Documentation of Staging
(Analytical File Creation and Validation)
for Local Control (LC) Analysis of data from Chay, Dobkin and Greenstone, *J Risk Uncertainty*, 2003.

[1] analysis_long.jmp = Maximal Dataset for Chay et al. 2003 paper (19,283 x 69 table); provided to us by Carlos Dobkin in CSV Format.

Delete the 15,923 rows with CAAA compliance flag: XTSP71 = missing or blank.

This leaves 3,360 rows with XTSP71 = 0 (attainment) or 1 (nonattainment).

[2] XTSP71_Labeled.jmp = 3,360 x 70 table (rows: 560 counties x 6 years) with ST_County labels for FIPS locations. Labels can be Table > Joined from either file FIPS_ST_County.jmp (560 rows, two "unknown" county labels) or FIPS_Labels.jmp (3,142 rows, with blanks for "locations" 5097 and 12025.)

XTSP71 = 0 or 1 is the Indicator (FLAG) variable for either:
> [a] XTSPGM > 75 in 1971 (CAAA attainment threshold), or
> [b] Second highest hourly reading is over 260 in 1971.

286 counties have XTSP71 = 1 => CAAA COMPLIANCE NONATTAINMENT.
274 counties have XTSP71 = 0 => CAAA COMPLIANCE ATTAINMENT.

Available X-Confounder Variables (Columns):
21 County level economic measures from REIS: EARN EMPLOY EPOP FAMASS FSTAMP IMAINT MANF MANFR MCARE MILMED MTSPAR MTSPGM OTHINC PCINC POP PUBMED RETIRE SSI TMED TRANSF or UI.
REIS_OK: indicator variable that REIS data is not missing for 1969-1974 (2,424 county-years)
REIS_OK_71_72: indicator variable that REIS data is not missing for 1971-1972 (2,748 county-years)

[3] LC_Vars.jmp = 25 Columns from XTSP71_Labeled.jmp selected for use as basis for LC Analyses.
Missing (blank) values in ST_County label for 2 locations are reset to either "AR_Montgomery" (location = 5097) or "FL_Unknown" (location = 12025).
First 4 columns: FIPS location code, ST_County label, year (69 to 74) and XTSP71 Flag (0 or 1).

Four TSP-exposure or Y-mortality outcome Columns:
XTSPAR: Since counties can have > 1 monitor, max of arithmetic means of TSP
      monitors in county.
XTSPGM: Since counties can have > 1 monitor, max of geometric means of TSP
      monitors in county.
deaths_10k_age50_00: Deaths in year per 10,000 residents over age 50.
deaths_10k_age65_84: Deaths in year per 10,000 residents between ages 65 and 84.
NOTE: It is currently more common to report mortality rates in deaths per 100K
      population.

Seventeen Confounder Columns:
Four of the 21 possible vars in XTSP71_Labeled.jmp (EMPLOY, MANF, PCINC &
      POP) are excluded for being "absolute" counts (county "size" measures)
      rather than "relative" rates.
Information From: Regional Economic Information System (REIS) from the US
      Bureau of Economic Analysis (BEA).

[4] OutcomePeriods.jmp = 3,360 x 7 table created primarily by deleting columns from
      LC_Vars.jmp.

      INCLUDED: location and year from LC_Vars.jmp
        New variable: period = 1 for years 72 to 74, or
                          = 0 for years 69 to 71.
        Plus 4 outcome variables from LC_Vars.jmp are kept.
        (MTSPAR, MTSPGM, deaths_10k_age50_00 and deaths_10k_age65_84)

[5] OutcomeChange1.jmp = 1,120 x 6 table created from OutcomePeriods.jmp using Table
      > Summary to compute Outcome Means by Location and Period (0 or 1)

[6] OutcomeChange2.jmp = 560 x 13 table created from OutcomeChange1.jmp using Table
      > Split on Period

    Next, Create 4 new columns for Change in Outcome (2 Air Quality and 2 Mortality) of
    the form Period 1 (72 to 74) minus Period 0 (69 to 71). All four change measures are
    to be interpreted as "the Lower (the more negative) the value, the Better" the
    outcome.

[7] ConfoundCentroids.jmp = 560 x 20 table created from LC_Vars.jmp using Table >
      Summary to compute Within-Location Confounder Mean Values.

      INCLUDED Vars: Location, ST_County, XTSP71 and 17 REIS variables: EARN
        EPOP FAMASS FSTAMP IMAINT MANFR MCARE MILMED MTSPAR
        MTSPGM OTHINC PUBMED RETIRE SSI TMED TRANSF & UI.

[8] LC_Staged.jmp = 560 x 24 table created by a <mark>Table > Join</mark> operation with tables matched on Location.

The tables joined are 4 columns of **OutcomeChange2.jmp** with all 20 columns of **ConfoundCentroids.jmp**.

Names and Definitions of the 24 Variables included in table LC_Staged.jmp:

No.	Name	Description *** Current Guesstimate ***
1	Location	FIPS (State Code) x 1000 + County Code
2	ST_County	Character (Label) for FIPS codes
3	XTSP71	Binary Treatment Indicator (1 => Nonattainment, 0 => Attainment of CAAA Compliance in 1971)
4	ChgTSPAR	Change in MTSPAR [(Mean 72-74) minus (Mean 69-71)]
5	ChgTSPGM	Change in MTSPGM (Geometric Mean in $\mu g/m^3$)
6	ChgMortAdult	Change in deaths_10k_age50_00
7	ChgMortElderly	Change in deaths_10k_age65_84
8	Mean(EARN)	Location Centroid for Earnings / Person / Year (69 to74) in $
9	Mean(EPOP)	Location Centroid (69 to 74) for Fraction Employed Persons
10	Mean(FAMASS)	Location Centroid (69 to 74) for Family Assistance
11	Mean(FSTAMP)	Location Centroid (69 to 74) for Food Stamps
12	Mean(IMAINT)	Location Centroid (69 to 74) for Income Maintenance
13	Mean(MANFR)	Location Centroid (69 to 74) for Fraction Manufacturing
14	Mean(MCARE)	Location Centroid (69 to 74) for Medicare Payments
15	Mean(MILMED)	Location Centroid (69 to 74) for Military Medical Benefits
16	Mean(MTSPAR)	Location Centroid (69 to 74) for TSP Arithmetic Mean $\mu g/m^3$
17	Mean(MTSPGM)	Location Centroid (69 to 74) for TSP Geometric Mean $\mu g/m^3$
18	Mean(OTHINC)	Location Centroid (69 to 74) for Other Income Benefits
19	Mean(PUBMED)	Location Centroid (69 to 74) for Public Medical Assistance
20	Mean(RETIRE)	Location Centroid (69 to 74) for Retirement Benefits
21	Mean(SSI)	Location Centroid (69 to 74) for Social Security Payments
22	Mean(TMED)	Location Centroid (69 to 74) for Total Medical Payments
23	Mean(TRANSF)	Location Centroid (69 to 74) for Transfer Payments
24	Mean(UI)	Location Centroid (69 to 74) for Unemployment Insurance

NOTE: File LC_Staged.JMP was first created and validated in February 2015.

Modified Variable NAMES: File CAAA_Comply_Cohorts.JMP was created and validated in June 2015 to greatly <u>simplify the variable names</u> originally used in file LC_Staged.JMP.

24 Variables in table CAAA_Comply_Cohorts.jmp:

No.	Name	Description *** Current Guesstimate ***
1	Location	FIPS (State Code) x 1000 + County Code
2	ST_County	Character (Label) for FIPS codes
3	CAAA_Comply	Labels: Nonattainment or Attainment of CAAA Compliance in 1971)
4	ChgTSPAR	Change in MTSPAR [(Mean 72-74) minus (Mean 69-71)]
5	ChgTSPGM	Change in MTSPGM (Geometric Mean in $\mu g/m^3$)
6	ChgMortAdult	Change in deaths_10k_age50_00
7	ChgMortElderly	Change in deaths_10k_age65_84
8	EARN	Location Centroid for Earnings / Person / Year (69 to74) in $
9	EPOP	Location Centroid (69 to 74) for Fraction Employed Persons
10	FAMASS	Location Centroid (69 to 74) for Family Assistance
11	FSTAMP	Location Centroid (69 to 74) for Food Stamps
12	IMAINT	Location Centroid (69 to 74) for Income Maintenance
13	MANFR	Location Centroid (69 to 74) for Fraction Manufacturing
14	MCARE	Location Centroid (69 to 74) for Medicare Payments
15	MILMED	Location Centroid (69 to 74) for Military Medical Benefits
16	MTSPAR	Location Centroid (69 to 74) for TSP Arithmetic Mean $\mu g/m^3$
17	MTSPGM	Location Centroid (69 to 74) for TSP Geometric Mean $\mu g/m^3$
18	OTHINC	Location Centroid (69 to 74) for Other Income Benefits
19	PUBMED	Location Centroid (69 to 74) for Public Medical Assistance
20	RETIRE	Location Centroid (69 to 74) for Retirement Benefits
21	SSI	Location Centroid (69 to 74) for Social Security Payments
22	TMED	Location Centroid (69 to 74) for Total Medical Payments
23	TRANSF	Location Centroid (69 to 74) for Transfer Payments
24	UI	Location Centroid (69 to 74) for Unemployment Insurance