1=size is <=2)
ps_x_f_hh_size3	float	%9.0g		pseudo obs x_f_hh_size3 - At baseline hhld size is 3 (1=size is 3)
ps_x_f_hh_size4	float	%9.0g		pseudo obs x_f_hh_size4 - At baseline hhld size is 4 (1=size is 4)
ps_x_f_hous_fndapt	float	%9.0g		pseudo obs x_f_hous_fndapt - At baseline, hhhead very sure of finding apt (flag)
ps_x_f_hous_mov3tm	float	%9.0g		pseudo obs x_f_hous_mov3tm - At baseline, hhhead had moved >3x in 5 yrs (1=flag)
ps_x_f_hous_sec8bef	float	%9.0g		pseudo obs x_f_hous_sec8bef - At baseline, hhhead applied for Section 8 before
ps_x_f_hous_movdrgs	float	%9.0g		pseudo obs x_f_hous_movdrgs - At baseline 1st/2nd reason want to move=drug/crime
ps_x_f_hous_movschl	float	%9.0g		pseudo obs x_f_hous_movschl - At baseline 1st/2nd reason want to move: schools
ps_x_f_release1	float	%9.0g		pseudo obs x_f_release1 - Release 1 Sample AD for Final Survey (1=release 1)
ps_cov_hous_movjob	float	%9.0g		pseudo obs cov_hous_movjob - At baseline 1st/2nd reason want to move: find job
ps_cov_hous_movapt	float	%9.0g		pseudo obs cov_hous_movapt - At baseline 1st/2nd reason want to move: better apt
ps_cov_hh_femhoh	float	%9.0g		pseudo obs cov_hh_femhoh - At baseline, head of household was female
ps_cov_hh_fdstmp	float	%9.0g		pseudo obs cov_hh_fdstmp - At baseline, AD receiving food stamp
ps_cov_hh_wic	float	%9.0g		pseudo obs cov_hh_wic - At baseline, AD receiving WIC
ps_cov_hh_medicd	float	%9.0g		pseudo obs cov_hh_medicd - At baseline, AD receiving Medicaid
ps_cov_hood_store	float	%9.0g		pseudo obs cov_hood_store - Baseline addr 30+ mins away from grocery store
ps_cov_hood_doctor	float	%9.0g		pseudo obs cov_hood_doctor - Baseline addr 30+ mins away from doctor
ps_cov_hh_ssi	float	%9.0g		pseudo obs cov_hh_ssi - At baseline, AD receiving SSI

Appendix C – Program Code Showing the Collapse and Expansion of the Data

Code Snippets (from 01_nejm_cellpuf_colexp_20131025.do):

```
*****************************
**** Step 2. Set variable lists ****
*****************************

* set weight variable *
global wtvar f_wt_totsvy

* set compliance var *
global movevar f_svy_cmove

* site interacted with the two treatment groups *
global sgx_ex sgx_rasite_ex_all_bal sgx_rasite_ex_all_bos sgx_rasite_ex_all_chi
sgx_rasite_ex_all_la sgx_rasite_ex_all_nyc
global sgx_s8 sgx_rasite_s8_all_bal sgx_rasite_s8_all_bos sgx_rasite_s8_all_chi
sgx_rasite_s8_all_la sgx_rasite_s8_all_nyc

* site covariates *
global sitecovs x_f_site_balt x_f_site_bos x_f_site_chi x_f_site_la
sgx_rasite_3g_all_nyc

* categorical site and treatment variables *
global categvar ra_site ra_group

* treatment group flags *
global treatvar ra_grp_exp ra_grp_s8 ra_grp_control

* OUTCOME/MEDIATOR measures *
* share poor in tract
global povvars rad_c9010t_perpov_yr1 rad_c9010t_perpov_yr5 f_c9010t_perpov_dw
* variables to be created
global tocreate rad_c9010t_perpov_bl rad_c9010t_perpov_yr10
* share minority in tract
global minovars f_c9010t_pminorty_dw

* miscellaneous mobility, neighborhood, housing, and social network measures
global miscvars f_spl_moves_n f_sn_monit_graffiti_ad f_sn_net_anyfrndgrad_ad

* physical health
global phvars f_ph_bmi_obese_srm_ad

* Outcome variables need a shorter labels to be created
global no_labels f_c9010t_pfsfem_dw f_c9010t_pcolldeg_dw
f_ph_care_place2go_noner_ad f_nb_safe_safday_ad f_ph_bmi_obese2_srm_ad
f_ph_bmi_obese3_srm_ad f_db_hba1c_diab_final


* create COMBINED list of outcome and mediator variables needed for NEJM
global outcvar $povvars $tocreate $minovars $miscvars $phvars $no_labels

* variables that need a dummy version
```

```
global dummyvar f_sn_monit_graffiti_ad f_nb_safe_safday_ad f_sn_net_anyfrndgrad_ad
f_ph_care_place2go_noner_ad f_ph_bmi_obese_srm_ad f_ph_bmi_obese2_srm_ad
f_ph_bmi_obese3_srm_ad f_db_hba1c_diab_final

* covariates used for NEJM paper (other than site)
global xcov x_rad_ad_male x_rad_ad_le_35 x_rad_ad_36_40 x_rad_ad_41_45
x_rad_ad_46_50 x_rad_ad_ethrace_black_nh x_rad_ad_ethrace_hisp ///
            x_f_ad_nevmarr x_f_ad_parentu18 x_f_ad_working x_f_ad_edinsch
x_f_ad_edgradhs x_f_ad_edged x_f_ad_edgradhs_miss ///
          x_f_hood_5y x_f_hood_chat x_f_hood_nbrkid x_f_hood_nofamily
x_f_hood_nofriend x_f_hood_unsafenit x_f_hood_verydissat ///
          x_f_hh_car x_f_hh_disabl x_f_hh_noteens x_f_hh_afdc x_f_hh_victim
x_f_hh_size2 x_f_hh_size3 x_f_hh_size4 ///
          x_f_hous_fndapt x_f_hous_mov3tm x_f_hous_sec8bef x_f_hous_movdrgs
x_f_hous_movschl x_f_release1

* additional baseline characteristics
* already have a shorter label
global blvar_yes cov_hous_movjob cov_hous_movapt
* not have a shorter label
global blvar_no cov_hh_femhoh cov_hh_fdstmp cov_hh_wic cov_hh_medicd cov_hood_store
cov_hood_doctor cov_hh_ssi
* combine blvar
global blvar $blvar_yes $blvar_no

* Variables that already have a shorter labels
global yes_labels $wtvar $movevar $sgx_ex $sgx_s8 $sitecovs $categvar $treatvar
$povvars $minovars $miscvars $phvars $xcov $blvar_yes...


************************************************************
**** Step 7. Prepare data for collapsing by creating additional variables ****
************************************************************

/*
This step uses cell_id_ad to aggregate the data.  The resulting cell-level file
will contain:

cell_id_ad,
raw count of obs,
sum of cell weights,
mean and standard deviation and sum of weights for outcome and mediator variables,
mean for African-American Non-Hispanic, and
means for key analysis variables such as ra_site, ra_group, core move, and site
dummies.

As cells should be homogenous by site and group, so the means of these key vars
should only take the usual values (1-5 and 1-3).
We need the sum of weights for outcomes and mediators since these variables can
have missing values.
We need the std dev for variables of interest such as the outcomes, mediators, and
perhaps baseline characteristics because they will be used to estimate
  the std errors.
*/

**** 7a. Create copies of variables with new prefixes to save mean, sd, and sum of
weights for raech variable ****
```

```
* FOR EACH OUTCOME, create 3 new variables to use in the collapse:
* wt_{outcomename} contains the weights for obs that have non-missing values for
the outcome. wt_* will be rawsum'd in the collapse.
* sd_{outcomename} contains copy of outcome values. In the collapse these will be
replaced with the weighted standard deviation of the outcome within the cell.
* mn_{outcomename} contains copy of outcome values. In the collapse these will be
replaced with the weighted mean value of the outcome within the cell.

foreach X of varlist $outcvar $xcov $blvar $movevar {
    gen wt_`X' = $wtvar if missing(`X') ~= 1
    gen sd_`X' = `X'
    gen mn_`X' = `X'
}



**** 7b.  Save full variable labels ****

foreach var of varlist $outcvar $xcov $blvar $sitecovs $categvar $treatvar
f_svy_cmove {
      local lb`var': var label `var'
}

...


*******************************************************************************
***********************************************
**** Step 9. Collapse the data by cell_id_ad and save weighted means and standard
deviations, sum of weights, and obseration counts ****
*******************************************************************************
***********************************************

* collapse data to create cell means, std devs, and sum of wgts for each variable
and for each cell a count of obs and raw sum of weights
* note that $sitecovs $categvar and $treatvar are homog by cell so these will be
preserved with original name (even though created using "mean")
#delimit;
collapse
(mean) mn_* $sitecovs $categvar $treatvar
(sd) sd_*
(count) cell_numobs = $wtvar
(rawsum) cell_sumwgt = $wtvar wt_* [aw=$wtvar], by(cell_id_ad);
#delimit cr;


*****************************************
**** Step 10. Relabel the new variables and values ****
*****************************************

label var cell_numobs "cell_numobs - # of observations contributing to cell"

label var cell_sumwgt "cell_sumwgt - sum of weights for the cell"
label var cell_id_ad "cell_id_ad - number id assigned to cell"

foreach var in $outcvar $xcov $blvar f_svy_cmove {
```

```
      global vname = "lb`var'"
      label var mn_`var' "mean of `${vname}'"
      label var sd_`var' "std dev of `${vname}'"
      label var wt_`var' "sum of wts `${vname}'"
}

foreach var of varlist $sitecovs $categvar $treatvar {
      global vname = "lb`var'"
      label var `var' "`${vname}'"
}


***********************************************
**** Step 11. Create Average Weight Variable ****
***********************************************

* generate average weight value for each cell = raw sum of weights for the cell
divided by number of obs in the cell
gen mn_$wtvar = cell_sumwgt  / cell_numobs
label variable mn_$wtvar "mn_$wtvar - average of $wtvar for obs in cell"

...

***********************************************
**** Step 13. Save final clean version of collapsed file ****
***********************************************

des, fullname
sum
saveold ${puf_cell}, replace



*********************************************************************************
*********************************
**** Step 14. Create Variables Needed to Expand File: average weight, observation
counts, and variance measures ****
*********************************************************************************
*********************************

**** create variables needed to expand data: avg weight, # of observ for a specific
outcome, variance of population, and VAR2
* for each outcome, calculate approx # of observations to use and convert stdev to
population variable
foreach X in $outcvar $xcov $blvar $movevar {

    * set number of observations for the outcome within the cell to yield sum of
weights as close to actual total as possible
    gen ob_`X' = round(cell_numobs * (wt_`X'/(mn_$wtvar*cell_numobs)))
    label variable ob_`X' "ob_`X' - approx. # of obs for variable"

    * convert stdev of sample to variance of population for each outcome
    gen vr_`X' = (sd_`X')^2*((ob_`X' -1)/ob_`X' )
    label variable vr_`X' "vr_`X' - variance of popul for the outcome"

    * set VAR2
    gen v2_`X' = ((vr_`X' * ob_`X' )/2)^0.5
```

```
    label variable v2_`X' "v2_`X' - var2 for the outcome"
}

*******************************************************************************
*****************************************************************
**** Step 15. Expand Collapsed PUF file to Pseudo-Individual Data and Reset Values
for Outcomes to Mimic Mean, SD, and N of Individual-Level Data ****

**** 15a. Expand the cell-level data to pseudo-individual
* expand data by the number of observations in each cell (the number of
observations in each cell)
expand cell_numobs
sum mn_$wtvar

**** 15b.  create a variable that is the record number (order) of each expanded
observation within that cell
sort cell_id_ad
by cell_id_ad: gen tmp_order_incell = _n
label var tmp_order_incell "tmp_order_incell - order of pseudo records within cell"

**** 15c.  Loop through each outcome and create and label a new pseudo individual
outcome (ps_*) that has same approx n, mean, and sd as actual data
****      by replacing first obs of cell with higher value, 2nd obs with lower
value, and blank out any cells beyond approx cell count.

foreach X in $outcvar $xcov $blvar $movevar {
    rename mn_`X' ps_`X'

    * relabel variables
    global vname = "lb`X'"
    label var ps_`X' "pseudo obs `${vname}'"

    * set first observation within cell to a value that is VAR2 distance ABOVE the
mean
    replace ps_`X' = ps_`X' + v2_`X' if tmp_order_incell == 1

    * set 2nd observation within cell to a value that is VAR2 distance BELOW the
mean
    replace ps_`X' = ps_`X' - v2_`X' if tmp_order_incell == 2

    * blank out values beyond the number of obs in the cell that represent valid
data for that outcome (e.g., if only 14 records had DBS data and the cell has 16
records, set records 15 and 16 to blank on dbs)
    replace ps_`X' = . if tmp_order_incell > ob_`X'
}

**** Step 15d.  Create pseudo dummy variables to achieve similar means to others
****

foreach X in $dummyvar {

    * first generate the mean of the variable. We sort by cell_id_ad and by the
variable so that missing values are at the bottom
    bysort cell_id_ad (ps_`X') : egen tmp_cellmean = mean(ps_`X')
    label var tmp_cellmean "tmp_cellmean - temporary variable for the cell mean"
```

```
    * assuming ob_`X' is the count of non-missing values in a cell, then generate
the number of '1' values we should have
    * which is determined by the mean since mean = count of '1's / total count of
non-missings
    by cell_id_ad : gen tmp_yescnt = round(tmp_cellmean * ob_`X')
    label var tmp_yescnt "tmp_yescnt - temporary variable for the number of records
that should have a 1 in the cell"

    * generate the new dummy variable with values equal to 1 or 0. The first `x'
number of records are set to 1 and the rest set to 0
    by cell_id_ad : gen psbi_`X' = (_n <= tmp_yescnt)
    global vname = "lb`X'"
    label var psbi_`X' "pseudo dum `${vname}'"

    * now replace values that should be missing with missings
    replace psbi_`X' = . if ps_`X'==.

    * drop temporary variables
    capture drop tmp_cellmean tmp_yescnt

    * summarize the dummy and continous versions and compare
    sum ps_`X' [aw=mn_$wtvar]
    scalar ps_mean = r(mean)
    scalar ps_sd   = r(sd)
    sum psbi_`X' [aw=mn_$wtvar]
    scalar psbi_mean = r(mean)
    scalar psbi_sd   = r(sd)

    di "The difference of means between ps and ps_dum is:   " ps_mean - psbi_mean
    di "The difference of std dev between ps and ps_dum is: " ps_sd - psbi_sd
}

************************************************************
**** Step 16. Keep Only the Needed Variables (ps_*, site, and group vars). ****
************************************************************

keep cell_id_ad tmp_order_incell ps_* psbi_* mn_$wtvar $categvar $treatvar
$sitecovs

* drop non-binary versions of binary outcomes
foreach var in $dummyvar {
      drop ps_`var'
}

*****************************************************
**** Step 17. Save expanded file as STATA and SAS datasets ****
*****************************************************
des, fullname
sum
saveold ${puf_pseudo}, replace
```

Appendix D – Stata Code for Estimating Group Means Using the *NEJM* PUF

```
/*
PROGRAM: 02_nejm_psuedopuf_mean_20131118.do
AUTHOR: Ray Yun Gou

Purpose: This program provides code for replicating the Moving to Opportunity final evaluation baseline
treatment and control sample sizes and weighted means that are presented in Table 1 of the New England Journal
of Medicine (NEJM) article by J. Ludwig et al. titled "Neighborhoods, Obesity, and Diabetes - A Randomized
Social Experiment" (October 20, 2011), using a public use file (PUF) with pseudo-individual data.

The PUF sample consists of MTO adults interviewed for the long-term (10- to 15-year) evaluation of MTO (n = 3273).
Researchers can access the data at http://www.nber.org/mtopuf/ and eventually at ICPSR (http://icpsr.umich.edu).
The PUF sample differs from the sample analyzed for the NEJM article, which consisted of female adults interviewed
as part of the long-term MTO evaluation with valid obesity or diabetes outcome measure values (n = 3186). For
confidentiality reasons, the PUF sample includes all adults (males and females) interviewed as part of the
long-term MTO evaluation.

The program loops through a list of baseline variables covering age, race and ethnicity, marital status, age at
birth of first child, employment, enrollment in school, education, and aid status. It calculates the overall
(unweighted) sample size for each treatment group and then the weighted percentage and an unweighted sample size for
each baseline variable by treatment group.

For each of the baseline variables in Table 1 of the NEJM article, the program creates the following outputs, which
are logged to a CSV file:

**** Control Group unweighted N and weighted mean:
The control N is the unweighted number of respondents in the control group (ra_group=3) who have a valid value
for the baseline variable. The weigthed control mean is the percentage of the sample weighted by mn_f_wt_totsvy who
have a valid value for the baseline variable.

**** Low-Poverty Voucher (LPV) Group unweighted N and weighted mean:
The LPV N is the unweighted number of respondents in the LPV treatment group, also known as the Experimental group
(ra_group=1/ra_grp_exp=1), who have a valid value for the baseline variable. The weighted LPV mean is the percentage
of the sample weighted by mn_f_wt_totsvy who have a valid value for the baseline variable. The LPV treatment group
```

```
received an MTO low-poverty rent voucher.

**** Traditional Voucher (TRV) Group unweighted N and weighted mean:
The TRV N is the unweighted number of respondents in the TRV treatment group, also
known as the Section 8 group
(ra_group=2/ra_grp_s8=1), who have a valid value for the baseline variable. The
weigthed TRV mean is the percentage
of the sample weighted by mn_f_wt_totsvy who have a valid value for the baseline
variable. The TRV treatment group
received a traditional Section 8 rent voucher.


STEPS:
1. Set options, program directories, and names of the input and output files
      1a. Set STATA Version
      1b. Set options
      1c. Set data file to use
      1d. Set output file
2. Define globals for weight variable and list of baseline measures to calculate
means for
      2a. Define weight global
      2b. Define baseline variable list
3. Open expanded pseudo-individual public use file and create omitted category
variables
4. Start results file by printing column header
5. Calculate overall sample sizes by treatment group
      5a. Experimental Group
      5b. Section 8 Group
      5c. Control Group
      5d. Write results to CSV results file
6. Calculate N and mean for each variable
      6a. Calculate Control Group unweighted sample size and weighted means
      6b. Calculate LPV Group unweighted sample size and weighted means
      6c. Calculate TRV Group unweighted sample size and weighted means
      6d. Print N, weighted mean, and percentage to CSV results file
7. Close results file
8. Print date and time.

*/

********************************************************************************
********
**** Step 1. Set options, program directories, and names of the input and output
files ****
********************************************************************************
********

**** 1a. Set STATA version ****
* Program is written for STATA version 12. Older versions may not run correctly
version 12


**** 1b. Set options ****
clear all
set more off
set maxvar 10000
set linesize 255
```

```
**** 1c. Set dataset to use ****
global data http://www.nber.org/mtopuf/mto_nejm_puf_pseudo_20131025.dta


* !!!! USER MUST UPDATE DIRECTORY OF OUTPUT FILE AND DATE !!!!
**** 1d. Set output file ****
* Set today's date for dating the log
global todaydt 20131118

* Set name of log file
global pgmdir ~/mtoproj/m10_pgm/papers/icpsr_archive/nejm/outputs/
global logname "${pgmdir}02_nejm_pseudo_mean_${todaydt}.csv"

* !!!! END NECESSARY USER-SET OPTIONS !!!!


********************************************************************************
**********************
**** Step 2. Define globals for weight variable and list of baseline measures to
calculate means for ****
********************************************************************************
**********************

**** 2a. Define weight global ****
global weight mn_f_wt_totsvy


**** 2b. Define baseline variable list ****
global grpmean ps_x_rad_ad_le_35 ps_x_rad_ad_36_40 ps_x_rad_ad_41_45
ps_x_rad_ad_46_50 tmp_age_gt_50 ///
               ps_x_rad_ad_ethrace_black_nh ps_x_rad_ad_ethrace_hisp
tmp_ethrace_other ///
               ps_x_f_ad_nevmarr ps_x_f_ad_parentu18 ps_x_f_ad_working
ps_x_f_ad_edinsch ps_x_f_ad_edgradhs ps_x_f_ad_edged ps_cov_hh_ssi



********************************************************************************
********************
**** Step 3. Open expanded pseudo-individual public use file and create omitted
category variables ****
********************************************************************************
********************

* Open dataset
use ${data}, clear

* Create omitted category variables for age group and race/ethnicity
gen tmp_age_gt_50=1-ps_x_rad_ad_le_35-ps_x_rad_ad_36_40-ps_x_rad_ad_41_45-
ps_x_rad_ad_46_50
gen tmp_ethrace_other=1-ps_x_rad_ad_ethrace_black_nh-ps_x_rad_ad_ethrace_hisp


************************************************************
**** Step 4. Start results file by printing column header ****
```

```
*************************************************************

file open csvlog using "$logname", write replace
file write csvlog "Variable,c_n,c_per,lpv_n,lpv_per,trv_n,trv_per" _n


*****************************************************************
**** Step 5. Calculate overall sample sizes by treatment group ****
*****************************************************************

**** 5a. LPV Group ****
sum ra_group [aw=${weight}] if ra_grp_exp==1
scalar all_lpv_n = r(N)

* check that sample size is correct
assert all_lpv_n == 1456


**** 5b. TRV Group ****
sum ra_group [aw=${weight}] if ra_group==2
scalar all_trv_n = r(N)

* check that sample size is correct
assert all_trv_n == 678


**** 5c. Control Group ****
sum ra_group [aw=${weight}] if ra_group==3
scalar all_contr_n = r(N)

* check that sample size is correct
assert all_contr_n == 1139


**** 5d. Write results to CSV results file ****
file write csvlog "Total Sample," %20.0f (all_contr_n) ",1,"
file write csvlog %20.0f (all_lpv_n) ",1,"
file write csvlog %20.0f (all_trv_n) ",1," _n



*****************************************************
**** Step 6. Calculate N and mean for each variable ****
*****************************************************

foreach X in $grpmean {

**** 6a. Calculate Control Group sample size and mean for the variable stored in
local `X' ****

    summarize `X' if ra_group==3
    scalar contr_n    = round(r(N)*r(mean))

    summarize `X' [aw=${weight}] if ra_group==3
    scalar contr_mean = r(mean)*100
```

```
**** 6b. Calculate LPV Group sample size and mean ****
    summarize `X' if ra_group==1
    scalar lpv_n    = round(r(N)*r(mean))

    summarize `X' [aw=${weight}] if ra_group==1
    scalar lpv_mean   = r(mean)*100


**** 6c. Calculate Section 8 Treatment Group sample size and weighted sample size
for interviewed adults ****
    summarize `X' if ra_group==2
    scalar trv_n    = round(r(N)*r(mean))

    summarize `X' [aw=${weight}] if ra_group==2
    scalar trv_mean    = r(mean)*100


**** 6d. Print N, weighted N, and percentage to CSV results file ****
    file write csvlog "`X'," %20.0f (contr_n) "," %20.3f (contr_mean) ","
    file write csvlog %20.0f (lpv_n) "," %20.3f (lpv_mean) ","
    file write csvlog %20.0f (trv_n) "," %20.3f (lpv_mean) "," _n

}

*************************
**** Step 7. Close log ****
*************************
file close csvlog

***********************************
**** Step 8. Print date and time ****
***********************************
disp "Program End on " c(current_date) " at " c(current_time)
```

Appendix E – Stata Code for Estimating Intention-to-Treat Effects Using the *NEJM* PUF

```
/*
PROGRAM: 03_nejm_pseudopuf_itt_20131118.do
AUTHOR: Ray Yun Gou

Purpose: This program provides code for replicating the Moving to Opportunity final
evaluation intention-to-treat (ITT)
effects and associated statistics for the outcomes shown in Tables 2 and 3 as well
as Figure 2
in the New England Journal of Medicine (NEJM) article by J. Ludwig et al. titled
"Neighborhoods, Obesity, and Diabetes -
A Randomized Social Experiment" (October 20, 2011), using a public use file (PUF)
with pseudo-individual data.

The PUF sample consists of MTO adults interviewed for the long-term (10- to 15-
year) evaluation of MTO (n = 3273).
Researchers can access the data at http://www.nber.org/mtopuf/ and eventually at
ICPSR (http://icpsr.umich.edu).
The PUF sample differs from the sample analyzed for the NEJM article, which
consisted of female adults interviewed
as part of the long-term MTO evaluation with valid obesity or diabetes outcome
measure values (n = 3186). For
confidentiality reasons, the PUF sample includes all adults (males and females)
interviewed as part of the
long-term MTO evaluation.

The program loops through the outcomes in Tables 2 and 3 of the NEJM article:
number of moves;
census tract poverty rate at baseline and 1, 5, and 10 years after random
assignment; duration-weighted census tract
share poor, share minority, share households with a single female-headed family,
and share college graduates;
collective efficacy; self-reported neighborhood safety; has at least one friend who
graduated from college; and access
to local health care services. For each of these measures the program calculates
the control group mean and then for
each of the two MTO treatemnt groups, the low-poverty voucher (LPV) group (also
known as the Experimental group) and
the traditional voucher (TRV) group (also known as the Section 8 group), it
calculates the ITT estimate and its 95%
confidence interval and p-value as well as the group mean.

The program creates the following outputs, which are logged to a CSV file:

**** Control Mean:
The control mean is calculated as a weighted mean using the weight mn_f_wt_totsvy
for particpants in the control group (ra_group=3).

**** Treatment Group Mean:
The mean for both the LPV group (ra_group=1/ra_grp_exp=1) and the TRV group
(ra_group=2/ra_grp_s8=1) is calculated
as the weighted mean using the weight mn_f_wt_totsvy.
```

```
**** Intention-to-treat effect, 95% confidence interval, Standard Error, T-
statistic, and P-value
The intention-to-treat (ITT) effect and its associated measures are calculated
using an ordinary least squares (OLS)
regression for outcomes that are non-binary. The ITT effects for binary outcomes
are caulcated using a logistic
regression model. In both cases the model controls for MTO randomization site and
is p-weighted (using mn_f_wt_totsvy).


STEPS:
1. Set options, program directories, and names of the input and output files
      1a. Set STATA Version
      1b. Set options
      1c. Set data file to use
      1d. Set output file
2. Define globals for weight variable and list of baseline measures to calculate
means for
      2a. Define weight global
      2b. Define OLS regression variables that are continuous (All used in Table 2)
      2c. Define logit variable list for table 3
3. Open Expanded Pseudo Individual Public Use File
4. Open CSV Results file and write column header names
      4a. Create file link to CSV results file for OLS and Logit ITT results
      4b. Write column headers to CSV results file for OLS and logit ITT results
5. Calculate Table 2 Values. OLS ITTs and associated statistics for each outcome
      5a. Open for loop to loop through each outcome defined in the global
$outcomes
      5b. Calculate weighted control mean for the outcome and store value.
      5c. calculate the weighted mean of the Experimental treatment group for the
outcome
      5d. Calculate ITT impact of the outcome on the Experimental treatment group
using linear regression model
      5e. Store Experimental Group ITT results as scalars or local variables,
calculate p-value,
            and set significance symbols, save results as scalars
      5f. Calculate the weighted mean of the Experimental treatment group for the
outcome
      5g. Calculate ITT impact of the outcome on the Section 8 treatment group
using linear regression model
      5h. Store Section 8 Group ITT results as scalars or local variables,
calculate p-value,
            and set significance symbols, save results as scalars
      5i. Write estimated results for the outcome/mediator to comma-separated file
6. Calculate Table 3 values. Logit with marginal effects
      6a. Open for loop to loop through each outcome defined in the global
$outcomes
      6b. Calculate weighted control mean for the outcome and store value.
      6c. calculate the weighted mean of the Experimental treatment group for the
outcome
      6d. Calculate ITT impact of the outcome on the Experimental treatment group
using logit regression model
      6e. Calculate marginal effects
      6f. Store Experimental Group ITT results as scalars or local variables,
calculate p-value,
            and set significance symbols, save results as scalars
```

```
      6g. Calculate the weighted mean of the Experimental treatment group for the
outcome
      6h. Calculate marginal effects
      6i. Calculate ITT impact of the outcome on the Section 8 treatment group
using logit regression model
      6j. Store Section 8 Group ITT results as scalars or local variables,
calculate p-value,
            and set significance symbols, save results as scalars
      6k. Write estimated results for the outcome/mediator to comma-separated file
7. Close log
8. Print date and time.
*/

********************************************************************************
********
**** STEP 1. Set options, program directories, and names of the input and output
files ****
********************************************************************************
********

**** 1a. Set STATA version ****
version 12

**** 1b. Set options ****
clear all
set more off
set maxvar 10000
set linesize 255


**** 1c. Set dataset to use ****
global data http://www.nber.org/mtopuf/mto_nejm_puf_pseudo_20131025.dta


* !!!! USER MUST UPDATE DIRECTORY OF OUTPUT FILE AND DATE !!!!

**** 1d. Set output file ****
* Set today's date for dating the log
global todaydt 20131118

* Set name of log file
global pgmdir ~/mtoproj/m10_pgm/papers/icpsr_archive/nejm/outputs/
global logname "${pgmdir}03_nejm_pseudo_itt_${todaydt}.csv"

* !!!! END NECESSARY USER-SET OPTIONS !!!!


********************************************************************************
************************
**** STEP 2. Define globals for weight variable and list of baseline measures to
calculate means for ******
********************************************************************************
************************

**** 2a. Define weight global ****
global weight mn_f_wt_totsvy
```

```
**** 2b. Define OLS regression variables that are continuous (All used in Table 2)
****
global ols_outcomes ps_f_spl_moves_n ///
       ps_rad_c9010t_perpov_bl ps_rad_c9010t_perpov_yr1 ps_rad_c9010t_perpov_yr5
ps_rad_c9010t_perpov_yr10 ///
       ps_f_c9010t_perpov_dw ps_f_c9010t_pminorty_dw ps_f_c9010t_pfsfem_dw
ps_f_c9010t_pcolldeg_dw


**** 2c. Define logit variable list for Table 2 and Table 3 ****
global logit_outcomes ///
       /*Table 2 Outcomes*/ ///
       psbi_f_sn_monit_graffiti_ad psbi_f_nb_safe_safday_ad
psbi_f_sn_net_anyfrndgrad_ad psbi_f_ph_care_place2go_noner_ad ///
       /*Table 3 Outcomes*/ ///
       psbi_f_ph_bmi_obese_srm_ad psbi_f_ph_bmi_obese2_srm_ad
psbi_f_ph_bmi_obese3_srm_ad psbi_f_db_hba1c_diab_final



*****************************************************************
**** STEP 3. Open Expanded Pseudo Individual Public Use File ****
*****************************************************************
use $data, clear



*********************************************************************
**** STEP 4. Open CSV Results files and write column header names ****
*********************************************************************

**** 4a. Create file link to CSV results file for OLS and Logit ITT results ****
file open csvlog using "$logname", write replace


**** 4b. Write column headers to CSV results file for OLS and logit ITT results
****
file write csvlog
"lookup,variable,abbr,c_mean,ex_n,ex_itt,ex_se,ex_pv,ex_ittlo,ex_itthi,ex_mean,"
file write csvlog
"s8_n,s8_itt,s8_se,s8_pv,s8_ittlo,s8_itthi,s8_mean" _n



*************************************************************************************
************
**** STEP 5. Calculate Table 2 Values: OLS ITTs and associated statistics for each
outcome ****
*************************************************************************************
************

**** 5a. Open for loop to loop through each outcome defined in the global $outcomes
****
```

```
** multiply mean. itt, and se by100 because we want to show means, ITT impacts, and
standard errors in percentage points in the final tables so that they match with
the original NEJM tables

foreach outcome of varlist $ols_outcomes ps_f_spl_moves_n {

**** 5b.  Calculate weighted control mean for the outcome and store value ****
    sum `outcome' [aw=${weight}] if ra_group==3
    * set scalar to control mean
    if "`outcome'" == "ps_f_spl_moves_n" scalar cmean = r(mean)
    else scalar cmean = r(mean)*100


**** 5c. calculate the weighted mean of the Experimental treatment group for the
outcome ****
    sum `outcome' [aw=${weight}] if ra_group==1
    if "`outcome'"=="ps_f_spl_moves_n" scalar ex_mean = r(mean)
    else scalar ex_mean = r(mean)*100


**** 5d. Calculate ITT impact of the outcome on the Experimental treatment group
using linear regression model ****
    * regress variable against the treatment group indicator and controlling for
randomization site
    regress `outcome' ra_grp_exp x_f_site_balt x_f_site_bos x_f_site_chi
x_f_site_la [pw=${weight}] if ra_group~=2


**** 5e. Store Experimental Group ITT results as scalars or local variables,
calculate p-value, and set significance symbols, save results as scalars ****

    * save ITT estimate
    if "`outcome'"=="ps_f_spl_moves_n" scalar ex_itt = _b[ra_grp_exp]
    else scalar ex_itt = _b[ra_grp_exp]*100

    * save robust standard error (for ITT)
    if "`outcome'"=="ps_f_spl_moves_n" scalar ex_ittse = _se[ra_grp_exp]
    else scalar ex_ittse = _se[ra_grp_exp]*100

    * calculat the t-statistic
    scalar ex_tsta  = ex_itt / ex_ittse

    * analytic N
    scalar ex_n = e(N)

    * degrees of freedom
    scalar ex_df_r = e(df_r)

    * calculate confidence interval
    scalar ex_itthi = (ex_itt+invnorm(.975)*ex_ittse)
    scalar ex_ittlo = (ex_itt-invnorm(.975)*ex_ittse)

    * Calculate p value
    scalar ex_ittpv = ttail( ex_df_r, abs( ex_tsta ))*2
```

```
**** 5f. calculate the weighted mean of the Section 8 treatment group for the
outcome ****
    sum `outcome' [aw=${weight}] if ra_group==2
    if "`outcome'"=="ps_f_spl_moves_n" scalar s8_mean = r(mean)
    else scalar s8_mean = r(mean)*100


**** 5g.  Calculate ITT impact of the outcome on the Section 8 treatment group
using linear regression model ****
    * regress variable against the treatment group indicator and controlling for
randomization site
    regress `outcome' ra_grp_s8 x_f_site_balt x_f_site_bos x_f_site_chi x_f_site_la
[pw=${weight}] if ra_group~=1


**** 5h.  Store Section 8 Group ITT results as scalars or local variables,
calculate p-value, and set significance symbols, save results as scalars ****

    * save ITT estimate
    if "`outcome'"=="ps_f_spl_moves_n" scalar s8_itt   = _b[ra_grp_s8]
    else scalar s8_itt   = _b[ra_grp_s8]*100

    * save robust standard error (for ITT)
    if "`outcome'"=="ps_f_spl_moves_n" scalar s8_ittse = _se[ra_grp_s8]
    else scalar s8_ittse = _se[ra_grp_s8]*100

    * calculat the t-statistic
    scalar s8_tsta  = s8_itt / s8_ittse

    * analytic N
    scalar s8_n = e(N)

    * degrees of freedom
    scalar s8_df_r = e(df_r)

    * calculate confidence interval
    scalar s8_itthi = (s8_itt+invnorm(.975)*s8_ittse)
    scalar s8_ittlo = (s8_itt-invnorm(.975)*s8_ittse)

    * Calculate p value
    scalar s8_ittpv = ttail( s8_df_r, abs( s8_tsta ))*2

**** 5i.  Write estimated results for the outcome/mediator to comma-separated file
****
    file write csvlog "`outcome'_ols,`outcome',_ols, " %20.3f (cmean) "," ///
    %20.0f (ex_n) "," %20.3f (ex_itt) "," %20.3f (ex_ittse) "," %20.3f (ex_ittpv)
"," ///
    %20.3f (ex_ittlo) "," %20.3f (ex_itthi) "," %20.3f (ex_mean) "," ///
    %20.0f (s8_n) "," %20.3f (s8_itt) "," %20.3f (s8_ittse) "," %20.3f (s8_ittpv)
"," ///
    %20.3f (s8_ittlo) "," %20.3f (s8_itthi) "," %20.3f (s8_mean) "," _n

}


*********************************************************************
**** STEP 6. Calculate Table 3 values: Logit with marginal effects ****
```

```
*********************************************************************

**** 6a. Open for loop to loop through each outcome defined in the global $outcomes
****

** multiply mean. itt, and se by100 because we want to show means, ITT impacts, and
standard errors in percentage points in the final tables so that they match with
the original NEJM tables

foreach outcome in $logit_outcomes {

**** 6b.  Calculate weighted control mean for the outcome and store value ****
    sum `outcome' [aw=${weight}] if ra_group==3

    * set scalar to control mean
    scalar cmean = r(mean)*100


**** 6c. Calculate the weighted mean of the Experimental treatment group for the
outcome ****
    sum `outcome' [aw=${weight}] if ra_group==1

    scalar ex_mean = r(mean)*100


**** 6d. Calculate ITT impact of the outcome on the Experimental treatment group
using logit regression model ****
    * regress variable against the treatment group indicator and controlling for
randomization site
    logit `outcome' ra_grp_exp x_f_site_balt x_f_site_bos x_f_site_chi x_f_site_la
[pw=${weight}] if ra_group~=2

    * get the degrees of freedom
    scalar ex_df_r = e(N) - e(df_m)


**** 6e. Calculate marginal effects ****
    noisily margins, dydx(ra_grp_exp) post


**** 6f. Store Experimental Group ITT results as scalars or local variables,
calculate p-value, and set significance symbols, save results as scalars ****

    * analytic N
    scalar ex_n = r(N)

    * save ITT estimate
    scalar ex_itt   = _b[ra_grp_exp] * 100

    * save robust standard error (for ITT)
    scalar ex_ittse = _se[ra_grp_exp] * 100

    * calculat the t-statistic
    scalar ex_tsta  = ex_itt / ex_ittse

    * calculate confidence interval
    scalar ex_itthi = (ex_itt+invnorm(.975)*ex_ittse)
```

```
    scalar ex_ittlo = (ex_itt-invnorm(.975)*ex_ittse)

    * Calculate p value
    scalar ex_ittpv = ttail( ex_df_r, abs( ex_tsta ))*2


**** 6g. calculate the weighted mean of the Section 8 treatment group for the
outcome ****
    sum `outcome' [aw=${weight}] if ra_group==2

    scalar s8_mean = r(mean)*100

**** 6h.  Calculate ITT impact of the outcome on the Section 8 treatment group
using linear regression model ****
    * regress variable against the treatment group indicator and controlling for
randomization site
    logit `outcome' ra_grp_s8 x_f_site_balt x_f_site_bos x_f_site_chi x_f_site_la
[pw=${weight}] if ra_group~=1

    * set the sample
    capture drop s8_logitsample
    gen s8_logitsample = e(sample)

    * get the degrees of freedom
    scalar s8_df_r = e(N) - e(df_m)


**** 6i.  calculate marginal effects ****
    noisily margins, dydx(ra_grp_s8) post


**** 6j.  Store Section 8 Group ITT results as scalars or local variables,
calculate p-value, and set significance symbols, save results as scalars ****

    * analytic N
    scalar s8_n = r(N)

    * save ITT estimate
    scalar s8_itt   = _b[ra_grp_s8] * 100

    * save robust standard error (for ITT)
    scalar s8_ittse = _se[ra_grp_s8] * 100

    * calculat the t-statistic
    scalar s8_tsta  = s8_itt / s8_ittse

    * calculate confidence interval
    scalar s8_itthi = (s8_itt+invnorm(.975)*s8_ittse)
    scalar s8_ittlo = (s8_itt-invnorm(.975)*s8_ittse)

    * Calculate p value
    scalar s8_ittpv = ttail( s8_df_r, abs( s8_tsta ))*2


**** 6k.  Write estimated results for the outcome/mediator to comma-separated file
****
    file write csvlog "`outcome'_logit,`outcome',_logit, " %20.3f (cmean) "," ///
```

```
    %20.0f (ex_n) "," %20.3f (ex_itt) "," %20.3f (ex_ittse) "," %20.3f (ex_ittpv)
"," ///
    %20.3f (ex_ittlo) "," %20.3f (ex_itthi) "," %20.3f (ex_mean) "," ///
    %20.0f (s8_n) "," %20.3f (s8_itt) "," %20.3f (s8_ittse) "," %20.3f (s8_ittpv)
"," ///
    %20.3f (s8_ittlo) "," %20.3f (s8_itthi) "," %20.3f (s8_mean) "," _n

}



*************************
**** STEP 7. Close log ****
*************************

file close csvlog


***********************************
**** STEP 8. Print date and time ****
***********************************
disp "Program End on " c(current_date) " at " c(current_time)
```

Appendix F – Detailed Coding of Selected Variables Used for the *NEJM* Article (separate document)

This appendix is available at www.nber.org/mtopuf/mto_nejm_puf_docu_memo_apndxF_20131209.pdf.