Documentation of the MTO Public Use Datasets

for the *Science* Article:

"Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults"

by

Jens Ludwig, Greg J. Duncan, Lisa A. Gennetian, Lawrence F. Katz,

Ronald C. Kessler, Jeffrey R. Kling, and Lisa Sanbonmatsu

Science Vol. 337 no. 6101 pp. 1505-1510

Published September 21, 2012

Documentation last updated:

October 8, 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 Nicholas Potter, research analysts at the National Bureau of Economic Research.

Contents

(click on any line below to skip to the selected section)

1. Overview 4

2. Background on the MTO Experiment 5

3. Sample, Data Sources, and Index Measures 7

3.1 Sample 7

3.2 Data Sources 7

3.3 Index Measures 8

4. Construction of the Cell-Level PUF 8

5. Construction of the Expanded Pseudo-Individual PUF 10

6. Replicating Results Using the PUFs 12

6.1 Estimation of Control and Treatment Means (Science Table 1) 12

6.2 Estimation of Intention-to-Treat Effects (Science Figure 1 and Table 2) 14

6.3 Estimation of the Relationship Between Outcomes and Specific Neighborhood Conditions Using Site Interactions with Treatment Group as Instruments (Science Figure 2) 21

6.4 Graphing the Relationship Between Outcomes and Neighborhood Characteristics Using Site-Group Means (Science Figure 2) 23

7. Summary 25

References 26

Appendix A – Variable List for the Cell-Level PUF File 27

Appendix B – Variable List for the Pseudo-Individual PUF File 36

Appendix C – Program Code Showing the Collapse and Expansion of the Data 39

Appendix D – Stata Code for Estimating Group Means Using the Science PUF 44

Appendix E – Stata Code for Estimating Intention-to-Treat Effects Using the Science PUF 50

Appendix F – Stata Code for Estimating the Relationship Between Outcomes and Neighborhood Conditions Using Site-Group Instruments Using the Science PUF 63

Appendix G – Stata Code for Graphing the Relationship Between Outcomes and Neighborhood Conditions Using the Science PUF 69

Appendix H – Stata Code Used to Create the Outcome Indices Used in the *Science* Article 93

Appendix I – Detailed Documentation for the Mental Health Disorder Outcomes Included in the *Science* Article 96

Appendix J – Detailed Coding of Selected Variables Used for the *Science* Article (separate document) 107

Also included are the following tables and figures (click on any line below to skip to the selected exhibit):

Table 1. Key Variables on the Cell-Level PUF Dataset 9

Table 2. Key Variables on the Expanded Pseudo-Individual PUF Dataset 11

Table 3. Individual-Level Data and Expanded Pseudo-Individual Data Comparisons of Baseline Characteristic Means by Treatment Group 13

Table 4. Individual-Level Data and Expanded Pseudo-Individual Data Comparisons of Pooled Low-Poverty and Traditional Voucher Group ITT/TOT Effects on the Science Article Outcomes (Plus Index Components and Additional Subjective Well-Being Outcomes) 15

Table 5. Site-Group IV Estimates of Effects of Share Poor and Minority on Subjective Well-Being, Comparison of Individual-Level and Expanded Pseudo-Individual Data 22

Figure 1. Replication of Figure 1 from the Science Article Using the Expanded Pseudo-Individual Data 20

Figure 2. Replication of Figure 2 from the Science Article Using the Cell-Level Data 24

DOCUMENTATION OF

MTO PUBLIC USE DATASETS

RELATED TO *SCIENCE* ARTICLE:

"Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults"

Science Vol. 337 no. 6101 pp. 1505-1510

1. Overview

The Moving to Opportunity (MTO) public use files (PUFs) are designed for a rough replication of Tables 1 and 2 and Figures 1 and 2 in the article "Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults" published in the journal *Science* on September 21, 2012 (available via www.mtoresearch.org/final.htm). The article shows that moving from a high-poverty to lower-poverty neighborhood leads to long-term improvements in adult physical and mental health and subjective well-being, despite not affecting economic self-sufficiency.

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) publication related archive (www.icpsr.umich.edu/icpsrweb/ICPSR/index.jsp) are:

- **mto_sci_puf_cells_20130206.dta** – To preserve the confidentiality of respondents, the data on 3,273 adults included in the Science analyses have been collapsed to 158 cells (ranging in size from 14 to 41 respondents). The cells are homogenous on MTO site and treatment group to allow for the estimation of impacts. When possible, the cells have also been divided by treatment compliance status and race [non-Hispanic African-Americans vs. all other races/ethnicities]. The file contains the mean, standard deviation, and sum of weights for all outcomes and mediators. There are 251 variables on the file (see Appendix A for a complete list of variables).

- **mto_sci_puf_pseudo_20130206.dta** – The cell-level data have been expanded to a pseudo-individual level dataset (n=3273) 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 94 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 outcomes including subjective well-being, economic self-sufficiency, physical and mental health, neighborhood poverty and minority concentration, neighborhood safety and housing conditions;
- demographic information like 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 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 and some continuous monetary measures were rounded and top-coded. Census tract characteristics have also been rounded.

The datasets allow one to estimate impact estimates for the combined voucher intervention groups and for each voucher group separately. However, due to the aggregation of the data, it is not possible to control for the full set of baseline covariates used for the estimates in the *Science* article. Using the aggregate data, only the site covariates can be included in the model. As Tables 3, 4, and 5 in this memo show, despite these limitations the estimates using the PUFs produce means, point estimates, and standard errors that are quite similar to those in the *Science* article. To more precisely estimate impacts and use a more complete set of covariates, the individual-level data are needed. These restricted access individual-level data will be archived shortly with ICPSR. To apply to use the restricted access data, search ICPSR's Moving to Opportunity holdings at www.icpsr.umich.edu/icpsrweb/content/ICPSR/access/restricted/index.html.

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 *Science* article.

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

3. Sample, Data Sources, and Index Measures

3.1 Sample

The PUF datasets consist of data on the 3,273 adults interviewed for 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 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 survey instrument is available at www.mtoresearch.org/instruments/final_hhold.pdf.

3.2 Data Sources

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

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. Subjective well-being was assessed using the survey question: "Taken all together, how would you say things are these days—would you say that you are very happy, pretty happy, or not too happy?"[1] The primary measure analyzed for the *Science* article is a scaled version of those responses (1 = not too happy, 2 = pretty happy, 3 = very happy) that is then transferred to a z-score by subtracting the control group mean and dividing by the control group standard deviation. Embedded within the survey was a structured diagnostic interview to assess mental health disorders such as depression and generalized anxiety, based on portions of the World Health Organization's Composite International Diagnostic Interview[2].

Biomeasures: Height and weight were measured during the long-term survey to allow for the calculation of body mass index (BMI; equal to weight in kilograms divided by height in meters squared) and to determine obesity (BMI $\geq$ 30). Hypertension was assessed using up to two blood pressure readings taken by the interviewers. Hypertension was defined as systolic pressure of 140 mmHg or higher or diastolic pressure of 90 mmHg or higher.

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 1 year after random assignment, 5 years after random assignment, and 10-15 years after

[1] This question was borrowed from the General Social Survey (Smith, Marsden, Hout and Kim, 2011).
[2] For more information, see Kessler and Üstün (2004).

random assignment (May 2008), 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 and Figure 2 in the *Science* article include measures of tract share poor (the fraction of tract residents living below the poverty level) and tract share minority (the fraction of tract residents who are members of racial or ethnic minority groups).

3.3 Index Measures

The data include index measures of mental health, physical health, and economic self-sufficiency. The elements of these outcome indices were pre-specified to the current study, based on what was constructed for the interim follow-up. Our economic self-sufficiency index is composed of: an indicator for whether the respondent is currently employed and not on Temporary Assistance to Needy Families (TANF), an indicator for currently employed, total annual earning, an indicator for currently on TANF, and total annual income from government programs. Our physical health index consists of: self-reported health is fair or poor, the respondent had an asthma attack the past year, obesity, hypertension, and trouble carrying groceries or climbing stairs. Our mental health index consists of: a psychological distress index score for the past month, depression in the past year, Generalized Anxiety Disorder in the past year, calm and peaceful during the past month, and normal sleep last night.

Each variable within these outcome indices is first re-scaled so that higher values equal "better" outcomes, then converted to z-scores by subtracting the control group mean and dividing by the control group standard deviation, then averaged across all individual outcomes within the domain, and then re-scaled again so that the index itself has a standard deviation of one. For people missing data on any element of the index, we impute the group average value of that variable, which yields estimates that are the equivalent of the average of the coefficients. See Appendix H for the Stata code used to generate the indices.

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 158 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. Cells were further subdivided at random into cells ranging from 14 to 41 observations. 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 **ra_poolgrp_exps8** – combined flag for the LPV and TRV groups
Compliance Status	**f_svy_cmove** –flag indicating that the family moved using an MTO housing voucher (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_ad_site_balt** – Baltimore site flag **x_f_ad_site_bos** – Boston site flag **x_f_ad_site_chi** – Chicago site flag **x_f_ad_site_la** – Los Angeles site flag (New York is the omitted category in the regression models)
Cell Information	**cell_id** – cell identification number, ranging from 1 to 158 **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)
Census Tract Characteristics Predicted Using Site-Group (for use in Instrumental Variable estimation)	**predsg_perpov_dw_z** – tract share poor (duration weighted and z-scored) predicted based on site X treatment group interactions **predsg_pminorty_dw_z** – tract share minority (duration weighted and z-scored) predicted based on site X treatment group interactions

†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_ for physical health, f_mh_ for mental health, f_ec_/f_em_/f_in_ for economic outcomes, f_nb_ for neighborhood outcomes, f_hc_ for housing consumption, f_sn_ for social networks, happy_ for subjective well-being, 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. However, 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. Within each cell, we artificially set one record 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)[3]. These new values are assigned to variables starting with the prefix: ps_*. The "pseudo individual-level" file roughly mimics the original data for the purposes of running impact estimates by site and treatment group and running instrumental variable estimates using the predicted share poor and share minority variables. However, they 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_* variables have been artificially set and while they capture the mean and the standard deviation for the variable, they DO NOT reflect the actual values in the original dataset. The ps_* 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.

[3] 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 **ra_poolgrp_exps8** – combined flag for the LPV and TRV groups
Compliance Status	**f_svy_cmove** –flag indicating that the family moved using an MTO housing voucher (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_ad_site_balt** – Baltimore site flag **x_f_ad_site_bos** – Boston site flag **x_f_ad_site_chi** – Chicago site flag **x_f_ad_site_la** – Los Angeles site flag (New York is the omitted category in the regression models)
Pseudo-Individual Observations 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. These variables include the z-scored happiness measure (ps_happy_scale123_z_ad) and index measures (ps_f_ec_idx_z_ad, ps_f_ph_idx_fix_z_ad, and ps_f_mh_idx_z_ad).
Census Tract Characteristics Predicted Using Site-Group (for use in IV estimation)	**predsg_perpov_dw_z** – tract share poor (duration-weighted and z-scored) predicted based on site X treatment group interactions **predsg_pminorty_dw_z** - tract share minority (duration-weighted and z-scored) predicted based on site X treatment group interactions

6. Replicating Results Using the PUFs

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

Table 1 in the *Science* article presents mean baseline characteristics for the control group and the MTO treatment (voucher) groups. The table also indicates the p-value level of the differences between the groups (no differences are statistically significant at the .05 level). These group means and tests of the differences can be roughly replicated using the expanded pseudo-individual file (except for the age of respondents, which is not included on the PUF). See Table 3 of this document for a comparison of the results published in the *Science* 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 the treatment group (*ra_poolgrp_exps8* = 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_poolgrp_exps8==1
```

To test the significance of the difference in means, we use a weighted regression of the treatment group dummy 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:

```
regress ps_x_f_ad_nevmarr ra_poolgrp_exps8 [pw=mn_f_wt_totsvy]

scalar sc_diff_pv = ttail(e(df_r), abs(_b[ra_poolgrp_exps8] / _se[ra_poolgrp_exps8])) * 2
```

where *e(N)* is the regression sample size, *e(df_m)* is the regression degrees of freedom, *_b[ra_poolgrp_exps8]* is the coefficient on the treatment dummy variable, and *_se[ra_poolgrp_exps8]* is the standard error of the treatment variable.

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

Table 3. Individual-Level Data and Expanded Pseudo-Individual Data Comparisons of Baseline Characteristic Means by Treatment Group

	Control group mean		Pooled LPV/TRV group mean	
	Individual-Level Data	Pseudo-Individual Data	Individual-Level Data	Pseudo-Individual Data
	N=1139	N=1139	N=2134	N=2134
Gender and age				
Female	0.978	0.978	0.984	0.984
Age as of December 31, 2007				
In whole years†	44.5	N/A	44.6	N/A
Age categories (age is…)				
Less than or equal 35	0.143	0.143	0.140	0.140
36 to 40	0.229	0.227	0.222	0.223
41 to 45	0.234	0.235	0.230	0.230
46 to 50	0.175	0.174	0.192	0.192
Race and ethnicity				
African-American††				
Any ethnicity	0.660	N/A	0.640	N/A
Non-Hispanic	N/A	0.635	N/A	0.623
Hispanic ethnicity (any race)	0.304	0.303	0.325	0.322
Other demographic characteristics				
Never married	0.637	0.637	0.623	0.623
Working	0.245	0.245	0.270	0.270
High school diploma	0.361	0.361	0.367	0.367
General Educational Development (GED)	0.199	0.199	0.169 *	0.169 *
Receiving Aid to Families with Dependent Children (AFDC)	0.763	0.763	0.752	0.752
Household characteristics				
Household income (2009 dollars)	$12,438.64	$12,445.28	$12,833.64	$12,830.74
Site				
Baltimore	0.135	0.135	0.136	0.136
Boston	0.205	0.205	0.203	0.203
Chicago	0.205	0.205	0.206	0.206
Los Angeles	0.226	0.226	0.225	0.225
New York	0.229	0.229	0.229	0.229
Neighborhood characteristics				
Household member was crime victim in last six months	0.416	0.416	0.425	0.425
Very dissatisfied with neighborhood	0.467	0.467	0.478	0.478
Primary or Secondary Reason for Wanting to Move				
To get away from gangs and drugs	0.779	0.779	0.770	0.770
Better schools for children	0.481	0.481	0.516 *	0.516 *
To get a bigger or better apartment	0.457	0.457	0.440	0.440
To get a job	0.069	0.069	0.058	0.058

Notes: * = p < .10, ** = p < .05, *** = p < .01.

† Age in whole years is not available in the pseudo-individual data, but the individual-level data means, which appeared in the *Science* article are included here for reference.

†† The individual-level data measure includes African-Americans of any ethnicity, i.e. including Hispanics, whereas the pseudo-individual data measures includes only non-Hispanic African-Americans (and those with missing ethnicity).

6.2 Estimation of Intention-to-Treat Effects (Science Figure 1 and Table 2)

The impacts presented in Figure 1 and Table 2 of the *Science* article are intention-to-treat (ITT) effects or the impacts of being offered an MTO housing voucher. For both Figure 1 and Table 2, the two voucher groups (LPV and TRV groups) are pooled. Using the expanded pseudo-individual file, the ITT effects are estimated using a linear regression with the outcome or mediator as the dependent variable and the key independent variable being the dummy variable indicating assignment to one of the two voucher groups (*ra_poolgrp_exps8* = 1). The regression also controls for randomization site (using the four *x_f_site_** indicators, with New York City as the omitted category)[4], applies a probability weight (*mn_f_wt_totsvy*), and generates Huber-White standard errors.

In Stata, the command line to estimate the impact on subjective well-being of being offered an MTO housing voucher is as follows:

```
regress ps_happy_scale123_z_ad ra_poolgrp_exps8 x_f_site_balt x_f_site_bos x_f_site_chi x_f_site_la
    [pw=mn_f_wt_totsvy]
```

where the ITT impact of the voucher will be the coefficient on the variable *ra_poolgrp_exps8* or *_b[ra_poolgrp_exps8]*.

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

Table 4 below compares three sets of impact estimates:

1. using the *individual-level data* and controlling for *all covariates* (these results match those presented in the *Science* article);
2. using the *individual-level data* and controlling *only for randomization site*; and
3. using the *expanded pseudo-individual data* 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 does not allow one to control for additional baseline characteristics in the regression model.

Note that in addition to the four outcomes presented in Figure 1, we have also included in this table and the data the component outcomes that feed into the three indices (economic self-sufficiency, physical health, and mental health). Also note that the table includes columns for the effects of treatment-on-the-treated (TOT), or the effect of the MTO treatment on those who actually move through MTO. Figure 1 and Table 2 in the *Science* article do not present TOT effects, but we include them here for completeness. The TOT estimate and standard error are calculated by dividing the ITT estimate and standard error, respectively, by the weighted compliance rate (the fraction of adults in the treatment groups who moved using their MTO vouchers).

The expanded pseudo-individual file can also be used to estimate MTO impacts for each site and for the two housing voucher groups separately. For example, to estimate impacts on happiness for only the Chicago site for LPV group, use Stata code such as the following, which restricts the data to the Chicago site (*x_f_site_chi* = 1) and excludes the TRV group (*ra_grp_s8* = 0) to compare the LPV and group to the control group:

```
regress ps_happy_scale123_z_ad ra_grp_exp if x_f_site_chi == 1 & ra_grp_s8 == 0 [pw=mn_f_wt_totsvy]
```

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

Table 4. Individual-Level Data and Expanded Pseudo-Individual Data Comparisons of Pooled Low-Poverty and Traditional Voucher Group ITT/TOT Effects on the Science Article Outcomes (Plus Index Components and Additional Subjective Well-Being Outcomes)

	Control mean		ITT			TOT			N	
			Full Covariates	Site Covariates		Full Covariates	Site Covariates			
	Individual-Level Data	Pseudo-Individual Data	Individual-Level Data	Individual-Level Data	Pseudo-Individual Data	Individual-Level Data	Individual-Level Data	Pseudo-Individual Data	Indiv.–Level Data	Pseudo-Individual Data
FIGURE 1										
Subjective well-being, 3-point scale, z-score	0.000	0.000	0.098 **	0.089 **	0.089 **	0.180 **	0.163 **	0.164 **	3266	3265
			(0.039)	(0.040)	(0.039)	(0.073)	(0.073)	(0.073)		
Absence of mental health problems index, z-score	0.000	0.000	0.070 *	0.081 **	0.081 **	0.129 *	0.149 **	0.149 **	3273	3273
			(0.041)	(0.040)	(0.039)	(0.075)	(0.074)	(0.073)		
Absence of physical health problems index, z-score	0.000	0.000	0.060	0.064	0.064	0.112	0.118	0.118	3273	3273
			(0.039)	(0.040)	(0.039)	(0.072)	(0.075)	(0.073)		
Economic self-sufficiency index, z-score	0.000	0.000	-0.061	-0.044	-0.044	-0.112	-0.081	-0.081	3271	3271
			(0.038)	(0.040)	(0.040)	(0.071)	(0.074)	(0.073)		

Table 4. (continued)

	Control mean		ITT			TOT			N	
			Full Covariates	Site Covariates		Full Covariates	Site Covariates			
	Individual-Level Data	Pseudo-Individual Data	Individual-Level Data	Individual-Level Data	Pseudo-Individual Data	Individual-Level Data	Individual-Level Data	Pseudo-Individual Data	Indiv.–Level Data	Pseudo-Individual Data
TABLE 2										
Census tract characteristics										
Share poor at different points in time										
1 year post-random assignment	0.499	0.499	-0.160 *** (0.007)	-0.155 *** (0.007)	-0.155 *** (0.006)	-0.295 *** (0.012)	-0.286 *** (0.012)	-0.287 *** (0.012)	3224	3224
5 years post-random assignment	0.399	0.399	-0.089 *** (0.007)	-0.087 *** (0.007)	-0.087 *** (0.007)	-0.165 *** (0.013)	-0.160 *** (0.013)	-0.160 *** (0.012)	3208	3209
10-15 years post-random assignment (May 2008)	0.311	0.311	-0.034 *** (0.007)	-0.035 *** (0.007)	-0.034 *** (0.006)	-0.063 *** (0.012)	-0.064 *** (0.012)	-0.063 *** (0.012)	3206	3205
Share poor for all addresses since random assignment										
Share poor	0.396	0.396	-0.082 *** (0.005)	-0.081 *** (0.005)	-0.081 *** (0.005)	-0.152 *** (0.009)	-0.149 *** (0.009)	-0.149 *** (0.009)	3270	3270
Share poor, z-score using U.S. tract poverty distribution	2.082	2.082	-0.666 *** (0.041)	-0.653 *** (0.041)	-0.653 *** (0.039)	-1.229 *** (0.076)	-1.205 *** (0.075)	-1.205 *** (0.073)	3270	3270
Share poor, z-score using MTO control group tract poverty distribution	0.000	0.000	-0.653 *** (0.040)	-0.640 *** (0.040)	-0.640 *** (0.039)	-1.205 *** (0.075)	-1.182 *** (0.073)	-1.182 *** (0.071)	3270	3270
Duration-weighted poverty rate										
Less than 20%	0.054	0.054	0.196 *** (0.013)	0.186 *** (0.013)	0.186 *** (0.012)	0.361 *** (0.024)	0.343 *** (0.023)	0.343 *** (0.023)	3270	3270
Less than 30%	0.242	0.242	0.237 *** (0.018)	0.231 *** (0.018)	0.231 *** (0.017)	0.438 *** (0.033)	0.426 *** (0.033)	0.426 *** (0.032)	3270	3270
Less than 40%	0.512	0.512	0.206 *** (0.018)	0.205 *** (0.018)	0.205 *** (0.018)	0.381 *** (0.034)	0.378 *** (0.034)	0.378 *** (0.033)	3270	3270

Table 4. (continued)

	Control mean		ITT			TOT			N	
			Full Covariates	Site Covariates		Full Covariates	Site Covariates			
	Individual-Level Data	Pseudo-Individual Data	Individual-Level Data	Individual-Level Data	Pseudo-Individual Data	Individual-Level Data	Individual-Level Data	Pseudo-Individual Data	Indiv.–Level Data	Pseudo-Individual Data
TABLE 2 (continued)										
Census tract characteristics										
Share minority										
10-15 years post-random assignment (May 2008)	0.844	0.844	-0.024 ** (0.009)	-0.027 *** (0.009)	-0.028 *** (0.009)	-0.044 ** (0.018)	-0.051 *** (0.017)	-0.051 *** (0.017)	3206	3205
All addresses since random assignment (duration-	0.880	0.880	-0.046 *** (0.006)	-0.043 *** (0.006)	-0.043 *** (0.006)	-0.085 *** (0.012)	-0.080 *** (0.011)	-0.080 *** (0.011)	3270	3270
Residential mobility										
Number of moves after random assignment	2.165	2.165	0.584 *** (0.068)	0.560 *** (0.069)	0.560 *** (0.069)	1.078 *** (0.125)	1.035 *** (0.128)	1.035 *** (0.128)	3273	3273
Self-reports on long-term (10- to 15-year) follow-up surveys about neighborhood and housing conditions										
Feel unsafe during day	0.196	0.196	-0.039 ** (0.015)	-0.040 ** (0.016)	-0.040 *** (0.015)	-0.071 ** (0.028)	-0.074 ** (0.029)	-0.074 *** (0.028)	3262	3261
Number of housing problems	2.051	2.051	-0.380 *** (0.076)	-0.374 *** (0.076)	-0.374 *** (0.075)	-0.700 *** (0.140)	-0.689 *** (0.140)	-0.691 *** (0.138)	3267	3266
Likely or very likely to report kids spraying graffiti (collective efficacy)	0.589	0.589	0.064 *** (0.020)	0.061 *** (0.020)	0.060 *** (0.020)	0.118 *** (0.036)	0.111 *** (0.037)	0.112 *** (0.036)	3255	3254
1+ friends with college degree	0.532	0.532	0.049 ** (0.020)	0.039 * (0.021)	0.039 * (0.020)	0.090 ** (0.037)	0.072 * (0.038)	0.072 * (0.037)	3203	3201

Table 4. (continued)

	Control mean		ITT			TOT			N	
			Full Covariates	Site Covariates		Full Covariates	Site Covariates			
	Individual-Level Data	Pseudo-Individual Data	Individual-Level Data	Individual-Level Data	Pseudo-Individual Data	Individual-Level Data	Individual-Level Data	Pseudo-Individual Data	Indiv.–Level Data	Pseudo-Individual Data
ADDITIONAL MEASURES										
(not shown in the article)										
Index components										
Mental health										
K6 raw score (range 0-24), past month	6.961	6.961	-0.527 *** (0.200)	-0.520 ** (0.203)	-0.520 *** (0.201)	-0.973 *** (0.370)	-0.960 ** (0.376)	-0.960 *** (0.371)	3273	3273
Depression, past year	0.138	0.138	-0.016 (0.014)	-0.015 (0.014)	-0.015 (0.013)	-0.029 (0.025)	-0.029 (0.025)	-0.029 (0.025)	3273	3273
Generalized anxiety disorder, past year	0.066	0.066	-0.003 (0.010)	-0.007 (0.009)	-0.007 (0.009)	-0.006 (0.018)	-0.014 (0.017)	-0.014 (0.017)	3273	3273
Calm and peaceful, past month	0.487	0.487	0.003 (0.020)	0.010 (0.020)	0.010 (0.020)	0.006 (0.038)	0.019 (0.038)	0.019 (0.037)	3272	3272
Normal sleep (7-8 hours) last night	0.291	0.291	0.017 (0.019)	0.018 (0.019)	0.018 (0.018)	0.031 (0.036)	0.034 (0.035)	0.034 (0.034)	3241	3241
Physical health										
Fair or poor self-reported	0.436	0.436	0.001 (0.019)	-0.004 (0.020)	-0.004 (0.020)	0.001 (0.036)	-0.007 (0.037)	-0.007 (0.036)	3269	3269
Asthma attack in past year	0.293	0.293	-0.026 (0.018)	-0.024 (0.018)	-0.024 (0.018)	-0.048 (0.034)	-0.043 (0.034)	-0.044 (0.033)	3267	3266
Currently obese (body mass index ≥ 30)	0.584	0.584	-0.012 (0.020)	-0.011 (0.020)	-0.011 (0.020)	-0.022 (0.038)	-0.020 (0.037)	-0.020 (0.037)	3221	3221
Currently has high blood	0.315	0.315	-0.004 (0.019)	-0.004 (0.019)	-0.004 (0.019)	-0.007 (0.035)	-0.007 (0.036)	-0.007 (0.035)	3102	3100
Health limits ability to climb several flights of stairs or lift or carry groceries	0.510	0.510	-0.040 ** (0.020)	-0.044 ** (0.020)	-0.044 ** (0.020)	-0.074 ** (0.036)	-0.082 ** (0.038)	-0.082 ** (0.037)	3270	3270

Table 4. (continued)

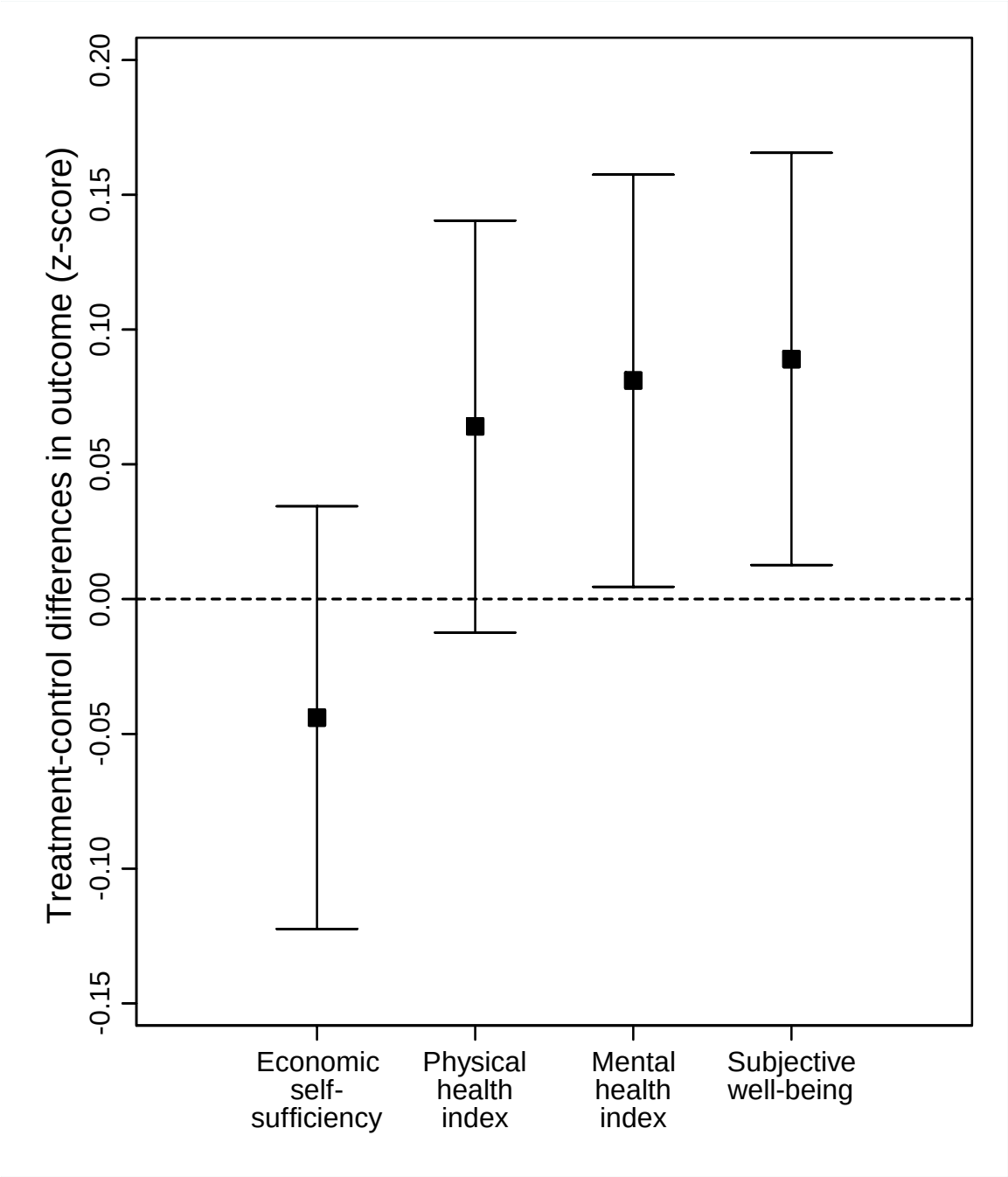
	Control mean		ITT			TOT			N	
			Full Covariates	Site Covariates		Full Covariates	Site Covariates			
	Individual-Level Data	Pseudo-Individual Data	Individual-Level Data	Individual-Level Data	Pseudo-Individual Data	Individual-Level Data	Individual-Level Data	Pseudo-Individual Data	Indiv.–Level Data	Pseudo-Individual Data
ADDITIONAL MEASURES										
(not shown in the article)										
Index components (continued)										
Economic self-sufficiency										
Currently employed	0.525	0.525	-0.031	-0.025	-0.025	-0.056	-0.045	-0.045	3264	3264
			(0.020)	(0.020)	(0.020)	(0.036)	(0.038)	(0.037)		
Currently receiving Temporary Assistance for Needy Families (TANF)	0.158	0.158	0.015	0.008	0.008	0.029	0.015	0.015	3262	3262
			(0.014)	(0.014)	(0.014)	(0.026)	(0.027)	(0.026)		
Currently working and household is not receiving	0.499	0.499	-0.037 *	-0.030	-0.030	-0.069 *	-0.055	-0.055	3260	3260
			(0.020)	(0.020)	(0.020)	(0.036)	(0.038)	(0.037)		
Annual individual earnings	$12,288.51	$12,292.16	10.62	142.40	135.57	19.61	262.99	250.16	3141	3140
			(558.21)	(593.82)	(582.78)	(1030.93)	(1096.69)	(1075.34)		
Annual household government income	$3,542.62	$3,542.57	271.08	213.82	212.59	496.98	392.01	391.26	3139	3139
			(204.35)	(215.00)	(207.57)	(374.64)	(394.16)	(382.02)		
Additional subjective well-being measures										
Happiness scale score, raw	1.953	1.953	0.069 **	0.063 **	0.063 **	0.127 **	0.116 **	0.116 **	3266	3265
			(0.028)	(0.028)	(0.028)	(0.051)	(0.052)	(0.051)		
Very happy (vs. pretty happy or not too happy)	0.228	0.228	0.026	0.024	0.024	0.047	0.044	0.044	3266	3265
			(0.017)	(0.017)	(0.017)	(0.032)	(0.032)	(0.031)		
Very happy or pretty happy (vs. not too happy)	0.725	0.725	0.044 **	0.039 **	0.039 **	0.080 **	0.072 **	0.072 **	3266	3265
			(0.017)	(0.018)	(0.017)	(0.032)	(0.032)	(0.032)		

Notes: * = p < .10, ** = p < .05, *** = p < .01.

Figure 1 below is a replication of Figure 1 in the *Science* article using the expanded pseudo-individual PUF. The figure presents the impacts of being offered a housing voucher (either low-poverty or traditional voucher) on economic self-sufficiency, physical health, mental health, and subjective well-being (outcomes shown on the x-axis). The squares represent the ITT point estimate, and the box whiskers represent the 95th percent confidence interval around the estimates.

The Stata code in Appendix E was used to generate Figure 1 below using the expanded pseudo-individual PUF.

Figure 1. Replication of Figure 1 from the Science Article Using the Expanded Pseudo-Individual Data



6.3 Estimation of the Relationship Between Outcomes and Specific Neighborhood Conditions Using Site Interactions with Treatment Group as Instruments (Science Figure 2)

The slopes of the graphs shown in Figure 2 of the *Science* article depict the relationship between subjective well-being and neighborhood conditions. In Figure 2 the relationship (represented by the slope) can also be generated via a two-stage least squares approach (2SLS), where in the first stage indicators for randomization site interacted with treatment assignment are used to predict census tract share poor and share minority[5] with additional controls for randomization site alone and applying probability weights. In the second stage, the instrumented neighborhood measure (share poor or share minority) is used to predict subjective well-being), with additional controls for randomization site as above. In addition to calculating the slope, the instrumental variable (IV) model also allows provides us with the standard error of the slope and allows us to test the difference between the slopes in panels C and D of Figure 2.

Using the expanded pseudo-individual PUF is the easiest way to roughly replicate the IV estimates reflected in Figure 2 and discussed in the text of the *Science* article. Also, because share poor and share minority cannot be properly predicted via the PUF, the PUF contains a predicted share poor and share minority value for all observations and therefore only the second stage of the 2SLS model needs to be estimated. A regression for the second stage of a 2SLS model estimating the relationship between duration-weighted share poor and subjective well-being would look like the following:

```
regress ps_happy_scale123_z_ad predsg_perpov_dw_z x_f_site_balt x_f_site_bos x_f_site_chi x_f_site_la
    [pw=mn_f_wt_totsvy]
```

See Appendix F for the Stata program that generates the rough equivalent of the IV estimates discussed in the *Science* article using the expanded pseudo-individual PUF.

As with the ITT estimates described above, the IV estimates produced using the PUF will not exactly replicate the estimates presented in the *Science* article, but as shown in Table 5 below, the estimates are generally very close. Table 5 compares four sets of IV estimates:

1. using the *individual-level data*, controlling for *all covariates*, and using a *standard* 2SLS approach (these results match those presented in the *Science* article);
2. using the *individual-level data*, controlling for *randomization site only*, and using a *standard* 2SLS approach;
3. using the *individual-level data*, controlling for *randomization site only*, and using a *manual* 2SLS approach; and
4. using the *expanded pseudo-individual data* from the PUF, controlling for *randomization site only*, and using a *manual* 2SLS approach.

[5] In Figure 2 these measures are duration-weighted for all addresses at which the adult respondent lived during the 10- to 15-year follow-up period.

Table 5. Site-Group IV Estimates of Effects of Share Poor and Minority on Subjective Well-Being, Comparison of Individual-Level and Expanded Pseudo-Individual Data

	IV Estimate (Two-Stage Least Squares)				N	
	Full Covariates	Site Covariates			Full/Site Covariates	Site Covariates
	Individual-Level Data	Individual-Level Data		Pseudo-Individual Data	Individual-Level Data	Pseudo-Individual Data
Mediator(s) Included in Model	Standard 2SLS	Standard 2SLS	Manual 2SLS	Manual 2SLS	Standard/ Manual 2SLS	Manual 2SLS
Share poor (duration-weighted)	-0.141 ***	-0.129 **	-0.129 **	-0.128 **	3263	3265
	(0.054)	(0.054)	(0.054)	(0.054)		
Share minority (duration-weighted)	-0.069	-0.054	-0.054	-0.048	3263	3265
	(0.098)	(0.098)	(0.098)	(0.098)		
Share poor, controlling for share minority (duration-weighted)	-0.261 ***	-0.255 ***	-0.255 ***	-0.252 ***	3263	3265
	(0.093)	(0.095)	(0.092)	(0.091)		
Share minority, controlling for share poor (duration-weighted)	0.279 *	0.289	0.289 *	0.287 *	3263	3265
	(0.169)	(0.176)	(0.167)	(0.163)		
P-value of test that coefficients are equal	0.030	0.036	0.028	0.026		

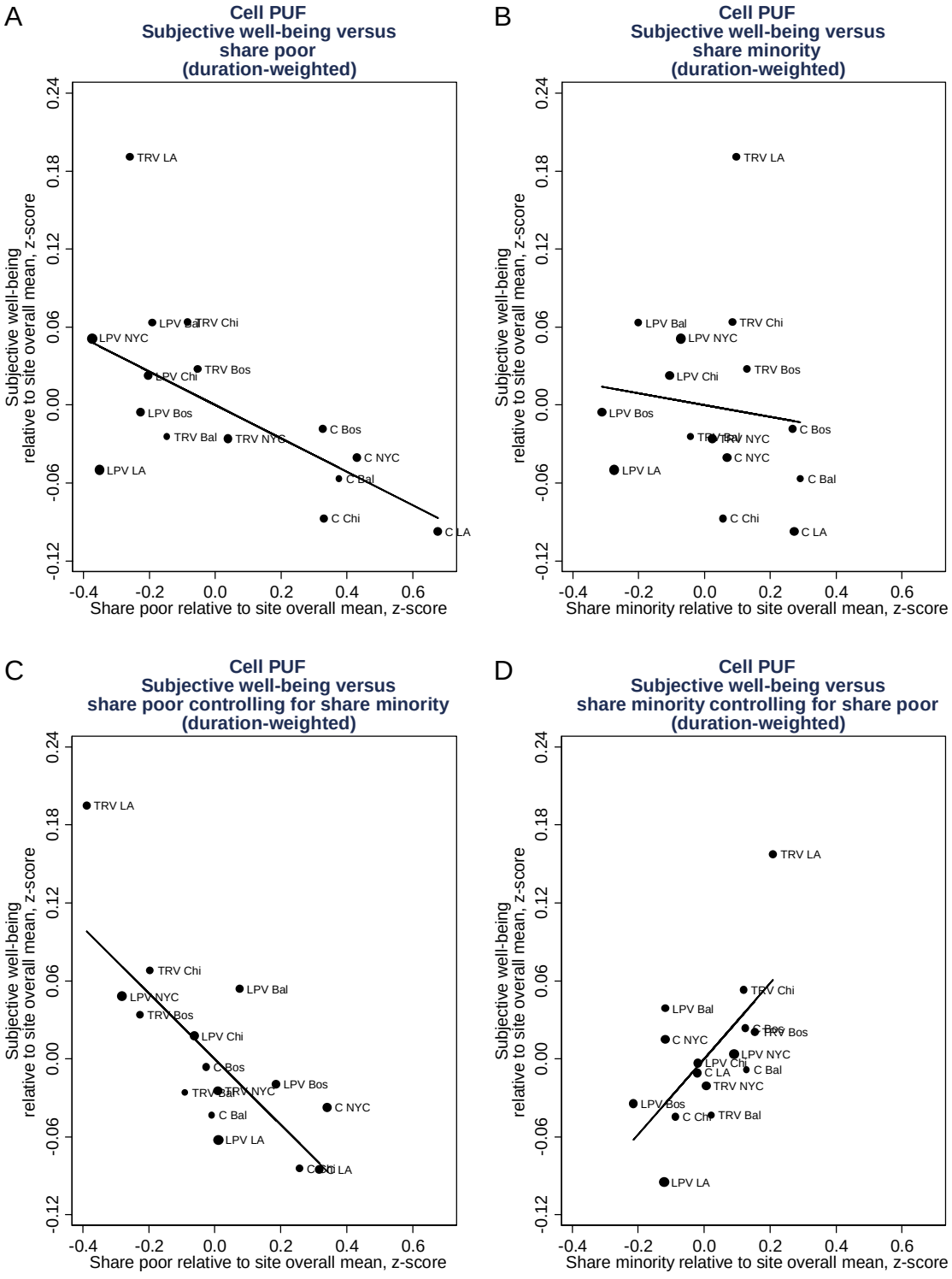
6.4 Graphing the Relationship Between Outcomes and Neighborhood Characteristics Using Site-Group Means (Science Figure 2)

Figure 2 of the *Science* article depicts the relationship between subjective well-being and average (duration-weighted) tract poverty rate [panel A], tract share minority [panel B], tract poverty controlling for minority share [panel c], and tract minority share controlling for tract poverty [panel D]. The y-axis is the three-point happiness scale described above (z-scored against the MTO control group). The data points represent the 15 site (Bal=Baltimore; Bos=Boston; Chi=Chicago; LA=Los Angeles; NY=New York City) and treatment group (LPV=low-poverty voucher; TRV=traditional voucher; C=control group) combinations. The slope of the line in each panel is equivalent to a 2SLS estimate of the relationship between subjective well-being and the mediator shown in each panel, using interactions of indicators for MTO treatment group assignment and controlling for site.. The model essentially estimates a "dose-response" relationship to determine whether the groups that experience relatively larger changes in census tract share poor or share minority as a result of the MTO treatment also experience larger changes in subjective well-being.

Because the data points in the plots in Figure 2 of the *Science* article are the site/treatment group means relative to the overall means for the site, the first step in replicating the figure is to calculate the overall site means. The cell-level data are homogenous by site and treatment group, but there are multiple cells for each site/treatment group combination, so the cells need to be collapsed to site/group level to generate site means. To express the site/treatment group means relative to each site's mean, we subtract the site mean from each of the averages and then use these demeaned averages to plot SWB versus the tract characteristic. The slope of the line matches the relationship between SWB and the tract characteristic estimated using an IV regression. Using panel C (subjective well-being versus share poor controlling for share minority) as an example, first regress demeaned subjective well-being against demeaned share minority and then regress demeaned share poor against demeaned share minority. Finally, use the predicted values from those two regressions to regress subjective well-being (demeaned and controlling for share minority) against share poor (also demeaned and controlling for share minority). For further details, please see Appendix G.

Figure 2 of this document replicates the *Science* article figure using the PUF data. See Appendix G for the Stata code to generate these plots.

Figure 2. Replication of Figure 2 from the Science Article Using the Cell-Level Data



7. Summary

The MTO Science PUFs are designed to allow the user to roughly replicate the results in the *Science* article "Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults". The data have been aggregated and a few variables were modified (prior to aggregating) to protect the confidentiality of the data, and 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

Kessler, R. C., and Merikangas, K. R (2004). The National Comorbidity Survey Replication (NCS-R): Background and aims. *Int. J. Methods Psychiatr. Res.* **13**, 60.

Kling, J. R., Liebman, J. B., and Katz, L. F. (2007) Experimental Analysis of Neighborhood Effects. *Econometrica* **75**, 83.

Orr, L., Feins, J.D., Jacob, R., Beecroft, E. , Sanbonmatsu, L. , Katz, L.F., Liebman, J.B., and Kling, J.R. (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.

Sanbonmatsu, L., Ludwig, J., Katz, L.F., Gennetian, L.A., Duncan, G.J., Kessler, R.C., Adam, E. , McDade, T.W. , and Tessler Lindau, S. (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, 2011); available at www.huduser.org/portal/publications/pubasst/MTOFHD.html.

Smith, T. W., Marsden, P. , Hout, M. , and Kim, J. (2011) General social surveys, 1972–2010 [machine-readable data file] Sponsored by National Science Foundation. NORC ed. Chicago: National Opinion Research Center [producer]; Storrs, CT: The Roper Center for Public Opinion Research, University of Connecticut.

Appendix A – Variable List for the Cell-Level PUF File
(mto_sci_puf_cells_20130206.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 J, which includes detailed documentation of the outcome and mediator measures.

obs: 158
vars: 251
size: 208,560

Variable Name	Type	Variable Label
cell_id	float %9.0g	cell_id - number id assigned to cell
cell_numobs	Long %10.0	cell_numobs - # of observations contributing to cell
cell_sumwgt	Double %10.0g	cell_sumwgt - sum of weights for the cell
mn_cov_hous_movapt	float %9.0g	mean of cov_Hous_movapt - 1st/2nd reason for wanting to move was better apt.
mn_cov_hous_movjob	float %9.0g	mean of cov_Hous_movapt - 1st/2nd reason for wanting to move was to find job
mn_f_c9010t_perpov_dw	float %9.0g	mean of f_c9010t_perpov_dw - Duration-weighted tract poverty
mn_f_c9010t_perpov_dw_und20	float %9.0g	mean of f_c9010t_perpov_dw_und20 - Duration-weighted tract poverty rate < 20%
mn_f_c9010t_perpov_dw_und30	float %9.0g	mean of f_c9010t_perpov_dw_und30 - Duration-weighted tract poverty rate < 30%
mn_f_c9010t_perpov_dw_und40	float %9.0g	mean of f_c9010t_perpov_dw_und40 - Duration-weighted tract poverty rate < 40%
mn_f_c9010t_perpov_dw_z	float %9.0g	mean of f_c9010t_perpov_dw_z - Duration-wgtd poverty, MTO control z-score
mn_f_c9010t_perpov_dw_zc00t	float %9.0g	mean of f_c9010t_perpov_dw_zc00t - Duration-wgtd poverty, US tracts z-score
mn_f_c9010t_pminorty_dw	float %9.0g	mean of f_c9010t_pminorty_dw - Duration-weighted tract share minority
mn_f_c9010t_pminorty_dw_z	float %9.0g	mean of f_c9010t_pminorty_dw_z - Duration-weighted share minority, z-score
mn_f_ec_idx_z_ad	float %9.0g	mean of f_ec_idx_z_ad - Economic Self-Sufficiency Index, z-score
mn_f_em_emp_ad	float %9.0g	mean of f_em_emp_ad - Adult was employed or temporarily absent last week
mn_f_hc_hsgprb_fix	float %9.0g	mean of f_hc_hsgprb_fix - Number of housing problems (0-7)
mn_f_in_selfsuf_ad	float %9.0g	mean of f_in_selfsuf_ad - Employed and not on TANF
mn_f_in_tanf_fam	float %9.0g	mean of f_in_tanf_fam - Adult or her/his children are currently on TANF
mn_f_mh_calm_ad	float %9.0g	mean of f_mh_calm_ad - Adult was calm/peaceful most of the time in past month
mn_f_mh_dep_y_ad	float %9.0g	mean of f_mh_dep_y_ad - DSM-IV Major Depressive Episode in past year

Appendix A, continued

Variable Name	Type	Variable Label
mn_f_mh_gad_y_ad	float %9.0g	mean of f_mh_gad_y_ad - DSM-IV Generalized Anxiety Disorder in past year
mn_f_mh_idx_z_ad	float %9.0g	mean of f_mh_idx_z_ad - Absence of Mental Health Problems Index, z-score
mn_f_ph_idx_fix_z_ad	float %9.0g	mean of f_ph_idx_z_ad - Absence of Physical Health Problems Index, z-score
mn_f_mh_k6_raw_ad	float %9.0g	mean of f_mh_k6_raw_ad - Psychological Distress Index (K6) raw score (0-24)
mn_f_nb_safe_unsafday_ad	float %9.0g	mean of f_nb_safe_safday_ad - Feels unsafe/very unsafe near home during day
mn_f_ph_asma_y_ad	float %9.0g	mean of f_ph_asma_y_ad - Asthma attack or wheezing in the past 12 months
mn_f_ph_bmi_obese_srm_ad	float %9.0g	mean of f_ph_bmi_obese_srm_ad - Obese adult (including self-reports): BMI>=30
mn_f_ph_bp_hi	float %9.0g	mean of f_ph_bp_hi - High blood pressure: Systolic>=140 or Diastolic>=90
mn_f_ph_habit_sleep_78hrs_ad	float %9.0g	mean of f_ph_habit_sleep_78hrs_ad - Adult slept 7-8 hours last night
mn_f_ph_hlth_fair_ad	float %9.0g	mean of f_ph_hlth_fair_ad - Adult has fair or poor health
mn_f_ph_limit_liftstair	float %9.0g	mean of f_ph_limit_liftstair - Health limits lift/stair climb a little/a lot
mn_f_sn_monit_graffiti_ad	float %9.0g	mean of f_sn_monit_graffiti_ad - Likely/very likely to report graffiti
mn_f_sn_net_anyfrndgrad_ad	float %9.0g	mean of f_sn_net_anyfrndgrad_ad - 1+ friends who graduated from college
mn_f_spl_moves_n	float %9.0g	mean of f_spl_moves_n - Number of moves according to spell file addresses
mn_f_svy_cmove	float %9.0g	mean of f_svy_cmove - Core Mover/Treatment Compliance Flag (1=Moved)
mn_happy_scale123_ad	float %9.0g	mean of happy_scale123_ad - Happiness scale score (1=not too/2=pretty/3=very)
mn_happy_scale123_z_ad	float %9.0g	mean of happy_scale123_z_ad - Happiness scale score, z-score
mn_happy_very_happy_ad	float %9.0g	mean of happy_very_happy_ad - Very (vs. pretty/not too) happy on 3-pt scale
mn_happy_verypretty_happy_ad	float %9.0g	mean of happy_verypretty_happy_ad - Very/pretty (vs. not too) happy 3pt scale
mn_rad_c9010t_perpov_m08	float %9.0g	mean of rad_c9010t_perpov_m08 - Tract poverty at address as of 5/31/08
mn_rad_c9010t_perpov_yr1	float %9.0g	mean of rad_c9010t_perpov_yr1 - Tract poverty 1 year post-randomization
mn_rad_c9010t_perpov_yr5	float %9.0g	mean of rad_c9010t_perpov_yr5 - Tract poverty 5 years post-randomization
mn_rad_c9010t_pminorty_m08	float %9.0g	mean of rad_c9010t_pminorty_m08 - Tract share minority as of 5/31/08
mn_rad_in_govt2009	float %9.0g	mean of rad_in_govt2009 - Gov't receipts income at final eval. (2009 dollars)
mn_rad_in_head2009	float %9.0g	mean of rad_in_head2009 - Adult indiv. earnings at final eval. (2009 dollars)
mn_rad_svy_bl_totincm_2009d	float %9.0g	mean of rad_svy_bl_totincm_2009d - Baseline household income (2009 dollars)
mn_x_f_ad_edged	float %9.0g	mean of x_f_ad_edged - At baseline, adult had a GED
mn_x_f_ad_edgradhs	float %9.0g	mean of x_f_ad_edgradhs - At baseline, adult had completed high school
mn_x_f_ad_edgradhs_miss	float %9.0g	mean of x_f_ad_edgradhs_miss - Missing flag for baseline GED/H.S. diploma

Appendix A, continued

Variable Name	Type	Variable Label
mn_x_f_ad_edinsch	float %9.0g	mean of x_f_ad_edinsch - At baseline, adult was enrolled in school
mn_x_f_ad_nevmarr	float %9.0g	mean of x_f_ad_nevmarr - At baseline, adult had never been married
mn_x_f_ad_parentu18	float %9.0g	mean of x_f_ad_parentu18 - Adult was less than age 18 at birth of 1st child
mn_x_f_ad_working	float %9.0g	mean of x_f_ad_working - At baseline, adult was working for pay
mn_x_f_hh_afdc	float %9.0g	mean of x_f_hh_afdc - At baseline, household was receiving AFDC/TANF
mn_x_f_hh_car	float %9.0g	mean of x_f_hh_car - At baseline, household owned a car
mn_x_f_hh_disabl	float %9.0g	mean of x_f_hh_disabl - At baseline, a household member had a disability
mn_x_f_hh_noteens	float %9.0g	mean of x_f_hh_noteens - No teens (ages 13-17) in baseline household
mn_x_f_hh_size2	float %9.0g	mean of x_f_hh_size2 - Baseline household size is 2 or smaller
mn_x_f_hh_size3	float %9.0g	mean of x_f_hh_size3 - Baseline household size equals 3
mn_x_f_hh_size4	float %9.0g	mean of x_f_hh_size4 - Baseline household size equals 4
mn_x_f_hh_victim	float %9.0g	mean of x_f_hh_victim - Household member victimized in 6 mos. before baseline
mn_x_f_hood_5y	float %9.0g	mean of x_f_hood_5y - At baseline, respondent had been living in nbhd. 5+ yrs
mn_x_f_hood_chat	float %9.0g	mean of x_f_hood_chat - At baseline, respondent chatted w/ neighbor >=1x/week
mn_x_f_hood_nbrkid	float %9.0g	mean of x_f_hood_nbrkid - At baseline, very likely to tell on neighbor's kid
mn_x_f_hood_nofamily	float %9.0g	mean of x_f_hood_nofamily - At baseline, no family living in neighborhood
mn_x_f_hood_nofriend	float %9.0g	mean of x_f_hood_nofriend - At baseline, no friends living in neighborhood
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, very dissatisfied with neighborhood
mn_x_f_hous_fndapt	float %9.0g	mean of x_f_hous_fndapt - Baseline respondent very sure of finding an apt.
mn_x_f_hous_mov3tm	float %9.0g	mean of x_f_hous_mov3tm - Respondent had moved >3x in 5 years before baseline
mn_x_f_hous_movdrgs	float %9.0g	mean of x_f_hous_movdrgs - 1st/2nd reason for wanting to move was drugs/crime
mn_x_f_hous_movschl	float %9.0g	mean of x_f_hous_movschl - 1st/2nd reason for wanting to move was good school
mn_x_f_hous_sec8bef	float %9.0g	mean of x_f_hous_sec8bef - At baseline, had applied for Section 8 before
mn_x_f_release1	float %9.0g	mean of x_f_release1 - Release 1 Sample Adult for Final Survey
mn_x_rad_ad_36_40	float %9.0g	mean of x_rad_ad_36_40 - Adult was age 36 to 40 as of 12/31/07
mn_x_rad_ad_41_45	float %9.0g	mean of x_rad_ad_41_45 - Adult was age 41 to 45 as of 12/31/07
mn_x_rad_ad_46_50	float %9.0g	mean of x_rad_ad_46_50 - Adult was age 46 to 50 as of 12/31/07
mn_x_rad_ad_ethrace_black_nh	float %9.0g	mean of x_rad_ad_ethrace_black_nh - Adult is African-American non-Hispanic

Variable Name	Type	Variable Label
mn_x_rad_ad_ethrace_hisp	float %9.0g	mean of x_rad_ad_ethrace_hisp - Adult is Hispanic [any race]
mn_x_rad_ad_le_35	float %9.0g	mean of x_rad_ad_le_35 - Adult was age 35 or younger as of 12/31/07
mn_x_rad_ad_male	float %9.0g	mean of x_rad_ad_male - Adult is male
mn_f_wt_totsvy	float %9.0g	mn_f_wt_totsvy - average of f_wt_totsvy for obs in cell
predsg_perpov_dw_z	float %9.0g	predsg_perpov_dw_z - predicted (on site-group) dw tract share poor
predsg_pminorty_dw_z	float %9.0g	predsg_pminorty_dw_z - predicted (on site-group) dw tract share minority
ra_group VALUE LABEL IS: lbl_group	float %21.0g	ra_group - Randomization group (1=Low-PovVouch/Exp, 2=TradVouch/Sec8, 3=Control)
ra_grp_control	float %8.0g	ra_grp_control - Control group flag
ra_grp_exp	float %8.0g	ra_grp_exp - Low-Poverty Voucher (Experimental) group flag
ra_grp_s8	float %8.0g	ra_grp_s8 - Traditional Voucher (Section 8) group flag
ra_poolgrp_exps8	float %9.0g	ra_poolgrp_exps8 - Low-Pov. (Exp)/Traditional Voucher (Sec8) group flag
ra_site VALUE LABEL IS: lbl_site	float %15.0g	ra_site - Site (1=Baltimore/2=Boston/3=Chicago/4=Los Angeles/5=New York City)
sgx_rasite_3g_all_nyc	float %8.0g	sgx_rasite_3g_all_nyc - Site-group interaction flag - New York site
sd_cov_hous_movapt	float %9.0g	std dev of cov_Hous_movapt - 1st/2nd reason for wanting to move was better apt.
sd_cov_hous_movjob	float %9.0g	std dev of cov_Hous_movapt - 1st/2nd reason for wanting to move was to find job
sd_f_c9010t_perpov_dw	float %9.0g	std dev of f_c9010t_perpov_dw - Duration-weighted tract poverty
sd_f_c9010t_perpov_dw_und20	float %9.0g	std dev of f_c9010t_perpov_dw_und20 - Duration-weighted tract poverty rate < 20%
sd_f_c9010t_perpov_dw_und30	float %9.0g	std dev of f_c9010t_perpov_dw_und30 - Duration-weighted tract poverty rate < 30%
sd_f_c9010t_perpov_dw_und40	float %9.0g	std dev of f_c9010t_perpov_dw_und40 - Duration-weighted tract poverty rate < 40%
sd_f_c9010t_perpov_dw_z	float %9.0g	std dev of f_c9010t_perpov_dw_z - Duration-wgtd poverty, MTO control z-score
sd_f_c9010t_perpov_dw_zc00t	float %9.0g	std dev of f_c9010t_perpov_dw_zc00t - Duration-wgtd poverty, US tracts z-score
sd_f_c9010t_pminorty_dw	float %9.0g	std dev of f_c9010t_pminorty_dw - Duration-weighted tract share minority
sd_f_c9010t_pminorty_dw_z	float %9.0g	std dev of f_c9010t_pminorty_dw_z - Duration-weighted share minority, z-score
sd_f_ec_idx_z_ad	float %9.0g	std dev of f_ec_idx_z_ad - Economic Self-Sufficiency Index, z-score
sd_f_em_emp_ad	float %9.0g	std dev of f_em_emp_ad - Adult was employed or temporarily absent last week
sd_f_hc_hsgprb_fix	float %9.0g	std dev of f_hc_hsgprb_fix - Number of housing problems (0-7)
sd_f_in_selfsuf_ad	float %9.0g	std dev of f_in_selfsuf_ad - Employed and not on TANF
sd_f_in_tanf_fam	float %9.0g	std dev of f_in_tanf_fam - Adult or her/his children are currently on TANF
sd_f_mh_calm_ad	float %9.0g	std dev of f_mh_calm_ad - Adult was calm/peaceful most of the time in past month

Variable Name	Type	Variable Label
sd_f_mh_dep_y_ad	float %9.0g	std dev of f_mh_dep_y_ad - DSM-IV Major Depressive Episode in past year
sd_f_mh_gad_y_ad	float %9.0g	std dev of f_mh_gad_y_ad - DSM-IV Generalized Anxiety Disorder in past year
sd_f_mh_idx_z_ad	float %9.0g	std dev of f_mh_idx_z_ad - Absence of Mental Health Problems Index, z-score
sd_f_ph_idx_fix_z_ad	float %9.0g	std dev of f_mh_idx_z_ad - Absence of Physical Health Problems Index, z-score
sd_f_mh_k6_raw_ad	float %9.0g	std dev of f_mh_k6_raw_ad - Psychological Distress Index (K6) raw score (0-24)
sd_f_nb_safe_unsafday_ad	float %9.0g	std dev of f_nb_safe_safday_ad - Feels unsafe/very unsafe near home during day
sd_f_ph_asma_y_ad	float %9.0g	std dev of f_ph_asma_y_ad - Asthma attack or wheezing in the past 12 months
sd_f_ph_bmi_obese_srm_ad	float %9.0g	std dev of f_ph_bmi_obese_srm_ad - Obese adult (including self-reports): BMI>=30
sd_f_ph_bp_hi	float %9.0g	std dev of f_ph_bp_hi - High blood pressure: Systolic>=140 or Diastolic>=90
sd_f_ph_habit_sleep_78hrs_ad	float %9.0g	std dev of f_ph_habit_sleep_78hrs_ad - Adult slept 7-8 hours last night
sd_f_ph_hlth_fair_ad	float %9.0g	std dev of f_ph_hlth_fair_ad - Adult has fair or poor health
sd_f_ph_limit_liftstair	float %9.0g	std dev of f_ph_limit_liftstair - Health limits lift/stair climb a little/a lot
sd_f_sn_monit_graffiti_ad	float %9.0g	std dev of f_sn_monit_graffiti_ad - Likely/very likely to report graffiti
sd_f_sn_net_anyfrndgrad_ad	float %9.0g	std dev of f_sn_net_anyfrndgrad_ad - 1+ friends who graduated from college
sd_f_spl_moves_n	float %9.0g	std dev of f_spl_moves_n - Number of moves according to spell file addresses
sd_f_svy_cmove	float %9.0g	std dev of f_svy_cmove - Core Mover/Treatment Compliance Flag (1=Moved)
sd_happy_scale123_ad	float %9.0g	std dev of happy_scale123_ad - Happiness scale score (1=not too/2=pretty/3=very)
sd_happy_scale123_z_ad	float %9.0g	std dev of happy_scale123_z_ad - Happiness scale score, z-score
sd_happy_very_happy_ad	float %9.0g	std dev of happy_very_happy_ad - Very (vs. pretty/not too) happy on 3-pt scale
sd_happy_verypretty_happy_ad	float %9.0g	std dev of happy_verypretty_happy_ad - Very/pretty (vs. not too) happy 3pt scale
sd_rad_c9010t_perpov_m08	float %9.0g	std dev of rad_c9010t_perpov_m08 - Tract poverty at address as of 5/31/08
sd_rad_c9010t_perpov_yr1	float %9.0g	std dev of rad_c9010t_perpov_yr1 - Tract poverty 1 year post-randomization
sd_rad_c9010t_perpov_yr5	float %9.0g	std dev of rad_c9010t_perpov_yr5 - Tract poverty 5 years post-randomization
sd_rad_c9010t_pminorty_m08	float %9.0g	std dev of rad_c9010t_pminorty_m08 - Tract share minority as of 5/31/08
sd_rad_in_govt2009	float %9.0g	std dev of rad_in_govt2009 - Gov't receipts income at final eval. (2009 dollars)
sd_rad_in_head2009	float %9.0g	std dev of rad_in_head2009 - Adult indiv. earnings at final eval. (2009 dollars)
sd_rad_svy_bl_totincm_2009d	float %9.0g	std dev of rad_svy_bl_totincm_2009d - Baseline household income (2009 dollars)
sd_x_f_ad_edged	float %9.0g	std dev of x_f_ad_edged - At baseline, adult had a GED
sd_x_f_ad_edgradhs	float %9.0g	std dev of x_f_ad_edgradhs - At baseline, adult had completed high school

Variable Name	Type	Variable Label
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_ad_edinsch	float %9.0g	std dev of x_f_ad_edinsch - At baseline, adult was enrolled in school
sd_x_f_ad_nevmarr	float %9.0g	std dev of x_f_ad_nevmarr - At baseline, adult had never been married
sd_x_f_ad_parentu18	float %9.0g	std dev of x_f_ad_parentu18 - Adult was less than age 18 at birth of 1st child
sd_x_f_ad_working	float %9.0g	std dev of x_f_ad_working - At baseline, adult was working for pay
sd_x_f_hh_afdc	float %9.0g	std dev of x_f_hh_afdc - At baseline, household was receiving AFDC/TANF
sd_x_f_hh_car	float %9.0g	std dev of x_f_hh_car - At baseline, household owned a car
sd_x_f_hh_disabl	float %9.0g	std dev of x_f_hh_disabl - At baseline, a household member had a disability
sd_x_f_hh_noteens	float %9.0g	std dev of x_f_hh_noteens - No teens (ages 13-17) in baseline household
sd_x_f_hh_size2	float %9.0g	std dev of x_f_hh_size2 - Baseline household size is 2 or smaller
sd_x_f_hh_size3	float %9.0g	std dev of x_f_hh_size3 - Baseline household size equals 3
sd_x_f_hh_size4	float %9.0g	std dev of x_f_hh_size4 - Baseline household size equals 4
sd_x_f_hh_victim	float %9.0g	std dev of x_f_hh_victim - Household member victimized in 6 mos. before baseline
sd_x_f_hood_5y	float %9.0g	std dev of x_f_hood_5y - At baseline, respondent had been living in nbhd. 5+ yrs
sd_x_f_hood_chat	float %9.0g	std dev of x_f_hood_chat - At baseline, respondent chatted w/ neighbor >=1x/week
sd_x_f_hood_nbrkid	float %9.0g	std dev of x_f_hood_nbrkid - At baseline, very likely to tell on neighbor's kid
sd_x_f_hood_nofamily	float %9.0g	std dev of x_f_hood_nofamily - At baseline, no family living in neighborhood
sd_x_f_hood_nofriend	float %9.0g	std dev of x_f_hood_nofriend - At baseline, no friends living in neighborhood
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, very dissatisfied with neighborhood
sd_x_f_hous_fndapt	float %9.0g	std dev of x_f_hous_fndapt - Baseline respondent very sure of finding an apt.
sd_x_f_hous_mov3tm	float %9.0g	std dev of x_f_hous_mov3tm - Respondent had moved >3x in 5 years before baseline
sd_x_f_hous_movdrgs	float %9.0g	std dev of x_f_hous_movdrgs - 1st/2nd reason for wanting to move was drugs/crime
sd_x_f_hous_movschl	float %9.0g	std dev of x_f_hous_movschl - 1st/2nd reason for wanting to move was good school
sd_x_f_hous_sec8bef	float %9.0g	std dev of x_f_hous_sec8bef - At baseline, had applied for Section 8 before
sd_x_f_release1	float %9.0g	std dev of x_f_release1 - Release 1 Sample Adult for Final Survey
sd_x_rad_ad_36_40	float %9.0g	std dev of x_rad_ad_36_40 - Adult was age 36 to 40 as of 12/31/07
sd_x_rad_ad_41_45	float %9.0g	std dev of x_rad_ad_41_45 - Adult was age 41 to 45 as of 12/31/07
sd_x_rad_ad_46_50	float %9.0g	std dev of x_rad_ad_46_50 - Adult was age 46 to 50 as of 12/31/07

Appendix A, continued

Variable Name	Type	Variable Label
sd_x_rad_ad_ethrace_black_nh	float %9.0g	std dev of x_rad_ad_ethrace_black_nh - Adult is African-American non-Hispanic
sd_x_rad_ad_ethrace_hisp	float %9.0g	std dev of x_rad_ad_ethrace_hisp - Adult is Hispanic [any race]
sd_x_rad_ad_le_35	float %9.0g	std dev of x_rad_ad_le_35 - Adult was age 35 or younger as of 12/31/07
sd_x_rad_ad_male	float %9.0g	std dev of x_rad_ad_male - Adult is male
wt_cov_hous_movapt	double %9.0g	sum of wts cov_Hous_movapt - 1st/2nd reason for wanting to move was better apt.
wt_cov_hous_movjob	double %9.0g	sum of wts cov_Hous_movapt - 1st/2nd reason for wanting to move was to find job
wt_f_c9010t_perpov_dw	double %9.0g	sum of wts f_c9010t_perpov_dw - Duration-weighted tract poverty
wt_f_c9010t_perpov_dw_und20	double %9.0g	sum of wts f_c9010t_perpov_dw_und20 - Duration-weighted tract poverty rate < 20%
wt_f_c9010t_perpov_dw_und30	double %9.0g	sum of wts f_c9010t_perpov_dw_und30 - Duration-weighted tract poverty rate < 30%
wt_f_c9010t_perpov_dw_und40	double %9.0g	sum of wts f_c9010t_perpov_dw_und40 - Duration-weighted tract poverty rate < 40%
wt_f_c9010t_perpov_dw_z	double %9.0g	sum of wts f_c9010t_perpov_dw_z - Duration-wgtd poverty, MTO control z-score
wt_f_c9010t_perpov_dw_zc00t	double %9.0g	sum of wts f_c9010t_perpov_dw_zc00t - Duration-wgtd poverty, US tracts z-score
wt_f_c9010t_pminorty_dw	double %9.0g	sum of wts f_c9010t_pminorty_dw - Duration-weighted tract share minority
wt_f_c9010t_pminorty_dw_z	double %9.0g	sum of wts f_c9010t_pminorty_dw_z - Duration-weighted share minority, z-score
wt_f_ec_idx_z_ad	double %9.0g	sum of wts f_ec_idx_z_ad - Economic Self-Sufficiency Index, z-score
wt_f_em_emp_ad	double %9.0g	sum of wts f_em_emp_ad - Adult was employed or temporarily absent last week
wt_f_hc_hsgprb_fix	double %9.0g	sum of wts f_hc_hsgprb_fix - Number of housing problems (0-7)
wt_f_in_selfsuf_ad	double %9.0g	sum of wts f_in_selfsuf_ad - Employed and not on TANF
wt_f_in_tanf_fam	double %9.0g	sum of wts f_in_tanf_fam - Adult or her/his children are currently on TANF
wt_f_mh_calm_ad	double %9.0g	sum of wts f_mh_calm_ad - Adult was calm/peaceful most of the time in past month
wt_f_mh_dep_y_ad	double %9.0g	sum of wts f_mh_dep_y_ad - DSM-IV Major Depressive Episode in past year
wt_f_mh_gad_y_ad	double %9.0g	sum of wts f_mh_gad_y_ad - DSM-IV Generalized Anxiety Disorder in past year
wt_f_mh_idx_z_ad	double %9.0g	sum of wts f_mh_idx_z_ad - Absence of Mental Health Problems Index, z-score
wt_f_ph_idx_fix_z_ad	double %9.0g	sum of wts f_mh_idx_z_ad - Absence of Physical Health Problems Index, z-score
wt_f_mh_k6_raw_ad	double %9.0g	sum of wts f_mh_k6_raw_ad - Psychological Distress Index (K6) raw score (0-24)
wt_f_nb_safe_unsafday_ad	double %9.0g	sum of wts f_nb_safe_safday_ad - Feels unsafe/very unsafe near home during day
wt_f_ph_asma_y_ad	double %9.0g	sum of wts f_ph_asma_y_ad - Asthma attack or wheezing in the past 12 months
wt_f_ph_bmi_obese_srm_ad	double %9.0g	sum of wts f_ph_bmi_obese_srm_ad - Obese adult (including self-reports): BMI>=30
wt_f_ph_bp_hi	double %9.0g	sum of wts f_ph_bp_hi - High blood pressure: Systolic>=140 or Diastolic>=90

Appendix A, continued

Variable Name	Type	Variable Label
wt_f_ph_habit_sleep_78hrs_ad	double %9.0g	sum of wts f_ph_habit_sleep_78hrs_ad - Adult slept 7-8 hours last night
wt_f_ph_hlth_fair_ad	double %9.0g	sum of wts f_ph_hlth_fair_ad - Adult has fair or poor health
wt_f_ph_limit_liftstair	double %9.0g	sum of wts f_ph_limit_liftstair - Health limits lift/stair climb a little/a lot
wt_f_sn_monit_graffiti_ad	double %9.0g	sum of wts f_sn_monit_graffiti_ad - Likely/very likely to report graffiti
wt_f_sn_net_anyfrndgrad_ad	double %9.0g	sum of wts f_sn_net_anyfrndgrad_ad - 1+ friends who graduated from college
wt_f_spl_moves_n	double %9.0g	sum of wts f_spl_moves_n - Number of moves according to spell file addresses
wt_f_svy_cmove	double %9.0g	sum of wts f_svy_cmove - Core Mover/Treatment Compliance Flag (1=Moved)
wt_happy_scale123_ad	double %9.0g	sum of wts happy_scale123_ad - Happiness scale score (1=not too/2=pretty/3=very)
wt_happy_scale123_z_ad	double %9.0g	sum of wts happy_scale123_z_ad - Happiness scale score, z-score
wt_happy_very_happy_ad	double %9.0g	sum of wts happy_very_happy_ad - Very (vs. pretty/not too) happy on 3-pt scale
wt_happy_verypretty_happy_ad	double %9.0g	sum of wts happy_verypretty_happy_ad - Very/pretty (vs. not too) happy 3pt scale
wt_rad_c9010t_perpov_m08	double %9.0g	sum of wts rad_c9010t_perpov_m08 - Tract poverty at address as of 5/31/08
wt_rad_c9010t_perpov_yr1	double %9.0g	sum of wts rad_c9010t_perpov_yr1 - Tract poverty 1 year post-randomization
wt_rad_c9010t_perpov_yr5	double %9.0g	sum of wts rad_c9010t_perpov_yr5 - Tract poverty 5 years post-randomization
wt_rad_c9010t_pminorty_m08	double %9.0g	sum of wts rad_c9010t_pminorty_m08 - Tract share minority as of 5/31/08
wt_rad_in_govt2009	double %9.0g	sum of wts rad_in_govt2009 - Gov't receipts income at final eval. (2009 dollars)
wt_rad_in_head2009	double %9.0g	sum of wts rad_in_head2009 - Adult indiv. earnings at final eval. (2009 dollars)
wt_rad_svy_bl_totincm_2009d	double %9.0g	sum of wts rad_svy_bl_totincm_2009d - Baseline household income (2009 dollars)
wt_x_f_ad_edged	double %9.0g	sum of wts x_f_ad_edged - At baseline, adult had a GED
wt_x_f_ad_edgradhs	double %9.0g	sum of wts x_f_ad_edgradhs - At baseline, adult had completed high school
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_ad_edinsch	double %9.0g	sum of wts x_f_ad_edinsch - At baseline, adult was enrolled in school
wt_x_f_ad_nevmarr	double %9.0g	sum of wts x_f_ad_nevmarr - At baseline, adult had never been married
wt_x_f_ad_parentu18	double %9.0g	sum of wts x_f_ad_parentu18 - Adult was less than age 18 at birth of 1st child
wt_x_f_ad_working	double %9.0g	sum of wts x_f_ad_working - At baseline, adult was working for pay
wt_x_f_hh_afdc	double %9.0g	sum of wts x_f_hh_afdc - At baseline, household was receiving AFDC/TANF
wt_x_f_hh_car	double %9.0g	sum of wts x_f_hh_car - At baseline, household owned a car
wt_x_f_hh_disabl	double %9.0g	sum of wts x_f_hh_disabl - At baseline, a household member had a disability
wt_x_f_hh_noteens	double %9.0g	sum of wts x_f_hh_noteens - No teens (ages 13-17) in baseline household

Variable Name	Type	Variable Label
wt_x_f_hh_size2	double %9.0g	sum of wts x_f_hh_size2 - Baseline household size is 2 or smaller
wt_x_f_hh_size3	double %9.0g	sum of wts x_f_hh_size3 - Baseline household size equals 3
wt_x_f_hh_size4	double %9.0g	sum of wts x_f_hh_size4 - Baseline household size equals 4
wt_x_f_hh_victim	double %9.0g	sum of wts x_f_hh_victim - Household member victimized in 6 mos. before baseline
wt_x_f_hood_5y	double %9.0g	sum of wts x_f_hood_5y - At baseline, respondent had been living in nbhd. 5+ yrs
wt_x_f_hood_chat	double %9.0g	sum of wts x_f_hood_chat - At baseline, respondent chatted w/ neighbor >=1x/week
wt_x_f_hood_nbrkid	double %9.0g	sum of wts x_f_hood_nbrkid - At baseline, very likely to tell on neighbor's kid
wt_x_f_hood_nofamily	double %9.0g	sum of wts x_f_hood_nofamily - At baseline, no family living in neighborhood
wt_x_f_hood_nofriend	double %9.0g	sum of wts x_f_hood_nofriend - At baseline, no friends living in neighborhood
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, very dissatisfied with neighborhood
wt_x_f_hous_fndapt	double %9.0g	sum of wts x_f_hous_fndapt - Baseline respondent very sure of finding an apt.
wt_x_f_hous_mov3tm	double %9.0g	sum of wts x_f_hous_mov3tm - Respondent had moved >3x in 5 years before baseline
wt_x_f_hous_movdrgs	double %9.0g	sum of wts x_f_hous_movdrgs - 1st/2nd reason for wanting to move was drugs/crime
wt_x_f_hous_movschl	double %9.0g	sum of wts x_f_hous_movschl - 1st/2nd reason for wanting to move was good school
wt_x_f_hous_sec8bef	double %9.0g	sum of wts x_f_hous_sec8bef - At baseline, had applied for Section 8 before
wt_x_f_release1	double %9.0g	sum of wts x_f_release1 - Release 1 Sample Adult for Final Survey
wt_x_rad_ad_36_40	double %9.0g	sum of wts x_rad_ad_36_40 - Adult was age 36 to 40 as of 12/31/07
wt_x_rad_ad_41_45	double %9.0g	sum of wts x_rad_ad_41_45 - Adult was age 41 to 45 as of 12/31/07
wt_x_rad_ad_46_50	double %9.0g	sum of wts x_rad_ad_46_50 - Adult was age 46 to 50 as of 12/31/07
wt_x_rad_ad_ethrace_black_nh	double %9.0g	sum of wts x_rad_ad_ethrace_black_nh - Adult is African-American non-Hispanic
wt_x_rad_ad_ethrace_hisp	double %9.0g	sum of wts x_rad_ad_ethrace_hisp - Adult is Hispanic [any race]
wt_x_rad_ad_le_35	double %9.0g	sum of wts x_rad_ad_le_35 - Adult was age 35 or younger as of 12/31/07
wt_x_rad_ad_male	double %9.0g	sum of wts x_rad_ad_male - Adult is male
x_f_site_balt	float %8.0g	x_f_site_balt - Baltimore Site
x_f_site_bos	float %8.0g	x_f_site_bos - Boston Site
x_f_site_chi	float %8.0g	x_f_site_chi - Chicago Site
x_f_site_la	float %8.0g	x_f_site_la - LA Site

Appendix B – Variable List for the Pseudo-Individual PUF File
(mto_sci_puf_pseudo_20130206.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 J, which includes detailed documentation of the outcome and mediator measures.

obs: 3,273
Vars: 94
size: 1,230,648

Variable Name	Type	Variable Label
cell_id	float %9.0g	cell_id - number id assigned to cell
mn_f_wt_totsvy	float %9.0g	mn_f_wt_totsvy - average of f_wt_totsvy for obs in cell
predsg_perpov_dw_z	float %9.0g	predsg_perpov_dw_z - predicted (on site-group) dw tract share poor
predsg_pminorty_dw_z	float %9.0g	predsg_pminorty_dw_z - predicted (on site-group) dw tract share minority
ps_cov_hous_movapt	float %9.0g	pseudo obs cov_Hous_movapt - 1st/2nd reason for wanting to move was better apt.
ps_cov_hous_movjob	float %9.0g	pseudo obs cov_Hous_movapt - 1st/2nd reason for wanting to move was to find job
ps_f_c9010t_perpov_dw	float %9.0g	pseudo obs f_c9010t_perpov_dw - Duration-weighted tract poverty
ps_f_c9010t_perpov_dw_und20	float %9.0g	pseudo obs f_c9010t_perpov_dw_und20 - Duration-weighted tract poverty rate < 20%
ps_f_c9010t_perpov_dw_und30	float %9.0g	pseudo obs f_c9010t_perpov_dw_und30 - Duration-weighted tract poverty rate < 30%
ps_f_c9010t_perpov_dw_und40	float %9.0g	pseudo obs f_c9010t_perpov_dw_und40 - Duration-weighted tract poverty rate < 40%
ps_f_c9010t_perpov_dw_z	float %9.0g	pseudo obs f_c9010t_perpov_dw_z - Duration-wgtd poverty, MTO control z-score
ps_f_c9010t_perpov_dw_zc00t	float %9.0g	pseudo obs f_c9010t_perpov_dw_zc00t - Duration-wgtd poverty, US tracts z-score
ps_f_c9010t_pminorty_dw	float %9.0g	pseudo obs f_c9010t_pminorty_dw - Duration-weighted tract share minority
ps_f_c9010t_pminorty_dw_z	float %9.0g	pseudo obs f_c9010t_pminorty_dw_z - Duration-weighted share minority, z-score
ps_f_ec_idx_z_ad	float %9.0g	pseudo obs f_ec_idx_z_ad - Economic Self-Sufficiency Index, z-score
ps_f_em_emp_ad	float %9.0g	pseudo obs f_em_emp_ad - Adult was employed or temporarily absent last week
ps_f_hc_hsgprb_fix	float %9.0g	pseudo obs f_hc_hsgprb_fix - Number of housing problems (0-7)
ps_f_in_selfsuf_ad	float %9.0g	pseudo obs f_in_selfsuf_ad - Employed and not on TANF
ps_f_in_tanf_fam	float %9.0g	pseudo obs f_in_tanf_fam - Adult or her/his children are currently on TANF
ps_f_mh_calm_ad	float %9.0g	pseudo obs f_mh_calm_ad - Adult was calm/peaceful most of the time in past month
ps_f_mh_dep_y_ad	float %9.0g	pseudo obs f_mh_dep_y_ad - DSM-IV Major Depressive Episode in past year
ps_f_mh_gad_y_ad	float %9.0g	pseudo obs f_mh_gad_y_ad - DSM-IV Generalized Anxiety Disorder in past year
ps_f_mh_idx_z_ad	float %9.0g	pseudo obs f_mh_idx_z_ad - Absence of Mental Health Problems Index, z-score
ps_f_ph_idx_fix_z_ad	float %9.0g	pseudo obs f_mh_idx_z_ad - Absence of Physical Health Problems Index, z-score
ps_f_mh_k6_raw_ad	float %9.0g	pseudo obs f_mh_k6_raw_ad - Psychological Distress Index (K6) raw score (0-24)
ps_f_nb_safe_unsafday_ad	float %9.0g	pseudo obs f_nb_safe_safday_ad - Feels unsafe/very unsafe near home during day
ps_f_ph_asma_y_ad	float %9.0g	pseudo obs f_ph_asma_y_ad - Asthma attack or wheezing in the past 12 months
ps_f_ph_bmi_obese_srm_ad	float %9.0g	pseudo obs f_ph_bmi_obese_srm_ad - Obese adult (including self-reports): BMI>=30
ps_f_ph_bp_hi	float %9.0g	pseudo obs f_ph_bp_hi - High blood pressure: Systolic>=140 or Diastolic>=90
ps_f_ph_habit_sleep_78hrs_ad	float %9.0g	pseudo obs f_ph_habit_sleep_78hrs_ad - Adult slept 7-8 hours last night
ps_f_ph_hlth_fair_ad	float %9.0g	pseudo obs f_ph_hlth_fair_ad - Adult has fair or poor health
ps_f_ph_limit_liftstair	float %9.0g	pseudo obs f_ph_limit_liftstair - Health limits lift/stair climb a little/a lot

Appendix B, continued

Variable Name	Type	Variable Label
ps_f_sn_monit_graffiti_ad	float %9.0g	pseudo obs f_sn_monit_graffiti_ad - Likely/very likely to report graffiti
ps_f_sn_net_anyfrndgrad_ad	float %9.0g	pseudo obs f_sn_net_anyfrndgrad_ad - 1+ friends who graduated from college
ps_f_spl_moves_n	float %9.0g	pseudo obs f_spl_moves_n - Number of moves according to spell file addresses
ps_f_svy_cmove	float %9.0g	pseudo obs f_svy_cmove - Core Mover/Treatment Compliance Flag (1=Moved)
ps_happy_scale123_ad	float %9.0g	pseudo obs happy_scale123_ad - Happiness scale score (1=not too/2=pretty/3=very)
ps_happy_scale123_z_ad	float %9.0g	pseudo obs happy_scale123_z_ad - Happiness scale score, z-score
ps_happy_very_happy_ad	float %9.0g	pseudo obs happy_very_happy_ad - Very (vs. pretty/not too) happy on 3-pt scale
ps_happy_verypretty_happy_ad	float %9.0g	pseudo obs happy_verypretty_happy_ad - Very/pretty (vs. not too) happy 3pt scale
ps_rad_c9010t_perpov_m08	float %9.0g	pseudo obs rad_c9010t_perpov_m08 - Tract poverty at address as of 5/31/08
ps_rad_c9010t_perpov_yr1	float %9.0g	pseudo obs rad_c9010t_perpov_yr1 - Tract poverty 1 year post-randomization
ps_rad_c9010t_perpov_yr5	float %9.0g	pseudo obs rad_c9010t_perpov_yr5 - Tract poverty 5 years post-randomization
ps_rad_c9010t_pminorty_m08	float %9.0g	pseudo obs rad_c9010t_pminorty_m08 - Tract share minority as of 5/31/08
ps_rad_in_govt2009	float %9.0g	pseudo obs rad_in_govt2009 - Gov't receipts income at final eval. (2009 dollars)
ps_rad_in_head2009	float %9.0g	pseudo obs rad_in_head2009 - Adult indiv. earnings at final eval. (2009 dollars)
ps_rad_svy_bl_totincm_2009d	float %9.0g	pseudo obs rad_svy_bl_totincm_2009d - Baseline household income (2009 dollars)
ps_x_f_ad_edged	float %9.0g	pseudo obs x_f_ad_edged - At baseline, adult had a GED
ps_x_f_ad_edgradhs	float %9.0g	pseudo obs x_f_ad_edgradhs - At baseline, adult had completed high school
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_edinsch	float %9.0g	pseudo obs x_f_ad_edinsch - At baseline, adult was enrolled in school
ps_x_f_ad_nevmarr	float %9.0g	pseudo obs x_f_ad_nevmarr - At baseline, adult had never been married
ps_x_f_ad_parentu18	float %9.0g	pseudo obs x_f_ad_parentu18 - Adult was less than age 18 at birth of 1st child
ps_x_f_ad_working	float %9.0g	pseudo obs x_f_ad_working - At baseline, adult was working for pay
ps_x_f_hh_afdc	float %9.0g	pseudo obs x_f_hh_afdc - At baseline, household was receiving AFDC/TANF
ps_x_f_hh_car	float %9.0g	pseudo obs x_f_hh_car - At baseline, household owned a car
ps_x_f_hh_disabl	float %9.0g	pseudo obs x_f_hh_disabl - At baseline, a household member had a disability
ps_x_f_hh_noteens	float %9.0g	pseudo obs x_f_hh_noteens - No teens (ages 13-17) in baseline household
ps_x_f_hh_size2	float %9.0g	pseudo obs x_f_hh_size2 - Baseline household size is 2 or smaller
ps_x_f_hh_size3	float %9.0g	pseudo obs x_f_hh_size3 - Baseline household size equals 3
ps_x_f_hh_size4	float %9.0g	pseudo obs x_f_hh_size4 - Baseline household size equals 4
ps_x_f_hh_victim	float %9.0g	pseudo obs x_f_hh_victim - Household member victimized in 6 mos. before baseline
ps_x_f_hood_5y	float %9.0g	pseudo obs x_f_hood_5y - At baseline, respondent had been living in nbhd. 5+ yrs
ps_x_f_hood_chat	float %9.0g	pseudo obs x_f_hood_chat - At baseline, respondent chatted w/ neighbor >=1x/week
ps_x_f_hood_nbrkid	float %9.0g	pseudo obs x_f_hood_nbrkid - At baseline, very likely to tell on neighbor's kid
ps_x_f_hood_nofamily	float %9.0g	pseudo obs x_f_hood_nofamily - At baseline, no family living in neighborhood
ps_x_f_hood_nofriend	float %9.0g	pseudo obs x_f_hood_nofriend - At baseline, no friends living in neighborhood
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, very dissatisfied with neighborhood
ps_x_f_hous_fndapt	float %9.0g	pseudo obs x_f_hous_fndapt - Baseline respondent very sure of finding an apt.
ps_x_f_hous_mov3tm	float %9.0g	pseudo obs x_f_hous_mov3tm - Respondent had moved >3x in 5 years before baseline
ps_x_f_hous_movdrgs	float %9.0g	pseudo obs x_f_hous_movdrgs - 1st/2nd reason for wanting to move was drugs/crime
ps_x_f_hous_movschl	float %9.0g	pseudo obs x_f_hous_movschl - 1st/2nd reason for wanting to move was good school
ps_x_f_hous_sec8bef	float %9.0g	pseudo obs x_f_hous_sec8bef - At baseline, had applied for Section 8 before
ps_x_f_release1	float %9.0g	pseudo obs x_f_release1 - Release 1 Sample Adult for Final Survey
ps_x_rad_ad_36_40	float %9.0g	pseudo obs x_rad_ad_36_40 - Adult was age 36 to 40 as of 12/31/07
ps_x_rad_ad_41_45	float %9.0g	pseudo obs x_rad_ad_41_45 - Adult was age 41 to 45 as of 12/31/07
ps_x_rad_ad_46_50	float %9.0g	pseudo obs x_rad_ad_46_50 - Adult was age 46 to 50 as of 12/31/07
ps_x_rad_ad_ethrace_black_nh	float %9.0g	pseudo obs x_rad_ad_ethrace_black_nh - Adult is African-American non-Hispanic

Appendix B, continued

Variable Name	Type	Variable Label
ps_x_rad_ad_ethrace_hisp	float %9.0g	pseudo obs x_rad_ad_ethrace_hisp - Adult is Hispanic [any race]
ps_x_rad_ad_le_35	float %9.0g	pseudo obs x_rad_ad_le_35 - Adult was age 35 or younger as of 12/31/07
ps_x_rad_ad_male	float %9.0g	pseudo obs x_rad_ad_male - Adult is male
ra_group VALUE LABEL IS: lbl_group	float %21.0	ra_group - Randomization group (1=Low-PovVouch/Exp, 2=TradVouch/Sec8, 3=Control)
ra_grp_control	float %8.0g	ra_grp_control - Control group flag
ra_grp_exp	float %8.0g	ra_grp_exp - Low-Poverty Voucher (Experimental) group flag
ra_grp_s8	float %8.0g	ra_grp_s8 - Traditional Voucher (Section 8) group flag
ra_poolgrp_exps8	float %8.0g	ra_poolgrp_exps8 - Low-Pov. (Exp)/Traditional Voucher (Sec8) group flag
ra_site VALUE LABEL IS: lbl_site	float %15.0g	ra_site - Site (1=Baltimore/2=Boston/3=Chicago/4=Los Angeles/5=New York City)
sgx_rasite_3g_all_nyc	float %8.0g	sgx_rasite_3g_all_nyc - Site-group interaction flag - New York site
tmp_order_incell	float %9.0g	tmp_order_incell - order of pseudo records within cell
x_f_site_balt	float %8.0g	x_f_site_balt - Baltimore Site
x_f_site_bos	float %8.0g	x_f_site_bos - Boston Site
x_f_site_chi	float %8.0g	x_f_site_chi - Chicago Site
x_f_site_la	float %8.0g	x_f_site_la - LA Site

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

Code Snippets (from 03_sci_cellpuf_colexp_20130206.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 ra_poolgrp_exps8

* OUTCOME/MEDIATOR measures *
* subjective well-being measures
global hpyvars happy_scale123_z_ad happy_scale123_ad happy_verypretty_happy_ad happy_very_happy_ad

* share poor in tract
global povvars rad_c9010t_perpov_yr1 rad_c9010t_perpov_yr5 rad_c9010t_perpov_m08 f_c9010t_perpov_dw
f_c9010t_perpov_dw_zc00t f_c9010t_perpov_dw_z f_c9010t_perpov_dw_und20 f_c9010t_perpov_dw_und30
f_c9010t_perpov_dw_und40

* share minority in tract
global minovars rad_c9010t_pminorty_m08 f_c9010t_pminorty_dw f_c9010t_pminorty_dw_z
* miscellaneous mobility, neighborhood, housing, and social network measures
global miscvars f_spl_moves_n f_nb_safe_unsafday_ad f_hc_hsgprb_fix f_sn_monit_graffiti_ad f_sn_net_anyfrndgrad_ad
* index measures
global idxvars f_ec_idx_z_ad f_ph_idx_fix_z_ad f_mh_idx_z_ad
* economic index components
global ecvars f_in_selfsuf_ad f_em_emp_ad rad_in_head2009 f_in_tanf_fam rad_in_govt2009
* physical health index components
global phvars f_ph_hlth_fair_ad f_ph_asma_y_ad f_ph_bmi_obese_srm_ad f_ph_bp_hi f_ph_limit_liftstair
* mental health index components
global mhvars f_mh_k6_raw_ad f_mh_dep_y_ad f_mh_gad_y_ad f_mh_calm_ad f_ph_habit_sleep_78hrs_ad
* create COMBINED list of outcome and mediator variables needed for Science
global outcvar $hpyvars $povvars $minovars $miscvars $idxvars $ecvars $phvars $mhvars

* site-group predicted variables
```

```
global predsgvar predsg_perpov_dw_z predsg_pminorty_dw_z

* additional baseline measures *
global blvar cov_hous_movapt cov_hous_movjob rad_svy_bl_totincm_2009d

* covariates 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_working x_f_ad_edged x_f_ad_edgradhs x_f_ad_edgradhs_miss x_f_ad_edinsch x_f_ad_nevmarr
x_f_ad_parentu18 ///
                  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

...


*****************************
**** Step 4. Prepare data for collapsing by creating additional variables
*****************************

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

cell_id,
raw count of obs,
sum of cell weights,
mean and standard deviation and sum of weights for outcome and mediator variables,
means for site-group predicted neighborhood characteristics (such as poverty and share minority),
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.

*/

**** 4a. 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'
}


**** 4c. Predict share poor and shaer minority using site-group **

** poverty prediction
regress f_c9010t_perpov_dw_z $sgx_ex $sgx_s8 x_f_site_balt x_f_site_bos x_f_site_chi x_f_site_la [pw = $wtvar]
predict predsg_perpov_dw_z if e(sample)
label var predsg_perpov_dw_z "predsg_perpov_dw_z - predicted (on site-group) dw tract share poor"

** minority prediction
regress f_c9010t_pminorty_dw_z $sgx_ex $sgx_s8 x_f_site_balt x_f_site_bos x_f_site_chi x_f_site_la [pw = $wtvar]
predict predsg_pminorty_dw_z if e(sample)
label var predsg_pminorty_dw_z "predsg_pminorty_dw_z  - predicted (on site-group) dw tract share minority"

**** 4d.  Save full variable labels

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


...

*******************************
**** Step 6. Collapse the data by cell_id 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 $predsgvar $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_* $predsgvar $sitecovs $categvar $treatvar
(sd) sd_*
(count) cell_numobs = $wtvar
(rawsum) cell_sumwgt = $wtvar wt_* [aw=$wtvar], by(cell_id);
#delimit cr;


*******************************
**** Step 7. 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 "cell_id - 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 $predsgvar {
global vname = "lb`var'"
label var `var' "`${vname}'"
}


* create value labels for Site
label define lbl_site 1 "1=Baltimore" 2 "2=Boston" 3 "3=Chicago" 4 "4=Los Angeles" 5 "5=New York City"
label values ra_site lbl_site
tab ra_site

* create value labels for Treatment Group
label define lbl_group 1 "1=Low Pov Voucher" 2 "2=Traditional Voucher" 3 "3=Control Group"
label values ra_group lbl_group
tab ra_group


*****************************
**** Step 8. 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 10. Save final clean version of collapsed file
*****************************

des, fullname
sum
save $puf_cell, replace


*********************************
**** Step 11. Create Variables Needed to Expand File:  average weight, observation counts, and variance measures that will be needed for Expanding the File
*********************************

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

**** 12a. 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

**** 12b.  create a variable that is the record number (order) of each expanded observation within that cell
sort cell_id
by cell_id: gen tmp_order_incell = _n
label var tmp_order_incell "tmp_order_incell - order of pseudo records within cell"

**** 12c.  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 13. Keep Only the Needed Variables (ps_* and site, group, and predsg vars).
*********************************

keep cell_id tmp_order_incell ps_* mn_$wtvar $categvar $treatvar $predsgvar $sitecovs

*********************************
**** Step 14. Save expanded file as STATA and SAS datasets
*********************************
des, fullname
sum
save $puf_pseudo, replace
```

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

/*

PROGRAM: 10_sci_pseudopuf_means_20130206.do

AUTHOR: Nicholas Potter

Purpose: This program provides code for replicating the Moving to Opportunity final evaluation baseline treatment and control means that are presented in Table 1 of the Science Magazine article by J. Ludwig et al. titled "Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults" (Sept 21, 2012), using a public use pseudodata file access file.

The sample consists of MTO adults interviewed for the 10-15 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 program loops through a list of baseline variables covering: gender, age, race and ethnicity, marital status, education, aid status, income, project site, crime victimization, neighborhood satisfaction, reasons for wanting to move, and whether or not the participant moved.

**** Control Mean:

Control mean is calculated as a weighted mean using the weight f_wt_totsvy for particpants in the control group (ra_group=3).

**** Treatment mean:

Treatment mean is calculated as the weighted mean using f_wt_totsvy as a weight for those participants in the treatment

groups (ra_group=1 or ra_group=2). Treatment group 1 (ra_group=1) received the low poverty voucher, while treatment

group 2 (ra_group=2) received the traditional Section 8 voucher.

**** Significance of the difference:

To estimate the significance of the difference, we ran a probability weighted regression of treatment status on the baseline characteristic.

STEPS:

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

1a. Set options

1b. Set directories for data files and output logs

1c. Set date and identify date of latest files

1d. Set dataset date and name

1e. Set log name

1f. Set line size for output

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

2a. Define weight global

2b. Define baseline variable list

STEP 3: Open Expanded Pseudo Individual Public Use File and recode female

3a. Open data set

3b. recode male covariate to tmp_female

STEP 4: Start results file by printing column header

STEP 5: Start loop to calculate mean

STEP 6: Calculate mean for each variable

6a. Calculate control mean for interviewed adults

6b. Calculate control group standard deviation

6c. Calculate treatment mean for interviewed adults

6d. Calculate treatment group standard deviation

6e. Use mean and standard deviation to calculate difference and p-value

6f. Set significance indicator (P<.10 => *, P<.05 => **, P<.01 => ***)

6g. Print means and signifiance to results file

STEP 7: Close loop

STEP 8: Close results file

STEP 9: Print date and time.

*/

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

*   1a. Set options
set more off
set maxvar 10000

*           1b. Set directories for data files and output logs
* UNIX
global dir ~/mtoproj/m10_data/papers/icpsr_archive/science/puf/
global pgmdir ~/mtoproj/m10_pgm/papers/
global slash /

/* PC
global dir E:\m10_data\papers\science
global pgmdir E:\m10_pgm\papers\
global slash \*/

*           1c. Set date and last data date
** Today's Date (will be appended to results file)
global todaydt 20130206

* data date
global datadt = 20130206

*           1d. Set dataset date and name
global data ${dir}mto_sci_puf_pseudo_${datadt}
```

```
*   1e. Set log name
global logname ${pgmdir}icpsr_archive${slash}science${slash}outputs${slash}10_sci_pseudopuf_means_${todaydt}.csv

*           1f. Set line size for output
set linesize 255

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

*    2a. Define weight global
global wt mn_f_wt_totsvy

*    2b. Define baseline variable list
global grpmean tmp_female ps_x_rad_ad_le_35 ps_x_rad_ad_36_40 ps_x_rad_ad_41_45 ps_x_rad_ad_46_50 ///
   ps_x_rad_ad_ethrace_black_nh ps_x_rad_ad_ethrace_hisp ///
   ps_x_f_ad_nevmarr  ps_x_f_ad_working  ps_x_f_ad_edgradhs  ps_x_f_ad_edged  ps_x_f_hh_afdc ///
   ps_rad_svy_bl_totincm_2009d ///
   x_f_site_balt x_f_site_bos x_f_site_chi x_f_site_la sgx_rasite_3g_all_nyc ///
   ps_x_f_hh_victim  ps_x_f_hood_verydissat ///
   ps_x_f_hous_movdrgs  ps_x_f_hous_movschl  ps_cov_hous_movapt  ps_cov_hous_movjob ///
          ps_f_svy_cmove

* ************************************************************************
* STEP 3: Open Expanded Pseudo Individual Public Use File and recode female
* ************************************************************************
```

```
*   3a. Open data set
use ${data}, clear

*   3b. Recode female from male covariate
gen tmp_female = 1 - ps_x_rad_ad_male

* ************************************************************************
* STEP 4: Start results file by printing column header
* ************************************************************************
file open csvlog using "$logname", write replace
file write csvlog "Variable, c_mean, c_n, exps8_mean, exps8_n, exps8_pv, exps8_sig" _n

* ************************************************************************
* STEP 5: Start loop to calculate mean
* ************************************************************************
foreach X in $grpmean {

* ************************************************************************
* STEP 6: Calculate mean for each variable
* ************************************************************************

*    6a. Calculate control mean for interviewed adults
   summarize `X' [aw=$wt] if ra_group==3
   scalar sc_contr_mean   = r(mean)
   scalar sc_contr_n      = r(N)

*    6c. Calculate treatment mean for interviewed adults
   summarize `X' [aw=$wt] if ra_group!=3
```

```
   scalar sc_exps8_mean   = r(mean)

   scalar sc_exps8_n      = r(N)


*    6e. Use mean and standard deviation to calculate difference and p-value

   regress `X' ra_poolgrp_exps8 [pw=$wt]

   scalar sc_diff_pv = (ttail(e(N) - e(df_m),abs(_b[ra_poolgrp_exps8]/_se[ra_poolgrp_exps8])))*2


*    6f. Set significance indicator (P<.10 => *, P<.05 => **, P<.01 => ***)

   local lc_diff_sig = (cond((sc_diff_pv < .10),"*"," "))

   local lc_diff_sig = (cond((sc_diff_pv < .05),"**","`lc_diff_sig'"))

   local lc_diff_sig = (cond((sc_diff_pv < .01),"***","`lc_diff_sig'"))


*    6g. Print means and signifiance to results file

   file write csvlog "`X' ," %20.3f (sc_contr_mean) "," %20.3f (sc_contr_n) "," %20.3f (sc_exps8_mean) "," %20.3f
(sc_exps8_n) ","

   file write csvlog %20.3f (sc_diff_pv) ", `lc_diff_sig' ," _n

* ****************************************************************************

* STEP 7: Close loop

* ****************************************************************************

}

* ****************************************************************************

* STEP 8: Close log

* ****************************************************************************

file close csvlog

* ****************************************************************************

* STEP 9: 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 Science PUF

/*

PROGRAM: 11_sci_pseudopuf_itttot_20130206.do

AUTHOR: NICHOLAS POTTER

Purpose: This program provides code for roughly replicating the Moving to Opportunity final evaluation impact estimates that were presented in Table 2 and Figure 1 of the Science Magazine article by J. Ludwig et al. titled "Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults" (Sept 21, 2012) using a public use file. As the public use file is derived from cell-level data, the results from Science cannot be reproduced precisely because the only baseline covariates that can be included in the model are the dummies for the MTO sites: Baltimore, Boston, Chicago, and Los Angeles (New York is omitted). In addition, the collapsing of the data means that in some cases the weight applied or the number of observations is slightly off. To reproduce the results more precisely, one can apply to use the MTO individual level restricted access data archived with ICPSR (http://www.icpsr.umich.edu/icpsrweb/ICPSR/index.jsp).

The sample consists of MTO adults interviewed for the 10-15 year evaluation of MTO (n = 3273). The pseudo-individual public use dataset for running this program can be obtained at http://www.nber.org/mtopuf (and later at ICPSR).

The program loops through a list of outcome and mediator variables covering: census tract share poor and share minority at different time points, number of moves, and self-reports about neighborhood and housing conditions. In addition to the items reported in Table 2 of the Science article, the program also includes each component of the outcome indices and additional happiness cutpoints.

After calculating estimates for each outcome/mediator, the program creates a confidence interval graph using the estimates for four indices: happiness, physical health, mental health, and an economic index.

For each outcome/mediator the program estimates:

**** Control Mean:

mean of the independent variable, weighted by mn_f_wt_totsvy for the MTO control group (ra_group = 3)

**** Intent to Treat (ITT) effect:

The ITT effect or the impact of being offered an MTO housing voucher (either a low-poverty voucher or traditional voucher) is estimated using a linear regression with the outcome/mediator as the dependent variable and the key independent variable being the dummy variable indicating assignment to a housing voucher group (ra_group = 1 or 2) and controlling for randomization site (x_f_site*) and applying a probability weight (mn_f_wt_totsvy).

The ITT standard error is a robust standard error (Huber-White).

**** Treatment on the treated (TOT) effects estimated using the Bloom method:

TOT = ITT / share of compliers

TOT Standard Error = ITT SE / share of compliers

Share of compliers is weighted average of the compliance variable: ps_f_svy_cmove.

Note that the TOT calculation is described in the Science paper's supplemental materials but TOT effects are not shown in the paper.

**** ITT p-value and significance:

P-values are calculated using a two-sided T-test of the ITT estimate using the standard error and the degrees of freedom of the regression.

Significance is denoted: * = p < .10 , ** = p < .05, *** = p <.01

**** Analytic N

Reported Ns are the count of the number of adults included in the ITT estimate for each dependent variable (calculated as the number of records with non-missing data for that outcome on the pseudo file).

STEPS:

STEP 1: Set options, program directories, and names of input file and output file.

STEP 2: Set globals with the weight variable and a list of outcome variables to analyze

STEP 3: Open MTO Science paper pseudo-individual dataset

STEP 4: Start file for results by writing column headers.

STEP 5: Create ITT, TOT, and control mean and save results to file.

5a. Start loop through list of outcome/mediator variables using "foreach"

5b. Calculate ITT impact using linear regression model

5c. Store ITT results as scalars or local variables, calculate p-value, and set significance symbols

5d. Calculate weighted control mean for the outcome and store value.

5e. Calculate the TOTs using the BLOOM method

5f. Write estimated results for the outcome/mediator to file

5g. drop sample and scalars (locals will clear with next loop)

5h. End variable loop with closing bracket

STEP 6: Close result output file.

STEP 7: Import result output file as dataset and code variables

7a. Import results file

7b. Keep only ITT results to be graphed: happiness, mental health, physical health, and economic indices

7c. Create counter for graphing each outcome

7d. Create high and low confidence interval values

7e. Print graphing values

STEP 8: Generate Confidence Inverval Graph (Figure 1)

8a. Set font Helvetica

8b. Set globals to define line widths

8c. Create graph

8d. Save graph

STEP 9: Display program run date.

*/

```
* ************************************************************************
* STEP 1: Set options, program directories, and names of input file and output comma separated (CSV) file.
* ************************************************************************
clear
set maxvar 10000
set linesize 255
set more off
```

```
* !!!! USER MUST UPDATE DIRECTORY OF DATA AND OUTPUT FILE AND DATE.

/*
* PC version
global pgmroot D:\
global datadir G:\papers\science\
*/

* Unix version
* Directory containing the Restricted Access individual data file:
global datadir ~/mtoproj/m10_data/papers/icpsr_archive/science/puf/

* Directory to write output file to:
global outdir ~/mtoproj/m10_pgm/papers/icpsr_archive/science/outputs/


* set global for dataset
global science  ${datadir}mto_sci_puf_pseudo_20130206

* Set today's date for dating the log
global todaydt 20130206

* log name
global logname ${outdir}/11_sci_pseudopuf_itttot_pool_${todaydt}.csv

* graph name
global graphname 11_sci_pseudopuf_fig1_ci_graph_${todaydt}
```

```
* ****************************************************************
* STEP 2: Set globals with the weight variable and a list of outcome
*       variables to analyze
* ****************************************************************

* set weight variable
global weight mn_f_wt_totsvy

* set global 'yvars' with outcome/mediator variables for which to calculate ITTs and TOTs

global yvars ps_f_c9010t_perpov_dw ps_f_c9010t_perpov_dw_zc00t ps_f_c9010t_perpov_dw_z ///
   ps_f_c9010t_perpov_dw_und20 ps_f_c9010t_perpov_dw_und30 ps_f_c9010t_perpov_dw_und40 ///
   ps_rad_c9010t_perpov_yr1 ps_rad_c9010t_perpov_yr5 ps_rad_c9010t_perpov_m08 ps_f_c9010t_pminorty_dw
ps_rad_c9010t_pminorty_m08 ///
   ps_f_spl_moves_n ps_f_hc_hsgprb_fix ps_f_sn_monit_graffiti_ad ///
   ps_f_sn_net_anyfrndgrad_ad ps_f_nb_safe_unsafday_ad ///
   ps_happy_scale123_z_ad ps_happy_very_happy_ad ps_happy_verypretty_happy_ad ps_happy_scale123_ad ///
   ps_f_mh_idx_z_ad ps_f_ph_idx_fix_z_ad ps_f_ec_idx_z_ad ///
   ps_f_mh_dep_y_ad ps_f_mh_gad_y_ad ps_f_mh_calm_ad ps_f_mh_k6_raw_ad   ps_f_ph_habit_sleep_78hrs_ad ///
   ps_f_ph_hlth_fair_ad ps_f_ph_asma_y_ad ps_f_ph_bmi_obese_srm_ad ps_f_ph_bp_hi ps_f_ph_limit_liftstair ///
   ps_f_em_emp_ad ps_rad_in_govt2009 ps_rad_in_head2009 ps_f_in_selfsuf_ad ps_f_in_tanf_fam

* display the variables assigned to the global yvars
di "$yvars"
```

```
* ****************************************************************
* STEP 3: Open MTO Science paper pseudo-individual dataset
* ****************************************************************

use $science, clear

* check that number of observations is 3273
count

* describe the data
desc

* ****************************************************************
* STEP 4: Start file for results by writing column headers.
* ****************************************************************
* open file for writing the results
file open csvlog using "${logname}", write replace

* write column headers to file
file write csvlog "lookup, variable, model, cmean, exps8_itt, exps8_ittsig, exps8_ittse, exps8_ittpv, exps8_tot, exps8_totsig, exps8_totse, exps8_totpv, N" _n

* replace mn_f_svy_cmove = (mn_f_svy_cmove>=0.5)
```

```
* ****************************************************************
* STEP 5: Create ITT, TOT, and control mean and save results to file.
* ****************************************************************

* 5a.  Start loop through list of outcome/mediator variables using 'foreach'
foreach y in $yvars {
    * where yvars is the list of outcome/mediator variables defined above
    * 5b.  Calculate ITT impact using linear regression model
    * regress variable against the treatment group indicator and controlling for randomization site
    regress `y' ra_poolgrp_exps8 x_f_site_balt x_f_site_bos x_f_site_chi x_f_site_la [pw=$weight]

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

    * save ITT estimate
    scalar sc_exps8_itt   = _b[ra_poolgrp_exps8]

    * save robust standard error (for ITT)
    scalar sc_exps8_ittse = _se[ra_poolgrp_exps8]

    * analytic N
    scalar sc_n = e(N)

    * degrees of freedom
    scalar sc_df_r = e(df_r)

    * identify analytic sample
    capture drop sample
    gen sample = e(sample)
```

```
*** Calculate p value
scalar sc_exps8_ittpv = ttail( sc_df_r, abs( sc_exps8_itt / sc_exps8_ittse ))*2

*** Set significance symbols
if sc_exps8_ittpv < .01 {
   local lc_exps8_ittsig ***
}
else if sc_exps8_ittpv < .05 {
        local lc_exps8_ittsig **
}
else if sc_exps8_ittpv < .1 {
        local lc_exps8_ittsig *
}
else {
        local lc_exps8_ittsig
}

* 5d.  Calculate weighted control mean for the outcome and store value.
sum `y' [aw=$weight] if ra_group==3

* set scalar to control mean
scalar sc_cmean = r(mean)

* 5e.  Calculate the TOTs using the BLOOM method
* first calculate the mean compliance value for respondents with valid values on the dependent variable
sum ps_f_svy_cmove [aw=$weight] if (ra_group==1 | ra_group==2) & sample==1
```

```
    * set scalar to compliance mean
    scalar sc_exps8_cmove = r(mean)

    * calculate TOT effect and standard error using Bloom method where TOT = impact/share compliers
    scalar sc_exps8_tot = sc_exps8_itt / sc_exps8_cmove
    scalar sc_exps8_totse = sc_exps8_ittse / sc_exps8_cmove

    * now get control mean
    sum `y' [aw=$weight] if ra_group==3 & sample==1

    * 5f.  Write estimated results for the outcome/mediator to comma-separated file
    file write csvlog "`y'_Adults_pool , `y', _Adults_pool , " %20.3f (sc_cmean) "," ///
    %20.3f (sc_exps8_itt) ", `lc_exps8_ittsig'," %20.3f (sc_exps8_ittse) "," %20.3f (sc_exps8_ittpv) "," ///
    %20.3f (sc_exps8_tot) ", `lc_exps8_ittsig'," %20.3f (sc_exps8_totse) "," %20.3f (sc_exps8_ittpv) ","  ///
    %20.0f (sc_n) "," _n

    * 5g.  drop sample and scalars (locals will clear with next loop)
    drop sample
    scalar drop _all

* 5h.  End variable loop with closing bracket
}

* ****************************************************************
* STEP 6: Close result output file.
* ****************************************************************
file close csvlog
```

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

* STEP 7: Import result output file as dataset and code variables

* ****************************************************************

* 7a.  Import results file

clear

insheet using "${logname}"

cd ${outdir}graphs/


* 7b.  Keep only ITT results to be graphed: happiness, mental health, physical health, and economic indices

keep if (lookup=="ps_f_ec_idx_z_ad_Adults_pool" | lookup=="ps_f_mh_idx_z_ad_Adults_pool" |
lookup=="ps_f_ph_idx_fix_z_ad_Adults_pool" | lookup=="ps_happy_scale123_z_ad_Adults_pool")


* list the data to check import

list


* 7c.  Create counter for graphing each outcome

gen outcome=1 if lookup == "ps_f_ec_idx_z_ad_Adults_pool"

replace outcome=2 if lookup == "ps_f_ph_idx_fix_z_ad_Adults_pool"

replace outcome=3 if lookup == "ps_f_mh_idx_z_ad_Adults_pool"

replace outcome=4 if lookup == "ps_happy_scale123_z_ad_Adults_pool"


* 7d.  Create high and low confidence interval values

rename exps8_itt itt

gen high = itt+1.96*exps8_ittse

gen low =  itt-1.96*exps8_ittse


* 7e.  Print graphing values
```

```
list lookup outcome itt exps8_ittse high low


* ****************************************************************

* STEP 8: Generate Confidence Inverval Graph (Figure 1)

* ****************************************************************

* 8a.  Set font Helvetica

graph set eps fontface Helvetica


* 8b.  Set globals to define line widths and label outcomes

capture label drop outcome_lab

label define outcome_lab 1 "Economic self-sufficiency" 2 "Physical health index" 3 "Mental health index" 4 "Subjective well-being"

label values outcome outcome_lab


global line_thick "lcolor(black) color(black) lwidth(medthin)"

global line_thin "lcolor(gray) color(gray) lwidth(medthin)"


global marker_thick "msymbol(square) msize(huge) mcolor(black) color(black) mfcolor(black)"

global marker_thin "msymbol(circle) msize(huge)  mcolor(gray) mfcolor(gray)"


global marker_thick2 "msymbol(square) msize(medium) mcolor(black) color(black) mfcolor(black)"

global marker_thin2 "msymbol(circle) msize(medium)  mcolor(gray) mfcolor(gray)"


* 8c.  Create graph

#delimit;

twoway

/* draw markers for the ITT coefficient point estimate*/

rcapsym itt itt outcome , ${marker_thick2}
```

```
/* draw confidence intervals */
|| rspike low high outcome , ${line_thick}
/* draw top caps on confid interval*/
|| rcap high high outcome , ${line_thick} msize(8)
/* draw bottom caps on confid interval*/
|| rcap low low outcome , ${line_thick} msize(8)
,
/* set graph options */
name(figure1, replace)
/* set size of graph in inches */
xsize(6) ysize(7)
/* tick marks for specific range of itt coefficients*/
ytick(-.15(.05).20, format(%3.2f) nogrid)
ytitle("Treatment-control differences in outcome (z-score)", height(4))
yline(0, lcolor(black) lpattern(shortdash))
ylabel(-.15(.05).20, format(%3.2f) labsize(small) nogrid)
xlabel(1 `" `"Economic"' `"self-"' `"sufficiency"' "' 2 `" `"Physical"' `"health"' `"index"' "' 3 `" `"Mental"' `"health"' `"index"' "'
        4 `" `"Subjective"' `"well-being"' "' , labsize(small)) xscale(range(0(1)5)) xtitle("")
legend(off)
graphregion(fcolor(white) lcolor(white))  plotregion(lcolor(black));

#delimit cr;

* 8d.  Save graph
graph save ${graphname}.gph, replace

* save eps version
graph export ${graphname}.eps, replace
```

```
* convert to pdf
!epstopdf ${graphname}.eps


* ****************************************************************
* STEP 9: Display program run date.
* ****************************************************************
disp "Program End on " c(current_date) " at " c(current_time)
```

Appendix F – Stata Code for Estimating the Relationship Between Outcomes and Neighborhood Conditions Using Site-Group Instruments Using the Science PUF

/*

PROGRAM: 12_sci_pseudopuf_ivs_20130206.do

AUTHOR: NICHOLAS POTTER

Purpose: This program provides code for roughly replicating the Moving to Opportunity final evaluation impact estimates that were presented in the Science Magazine article by J. Ludwig et al. titled "Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults" (Sept 21, 2012) using a public use file. As the public use file is derived from cell-level data, the results from Science cannot be reproduced precisely because the only baseline covariates that can be included in the model are the dummies for the MTO sites: Baltimore, Boston, Chicago, and Los Angeles (New York is omitted). In addition, the collapsing of the data means that in some cases the weight applied or the number of observations is slightly off. To reproduce the results more precisely, one can apply to use the MTO individual level restricted access data archived with ICPSR (http://www.icpsr.umich.edu/icpsrweb/ICPSR/index.jsp).

The sample consists of MTO adults interviewed for the 10-15 year evaluation of MTO (n = 3273). The dataset for running this program can be obtained at http://www.nber.org/mtopuf and eventually through ICPSR.

The program uses the predicted values of share poor and share minority that are the result of a regression of the mediator against site covariates interacted with treatment group. Using the predicted values, this program runs the 2nd stage of a Two-Stage Least Squares Model to roughly replicate the individual level instrumental variable regressions shown in the Science paper. We model share the relationship between subjective well-being and the two tract characteristics separately and then estimate the relationship to share poor controlling for share minority and then share minority controlling for share poor. We then estimate the significance level of the coefficients. The program also creates a csv table with the following results: mediator, coefficient, robust standard error, p-value, and significance indicator.

STEPS:

STEP 1: Set options, program directories, and names of input file and output file.

STEP 2: Set globals with the weight variable, covariates, and instruments

STEP 3: Open MTO Science paper pseudo-individual dataset

STEP 4: Start file for results by writing column headers.

STEP 5: Run Manual 2nd Stage of 2SLS Model with Site-Group as the Instruments of Relationship Between Nghd Poverty/Minority Share and Subjective Well-Being Characteristics and Subjective Well-Being

5a. 2nd Stage of 2SLS regression of share poor against subjective well-being

5b. Calculate poverty p-value and significance

```
        5c.  Write regression results to csv file

        5d.  2nd Stage of 2SLS regression of share minority against subjective well-being

        5e.  Calculate poverty p-value and significance

        5f.  Write regression results to csv file

                5g.  2nd Stage of 2SLS regression of share poor and share minority against subjective well-being

                5h.  Calculate poverty controlling for minority and minority controlling for poverty p-value and
                     significance

        5i.  Calculate Wald test of difference between poverty and minority

        5j.  Write regression results to csv file

    STEP 6: Close result output file.

    STEP 7: Display program run date.

*/

* ****************************************************************

*      STEP 1: Set options, program directories, and names of input file and output comma separated (CSV) file.

* ****************************************************************

clear

set maxvar 10000

set linesize 255

set more off

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

/*

* PC version

global pgmroot D:\

global datadir G:\papers\science\

*/

* Unix version

* Directory containing the Restricted Access individual data file:

global datadir ~/mtoproj/m10_data/papers/icpsr_archive/science/puf/
```

```
* Directory to write output file to:
global outdir ~/mtoproj/m10_pgm/papers/icpsr_archive/science/outputs/

* set global for dataset
global sci_pseudodata  ${datadir}mto_sci_puf_pseudo_20130206

* Set today's date for dating the log
global todaydt 20130206

* log name
global logname ${outdir}/12_sci_pseudopuf_ivs_${todaydt}.csv

* ****************************************************************
*       STEP 2: Set globals with the weight variable
* ****************************************************************
* weight variable
global wt mn_f_wt_totsvy

* ****************************************************************
*       STEP 3: Open MTO Science paper pseudo-individual dataset
* ****************************************************************
use $sci_pseudodata, clear

* ****************************************************************
*       STEP 4: Start file for results by writing column headers.
* ****************************************************************
* open file for writing the results
file open csvlog using "${logname}", write replace
```

```
* write column headers to file

file write csvlog "lookup, yvar, xvar, control_var, model, coef, se, pv, sig, test_diff_pv, n" _n


* ****************************************************************

*       STEP 5: Run Manual 2nd Stage of 2SLS Model with Site-Group as the Instruments of Relationship Between Nghd Poverty/Minority Share and Subjective Well-Being

* ****************************************************************

*           5a.  2nd Stage of 2SLS regression of share poor against subjective well-being

regress ps_happy_scale123_z_ad predsg_perpov_dw_z x_f_site_bal x_f_site_bos x_f_site_chi x_f_site_la [pw=$wt]


*           5b.  Calculate poverty p-value and significance

* p value

scalar sc_pov_pv = ttail( e(df_r), abs( _b[predsg_perpov_dw_z] / _se[predsg_perpov_dw_z] ))*2

* set significance

local lc_pov_sig = (cond((sc_pov_pv < .10),"*"," "))

local lc_pov_sig = (cond((sc_pov_pv < .05),"**","`lc_pov_sig'"))

local lc_pov_sig = (cond((sc_pov_pv < .01),"***","`lc_pov_sig'"))


*           5c.  Write regression results to csv file

file write csvlog "2sls_ps_happy_scale123_z_ad_predsg_perpov_dw_z__sitecov, ps_happy_scale123_z_ad, predsg_perpov_dw_z, none, sitecov ," ///

%20.3f (_b[predsg_perpov_dw_z]) "," %20.3f (_se[predsg_perpov_dw_z]) "," %20.3f (sc_pov_pv) ", `lc_pov_sig' ," "n/a," %20.0f (e(N)) _n


*           5d.  2nd Stage of 2SLS regression of share minority against subjective well-being

regress ps_happy_scale123_z_ad predsg_pminorty_dw_z x_f_site_bal x_f_site_bos x_f_site_chi x_f_site_la [pw=$wt]


*           5e.  Calculate minority p-value and significance
```

```
* p value

scalar sc_min_pv = ttail( e(df_r), abs( _b[predsg_pminorty_dw_z] / _se[predsg_pminorty_dw_z] ))*2

* set significance

local lc_min_sig = (cond((sc_min_pv < .10),"*"," "))

local lc_min_sig = (cond((sc_min_pv < .05),"**","`lc_min_sig'"))

local lc_min_sig = (cond((sc_min_pv < .01),"***","`lc_min_sig'"))


*       5f.  Write regression results to csv file

file write csvlog "2sls_ps_happy_scale123_z_ad_predsg_pminorty_dw_z__sitecov, ps_happy_scale123_z_ad,
predsg_pminorty_dw_z, none, sitecov ," ///

%20.3f (_b[predsg_pminorty_dw_z]) "," %20.3f (_se[predsg_pminorty_dw_z]) "," %20.3f (sc_min_pv) ", `lc_min_sig'","
"n/a," %20.0f (e(N)) _n


*       5g.  2nd Stage of 2SLS regression of share poor and share minority against subjective well-being

regress ps_happy_scale123_z_ad predsg_perpov_dw_z predsg_pminorty_dw_z x_f_site_bal x_f_site_bos x_f_site_chi
x_f_site_la [pw=$wt]


*       5h.  Calculate poverty controlling for minority and minority controlling for poverty p-value and significance

** poverty p value

scalar sc_pov_co_min_pv = ttail( e(df_r), abs( _b[predsg_perpov_dw_z] / _se[predsg_perpov_dw_z] ))*2

* set significance

local lc_pov_co_min_sig = (cond((sc_pov_co_min_pv < .10),"*"," "))

local lc_pov_co_min_sig = (cond((sc_pov_co_min_pv < .05),"**","`lc_pov_co_min_sig'"))

local lc_pov_co_min_sig = (cond((sc_pov_co_min_pv < .01),"***","`lc_pov_co_min_sig'"))


** minorty p value

scalar sc_min_co_pov_pv = ttail( e(df_r), abs( _b[predsg_pminorty_dw_z] / _se[predsg_pminorty_dw_z] ))*2

* set significance

local lc_min_co_pov_sig = (cond((sc_min_co_pov_pv < .10),"*"," "))
```

```
local lc_min_co_pov_sig = (cond((sc_min_co_pov_pv < .05),"**","`lc_min_co_pov_sig'"))

local lc_min_co_pov_sig = (cond((sc_min_co_pov_pv < .01),"***","`lc_min_co_pov_sig'"))


*       5i.  Wald test if coefficients are different and assign p-value

test predsg_perpov_dw_z = predsg_pminorty_dw_z


*       5j.  Write regression results to csv file

* poverty controlling for minority

file write csvlog "2sls_ps_happy_scale123_z_ad__predsg_perpov_dw_z__predsg_pminorty_dw_z__sitecov,
ps_happy_scale123_z_ad, predsg_perpov_dw_z, predsg_pminorty_dw_z, sitecov ," ///

%20.3f (_b[predsg_perpov_dw_z]) "," %20.3f (_se[predsg_perpov_dw_z]) "," %20.3f (sc_pov_co_min_pv) ",
`lc_pov_co_min_sig' ," %20.4f (r(p)) "," %20.0f (e(N)) _n


* minority controlling for poverty

file write csvlog "2sls_ps_happy_scale123_z_ad__predsg_pminorty_dw_z__predsg_perpov_dw_z__sitecov,
ps_happy_scale123_z_ad, predsg_pminorty_dw_z, predsg_perpov_dw_z, sitecov ," ///

%20.3f (_b[predsg_pminorty_dw_z]) "," %20.3f (_se[predsg_pminorty_dw_z]) "," %20.3f (sc_min_co_pov_pv) ",
`lc_min_co_pov_sig' ," %20.4f (r(p)) "," %20.0f (e(N)) _n


* ****************************************************************

*       STEP 6: Close result output file.

* ****************************************************************

file close csvlog


* ****************************************************************

*       STEP 7: Display program run date.

* ****************************************************************

disp "Program End on " c(current_date) " at " c(current_time)
```

Appendix G – Stata Code for Graphing the Relationship Between Outcomes and Neighborhood Conditions Using the Science PUF

```
/*
PROGRAM: 13_sci_cellpuf_ivplots_20130206.do
AUTHOR: NICHOLAS POTTER

Purpose: This program provides code for roughly replicating the Moving to Opportunity
final evaluation Figure 2 plots that were presented in the Science Magazine article
by J. Ludwig et al. titled "Neighborhood Effects on the Long-Term Well-Being of
Low-Income Adults" (Sept 21, 2012) using a collapsed cell-level public use file.
In the plots, the points represent a site-treatment group and the slope of the line
represents the estimated relationship (using site-group as instruments) between
subjective well-being and a census tract characteristic (either share poor or share minority).

The figure is a 4 panel plot showing the relationship of:
(A) share poor vs. subjective well-being,
(B) share minority vs. subjective well-being,
(C) share poor vs. subjective well-being controlling for share minority, and
(D) share minority vs. subjective well-being controlling for share poor.

The cell-level dataset used consists of 158 cells.  The cells are aggregated or collapsed data
from the sample of MTO adults interviewed for the 10-15 year evaluation of MTO (n = 3273).
The cells are constructed to be homogenous by site and treatment group. The cell-level file
contains the means (mn_*), standard deviations (sd_*), and weights for each outcome (wt_*).
The key variables used in constructing the plots are:
mn_happy_scale123_z_ad - the cell-average of the z-scored measure of subjective well being
```

mn_f_c9010t_perpov_dw_z - the cell-average of the z-scored measure of duration-weighted tract share poor,

mn_f_c9010t_pminorty_dw_z - the cell-average of the z-scored measure of duration-weighted tract share minority,

treatment group categories (ra_group: 1 = LPV, 2 = TRV, 3 = control),

site categories (ra_site: 1 = Balt, 2 = Bos, 3 = Chi, 4 = LA, 5 = NYC), and

wt cell_sumwgt (sum of weights).

The dataset for running this program can be obtained at http://www.nber.org/mtopuf and will

be available later via ICPSR (http://www.icpsr.umich.edu/icpsrweb/ICPSR/index.jsp).

The program generates the plots by:

i. collapsing the data to site X group (15 categories)

ii. demeaning the data by site (e.g., so the mean value for a site-group is expressed relative to each sites mean)

iii. plotting each panel

iv. combining the panels

More specifically, the STEPS are:

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

1a. Set options

1b. Set directories for data files and output logs

1c. Set date and last data date

1d. Set dataset date and name

1e. Set log name

1f. Set line size for output

1g. Set graph names

STEP 2: Set globals with the weight variable, covariates

STEP 3: Open MTO Science paper cell-level public use file dataset

STEP 4: Collapse data by site-group dummy variables

STEP 5: Calculate values for plotting

5a. Create overal site means for the outcomes and mediators using a regression

5b. Demean the outcomes and mediators by subtracting off the overall site mean

5c. Predict the slope for poverty (std errors will not be correct)

5d. Predict the slope for minority (std errors will not be correct)

5e. Predict the slope of 2-variable effects (poverty controlling for minority) - std errors are not correct

5f. Predict the slope of 2-variable effects (minority controlling for poverty) - std errors are not correct

5g. Create variables with 0 value for graphing at origin

STEP 6: Create scatter plot for poverty vs. subjective well-being

6a. Create label positions

6b. Change the end of line delimiter to allow a longer code for the scatter plot

6c. Create scatter with weighted symbols and titles/

6d. Add another scatter with labels for the points /

6e. Add scatter with a regression line/

6f. Plot the origin and add panel letter (NOTE)

6g. End the scatter command/

6h. Change the delimiter back to a newline/

6i. Save the graph /

STEP 7: Create scatter plot for minority vs. subjective well-being

7a. Create label positions

7b. Change the end of line delimiter to allow a longer code for the scatter plot

7c. Create scatter with weighted symbols and titles/

7d. Add another scatter with labels for the points /

7e. Add scatter with a regression line/

7f. Plot the origin and add panel letter (NOTE)/

7g. End the scatter command/

7h. Change the delimiter back to a newline/

7i. Save the graph /

STEP 8: Create scatter plot of poverty vs. subjective well-being controlling for minority

8a. Create label positions

8b. Change the end of line delimiter to allow a longer code for the scatter plot

8c. Create scatter with weighted symbols and titles/

8d. Add another scatter with labels for the points /

8e. Add scatter with a regression line/

8e1. Use black lines of medium width for regression line/

8f. Plot the origin and add panel letter (NOTE)/

8g. End the scatter command/

8h. Change the delimiter back to a newline/

8i. Save the graph

STEP 9: Create scatter plot of minority vs. subjective well-being, controlling for poverty

9a. Create label positions

9b. Change the end of line delimiter to allow a longer code for the scatter plot

9c. Create scatter with weighted symbols and titles/

9d. Add another scatter with labels for the points /

9e. Add scatter with a regression line/

9f. Plot the origin and add panel letter (NOTE)/

9g. End the scatter command/

9h. Change the delimiter back to a newline/

9i. Save the graph /

STEP 10: Join the 4 panels to create a 4-panel figure

10a. Combine 4 panels

10b. Export as eps

10c. Export as pdf

STEP 11: Open results file to write IV coefficients.

```
    STEP  12:  Write Slopes (IV coefficients) to output file.
                12a.  Print column names to log
                12b.  Write poverty coefficient
                12c.  Write minority coefficient
                12d.  Write poverty controlling for minority coefficient
                12e.  Write minority controlling for poverty coefficient
    STEP  13: Close result output file.
    STEP  14: Display program run date.
*/

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

*   1a. Set options
set more off
set maxvar 10000

*           1b. Set directories for data files and output logs
* UNIX
global dir ~/mtoproj/m10_data/papers/icpsr_archive/science/puf/
global pgmdir ~/mtoproj/m10_pgm/papers/
global slash /
/* PC
global dir E:\m10_data\papers\science
global pgmdir E:\m10_pgm\papers\
global slash \
*/
```

```
*          1c. Set date and last data date

** Today's Date (will be appended to results file)

global todaydt 20130206


* data date

global datadt = 20130206


*          1d. Set name of cell-level PUF dataset to use

global data ${dir}mto_sci_puf_cells_20130206


*   1e. Set log name

global logname ${pgmdir}icpsr_archive${slash}science${slash}outputs${slash}13_sci_cellpuf_ivplotvalues_${todaydt}.csv


*          1f. Set line size for output

set linesize 255


*   1g. Set graph names

global pov_graph
${pgmdir}icpsr_archive${slash}science${slash}outputs${slash}graphs${slash}13_cellpuf_fig2_swbz_vs_povz_dw_panel1_
${todaydt}

global min_graph
${pgmdir}icpsr_archive${slash}science${slash}outputs${slash}graphs${slash}13_cellpuf_fig2_swbz_vs_minz_dw_panel2_
${todaydt}

global pov_ctrl_min_graph
${pgmdir}icpsr_archive${slash}science${slash}outputs${slash}graphs${slash}13_cellpuf_fig2_swbz_vs_povz_dw_panel3_
${todaydt}

global min_ctrl_pov_graph
${pgmdir}icpsr_archive${slash}science${slash}outputs${slash}graphs${slash}13_cellpuf_fig2_swbz_vs_minz_dw_panel4_
${todaydt}
```

```
global final_graph
${pgmdir}icpsr_archive${slash}science${slash}outputs${slash}graphs${slash}13_cellpuf_fig2_4panel_swbz_vs_povz_dw_
${todaydt}


* ************************************************************************

*       STEP  2: Set globals with the weight variable, covariates

* ************************************************************************

* weight variable

global wt cell_sumwgt


* lhs variable

global yvar happy_scale123_z_ad


* site covariates

global xcovs x_f_site_bal x_f_site_bos x_f_site_chi x_f_site_la


* ************************************************************************

*       STEP  3: Open MTO Science paper cell-level public use file dataset

* ************************************************************************

*           3a. Open cell-level PUF data file

use $data, clear


*           3b. Create point labels

gen pt_label = "blank"

** replace for low poverty voucher group and site

replace pt_label = "LPV Bal" if ra_group==1 & ra_site==1

replace pt_label = "LPV Bos" if ra_group==1 & ra_site==2

replace pt_label = "LPV Chi" if ra_group==1 & ra_site==3

replace pt_label = "LPV LA"  if ra_group==1 & ra_site==4
```

```
replace pt_label = "LPV NYC" if ra_group==1 & ra_site==5

** replace for traditional poverty voucher group and site
replace pt_label = "TRV Bal" if ra_group==2 & ra_site==1
replace pt_label = "TRV Bos" if ra_group==2 & ra_site==2
replace pt_label = "TRV Chi" if ra_group==2 & ra_site==3
replace pt_label = "TRV LA"  if ra_group==2 & ra_site==4
replace pt_label = "TRV NYC" if ra_group==2 & ra_site==5

** replace for Control group and site
replace pt_label = "C Bal"   if ra_group==3 & ra_site==1
replace pt_label = "C Bos"   if ra_group==3 & ra_site==2
replace pt_label = "C Chi"   if ra_group==3 & ra_site==3
replace pt_label = "C LA"    if ra_group==3 & ra_site==4
replace pt_label = "C NYC"   if ra_group==3 & ra_site==5

* check that labels are correct for site and group
tab pt_label ra_site, mi
tab pt_label ra_group, mi

* Generate a sample counter
gen tmp_iv_sample = 1

* ************************************************************************
*       STEP  4: Collapse data by site-group dummy variables
* ************************************************************************
collapse ///
```

```
(mean) mn_happy_scale123_z_ad predsg_perpov_dw_z predsg_pminorty_dw_z mn_f_c9010t_perpov_dw_z
mn_f_c9010t_pminorty_dw_z ///

(max)  $xcovs ///

(sum)  tmp_sumwgt = tmp_iv_sample ///

(rawsum) tmp_sumcount = tmp_iv_sample [pw=$wt], by(ra_site ra_group pt_label)

* check total sum weight

sum tmp_sumwgt

total tmp_sumwgt


* ************************************************************************

*       STEP  5: Calculate values for plotting

* ************************************************************************

*           5a.  Create overal site means for the outcomes and mediators using a regression

** predict happiness based on site only.

regress mn_happy_scale123_z_ad $xcovs [pw=tmp_sumwgt]

predict tmp_iv_sitemean_y


** predict poverty based on site only.

regress mn_f_c9010t_perpov_dw_z $xcovs [pw=tmp_sumwgt]

predict tmp_iv_sitemean_pov


** predict minority based on site only.

regress mn_f_c9010t_pminorty_dw_z $xcovs [pw=tmp_sumwgt]

predict tmp_iv_sitemean_min


*           5b.  Demean the outcomes and mediators by subtracting off the overall site mean

*** subtract the site mean from the average outcome for that site-group

** Happiness
```

```
gen double tmp_iv_demean_y = mn_happy_scale123_z_ad - tmp_iv_sitemean_y


** Poverty

gen double tmp_iv_demean_pov = mn_f_c9010t_perpov_dw_z - tmp_iv_sitemean_pov


** Minority

gen double tmp_iv_demean_min = mn_f_c9010t_pminorty_dw_z - tmp_iv_sitemean_min



*          5c.  Predict the slope for a poverty (std errors will not be correct)

*** Poverty

regress tmp_iv_demean_y tmp_iv_demean_pov [pw=tmp_sumwgt]


** Create a scalar with the slope of the poverty graph

scalar mto_slope_pov = _b[tmp_iv_demean_pov]


** Predict outcome usng hte model to plot regression line

predict predict_ypov


** Predict the residuals

predict out_residpov, resid


*          5d.  Predict the slope for minority (std errors will not be correct)

** Minority

regress tmp_iv_demean_y tmp_iv_demean_min [pw=tmp_sumwgt]


** Create a scalar with the slope of the poverty graph

scalar mto_slope_min = _b[tmp_iv_demean_min]
```

```
** Predict outcome usng hte model to plot regression line
predict predict_ymin

** Predict the residuals
predict out_residmin, resid

*          5e.  Predict the slope of 2-variable effects (poverty controlling for minority) - std errors are not correct
*** Poverty controlling for minority
reg tmp_iv_demean_y tmp_iv_demean_min [pw=tmp_sumwgt]
predict resid_y_ctrl_min, resid

reg tmp_iv_demean_pov tmp_iv_demean_min [pw=tmp_sumwgt]
predict resid_pov_ctrl_min, resid

* Regress Outcome (less minority) on poverty (less minority)
* to generate slope of Poverty controlling for Minority
reg resid_y_ctrl_min resid_pov_ctrl_min [pw=tmp_sumwgt]
scalar mto_slope_pov_ctrl_min = _b[resid_pov_ctrl_min]
predict predict_y_ctrl_min

*          5f.  Predict the slope of 2-variable effects (minority controlling for poverty) - std errors are not correct
*** Minority controlling for Poverty
reg tmp_iv_demean_y tmp_iv_demean_pov [pw=tmp_sumwgt]
predict resid_y_ctrl_pov, resid

reg tmp_iv_demean_min tmp_iv_demean_pov [pw=tmp_sumwgt]
predict resid_min_ctrl_pov, resid
```

```
* Regress Outcome (less poverty) on minority (less poverty)
* to generate slope of minority controlling for poverty
reg resid_y_ctrl_pov resid_min_ctrl_pov [pw=tmp_sumwgt]
scalar mto_slope_min_ctrl_pov = _b[resid_min_ctrl_pov]
predict predict_y_ctrl_pov


*          5g.  Create variables with 0 value for graphing at origin
gen tmp_zero1 = 0
gen tmp_zero2 = 0

* ************************************************************************
*      STEP  6: Create scatter plot for poverty vs. subjective well-being
* ************************************************************************
*          6a.  Create label positions
gen tmp_mpos_pov = 3

*          6b.  Change the end of line delimiter to allow a longer code for the scatter plot
#delimit ;

/*          6c.  Create scatter with weighted symbols and titles*/
scatter tmp_iv_demean_y tmp_iv_demean_pov [aw=tmp_sumwgt],

/*             6c1.  Set main title*/
title("{bf: Cell PUF}" "{bf: Subjective well-being versus}" "{bf: share poor}" "{bf: (duration-weighted)}" "", size(medium)
pos(12) ring(1))
```

```
/*          6c2.  Set x and y axis titles*/
ytitle("Subjective well-being" "relative to site overall mean, z-score", height(8))
xtitle("Share poor relative to site overall mean, z-score")

/*          6c3.  Turn off legend*/
legend(off)
/*          6c4.  Set axis tick marks*/
xlabel(-0.4(.20)0.7, format(%2.1f))
ylabel(-0.12(.06)0.24, format(%3.2f) nogrid)
/*          6c5.  Set markers for site-groups*/
mcolor(black)
msymbol(O) msize(*.5)
/*          6c6.  White background with black border*/
graphregion(fcolor(white) lcolor(white)) plotregion(lcolor(black))
/*          6c7.  Set graph height and width in inches*/
ysize(8.25) xsize(6.25)
/*          6d.  Add another scatter with labels for the points */
|| scatter tmp_iv_demean_y tmp_iv_demean_pov [aw=tmp_sumwgt],
/*          6d1.  Set axis tick marks*/
xlabel(-0.4(.20)0.7, format(%2.1f))
ylabel(-0.12(.06)0.24, format(%3.2f) nogrid)

/*          6d2.  Set symbol to empty*/
msymbol(i)
/*          6d3.  Label points with group-site*/
mlabel(pt_label)
/*          6d4.  Set marker labels*/
mlabvposition(tmp_mpos_pov)
```

```
mlabcolor(black)
mlabsize(*0.85 *0.85)

/*          6e.  Add scatter with a regression line*/
|| scatter predict_ypov tmp_iv_demean_pov [aw=tmp_sumwgt],
/*              6e1.  Use black lines of medium width for regression line*/
msymbol(i) c(l)
`mto_clp'
clwidth(medium) clcolor(black) mcolor(black)

/*          6f.  Plot the origin and add panel letter (NOTE)
The purpose of this scatter is to add the note. the scatter of zero vs zero is not displayed */
|| scatter tmp_zero1 tmp_zero2, msymbol(i)
note("A", size(vlarge) pos(11) ring(1) span)

/*          6g.  End the scatter command*/
;

/*          6h.  Change the delimiter back to a newline*/
#delimit cr;

/*          6i.  Save the graph */
graph save "$pov_graph", replace


graph export "${pov_graph}.eps", replace
!epstopdf ${pov_graph}.eps
```

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

*      STEP  7: Create scatter plot for minority vs. subjective well-being

* ************************************************************************

*          7a.  Create label positions

gen tmp_mpos_min = 3


*          7b.  Change the end of line delimiter to allow a longer code for the scatter plot

#delimit ;


/*          7c.  Create scatter with weighted symbols and titles*/

scatter tmp_iv_demean_y tmp_iv_demean_min [aw=tmp_sumwgt],



/*              7c1.  Set main title*/

title("{bf: Cell PUF}" "{bf: Subjective well-being versus}" "{bf: share minority}" "{bf: (duration-weighted)}" "",
size(medium) pos(12) ring(1))


/*              7c2.  Set x and y axis titles*/

ytitle("Subjective well-being" "relative to site overall mean, z-score", height(8))

xtitle("Share minority relative to site overall mean, z-score")


/*              7c3.  Turn off legend*/

legend(off)

/*              7c4.  Set axis tick marks*/

xlabel(-0.4(.20)0.7, format(%2.1f))

ylabel(-0.12(.06)0.24, format(%3.2f) nogrid)

/*              7c5.  Set markers for site-groups*/
```

```
mcolor(black)
msymbol(O) msize(*.5)
/*          7c7.  White background with black border*/
graphregion(fcolor(white) lcolor(white)) plotregion(lcolor(black))
/*          7c7.  Set graph height and width in inches*/
ysize(8.25) xsize(6.25)
/*        7d.  Add another scatter with labels for the points */
|| scatter tmp_iv_demean_y tmp_iv_demean_min [aw=tmp_sumwgt],
/*          7d1.  Set axis tick marks*/
xlabel(-0.4(.20)0.7, format(%2.1f))
ylabel(-0.12(.06)0.24, format(%3.2f) nogrid)

/*          7d2.  Set symbol to empty*/
msymbol(i)
/*          7d3.  Label points with group-site*/
mlabel(pt_label)
/*          7d4.  Set marker labels*/
mlabvposition(tmp_mpos_min)
mlabcolor(black)
mlabsize(*0.85 *0.85)

/*        7e.  Add scatter with a regression line*/
|| scatter predict_ymin tmp_iv_demean_min [aw=tmp_sumwgt],
/*          7e1.  Use black lines of medium width for regression line*/
msymbol(i) c(l)
`mto_clp'
clwidth(medium) clcolor(black) mcolor(black)
```

```
/*          7f.  Plot the origin and add panel letter (NOTE)*/
/* The purpose of this scatter is to add the note. the scatter of zero vs zero is not displayed */
|| scatter tmp_zero1 tmp_zero2, msymbol(i)
/*e.g., something like: note("Z", size(vhuge) pos(11) ring(1) span)*/
note("B", size(vlarge) pos(11) ring(1) span)

/*          7g.  End the scatter command*/
;

/*          7h.  Change the delimiter back to a newline*/
#delimit cr;

/*          7i.  Save the graph */
graph save "$min_graph", replace


graph export "${min_graph}.eps", replace
!epstopdf ${min_graph}.eps


* ************************************************************************
*       STEP  8: Create scatter plot of poverty vs. subjective well-being controlling for minority
* ************************************************************************
*          8a.  Create label positions
gen tmp_mpos_povmin = 3

*          8b.  Change the end of line delimiter to allow a longer code for the scatter plot
#delimit ;
```

```
/*          8c.  Create scatter with weighted symbols and titles*/

scatter resid_y_ctrl_min resid_pov_ctrl_min [aw=tmp_sumwgt],


/*          8c1.  Set main title*/

title("{bf: Cell PUF}" "{bf: Subjective well-being versus}" "{bf: share poor controlling for share minority}" "{bf: (duration-weighted)}" "", size(medium) pos(12) ring(1))


/*          8c2.  Set x and y axis titles*/

ytitle("Subjective well-being" "relative to site overall mean, z-score", height(8))

xtitle("Share poor relative to site overall mean, z-score")


/*          8c3.  Turn off legend*/

legend(off)

/*          8c4.  Set axis tick marks*/

xlabel(-0.4(.20)0.7, format(%2.1f))

ylabel(-0.12(.06)0.24, format(%3.2f) nogrid)

/*          8c5.  Set markers for site-groups*/

mcolor(black)

msymbol(O) msize(*.5)

/*          8c6.  White background with black border*/

graphregion(fcolor(white) lcolor(white)) plotregion(lcolor(black))

/*          8c7.  Set graph height and width in inches*/

ysize(8.25) xsize(6.25)

/*          8d.  Add another scatter with labels for the points */

|| scatter resid_y_ctrl_min resid_pov_ctrl_min [aw=tmp_sumwgt],

/*          8d1.  Set axis tick marks*/

xlabel(-0.4(.20)0.7, format(%2.1f))
```

```
ylabel(-0.12(.06)0.24, format(%3.2f) nogrid)

/*          8d2.  Set symbol to empty*/
msymbol(i)
/*          8d3.  Label points with group-site*/
mlabel(pt_label)
/*          8d4.  Set marker labels*/
mlabvposition(tmp_mpos_povmin)
mlabcolor(black)
mlabsize(*0.85 *0.85)

/*        8e.  Add scatter with a regression line*/
|| scatter predict_y_ctrl_min resid_pov_ctrl_min [aw=tmp_sumwgt],
/*          8e1.  Use black lines of medium width for regression line*/
msymbol(i) c(l)
`mto_clp'
clwidth(medium) clcolor(black) mcolor(black)

/*        8f.  Plot the origin and add panel letter (NOTE)*/
/* The purpose of this scatter is to add the note. the scatter of zero vs zero is not displayed */
|| scatter tmp_zero1 tmp_zero2, msymbol(i)
note("C", size(vlarge) pos(11) ring(1) span)

/*        8g.  End the scatter command*/
;

/*        8h.  Change the delimiter back to a newline*/
#delimit cr;
```

```
*          8i.  Save the graph

graph save "$pov_ctrl_min_graph", replace


graph export "${pov_ctrl_min_graph}.eps", replace

!epstopdf ${pov_ctrl_min_graph}.eps


* ************************************************************************

*       STEP  9: Create scatter plot of minority vs. subjective well-being, controlling for poverty

* ************************************************************************

*          9a.  Create label positions

gen tmp_mpos_minpov = 3


*          9b.  Change the end of line delimiter to allow a longer code for the scatter plot

#delimit ;


/*          9c.  Create scatter with weighted symbols and titles*/

scatter resid_y_ctrl_pov resid_min_ctrl_pov [aw=tmp_sumwgt],


/*             9c1.  Set main title*/

title("{bf: Cell PUF}" "{bf: Subjective well-being versus}" "{bf: share minority controlling for share poor}" "{bf: (duration-
weighted)}" "", size(medium) pos(12) ring(1))


/*             9c2.  Set x and y axis titles*/

ytitle("Subjective well-being" "relative to site overall mean, z-score", height(8))

xtitle("Share minority relative to site overall mean, z-score")


/*             9c3.  Turn off legend*/
```

```
legend(off)
/*          9c4.  Set axis tick marks*/
xlabel(-0.4(.20)0.7, format(%2.1f))
ylabel(-0.12(.06)0.24, format(%3.2f) nogrid)
/*          9c5.  Set markers for site-groups*/
mcolor(black)
msymbol(O) msize(*.5)
/*          9c6.  White background with black border*/
graphregion(fcolor(white) lcolor(white)) plotregion(lcolor(black))
/*          9c7.  Set graph height and width in inches*/
ysize(8.25) xsize(6.25)
/*        9d.  Add another scatter with labels for the points */
|| scatter resid_y_ctrl_pov resid_min_ctrl_pov [aw=tmp_sumwgt],
/*          9d1.  Set axis tick marks*/
xlabel(-0.4(.20)0.7, format(%2.1f))
ylabel(-0.12(.06)0.24, format(%3.2f) nogrid)

/*          9d2.  Set symbol to empty*/
msymbol(i)
/*          9d3.  Label points with group-site*/
mlabel(pt_label)
/*          9d4.  Set marker labels*/
mlabvposition(tmp_mpos_minpov)
mlabcolor(black)
mlabsize(*0.85 *0.85)

/*        9e.  Add scatter with a regression line*/
|| scatter predict_y_ctrl_pov resid_min_ctrl_pov [aw=tmp_sumwgt],
```

```
/*              9e1.  Use black lines of medium width for regression line*/
msymbol(i) c(l)
`mto_clp'
clwidth(medium) clcolor(black) mcolor(black)

/*              9f.  Plot the origin and add panel letter (NOTE)*/
/* The purpose of this scatter is to add the note. the scatter of zero vs zero is not displayed */
|| scatter tmp_zero1 tmp_zero2, msymbol(i)
/*e.g., something like: note("Z", size(vhuge) pos(11) ring(1) span)*/
note("D", size(vlarge) pos(11) ring(1) span)

/*              9g.  End the scatter command*/
;

/*              9h.  Change the delimiter back to a newline*/
#delimit cr;

/*              9i.  Save the graph */
graph save "$min_ctrl_pov_graph", replace

graph export "${min_ctrl_pov_graph}.eps", replace
!epstopdf ${min_ctrl_pov_graph}.eps
```

```
* ************************************************************************
*       STEP  10: Join the 4 panels to create a 4-panel figure
* ************************************************************************
*           10a.  Combine 4 panels
graph combine "${pov_graph}.gph" "${min_graph}.gph" "${pov_ctrl_min_graph}.gph" "${min_ctrl_pov_graph}.gph", ///
altshrink row(3) saving("${final_graph}.gph", replace) xsize(6.25) ysize(8.25) graphregion(color(white))

*           10b.  Export as eps
graph export "${final_graph}.eps", replace

*           10c.  Export as pdf
!epstopdf ${final_graph}.eps

* ************************************************************************
*       STEP  11: Open results file to write IV coefficients.
* ************************************************************************
file open csvlog using "${logname}", write replace

* ************************************************************************
*       STEP  12:  Write Slopes (IV coefficients) to output file.
* ************************************************************************
*           12a.  Print column names to log
file write csvlog "lookup, yvar, endo1, endo2, slope" _n

*           12b.  Write poverty coefficient
file write csvlog "happy_scale123_z_ad__mn_f_c9010t_perpov_dw_z, happy_scale123_z_ad, mn_f_c9010t_perpov_dw_z,
none," ///
%20.3f (mto_slope_pov) _n
```

```
*          12c.  Write minority coefficient

file write csvlog "happy_scale123_z_ad__mn_f_c9010t_pminorty_dw_z, happy_scale123_z_ad,
mn_f_c9010t_pminorty_dw_z, none," ///

%20.3f (mto_slope_min) _n


*          12d.  Write poverty controlling for minority coefficient

file write csvlog "happy_scale123_z_ad__mn_f_c9010t_perpov_dw_z__mn_f_c9010t_pminorty_dw_z,
happy_scale123_z_ad, mn_f_c9010t_perpov_dw_z, mn_f_c9010t_pminorty_dw_z," ///

%20.3f (mto_slope_pov_ctrl_min) _n


*          12e.  Write minority controlling for poverty coefficient

file write csvlog "happy_scale123_z_ad__mn_f_c9010t_pminorty_dw_z__mn_f_c9010t_perpov_dw_z,
happy_scale123_z_ad, mn_f_c9010t_pminorty_dw_z, f_c9010t_perpov_dw_z," ///

%20.3f (mto_slope_min_ctrl_pov) _n


* ****************************************************************************

*       STEP  13: Close result output file.

* ****************************************************************************

file close csvlog


* ****************************************************************************

*       STEP  14: Display program run date.

* ****************************************************************************

di "Program End on " c(current_date) " at " c(current_time)
```

Appendix H – Stata Code Used to Create the Outcome Indices Used in the *Science* Article

The code below creates the "mkindex" program that is referenced in the code shown in the detailed coding for key variables in Appendix J.

```
capture program drop mkindex

program define mkindex

syntax varlist(max=30) [pw] [if] , [iname(name)]

* varlist is the variables to include in index

* accepts pweights

* accepts new name for index

* accepts "if" statement to identify the sample

* ra_group MUST BE DEFINED.


di "index to create: `iname'"

di "vars to put in index: `varlist'"

di "type of weight: `weight'"

di "actual weight: `exp'"

*di "sample restriction: " sub`if'


* mark records to use in creating the index *

tempvar touse

mark `touse' `if'

count if `touse'


* creating temporary variables for the index (y) and counter of # of non-missing outcomes (n)

 tempvar y n

 gen `y' = 0

 gen `n' = 0
```

```
* START LOOP THROUGH EACH OUTCOME TO INCLUDE IN INDEX
 foreach v of varlist `varlist' {
   di "`v'"

* standardize the outcome using the control groups mean and std deviation
* set temporary variable for mean value for that sample and treatment group
  tempvar mv
  gen `mv'=0

  * start RA group index
  forvalues i = 1/3 {

* set mean to that ra group's mean if it matches the person's ra group assignment
            sum `v' [aw `exp'] if ra_group==`i' & `touse'
            replace `mv' = r(mean) if ra_group==`i' & `touse'
            }
  * end RA group index
    sum `v' [aw `exp'] if ra_group==3 & `touse'
* if outcome is not missing, standardize the outcome, if missing then use mean for that treatment group
  capture drop chk_`v'
  gen chk_`v' = cond(          (inlist(`v', ., .d, .r) == 0) ,( (`v'-r(mean))/r(sd) ), ( (`mv'-r(mean))/r(sd) ) ) if `touse'
  qui replace `y' = `y' + cond( (inlist(`v', ., .d, .r) == 0) ,( (`v'-r(mean))/r(sd) ), ( (`mv'-r(mean))/r(sd) ) ) if `touse'

* update counter of outcomes in the index for an individual
  qui replace `n' = `n' + (inlist(`v', ., .d, .r) == 0) if `touse'
   }
```

```
 qui sum `n'

* set final index value to sum of the standardized values divided by # of non-missing items

capture drop `iname'

gen `iname' = `y'/r(max) if `n'>0 & `touse'

  end
```

* The format for calling the “mkindex” program is:

* mkindex my_outcome_vars_to_include_in_index [pw= my_weight_var] if my_sample_spec , iname (my_name_for_index)

Appendix I – Detailed Documentation for the Mental Health Disorder Outcomes Included in the *Science* Article

This appendix describes the construction of the major depression in the past year (*f_mh_dep_y_ad*) and Generalized Anxiety Disorder (GAD) in the past year (*f_mh_gad_y_ad*) variables that were components of the mental health index (*f_mh_idx_z_ad*) that was presented in the *Science* article. Also included below is information about the construction of bipolar and related disorder variables, which were used in the construction of the depression and GAD measures. General information about these two variables can be found in Appendix J.

DSM-IV Major Depressive Episode

A. Five(or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure. Note: DSM-IV states that children and adolescents may be "irritable rather than sad". This is not operationalized when examining adults who report symptoms from childhood.

Part 1 AND Part 2.

Part 1. Symptoms have been present during the same 2 week period and at least one of the symptoms is either(1) depressed mood or (2) loss of interest or pleasure.

(HDE6b_D22b >= 2 weeks OR HDE6e_D22d >= 2 weeks) AND
(HDE7a_D24a is Yes(1) OR HDE7b_D24c is Yes(1) OR HDE7c_D24e is Yes(1) OR HDE7d_D24f is Yes(1))

Note: D24b, D24d, D39 are deleted from NCSR.

Part 2. least five of the following symptoms must be present and represent a change from previous functioning:

Note: "change from previous functioning" is implicit in the item corresponding to each symptom (e.g. "more than usual", "less than usual").

1. depressed mood most of the day, nearly every day, as indicated by either subjective report(e.g., feels sad or empty) or observation made by others.

   HDE7a_D24a is Yes(1) OR HDE7b_D24c is Yes(1).

   Note D24b, D24d are deleted from NCSR.

2. markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day(as indicated by either subjective account or observation made by others)

   HDE7c_D24e is Yes(1) OR HDE7d_D24f is Yes(1).

3. significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day.

HDE9a_D26a is Yes(1) OR HDE9b_D26b is Yes(1) OR HDE9c_D26c is Yes(1) OR HDE9d_D26e is Yes(1).

Note D26d, D26f(weight gain, loss) deleted from NCSR.

4. insomnia or hypersomnia nearly every day.

   HDE9e_D26g is Yes(1) OR HDE9f_D26h is Yes(1).

5. psychomotor agitation or retardation nearly every day(observable by others, not merely subjective feelings of restlessness or being slowed down).

   HDE9i_D26m is Yes(1) OR HDE9k_D26o is Yes(1).

6. fatigue or loss of energy nearly every day.

   HDE9g_D26j is Yes(1).

7. feelings of worthlessness or excessive or inappropriate guilt(which may be delusional) nearly every day(not merely self-reproach or guilt about being sick)

   HDE9o_D26u is Yes(1).

8. diminished ability to think or concentrate, or indecisiveness, nearly every day(either by subjective account or as observed by others)

   HDE9l_D26p is Yes(1) OR HDE9m_D26r is Yes(1) OR HDE9n_D26s is Yes(1).

9. recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.

   HDE9p_D26aa is Yes(1) OR HDE9q_D26bb is Yes(1) OR HDE9r_D26cc is Yes(1).

Note: Respondents leave D24, D26 series after 5 of the following symptoms endorsed:

(1) IF HDE7a_D24a OR HDE7b_D24c IS CODED '1' (YES) INCREMENT COUNT BY ONE.
(2) IF HDE7c_D24e OR HDE7d_D24f IS CODED '1' (YES) INCREMENT COUNT BY ONE.
(3) IF HDE9a_D26a OR HDE9b_D26b IS CODED '1' (YES) INCREMENT COUNT BY ONE.
(4) If HDE9c_D26c OR HDE9d_D26d IS CODED '1' (YES) INCREMENT COUNT BY ONE
(5) IF HDE9e_D26g OR HDE9f_D26h IS CODED '1' (YES) INCREMENT COUNT BY ONE.
(6) IF HDE9g_D26j IS CODED '1' (YES) INCREMENT COUNT BY ONE.
(7) IF HDE9i_D26m OR HDE9k_D26o IS CODED '1' (YES) INCREMENT COUNT BY ONE.
(8) IF HDE9l_D26p OR HDE9m_D26r OR HDE9n_D26s IS CODED '1' (YES) INCREMENT COUNT BY ONE.
(9) IF HDE9o_D26u IS CODED '1' (YES) INCREMENT COUNT BY ONE.
(10) IF HDE9p_D26aa OR HDE9q_D26bb OR HDE9r_D26cc IS CODED '1' (YES) INCREMENT COUNT BY ONE.

If respondent has 4 of 9 DSM-IV Criteria A part 2 symptoms and
sum(HDE9a_D26a = 1 or HDE9b_D26b = 1,HDE9c_D26c = 1 or HDE9d_D26e = 1) = 2 and
HDE9p_D26aa is missing then respondent meets Criteria A Part 2.

B. The symptoms do not meet criteria for a Mixed Episode

Not operationalized.

C. Part 1 OR Part 2.

Part 1. The symptoms cause clinically significant distress.

HDE12_D17 is (2,3,4).

Note: D18, D19, D24b deleted from NCSR.

Part 2. The symptoms cause clinically significant impairment in social, occupational, or other important areas of functioning.

HDE11_D28 is (3,4,5)

Note D28a, D66a-d deleted from NCSR.

D. The symptoms are not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication), or are not due to a general medical condition. NOTE: D29b is used as an initial screener only. All open ended items are reviewed by a clinician to determine organic exclusion.

Note: D29a, D29b deleted from NCSR, criteria not operationalized.

E. Part 1 OR Part 2 OR Part 3

Part 1. The symptoms are not better accounted for by Bereavement.
Not operationalized.

Part 2. If the symptoms are associated with bereavement, they persist for longer than two months
Not operationalized

Part 3. If the symptoms are associated with bereavement, they are characterized by (a) marked functional impairment, (b) morbid preoccupation with worthlessness, (c) suicidal ideation, (d) psychotic symptoms, or (e) psychomotor retardation. At least one of a-e must be present.
Not operationalized

NOTE: D23 was deleted from the instrument therefore the bereavement criteria could not be operationalized.

GAD - DSM-IV Criteria

A. Part 1 AND Part 2 AND Part 3

Part 1. Excessive anxiety and worry(apprehensive expectation)

HSC12_SC26 is Yes(1) OR HSC13_SC26a is Yes(1) OR HSC14_SC26b is Yes(1).

Part 2. Anxiety occurring more days than not for at least 6 months

HGA5_G5 is at least 6 months.

Note G30, G31 from the NCSR are deleted.

Part 3. Anxiety about a number of events or activities(such as work or school performance).

At least 2 worries from HGA1_G1__S series.

B. The person finds it difficult to control the worry.

HGA3_G4 is (1,2).

C. The anxiety and worry are associated with three (or more) of the following six symptoms(with at least some symptoms present for more days than not for the past 6 months) (1. restlessness or feeling keyed up or edge, 2. being easily fatigued, 3. difficulty concentrating or mind going blank, 4. irritability, 5. muscle tension, 6. sleep disturbance.)

Note: "at least some symptoms present for more days than not" is not operationalized

At least 3 of (HGA7a_G9a, HGA7b_G9b, HGA7c_G9c, HGA7d_G9d, HGA7e_G9e, HGA7f_G9f) are Yes(1).

D. The focus of the anxiety and worry is not confined to features of an Axis 1 disorder.

At least 1 value of 1-10, 13, 20-32 in HGA1_G1__S series.

Note: DSM IV states "the anxiety and worry should not occur exclusively during PTSD". This part of criteria D is not operationalized.

E. Part 1 OR Part 2

Part 1. The anxiety, worry, or physical symptoms cause clinically significant distress.

HGA4_G4a is(1,2) OR HGA9_G15 is (3,4,5) .

Part 2. The anxiety, worry, or physical symptoms cause clinically significant impairment in social, occupational, or other important areas of functioning.

HGA10_G17 is (3,4,5).

F. Part 1 AND Part 2

Part 1. The disturbance is not due to the direct physiological effects of a substance(e.g. a drug of abuse, a medication) or due to a general medical condition(e.g., hyperthyroidism).

Note: G18a, G18b from NCSR are deleted, criteria not operationalized.

Part 2. The disturbance does not occur exclusively during a Mood Disorder, a Psychotic Disorder, or a Pervasive Developmental Disorder.

Note: Psychotic Disorder and Pervasive Developmental Disorder hierarchies are not operationalized.

(Major Depression = No(5) AND Minor Depression = No(5) AND Mania = No(5))

OR

((Major Depression = Yes(1) OR Minor Depression = Yes(1) OR Mania = Yes(1)) AND
(GAD onset < Mood onset) OR
(GAD recency > Mood recency)))

Manic Episode (Old Version) – DSM-IV Criteria (MH_MAN_OLD)

A. Part 1 AND Part 2

Part 1. A distinct period of abnormally and persistently elevated, expansive, or irritable mood.

HSC9_SC24 = Yes(1) OR HSC11_SC25a = Yes(1).

Part 2. A distinct period of abnormally and persistently elevated, expansive, or irritable mood lasting at least 1 week(or any duration if hospitalization is necessary).

(HMA2b_M3b >= 1 week) OR (HMA2e_M3d >= 1 week) OR (HMA6b_M6b >= 1 week) OR (HMA6e_M6d >= 1week).

Note: M20, M22, M48 from NCSR are deleted.

B. During the mood disturbance, three(or more) of the following symptoms have persisted (four if the mood is only irritable) and have been present to a significant degree:

Mood is only irritable: HSC11_SC25a is Yes(1) AND (HSC9_SC24 is NOT Yes(1))

1. inflated self-esteem or grandiosity

   HMA7n_M7n is Yes(1) OR HMA7o_M7o is Yes(1).

2. decreased need for sleep(e.g., feels rested after only 3 hours of sleep)

   HMA7j_M7j is Yes(1).

3. more talkative than usual or pressure to keep talking

   HMA7f_M7f is Yes(1).

4. flight of ideas or subjective experience that thoughts are racing

   HMA7i_M7i is Yes(1).

5. distractibility (i.e., attention too easily drawn to unimportant or irrelevant external stimuli)

   HMA7g_M7g is Yes(1) OR HMA7h_M7h is Yes(1).

6. increase in goal-oriented activity(either socially, at work or school, or sexually) or psychomotor agitation.

   HMA7a_M7a is Yes(1) OR HMA7b_M7b is Yes(1) OR HMA7c_M7c is Yes(1) OR HMA7e_M7e is Yes(1).

7. excessive involvement in pleasurable activities that have a high potential for painful consequences(e.g., engaging in unrestrained buying sprees, sexual indiscretions, or foolish business investments)

   HMA7k_M7k is Yes(1) OR HMA7l_M7l is Yes(1) OR HMA7m_M7m is Yes(1).

C. The symptoms do not meet criteria for a Mixed Episode

Not Operationalized

D. Part 1 OR Part 2 OR Part 3

Part 1. The mood disturbance is sufficiently severe to cause marked impairment in occupational functioning or in usual social activities or relationships with others.

HMA9_M9 is (4,5).

Note: M9a, M27a-d, M29, M33 from NCSR are deleted.

Part 2. The mood disturbance is sufficiently severe to necessitate hospitalization to prevent harm to self.

Note M48 from NCSR is deleted, criteria not operationalized.

Part 3. There are psychotic features

HMA7o_M7o is Yes(1).

E. The symptoms are not due to the direct physiological effects of a substance(e.g., a drug of abuse, a medication, or other treatment) or a general medical condition (e.g. hyperthyroidism)

Note M10a, M10b from NCSR are deleted, criteria not operationalized.

MTO Bipolar I Old (F_MH_BIPOLARI_OLD)

MH_MAN_OLD is Yes(1).

Hypomanic Episode (Old version) – DSM-IV Criteria (MH_HYP_OLD)

A. Part 1 AND Part 2

Part 1. A distinct period of abnormally and persistently elevated, expansive, or irritable mood.

HSC9_SC24 is Yes(1) OR HSC11_SC25a is Yes(1).

Part 2. A distinct period of abnormally and persistently elevated, expansive, or irritable mood lasting at least 4 days, that is clearly different from the usual nondepressed mood.

HSC9_SC24 is Yes(1) OR (HMA2b_M3b >= 4 days) OR (HMA2e_M3d >= 4 days) OR
(HMA6b_M6b >= 4 days) OR
HMA6e_(M6d >= 4 days).

Note: M20, M22 from NCSR are deleted.

B. During the mood disturbance, three(or more) of the following symptoms have persisted(four if the mood is only irritable) and have been present to a significant degree:

Mood is only irritable: HSC11_SC25a is Yes(1) and (HSC9_SC24 is NOT Yes(1))

1. inflated self-esteem or grandiosity

   HMA7n_M7n is Yes(1) OR HMA7o_M7o is Yes(1).

2. decreased need for sleep(e.g., feels rested after only 3 hours of sleep)

   HMA7j_M7j is Yes(1).

3. more talkative than usual or pressure to keep talking

   HMA7f_M7f is Yes(1).

4. flight of ideas or subjective experience that thoughts are racing

   HMA7i_M7i is Yes(1).

5. distractibility (i.e., attention too easily drawn to unimportant or irrelevant external stimuli)

   HMA7g_M7g is Yes(1) OR HMA7h_M7h is Yes(1).

6. increase in goal-oriented activity(either socially, at work or school, or sexually) or psychomotor agitation.

   HMA7a_M7a is Yes(1) OR HMA7b_M7b is Yes(1) OR HMA7c_M7c is Yes(1) OR HMA7e_M7e is Yes(1).

7. excessive involvement in pleasurable activities that have a high potential for painful consequences(e.g., engaging in unrestrained buying sprees, sexual indiscretions, or foolish business investments)

   HMA7k_M7k is Yes(1) OR HMA7l_M7l is Yes(1) OR HMA7m_M7m is Yes(1).

C. The episode is associated with an unequivocal change in functioning that is uncharacteristic of the person when not symptomatic.

HMA9_M9 is (3,4,5).

Note: M9a, M27a-d, M29, M33 from NCSR are deleted.

D. The disturbance in mood and the change in functioning are observable by others.

Not Operationalized

E. Part 1 AND Part 2 AND Part 3.

Note: By strict DSM criteria, those people who meet all criteria for mania but have a duration of 4 to 6 days without hospitalization are excluded from a diagnosis of hypomania. (See mania criterion A,D and hypomania criterion E). We have defined these people as meeting hypomania. This is implemented by suppressing Criterion E for those with a duration of 4 to 6 days and without hospitalization.

Part 1. The mood disturbance is not severe enough to cause marked impairment in occupational functioning or in usual social activities or relationships with others.

NOT(HMA9_M9 is (4,5)).

Note: M9a, M27a-d, M29, M33 from NCSR are deleted.

Part 2. The mood disturbance is not severe enough to necessitate hospitalization to prevent harm to self.

Note M48 from NCSR is deleted, criteria not operationalized.

Part 3. There are no psychotic features

HMA7o_M7o is No(5).

F. The symptoms are not due to the direct physiological effects of a substance(e.g., a drug of abuse, a medication, or other treatment) or a general medical condition (e.g. hyperthyroidism)

Note M10a, M10b from NCSR are deleted, criteria not operationalized.

MTO Adult Bipolar II Old(F_MH_BIPOLARII_OLD)

A. Presence (or history) of one or more Major Depressive Episodes

f_mh_mde_evr_ad is Yes(1)

B. Presence (or history) of at least one Hypomanic Episode

mh_hyp_old is Yes(1)

C. There has never been a Manic Episode or Mixed Episode

mh_man_old is NOT Yes(1)

D. Not operationalized.

E. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

HMA9_M9 is (3,4,5).

Note: M9a, M27a-d, M29, M33 from NCSR are deleted.

MTO Adult (Bipolar I/II/Sub), Mania, Hypomania

Bipolar I

mh_man_old is Yes (1) AND
at least 6 symptoms in the HMA7_M7 series(MH_MAN_OLD Criteria B1-B7) AND
at least two of the following symptoms: HMA7b_M7b, HMA7c_M7c, HMA7k_M7k, HMA7l_M7l, HMA7o_M7o is 1(yes).

Bipolar II

NOT Bipolar I AND

(f_mh_bipolar I old is Yes (1) AND f_mh_dep_evr_ad = 1 AND HMA1_M1 = 1 AND HMA7i_M7i = 1)
*** these are the people who meet criteria for our old bipolar I definition (mania) but no longer meet criteria with the new definition, and have a major depressive episode and euphoria and racing thoughts ***

OR

(f_mh_bipolarII_old is Yes(1) AND
(HMA3b_M3b >= 14 days OR HMA3d_M3d >= 14 days OR HMA6b_M6b >= 14 days OR HMA6e_M6d >= 14 days) AND
at least 2 of the following symptoms (HMA7b_M7b, HMA7c_M7c, HMA7k_M7k, HMA7l_M7l, HMA7o_M7o)
)

Note: M20, M22 from NCSR are deleted.

***This is our old definition of bipolar II tightened up to include a duration of at least 14 days and at least 2 of the "super" symptoms in terms of concordance.

Bipolar Sub

*** anyone left with old mania/hypomania who did not meet criteria for bipolar I and bipolar II above***

Not Bipolar I or Bipolar II as defined above AND (mh_man_old is Yes(1) OR mh_hyp_old is Yes(1)).

Mania (dsm_man)

Bipolar I is Yes(1).

Hypomania (dsm_hyp)

Bipolar II is Yes(1) OR (Bipolar Sub is Yes(1) AND mh_hyp_old is Yes(1)).

Sub-Hypomania (dsm_hypsub)

Bipolar Sub is Yes(1) AND (f_mh_bipolarIII Old is Yes(1) OR mh_hyp_old is Yes(1)).

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

This appendix is available at www.nber.org/mtopuf/mto_sci_puf_docu_memo_apndxJ_20131008.pdf.