README: HUD's Moving to Opportunity Metadata

THIS VERSION OF THE README FILE CAN BE MADE PUBLIC- ALL REFERENCES TO SERVER DIRECTORIES HAVE BEEN REMOVED.

FILES

1994_2010

- 1 data file with PIK, PVS verification flag, 3 alternate quick PIKs, year of birth, and flag indicating date of birth was imputed:

• mto1994_2010.sas7bdat

1994_2010_restricted

- 1 data file with full date of birth (DOB) (RESTRICTED ACCESS):

• mto1994_2010_fulldob.sas7bdat

1994_2010_name

- 1 data file with names (HIGHLY RESTRICTED ACCESS):

• mto1994_2010_name.sas7bdat

1994_2000

- 2 data files from baseline and canvass files for individuals and households:

• MTO_BASCNV_PERS.sas7bdat

• MTO_BASCNV_HHLD.sas7bdat

2002

- 10 data files from interim wave:

• MTO_INT_ANLY_CHILDYTH.sas7bdat

• MTO_INT_ANLY_ADULT.sas7bdat

• MTO_INT_RAWSVY_ADULT.sas7bdat

• MTO_INT_RAWSVY_YOUTH.sas7bdat

• MTO_INT_RAWSVY_CHILD.sas7bdat

• MTO_INT_RAWSVY_POCY.sas7bdat

• MTO_INT_SCH_HX.sas7bdat

• MTO_INT_ROSTER.sas7bdat

• MTO_INT_ADDRHX.sas7bdat

- tier1xwalk_donotpost.sas7bdat - crosswalk of actual study IDs to the masked IDs on interim files provided in the past by HUD/ICPSR

2008_2010

- 19 data files from final wave:

- MTO_FIN_ANALYSIS.sas7bdat

- MTO_FIN_RAWSVY_ADULT.sas7bdat

- MTO_FIN_RAWSVY_YOUTH.sas7bdat

- MTO_FIN_ADDRHX.sas7bdat

- MTO_FIN_RAW_DWELLING.sas7bdat

- MTO_FIN_RAW_DWELLMAP.sas7bdat

- MTO_FIN_LING_TOKEN.sas7bdat

- MTO_FIN_SCH_HX.sas7bdat

- MTO_FIN_ROSTER.sas7bdat

DOCUMENTATION

doc

- 1 file:

hud_mto_covermemo_20170627-F.docx - this document provides an overview of the data, data files, documentation files, and some technical notes. This cover memo can be in the public domain.

Note that the documentation references data file 1 that included PII (Personally Identifiable Information), which is not accessible to users. File 1 was replaced with 3 files:

 mto1994_2010.sas7bdat

 mto1994_2010_fulldob.sas7bdat   (RESTRICTED ACCESS)

 mto1994_2010_name.sas7bdat   (HIGHLY RESTRICTED ACCESS)

There is no data file 2; it does not exist.

6 SUBFOLDERS:

content

Codebook, SAS Proc Contents listings for the datasets, QC results, and 4 documents providing detailed information on construction of recode variables for interim and final.

Detailed documentation of recodes for select outcomes and mediators. Many of the outcomes and mediators constructed have detailed documentation showing the code used to create them from the raw variables.

- 30 files:

• hud_mto_confid_codebook20170420_0626edit-F.xlsx -- RESTRICTED CODEBOOK -- This codebook CANNOT be in the public domain- must remove the min/max and N Non-Missing.

• hud_mto_PUBLIC_codebook20170420_0626edit-F.xlsx -- PUBLIC CODEBOOK -- This codebook CAN be in the public domain- has removed 3 columns: min, max and N Non-Missing.

• metdt_mto1994_2010.pdf

• metdt_mto1994_2010_fulldob.pdf

• metdt_mto1994_2010_name.pdf

• contents_03_mto_bascnv_pers.pdf

• contents_04_mto_bascnv_hhld.pdf

• contents_05_mto_int_anly_childyth.pdf

• contents_06_mto_int_anly_adult.pdf

• contents_07_mto_int_rawsvy_adult.pdf

• contents_08_mto_int_rawsvy_youth.pdf

• contents_09_mto_int_rawsvy_child.pdf

• contents_10_mto_int_rawsvy_pocy.pdf

• contents_11_mto_int_sch_hx.pdf

• contents_12_mto_int_roster.pdf

• contents_13_mto_int_addrhx.pdf

• contents_14_mto_fin_analysis.pdf

• contents_15_mto_fin_rawsvy_adult.pdf

• contents_16_mto_fin_rawsvy_youth.pdf

• contents_17_mto_fin_addrhx.pdf

• contents_18_mto_fin_raw_dwelling.pdf

• contents_19_mto_fin_raw_dwellmap.pdf

• contents_20_mto_fin_ling_token.pdf

• contents_21_mto_fin_sch_hx.pdf

- contents_22_mto_fin_roster.pdf

- dtdic_hud_mto_interim_child_detailed_recodes.pdf

- dtdic_hud_mto_interim_adult_detailed_recodes.pdf

- hud_mto_fin_adult_detailed_recodes.docx

- hud_mto_fin_youthgc_detailed_recodes.docx

- MTO-dataQC-Results.pdf - Results of a cursory quality control review. File restricted- CANNOT be in the public domain.

SASprograms

- 2 files:

- hud_mto_final_fmts.sas – contains the "proc format" statements needed to create the formats listed in the codebook.

- hud_mto_macros_attfmts_rev.sas – contains the snippets of SAS code needed to assign formats to specific variables.

mto_survey_instruments

These are available on a public website: http://www.nber.org/mtopublic/instruments.html

- 11 files of survey instruments:

- mto_instrument_baseline.pdf – baseline survey instrument

- mto_instrument_interim_adult.pdf – adult/household instrument survey at interim.

- mto_instrument_interim_child.pdf – child survey instrument at interim

- mto_instrument_interim_youth.pdf – youth survey instrument at interim

- mto_instrument_interim_neighborhood_obs.pdf - neighborhood observations instrument at interim

- mto_instrument_interim_hhold_obs.pdf - household observations instrument at interim

- mto_instrument_final_adult.pdf – adult survey instrument at final.

- mto_instrument_final_adult_short.pdf – short version of adult survey at final. Administered to small number of respondents.

- mto_instrument_final_youth.pdf – youth survey instrument at final.

- mto_instrument_final_youth_short.pdf – short version of youth survey instrument at final. Administered to small number of respondents.

- mto_instrument_final_neighborhood_obs.pdf - instrument for conducting the neighborhood walk around (NWA) at final

mto_MQDS_documentation

Michigan Questionnaire Documentation System (MQDS). These documents contain extremely detailed information about the Blaise computer administration of the final surveys including the "universe" for each question.

MQDS is best used only when the simpler "mto_instrument" documentation cannot answer a question about the survey.

- 3 files:

- mto_final_mqds_adult_english.doc – MQDS document for the adult (household) final survey.

- mto_final_mqds_adult_spanish.rtf – MQDS document for the Spanish version of the adult final survey.

- mto_final_mqds_youth.doc – MQDS document for the final youth survey.

IOcodes

Publicly available codes for industry and occupation, crosswalks to the SIC/NAICS and SOC standards, and crosswalks for bridging to more recent classification standards.

At interim and final, 3-digit numeric codes for industry and occupation were coded by the Census Bureau's National Processing Center using 1990 SIC/1980 SOC with 100% verification.

At interim, HUD provided Census with Level of Education, Description of Industry, Description of Occupation, and Description of Duties. HUD did not have employer name or industry type available.

At final, HUD had industry type. Note that the older 1990 SIC and 1980 SOC codes were used at final for consistency with interim.

These code labels are provided in the file of formats (above): hud_mto_final_fmts.sas

See the Census Bureau's website for additional information: https://www.census.gov/people/io/methodology/

- 7 files

OCCUPATION

- SOCCEN80.xls - Crosswalk of 1980 Standard Occupational Classification to Census Occupation Codes

- SOCCEN90cross80.xls - Crosswalk of 1980 Standard Occupational Classification to 1990 Census Occupation Codes

- techpap65-tab02-90to00occ.pdf - Technical Paper 65, Table 2. 1990 Census Occupation Classification System and Its Redistribution into the Census 2000 Occupation Classification System

- 2002 Census Occupation Codes.xlx - 2002 Census Occupation Codes with 2000 Standard Occupational Classification Code

- 2010_OccCodeswithCrosswalkfrom2002-2011nov04.xls - 2010 Census Occupation Codes with Crosswalk from 2002, updated 11/04/2011

INDUSTRY

- 1990 Census_SIC codes.pdf - 1990 Industrial Classification System

- IndustryCrosswalk90-00-02-07-12.xls - 1990-2012 Census Industry Codes with Crosswalk

reports

Publicly available reports that have technical documentation

- 2 files

- MTOFHD_fullreport_v2-Final.pdf - Also at
https://www.huduser.gov/portal//publications/pdf/MTOFHD_fullreport_v2.pdf

- MTOFullReport-Interim.pdf - Also at https://www.huduser.gov/Publications/pdf/MTOfullreport.pdf

ADDITIONAL NOTES ON THE MTO DATA:

The Census Bureau's copy of HUD MTO data do NOT include arrest data (or other state and local data) due to restrictions imposed by agencies in releasing the data.

For qualitative information, the Census Bureau's copy of the data includes coded linguistic speech components. This information comes from audio recordings that Census Bureau did not receive.

GEOCODING

Geocoding includes both 1990 and 2000 tract information. The files contain a "final" tract (and its geocoding source) to use for analyses.

In addition, the files have the tracts from the earlier rounds of geocoding in case those tracts are more accurate (these were generated closer in time to when the person lived at the address).

The construction of the address histories used an algorithm that attempted to reconcile information from many sources to create a continuous address history for each person.

The data and algorithm are imperfect and so the files have the alternative tract, in case it is more complete for some addresses.

Depending on the research question, one might want to make a different choice about which set of geocodes to rely upon.

Address standardization and geocoding of tracts were conducted over time by different vendors (Lorton, Tele-atlas). During the final evaluation, NBER also sent the full set of addresses to HUD for re-geocoding.

The HUD geocoding process provides consistency and good data quality.

Currently (2017), HUD no longer geocodes for 1990 geographies. The HUD software knows if an address is valid (per USPS).

A non-extant address that existed in 1990 can still be geocoded if the structure number fits with the address range along that street segment/block.

Geocoding data quality before early 2000s is less accurate and rural areas are generally less accurate than urban areas.

Census tracts can change from census to census, though there is great interest in maintaining comparability and most changes are splits, combined tracts, or minor boundary changes.

Still, one should probably standardize to either 2000 or 2010 geographies.

Also, it is critical to use the data quality information to determine if one should use the output tract or not. Some may refer to a ZIP Code centroid tract (if the geocoder could not geocode address for various reasons).

A ZIP Code in urban areas can contain more than 20 full or partial tracts and the tract assigned by the geocoding software is likely incorrect in such cases.

GENERAL DISCLOSURE REVIEW NOTES

All results to be removed from the secure environment will be reviewed for disclosure risk, following HUD's disclosure protocols, as well as Census Bureau's and the rules of data providers of any linked data.

HUD's rule of eleven applies to all HUD datasets: The research results cannot disclose information about the characteristics of any group of individuals (households) numbering less than eleven.

The results also cannot show residential tract histories.

Examples:

1. One person with characteristics reported on the researcher file moved from tract 1 in the second quarter of 1995 to tract 2 in the third quarter of 1995. - Cannot be released, Violation of the rule of 11.

2. Results show 11 or more people moved from tract 1 to tract 2. - Non-violation of the rule of 11.

3. Classify tracts on characteristics such as above or below 10% poverty. Eleven or more people moved from a low-poverty to a different low-poverty tract. - Non-violation of the rule of 11.

4. Create geographically contiguous groupings of tracts on origin and destination. Eleven or more people moved from tract-group 1 to tract-group 2. - Non-violation of the rule of 11.

5. Results show 11 or more people live in tract 1 and work in tract 2. - Non-violation of the rule of 11.

Also, when crossing multiple characteristics, there must be 11 or more people (households) in each cell/combination of characteristics.