Steps and syntax files used in creating **X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS\_98061214\_final.xlsx** to process it further for the analytical data file.

|  |  |  |  |
| --- | --- | --- | --- |
| **Software** | **name of do/sas editor file** | **data used** | **permanent output** |
| SAS9.4 | X:\xDATA\NLSPHS 2014\Analysis\Github\NLSPHS\NLSPHS2014\_pop2013\_workingdata\_Lava\_final.sas | Calls X:\xDATA\NLSPHS 2014\Analysis\Github\NLSPHS\NLSPHS\_lava\_working\_final.sas |  |
| NLSPHS9 |  |
| U:\Data\NACCHO2013\NACCHO2013population.xlsx |  |
| X:\xDATA\NLSPHS 2014\Analysis\nlsphs2014population\_full.csv |  |
| X:\xDATA\NLSPHS 2014\Analysis\NLSPSH2014popmiss.xlsx | X:\xDATA\NLSPHS 2014\Analysis\NLSPHS2014popmiss.xlsx |
|  | X:\xDATA\NLSPHS 2014\Analysis\data\nlsphs2014population.csv |
|  | X:\xDATA\NLSPHS 2014\Analysis\NLSPHS2014\_LARGE.sas7bdat |
|  | X:\xDATA\NLSPHS 2014\Analysis\NLSPHS2014\_SMALL.sas7bdat |
|  | X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS2014\_LARGE.dta |
|  | X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS2014\_SMALL.dta |
| X:\xDATA\NLSPHS 2014\Analysis\Github\NLSPHS\modifyingwts.sas | X:\xDATA\NLSPHS 2014\Analysis\Nlsphs2014\_small | X:\xDATA\NLSPHS 2014\Analysis\WTSSMALL.dta |
| X:\xDATA\NLSPHS 2014\Contacts\Mail merger\nlsphswithsurveylinks\_Master\_Final1.xls |  |
| U:\NLSPHS2014\Sampled511a | X:\xDATA\NLSPHS 2014\Analysis\mispopsmalljuris.xls |
| X:\xDATA\NLSPHS 2014\Analysis\mispopsmalljuris1.xls |  |
| sashelp.US\_DATA | X:\xDATA\NLSPHS 2014\Analysis\NLSPHS\_small\_wts\_adj\_frm\_large.dta |
| Stata13.1 | X:\xDATA\NLSPHS 2014\Analysis\Github\NLSPHS\weights for large LHDs.do | U:\Data\NACCHO2013\2013 Profile\_id.dta | X:\xDATA\NLSPHS 2014\Analysis\NACCHO13\_large.dta |
|  |  |
| X:\xDATA\NLSPHS 2014\Contacts\Mail merger\nlsphswithsurveylinks\_Master\_Final1.dta |  |
| X:\xDATA\NLSPHS 2014\Analysis\data\SAMPFRAME\_MISPOP.dta |  |
|  |  |
| X:\xDATA\NLSPHS 2014\Analysis\USSTATEREGIONS.dta |  |
| nlsphsnaccho\_large | X:\xDATA\NLSPHS 2014\Analysis\NLSPHS\_large\_wts.dta |
| X:\xDATA\NLSPHS 2014\Analysis\NLSPHS\_large\_wts.dta |  |
| X:\xDATA\NLSPHS 2014\Analysis\WTSSMALL.dta |  |
| small |  |
| X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS2014\_large.dta | X:\xDATA\NLSPHS 2014\Analysis\NLSPHS\_large\_wts\_adj\_frm\_small.dta |
| X:\xDATA\NLSPHS 2014\Analysis\Github\NLSPHS\full NLSPHS with weights.do | X:\xDATA\NLSPHS 2014\Contacts\Mail merger\nlsphswithsurveylinks\_Master\_Final1.dta |  |
| X:\xDATA\NLSPHS 2014\Analysis\NLSPHS\_small\_wts\_adj\_frm\_large.dta |  |
| X:\xDATA\NLSPHS 2014\Analysis\NLSPHS\_large\_wts\_adj\_frm\_small.dta | X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS\_2014\_wts\_adj.dta |
| X:\xDATA\NLSPHS 2014\Analysis\Github\NLSPHS\manage previous waves of data.do | X:\xDATA\NLSPHS 2014\nlsphs\_tot.dta |  |
|  |  |
| nlsphs\_tot |  |
| nlsphs\_tot |  |
|  | X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS\_2012\_noresp.dta |
| X:\xDATA\NLSPHS 2014\Analysis\Github\NLSPHS\final.do | X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS\_2014\_wts\_adj.dta |  |
| X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS\_2012\_noresp.dta |  |
| X:\xDATA\NLSPHS 2014\nlsphs\_tot.dta | X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS\_98061214.dta |
| Stata13.1 | X:\xDATA\NLSPHS 2014\Analysis\Github\NLSPHS\Getting FIPS for Arm23.do | X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS\_98061214.dta |  |
|  |  |
|  |  |
|  |  |
| X:\xDATA\NLSPHS 2014\Analysis\data\NACCHO\_2013\_LHDBoundaries\_JurisdictionTable.xlsx | X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS14\_Arm23\_FIPS.dta |
| SAS9.4 U:\Data\AHRF\_2013-2014\DOC\AHRF2014.sas | U:\Data\AHRF\_2013-2014\DATA\ahrf2014.asc | X:\xDATA\NLSPHS 2014\Analysis\data\AHRF1314.dta |
| contd…  X:\xDATA\NLSPHS 2014\Analysis\Github\NLSPHS\Getting FIPS for Arm23.do |  | X:\xDATA\NLSPHS 2014\Analysis\data\AHRF1314\_trunc.dta |
| X:\xDATA\NLSPHS 2014\Analysis\data\AHRF1314\_trunc.dta |  |
| X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS14\_Arm23\_FIPS.dta | X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS14\_Arm23\_FIPS\_for\_Clustering.dta |
| U:\Data\NACCHO2013\2013 Profile\_id.dta |  |
| X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS14\_Arm23\_FIPS\_for\_Clustering.dta | X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS14\_Arm23\_FIPS\_for\_Clustering\_final.dta |
|  | X:\xDATA\NLSPHS 2014\Analysis\data\Arveen.dta |
| X:\xDATA\NLSPHS 2014\Analysis\data\Arveen\_Completed\_Final.xlsx |  |
| X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS14\_Arm23\_FIPS\_for\_Clustering\_final.dta | X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS\_Small\_Clustering.dta |
|  | X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS\_Small\_Clustering.xls |
| SAS9.4 | X:\xDATA\NLSPHS 2014\Analysis\Github\NLSPHS\Clustering\_Small\_NLSPHS\_556\_final.sas | X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS\_Small\_Clustering.xls |  |
|  |  |
|  |  |
|  |  |
| work.can6 |  |
|  |  |
|  |  |
|  | X:\xDATA\NLSPHS 2014\Analysis\data\SMALL\_PEER\_FINAL.dta |
| Stata13.1 | X:\xDATA\NLSPHS 2014\Analysis\Github\NLSPHS\peergrouping.do | X:\xDATA\NLSPHS 2014\Analysis\data\SMALL\_PEER\_FINAL.dta |  |
| X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS\_98061214.dta |  |
|  | X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS\_98061214\_final.dta |
|  | X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS\_98061214\_final.xlsx |

**Detail for the above flowchart**

**X:\xDATA\NLSPHS 2014\Analysis\Github\NLSPHS\NLSPHS\_lava\_working\_final. sas**

Purpose:

This program performs data corrections, transformations, and formatting on the NLSPHS 2014 datasets and generates the CSV datafiles

Input:

C:\Users\lrti222\Documents\My SAS Files\INSA.csv

C:\Users\lrti222\Documents\My SAS Files\INSB.csv

X:\xDATA\NLSPHS 2014\Contacts\Mail merger\nlsphswithsurveylinks\_Master\_Final1.xls

U:\NLSPHS2014\sampled511a.sas7bdat

Steps:

* Read in csv data imported from REDCap
* Use appropriate format and informat statements to import the cvs data
* Perform data transformations and corrections needed for the comparative reports
* Define all 3266 counties listed for the whole project to define counties fully within and partially within an LHD
* Rename variables

Temporary Output: NLSPHS9

Output: C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\NLSPHS2014\_formated.csv

**X:\xDATA\NLSPHS 2014\Analysis\Github\NLSPHS\NLSPHS2014\_pop2013\_workingdata\_Lava\_final.sas**

Purpose:

Input: Temporary output from the above .sas file: NLSPHS9

Steps:

* Creating NLSPHS2014 with population variable embedded in it
* Imputing missing populations using google & uscensus.gov
* Subsetting data into large and small jurisdictions

Output:

C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\NLSPHS2014\_LARGE.dta

C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\NLSPHS2014\_SMALL.dta

**X:\xDATA\NLSPHS 2014\Analysis\Github\NLSPHS\modifyingwts.sas**

Purpose:

Assigning weights to small size jurisdiction

Formula used: Weights=1/selection probability

(For detail refer to the .sas file)

Output: X:\xDATA\NLSPHS 2014\Analysis\NLSPHS\_small\_wts\_adj\_frm\_large.dta

**X:\xDATA\NLSPHS 2014\Analysis\Github\NLSPHS\weights for large LHDs.do**

Purpose:

To assign weights for large size jurisdictions.

Steps:

Compute selection probabilities for large size jurisdictions in each stratum (Region and population categories)

. tab nlsphs strata, m

| strata

nlsphs | 1 2 3 4 5 6 | Total

-----------+------------------------------------------------------------------+----------

1 | 25 84 23 18 34 15 | 497

. | 14 13 2 10 13 9 | 134

-----------+------------------------------------------------------------------+----------

Total | 39 97 25 28 47 24 | 631

| strata

nlsphs | 7 8 9 10 11 12 | Total

-----------+------------------------------------------------------------------+----------

1 | 26 130 51 4 57 29 | 497

. | 21 27 8 6 7 4 | 134

-----------+------------------------------------------------------------------+----------

Total | 47 157 59 10 64 33 | 631

| strata

nlsphs | . | Total

-----------+-----------+----------

1 | 1 | 497

. | 0 | 134

-----------+-----------+----------

Total | 1 | 631

Decision rules:

|  |
| --- |
| if strata==1 then SelecProb=25/39 |
| if strata==2 then SelecProb=84/97 |
| if strata==3 then SelecProb=23/25 |
| if strata==4 then SelecProb=18/28 |
| if strata==5 then SelecProb=34/47 |
| if strata==6 then SelecProb=15/24 |
| if strata==7 then SelecProb=26/47 |
| if strata==8 then SelecProb=130/15 |
| if strata==9 then SelecProb=51/59 |
| if strata==10 then SelecProb=4/10 |
| if strata==11 then SelecProb=57/64 |
| if strata==12 then SelecProb=29/33 |

Adjust the weights of large size jurisdictions whose population was <100000 in 2013

Output: U:\NLSPHS2014\Analysis\AnalyticalFiles\NLSPHS\_large\_wts\_adj\_frm\_small.dta

**X:\xDATA\NLSPHS 2014\Analysis\Github\NLSPHS\full NLSPHS with weights.do**

Purpose:

To normalize weights for national comparison

Formula: wt\_adj=pw/r(mean), where pw is the probability weights for large and small size jurisdictions calculated above and r(mean) is the mean for pw

Output: X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS\_2014\_wts\_adj.dta

**X:\xDATA\NLSPHS 2014\Analysis\Github\NLSPHS\manage previous waves of data.do**

Purpose

Impute missing nacchoids

Bringing in non-response observations for No response (and, not sampled in 2006, 2012)

Input:

U:\NLSPHS2014\Analysis\AnalyticalFiles\nlsphs\_tot.dta

Output:

X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS\_2012\_noresp.dta

**X:\xDATA\NLSPHS 2014\Analysis\Github\NLSPHS\final.do**

Purpose:

To get final all 4 waves nlsphs data

Output:

U:\NLSPHS2014\Analysis\AnalyticalFiles\NLSPHS\_98061214.dta

**X:\xDATA\NLSPHS 2014\Analysis\Github\NLSPHS\Getting FIPS for Arm23.do**

Input: U:\NLSPHS2014\Analysis\AnalyticalFiles\NLSPHS\_98061214.dta

Purpose:

To impute county-fips code for those LHDs in NLSPHS so that we can bring in ARF variables (and also NACCHO variables) to conduct cluster analysis for peer grouping

Steps:

* Bringing in previous waves of NLSPHS data
* Use 2013 NACCHO Boundary file to get county-fips code for those in NLSPHS sample

Output: X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS14\_Arm23\_FIPS.dta

X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS\_Small\_Clustering.xls

**X:\xDATA\NLSPHS 2014\Analysis\Github\NLSPHS\Clustering\_Small\_NLSPHS\_556\_final.sas**

Input: X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS\_Small\_Clustering.xls

Purpose:

To generate peer groupings for small size jurisdictions

Steps:

* Use proc fastclus for cluster analysis. The FASTCLUS procedure performs a disjoint cluster analysis on the basis of distances computed from one or more quantitative variables. The distance used in FASTCLUS is Euclidean distances, so the cluster centers are based on least squares estimation. This kind of clustering method is often called a *https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/images/statug_fastclus0001.png*-means model, since the cluster centers are the means of the observations assigned to each cluster when the algorithm is run to complete convergence.
* Variables used for clustering: var: pop13; f0453710 (%white); pctnonwh; f0978112 (per cap income); f1440808 (% below poverty); epi\_direct
* Use Cubic Clustering Criterion to decide on choosing the number of clusters (peers). Please refer to X:\xDATA\NLSPHS 2014\Analysis\AnalyticalFiles\Plots\Clustering decision.xlsx

Output: X:\xDATA\NLSPHS 2014\Analysis\data\SMALL\_PEER\_FINAL.dta

**X:\xDATA\NLSPHS 2014\Analysis\Github\NLSPHS\peergrouping.do**

Purpose: To create final nlsphs dataset with peer grouppings assigned to all NlSPHS sample

Note: **peerclus** is peer grouping obtained from cluster analysis for small size jurisdiction and these peer grouping are different than the peer grouping in variable **peer** for large size jurisdictions

Output: X:\xDATA\NLSPHS 2014\Analysis\data\NLSPHS\_98061214\_final.dta

**C:\Users\LRTI222\Dropbox\Data\NLSPHS\Creating analytical file for SNA.do**

Input: C:\Users\LRTI222\Dropbox\Data\NLSPHS\Analysis\data\NLSPHS\_98061214\_final.dta

Purpose: To create a working data file with unid and variables required to generate SNA variables for NLSPHS 2014 data file

Steps:

* Keep only those variables that will be used in SNA analysis for the survey year 2014
* Generate org1-org19 variables=rowtotal of all the agency dummy variables

Final Output: C:\Users\LRTI222\Dropbox\Data\NLSPHS\Analysis\data\nls4wvs\_forSNA\_14.xlsx

**C:\Users\LRTI222\Dropbox\Data\NLSPHS\NLS4WAVES\_SNA14.sas**

Input: C:\Users\LRTI222\Dropbox\Data\NLSPHS\Analysis\data\nls4wvs\_forSNA\_14.sas7bdat

Macro: prdcphfs1\netusers\lrti222\NLSPHS2014\Network Analysis\bcent.sas

Steps:

As explained in Dr. Mays’ SNA computing .sas file:

* Construct network measures at service level
* Construct density measures,
* Construct degree centrality measures for each service
* Construct degree centrality measure for overall network
* Construct network measures at the organizational level
* Construct density measure for org network
* Construct degree centrality measures for each org
* Construct degree centrality measure for overall network

Use the macro above to calculate betweeness centrality

Merge with NLSPHS data

Final Output: C:\Users\LRTI222\Dropbox\Data\NLSPHS\Analysis\data\nlsphs2014\_final.sas7bdat

**C:\Users\Lava\Dropbox\Data\NLSPHS\ARFNACCHO2NLSPHS\_14.sas**

Purpose:

To link NLSPHS 2014 with NACCHO 2013 and ARF 2013-2014

Steps:

* Read ARF 2014 data
* Splitting the dataset by type of jurisdictions
* Importing nacho 2013 boundary files
* Linking naccho data with the naccho boundary file to get FIPS code
* 1:m match merge data sets from naccho and nacchobound
* Using zip codes with county fips file to impute missing county\_fips in naccho file such that each nacchoid will have its county fips linked
* Based on county fips, linking naccho data with AHRF data
* For each nacchoid we identify max, minimum of cbsa, rucc and urbinfc variables. This will be useful is assigning rural/urban code for multicounties jurisdictions. Presence of at least one urban county makes the multicounty jurisdiction "Urban”
* Computing mean values for ARF variables for multicounties jurisdiction grouped by nachoid; mean(varname) will not account for the missing data when calculating mean
* Removing duplicates based on nacchoid. This will keep those observation at LHD level with mean ARF values for multicounty jurisdictions
* Output: C:\Users\LRTI222\Dropbox\Data\NLSPHS\Analysis\data\NACARFAVGunique.sas7bdat
* Now we need to bring in NLSPHS 2014 data (nlsphs2014\_final) and match-merge 1:1 with NACARFAVGunique
* This linked dataset will have missing for county fips and zip codes for those who did not respond to NACCHO 2013 and were the sample in our original cohort of NLSPHS
* Bringing in nacchoboundary 2010 files to impute missing fips code and ARF variables

Final Output:

**C:\Users\LRTI222\Dropbox\Data\NLSPHS\Analysis\data\nls14nac13arf1314\_final.sas7bdat**

**C:\Users\Lava\Dropbox\Data\NLSPHS\ARFNACCHO2NLSPHS\_12\_missingvalues.sas**

Purpose:

To bring in missing ARF variables for 2012 observations

Steps:

* Read NACCHO 2010 data
* Splitting the dataset by type of jurisdictions
* Importing naccho 2010 boundary files
* Link nacho boundary files with nacho profile and county fips data to impute missing fips code
* Read ARF 2014 data
* Based on county fips, linking naccho data with AHRF data
* For each nacchoid we identify max, minimum of cbsa, rucc and urbinfc variables. This will be useful is assigning rural/urban code for multicounties jurisdictions. Presence of at least one urban county makes the multicounty jurisdiction "Urban"
* Computing mean values for ARF variables for multicounties jurisdiction grouped by nachoid; mean(varname) will not account for the missing data when calculating mean
* Removing duplicates based on nacchoid. This will keep those observations at LHD level with mean ARF values for multicounty jurisdictions
* This dataset "LHD level information from NACCHO 2010 linked with ARF data at county level" is used later: C:\Users\LRTI222\Dropbox\Data\NLSPHS\Analysis\data\NAC10ARF14AVGunique.sas7bdat

**Final Output:**

**C:\Users\LRTI222\Dropbox\Data\NLSPHS\Analysis\data\NAC10ARF14AVGunique.sas7bdat**

**C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\nl980612with NACCHO10full.sas**

Purpose: To bring in NLSPHS\_TOT that has 98-12 data and link with naccho 2010 full dataset to bring in non-respondents & not in sample LHDs from nacho 2010

/\*

| **Year of Mays survey (1998 or 2006)** | | | | |
| --- | --- | --- | --- | --- |
| **yearsurvey** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| **1998** | 2597 | 49.19 | 2597 | 49.19 |
| **2006** | 2441 | 46.24 | 5038 | 95.43 |
| **2012** | 241 | 4.57 | 5279 | 100.00 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | |  | | --- | | **Frequency** | | | | **Table of yearsurvey by survresp** | | | | | --- | --- | --- | --- | | **yearsurvey** | **survresp** | | | | **0** | **1** | **Total** | | **1998** | |  | | --- | | 143 | | |  | | --- | | 354 | | |  | | --- | | 497 | | | **2006** | |  | | --- | | 118 | | |  | | --- | | 236 | | |  | | --- | | 354 | | | **2012** | |  | | --- | | 0 | | |  | | --- | | 241 | | |  | | --- | | 241 | | | **Total** | |  | | --- | | 261 | | |  | | --- | | 831 | | |  | | --- | | 1092 | | | **Frequency Missing = 4187** | | | | |

\*/

Steps:

* Read nlsphs\_tot
* Split nlsphs\_tot into nlsphs9806 and nlsphs12
* merge nlsphs12 with nacho 2012 using nacchoid and Concatenate it with nlsphs9806

Final output:

**C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\nlsphs\_tot\_NAC10.sas7bdat**

**C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\Combining all 4 waves2.sas**

Purpose: To combine nlsphs980612 with nlsphs14 data

Steps:

* Concatenate two datasets C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\Nls14nac13arf1314\_final.sas7bdat and, C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\Nlsphs\_tot\_nac10.sas7bdat

Final output:

**C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\all4wavesnacarfnls.dta**

**C:\Users\Lava\Dropbox\Data\NLSPHS\Creating Final Analytical file2\_rev.do**

Purpose:

To impute missing nacchoid

To create composite and Typology variables for the NACCHO-ARF-NLSPHS linked data

To assign weights and peer values to all NLSPHS respondents

Steps:

* Read C:\Users\LRTI222\Dropbox\Data\NLSPHS\Analysis\data\all4wavesnacarfnls.dta
* Impute missing nacchoid for 1998-2012 data
* Create composite variables using nlsphs data: av, eff, lhd, sta, pct,

\*assessment, policy development, assurance, total

\*sta`j’=maximum(sha`j’ and sao`j’) for those who responded to NLSPHS

\*composite variables for sta

* Create Typology variables: 2014🡪 It uses threshold created by logit/probit command in 1998 data. For better comparability across the waves of NLSPHS we will stick with this threshold for both small and large size jurisdictions.
* Bringing in the adjusted weights calculated in Stata "full NLSPHS with weights.do"
* Input: C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\NLSPHS\_2014\_wts\_adj.dta
* Output: C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\NLSPHSNACCHOARFAll4Waves\_wts.dta
* Bringing in peer values: peer for 1998 used in 2006, 2012, 2014 (large size jurisdiction) and for 2014 we use cluster analysis to create peer groupings

Final Output:

**C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\NLSPHSNACCHOARFAll4Waves\_wts\_peer\_rurb.dta**

**C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\making final data same as nlsphs\_tot.do**

Purpose:

To make the dataset look similar to the previous data in terms of variable names, orders.

To impute missing values for some additional variables

Steps:

* Read C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\NLSPHSNACCHOARFAll4Waves\_wts\_peer\_rurb.dta
* Order the variables as they were ordered in nlsphs\_tot (original data for 980612)
* Impute variables that were missing:
* boh, exp, prexp, emp, curfy, fiscal, msamicro, msametro, rucc, urbinf, incpcap, uninstot, unemp, popdens, expcap, mdpcap, bedcap, pctnonwh, pct65, povpct, collpct, fqhcnum, fqhcppov, admpcap
* Strategy to impute:
  + Use the data available from the most recent ARF variable.
  + For multi-county use the mean values.
* Use “C:\Users\Lava\Dropbox\Research Projects\Dissertation\NLSPHSData\CentralDecentral.xlsx” to impute missing values for central or mixed dummies for each year

Final Output:

**C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\NLSPHSNACCHOARFAll4Waves\_wts\_peer\_rurb2.dta**

**C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\making final data same as nlsphs\_tot\_with missingfor12.do**

(Needbased)

Purpose: To bring in missing data for the year 2012 as we progress further and identify missingness

Steps:

* Read C:\Users\LAVA\Dropbox\Data\NLSPHS\Analysis\NAC10ARF14AVGunique.dta
* As needed, add variables to pull missing data
* Merge with C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\NLSPHSNACCHOARFAll4Waves\_wts\_peer\_rurb2.dta
* Impute variables that were missing similar to the syntax used in C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\making final data same as nlsphs\_tot.do

Final Output:

**C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\NLSPHSNACCHOARFAll4Waves\_wts\_peer\_rurb3.dta**

**C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\bringing sna for previous waves.do**

Purpose:

To bring in some of the sna variables for the previous waves that was missing

To impute county-fips for some missing LHDs

Steps:

* Read C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\lphs\_sna2.dta
* Lookfor duplicates by nacchoid yr\_naccho and keep unique values with complete records for nlsphs variables
* By nacchoid and yr\_naccho merge with C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\NLSPHSNACCHOARFAll4Waves\_wts\_peer\_rurb3.dta
* Output: C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\NLSPHSNACCHOARFAll4Waves\_wts\_peer\_rurb4.dta
* generate dummy variable for metro
* Look for and take care of duplicates if any
* Correct nacchoid for some observations in 2012 based on id1998 id2006 & id2012
* Output: C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\NLSPHSNACCHOARFAll4Waves\_wts\_peer\_rurb5.dta
* Imputing missing county fips for 1997, 2005 and 2010 (n=10) using zipcodes

Final Output:

**C:\Users\Lava\Dropbox\Data\NLSPHS\Analysis\data\NLSPHSNACCHOARFAll4Waves\_wts\_peer\_rurb5.dta**

**X:\xDATA\NLSPHS 2014\Analysis\AnalyticalFiles\data\correcting clusttot and sna for 2014.do**

Observation: Some of the Public Health Typology for 2014 were not appropriately matching with GPM’s results. Need to use corrected typology and sna variables for 2014 and replace all sna and typology variables calculated previously using the corrected syntax.

Purpose: To correct typology and sna variables for 2014 in the dataset

Steps:

* Read X:\xDATA\NLSPHS 2014\Analysis\AnalyticalFiles\data\NLSPHSNACCHOARFAll4Waves\_wts\_peer\_rurb5.dta
* Import nlsphs 2014 raw data as downloaded from REDCap: C:\Users\LRTI222\Dropbox\Data\NLSPHS\Analysis\data\NLSPHS2014\_raw.xlsx

Then use **C:\Users\LRTI222\Dropbox\Data\NLSPHS\Analysis\nlsphs2014\_2\_withSNAandTBD\_UNID.sas**

The output from this file will be:

C:\Users\LRTI222\Dropbox\Data\NLSPHS\Glen\nlsphs2014\_SNAandCLUSTTOT.dta

This data will have TBD\_unid as merging variable

* Merge the two datasets and replace all variables in 2014 for PH typology and SNA

Final Output:

**X:\xDATA\NLSPHS 2014\Analysis\AnalyticalFiles\data\NLSPHSNACCHOARFAll4Waves\_wts\_peer\_rurb6.dta**

**X:\xDATA\NLSPHS 2014\Analysis\AnalyticalFiles\data\correcting zip using \_rurb6 data.do**

Purpose: To impute some of the missing zip code variables for those who were in the sample of NLSPHS

Steps:

* Read X:\xDATA\NLSPHS 2014\Analysis\AnalyticalFiles\data\NLSPHSNACCHOARFAll4Waves\_wts\_peer\_rurb6.dta
* Use naccho profile surveys, listofsample2014.xlsx and google to impute the missing zip codes

Final Output:

**X:\xDATA\NLSPHS 2014\Analysis\AnalyticalFiles\data\NLSPHSNACCHOARFAll4Waves\_wts\_peer\_rurb7.dta**

Addendum:

Steps in Importing files from REDCap

* Login to REDCap
* Select the project that you want to import data from
* Under Application, Click “Data Exports, Reports, and Stats”
* Click “Export Data” tab from All data
* Select the export format that you want and click “Export data”
* Wait while your data is being exported
* Click the icon (format of your choice) to download the data
* Import into SAS with appropriate format and informat commands

Address to Lizeth’s concern:

1. NACCHOID crosswalk
2. Can you please indicate the location of your excel sheets on the x drive

X:\PH PBRN National Coordinating Center\Research Projects\NLSPHS\2012 Data\NACCHO 2010 and 2013\NACCHO 2010and2013 ID walk\_updated.xlsx

b.      Can you create a table indicting the following:

                   i.      The number of observations with same ID from 2010 and 2013: 2095

                   ii.      The number of observation that did not match: 12

                   iii.      What criteria you used to make decisions:

Matched based on city, state name. For those that did not match, used nacho profile data from webscrapping and google to identify the name of the LHD

2)      ARF to NACCHO

a.       Can you please create a table indicting the following:

                       i.      How many observations in the sub-county category

                       ii.      How many observations in the single-county category

                       iii.      How many observations in the sub-county category

geographic |

jurisdiction | Freq. Percent Cum.

-------------+-----------------------------------

city | 272 13.60 13.60

city-county | 4 0.20 13.80

county | 1,467 73.35 87.15

multi-city | 74 3.70 90.85

multi-county | 183 9.15 100.00

-------------+-----------------------------------

Total | 2,000 100.00

                         iv.      How many dropped out

None, rather we need to add those that were in our 1998 sample but did not respond to NACCHO 2013 profile survey

                          v.      Decision rules you used

NACCHO 2013

City

City-county

County

Multi-County

Multi-City

Boundary

Boundary

Boundary

Boundary

Boundary

1:1 merge / 1:m merge based on nacchoid

NACCHO 2013 with constituent counties for multi-counties/cities and county-fips

ARF 2013-2014

1:1 merge based on county fips

City

1:1 match

City-county

1:1 match

County

1:1 match

Multi-County

Aggregate to mean

Multi-City

Aggregate to mean

Remove duplicates based on nacchoid

NACCHO-ARF linked data at LHD level

C:\Users\LRTI222\Dropbox\Data\NLSPHS\Analysis\data\NACARFAVGunique.dta

Description of the steps

* Splitting the dataset by type of jurisdictions
* Importing nacho 2013 boundary files
* Linking naccho data with the naccho boundary file to get FIPS code
* 1:m match merge data sets from naccho and nacchobound
* Using zip codes with county fips file to impute missing county\_fips in naccho file such that each nacchoid will have its county fips linked
* Based on county fips, linking naccho data with AHRF data
* For each nacchoid we identify max, minimum of cbsa, rucc and urbinfc variables. This will be useful is assigning rural/urban code for multicounties jurisdictions. Presence of at least one urban county makes the multicounty jurisdiction "Urban”
* Computing mean values for ARF variables for multicounties jurisdiction grouped by nachoid; mean(varname) will not account for the missing data when calculating mean
* Removing duplicates based on nacchoid. This will keep those observation at LHD level with mean ARF values for multicounty jurisdictions

1. Descriptive stats on ARF/NACCHO linkage

Do file: X:\xDATA\NLSPHS 2014\Analysis\AnalyticalFiles\data\Quality Checks.do

Please refer to the stata output file: X:\xDATA\NLSPHS 2014\Analysis\AnalyticalFiles\ QC\_Output\_NACARFAVG.pdf

c.       Please upload the DO files for the following to GitHub

                        i.      DO files used for merging the data to NACCHO or ARF

3)      NLSPHS

a.       Syntax for creating the NLSPHS from the raw data

Please refer to the flow-chart presented as table and its detail above.

1. Upload DO files used for creating the calculated variables in NLSPHS to GitHub

4)      For entire merge- zip code

a.       Table with the following:

                   i.      How many observations have a zip code

                   ii.      How many observations with missing zip code

Final data (NLSPHSNACCHOARFAll4Waves\_wts\_peer\_rurb7.dta) with updated zip code is located in

X:\xDATA\NLSPHS 2014\Analysis\AnalyticalFiles\data\

. count

9,307

count if zip\_rec=="" & survresp!=.

5

list nacchoid state2014 yearsurvey if zip\_rec=="" & survresp!=.

+--------------------------------+

| nacchoid sta~2014 yearsu~y |

|--------------------------------|

242. | 1998 |

1892. | VA004 VA 1998 |

4228. | 2006 |

7146. | 2012 |

8408. | VA004 VA 2014 |

+--------------------------------+

There is county\_fips code assigned to each LHDs.

. list nacchoid state2014 yearsurvey if county\_fips=="" & survresp!=.

+--------------------------------+

| nacchoid sta~2014 yearsu~y |

|--------------------------------|

242. | 1998 |

1892. | VA004 VA 1998 |

4228. | 2006 |

5202. | ZZ001 2012 |

7146. | 2012 |

|--------------------------------|

8408. | VA004 VA 2014 |

+--------------------------------+

                  iii.      Decision rules applied

For those LHDs with missing zip codes we used following datasets in this order to address missingness

NACCHOID for each waves

NLSPHS complete data for each waves:

If the same zip code is consistently occurring in any 3 waves then impute the missing zip code to be the same for the remaining wave

If the zip code is same for ’98 and ’14 then use the same zip code for ’06 and ‘12

If the zip code is same for ’98 and ’12 then use the same zip for ‘06

If the zip code is same for ’06 and ’14 then use the same zip for ‘12

Based on the Name of the LHD, city and state google for the remaining missing zip codes