MANAGING LONGITUDINAL RESEARCH STUDIES:

POPULATION MASTER FILES

By

Linda L Remy, MSW PhD

Ted Clay, MS

Rita Shiau, MPH

Geraldine Oliva, MD MPH, Director
Jennifer Rienks, PhD, Associate Director
Linda L Remy, MSW PhD, Research Director

UCSF Family Health Outcomes Project
500 Parnassus Ave. Room MU-337
San Francisco, California 94143-0900
Phone: 415-476-5283
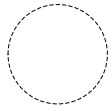Fax: 415-476-6051
Web: fhop.ucsf.edu

October 2020

TABLE OF CONTENTS

Table of Legends

From program creating incoming file or name of next program using file

Excel file input or output
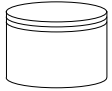
SAS file input or output
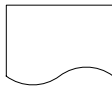
Flat or text file input or output

PROGRAM NAME SAS program with brief description of steps

ACRONYMS

ACS	American Community Survey
AIAN	American Indian/Alaska Native
API	Asian/Pacific Island
CDC	Centers for Disease Control
DOF	California Department of Finance
FHOP	Family Health Outcomes Project
FIPS	Federal Information Processing Standards.
LST	Listing file SAS produces to show results of a program
NCHS	National Center for Health Statistics
UCSF	University of California, San Francisco
WHO	World Health Organization
ZCTA	ZIP-Code Tabulation Area

POPULATION MASTER FILES

This document describes methods the UCSF Family Health Outcomes Project (FHOP) uses to prepare longitudinal population master files to use as denominators. We assume that the user of this document has read other documents describing the foundation of our methodology [1-3].

The Demographic Research Unit of the California Department of Finance (DOF) distributes county-level population estimates and projections. The US Census Bureau distributes decennial multi-level population estimates. In cooperation with the Census Bureau, the National Center for Health Statistics (NCHS) distributes longitudinal bridged-race county-level population files. We population data from GEOLYTICS, ESRI and others.

OVERVIEW

To support its longitudinal research, FHOP maintains a series of non-confidential population files distributed by the Census Bureau, NCHS, the DOF, GEOLYTICS, and ESRI. At the writing of this document, we have: county-level DOF files from 1970 through 2060 projections, multi-level 1970-2010 Census files, NCHS 1990-2015 bridged race, county-level population files, 1970-2010 and 2016 multi-level, longitudinally normalized census files from GEOLYTICS [4] and various intercensal ZIP-level population estimates from commercial vendors. For example, ESRI distributes small area estimates as part of ArcMap.

Processes and resources described here will help users of other population files. We are making this basic methodology and its associated software public to help researchers understand the nature of data management for complex longitudinal research. This also will provide a background to users of our longitudinal DataBook products and readers of FHOP studies using external population denominators.

Contracts with various departments in the State of California require us to provide funding agencies with an annual backup of all programs, logs, listings, and output files. This creates an audit trail of our work. Since we do not know where the programs and/or resulting files will be used, we try to write code that will run in any environment and provide as much documentation as possible. All work is in SAS, assisted by Microsoft Excel and Visio. Because public funding supports development of these programs, they are in the public domain. This is why we are making them available.

DEPARTMENT OF FINANCE COUNTY-LEVEL POPULATION

Obtaining the Data

The California State Legislature designated the DOF Demographic Research Unit as the single official source of demographic data for state planning and budgeting. When FHOP contracts with state agencies to produce data products, we are required to use DOF population estimates. These data are not confidential. Using them does not require an approved research protocol.

DOF produces total population and housing estimates as of January each year, at the state, county, and city level. July population estimates are county-level with annual estimates by age, sex, and race/ethnicity. Analyses needing lower-level population estimates must use other data sources. Race/ethnic groupings changed over time, and preparing for longitudinal analyses must address these changes.

DOF distributes the 1970-1989 files as annual text files [5]. It distributes the 1990-1999 files as annual excel files by county [6]. The 2000-2009 data originally was distributed in excel files by year, but now it is distributed as a text file that is too large to import into Excel [7]. Population projections currently are available for 2010 through 2060 [8]. Again, the projection file is too large to import into Excel.

Be aware that DOF population estimates sometimes have been inaccurate. On occasion, we have identified problems with their estimates and they have corrected them. Sometimes DOF has updated a given sequence of years without changing the publication date on the web. Be sure to check the file creation date internal to the ZIP file. DOF issued three population projection updates in 2017, with the same external ZIP file name but different creation dates in the internal files. For these and other reasons, we visit the DOF website every time we need to use their population estimates, to confirm that we have the most current versions.

Preparing the Population Files

Since beginning to work with DOF data in 1999, we have had to update source files numerous times. For example, in 2010, we updated the period 1970-2010. In 2013, DOF updated the 2000-2010 intercensal estimates and released its first post-censal projections based on the 2010 Census. Later that year, DOF removed 2010 from the intercensal estimates and moved 2010 into the projections. In late 2014 and 2017, DOF again updated projections. We now feel it is likely that the 1970-2009 population estimates will remain stable.

Figure 1.   Preparing 1970-2007 DOF Data
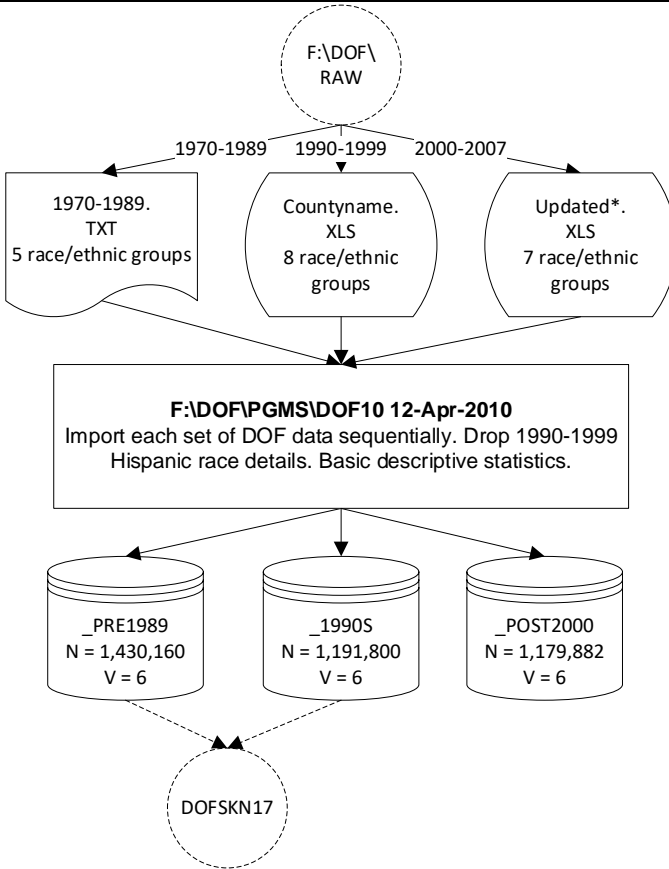
Figure 1 summarizes the program to import the 1970-2007 DOF population data. The program DOF10 (we did this in 2010) sequentially reads each set of files, creates and labels standard variables. At the end of importing each set of files, basic descriptive statistics are reported.

Note that the three sets of files have different numbers of race/ethnic categories. The program standardizes the race/ethnic variables, defining them consistently over time.

We address problems and solutions associated with longitudinal changes in race/ethnic categories elsewhere [9].

Figure 2. Revising 2000-forward DOF Data

F:\DOF\ RAW

Intercensal_2000-2010_DBInput_csv.txt

P3_Complete.csv
Last update 20170620

F:\DOF\PGMS\INTERCENS0010
26-Apr-2013 Rev. 12-Dec-2017
This replaces DOF10_POST2000. We were having problems with bridging. In searching out the cause, found that DOF updated the intercensal file without changing the external documentation. Basic descriptive statistics. Had to revise and re-run again with 2017 update

F:\DOF\PGMS\PROJ1060_2017
03-Dec-2017
As of this date, DOF has updated the 2010-forward projections 5 times. This reflects the most recent iteration. Race/ethnicity codes differ from previous years. Uses FIPS code rather than County. We imported the revised text file and made standard variables. Basic descriptive statistics

INTERCENS0010
N = 1,001,112
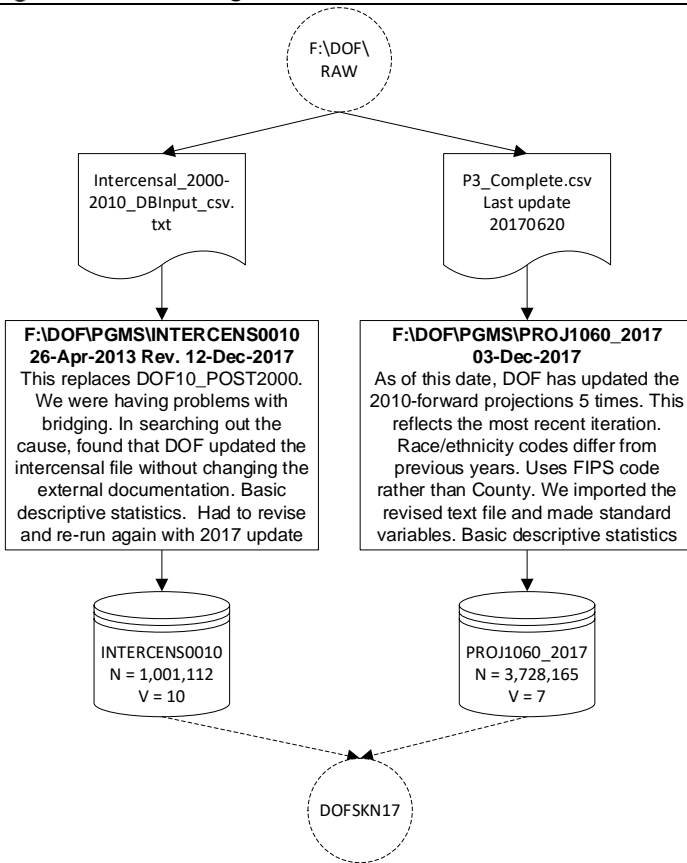V = 10

PROJ1060_2017
N = 3,728,165
V = 7

DOFSKN17

Figure 2 summarizes the programs to re-import the 2000-forward DOF population data. INTERCENS0010 reads replaced files into SAS, creates and labels standard variables, and reports basic descriptive statistics.

We thought this file finally was stable when we ran it in Apr-2013, but discovered in Dec-2017 that DOF updated it a few weeks after we downloaded the file from the website. The ZIP file containing the data had same name, but the text file in the ZIP file had a different creation date and different structure for the year variable. These changes required yet another program re-write. This time we kept the same name.

We have updated the projection program as DOF makes updates, currently known 5 times including 3 times in 2017. Projections are for the years 2010-2060. This program creates and labels standard variables, and reports basic descriptive statistics.

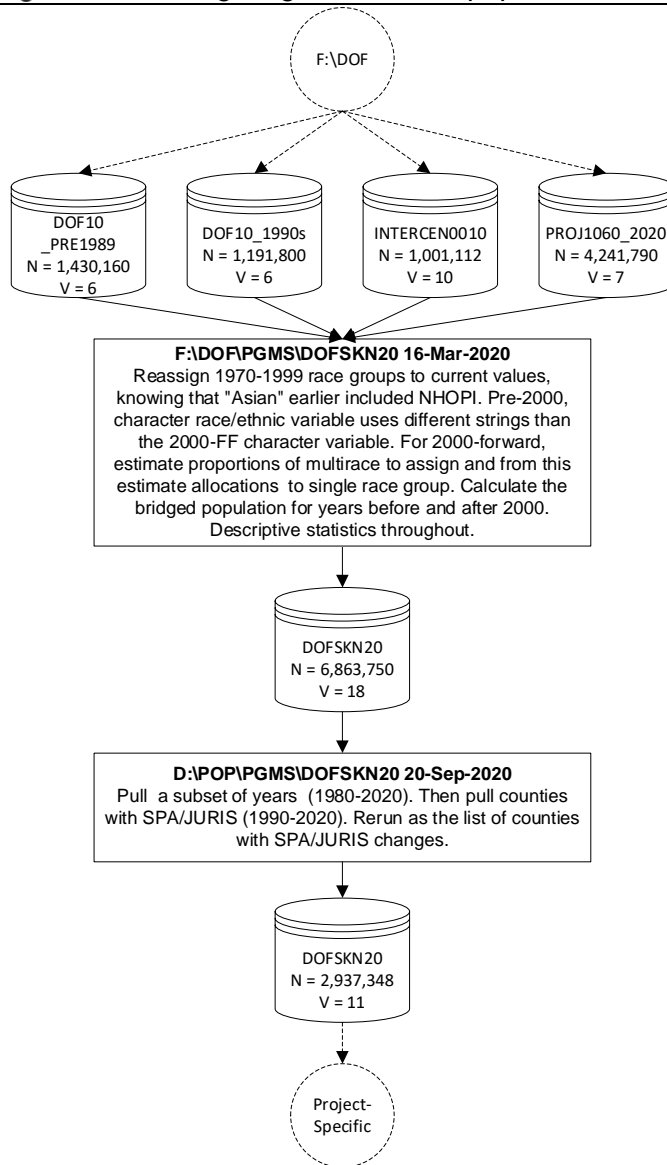## Figure 3.   Making longitudinal DOF population file

Figure 3 summarizes the last steps to make a county-level population file with longitudinally consistent groups.

Within each file, we convert incoming character variables for sex and race/ethnicity into numeric variables, cognizant of when definitions change.

From 2000 forward, we make a single race/ethnic variable consistent with the pre-2000 race/ethnic variable. Drawing on methods described elsewhere [10], we proportionally allocate multi-race to five defined groups (White, Black, Hispanic, Asian/Pacific Island (API), American Indian/Alaska Native (AIAN)) that existed before 2000.

Next, we move from the drive/directory where we make the DOF population files to the drive/directory where we make population files that we use for our work. Here, we pull records from 1980 through 2020. We pull 1980 forward because 1980 is the first year we have longitudinal population health data.

Los Angeles County divides itself into eight Service Planning Areas (SPA). Long Beach, Pasadena, and Berkeley are state-designated Health Jurisdictions. Some counties have asked us to provide smaller-area products based on supervisor districts or other locally-defined geographies. We pull the county-level population for these areas, which we use to interpolate intercensal small-area populations, discussed later in this document.

NATIONAL POPULATION ESTIMATES AND PROJECTIONS

FHOP has the 1970 through 2010 Census files. We have all census years for the State of California, and additional states for some years. Here, we focus on the 1990 through 2010 California Census SF1 files. In 1990, the Census Bureau added a new summary level, the ZIP-

code, which it named ZCTA. Unlike other Census summary levels, ZIP-codes do not have defined geographic areas, because the US Postal Service (USPS) defines them administratively. From time to time, the USPS changes their boundaries, splits them, or combines them to facilitate mail delivery. ZIP-codes also cross county boundaries. For an in-depth discussion of this, we suggest the document that describes how we make our geography master [11]. For purposes of our research, we focus on state, county, place, census tract, and ZIP-code population levels.

From Geolytics, we have 1970 to 2010 multi-level census files with geography standardized longitudinally. We also have their 2016 intercensal population estimates, which we address here, and earlier intercensal estimates from other commercial vendors.

Additionally, we have the county-level NCHS bridged race files for 1990 through 2019, as well as the NCHS county-level crosswalk file, which we use to bridge race/ethnicity [12].

Obtaining Census Data

In the early 1990s, we purchased 1990 California Census data directly from the Census Bureau, distributed at that time on a CD as dBase files. These data now are available on the web [13,14]. Traditionally, the SF1 file has 100 percent population counts at all Summary Levels, with the SF3 files based on samples. For the 1990 Census, the Bureau distributed ZIP-level data only in the SF3 file, with 100 percent population estimates and samples for the other variable sets.

The 2000 Census data are available on the web as text files [15,16]. Programs to read the 2000 data into SAS also are available [17,18], which we modified for our purposes. The Census Bureau changed race/ethnic group definitions with the 2000 Census.

For 2010, the Census Bureau only distributed the SF1 file, and switched the equivalent of the SF3 file to the American Community Survey (ACS). The 2010 SF1 data is available on the web [19], and we found a program to read it into SAS [20], again modified for our purposes.

We concentrate this discussion on reading the SF1 files into SAS, subsetting the population variables, and preparing the population variables for further use. The Census Bureau distributes data as a series of upwards of 80 distinct files, depending on the year. When we read Census files into SAS, we first pull in all files and all summary levels, and do not label variables. When we subset population variables, we only keep the state, county, place, tract, and ZIP levels.

Preparing the Census Population

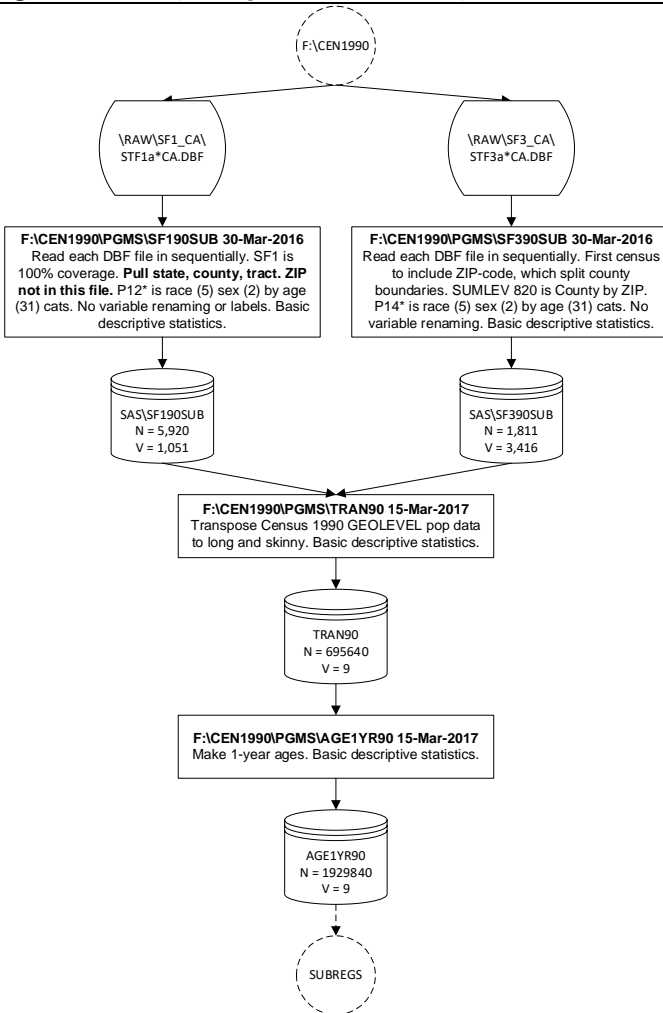Figure 4. Preparing the Census Population

Figure 4 summarizes steps to prepare a multi-level population file, using the 1990 Census as an example.

The 1990 data we received from the Census Bureau only included the geographic levels that we requested. In this instance, we pulled all files directly into our "subset" file, without a first step of reading all Summary Levels into SAS. As mentioned earlier, the 1990 SF3 only had ZIP-level data.

The program TRANyy (e.g., TRAN90) pulls each census set of race by age by sex variables and transposes the structure to coincide with the structure of the DOF files: one record per each year, geography level, sex, age, and race/ethnicity combination.

The program AGE1YRyy converts multi-age groups into one-year estimates. For example, if the age group is 1 to 4, the program assigns an equal proportion of each age to the one-year estimate.

We repeat this sequence of programs for each census period from 1970 forward. We use SF3 files as needed for specific projects and do not discuss its use in this document.

NCHS Bridged Race Population

In a collaborative arrangement with the Census Bureau, NCHS releases county-level bridged-race population that it re-estimates annually for the United States to use in calculating national vital rates. Bridged data are available in single-year age from 1990 forward for all counties in the US [12]. Where possible, we prefer to use these files rather than DOF files because of their national comparability.

Race bridging refers to making data collected using one set of race categories consistent with data collected using a different set of race categories, to permit estimation and comparison of

race-specific statistics at a point in time or over time. More specifically, race bridging is a method used to make multiple-race and single-race data collection systems sufficiently comparable to permit longitudinal estimation and analysis of race-specific statistics. The goal of bridging is to approximate the size of single-race groups rather than to approximate how each individual would have responded to the traditional single-race question [21,22]. The bridging methodology is described elsewhere [10].

NCHS estimates result from bridging the 31 race categories used in Census 2000 and Census 2010, as specified in the 1997 Office of Management and Budget (OMB) standards for the collection of data on race and ethnicity [23], to the four race categories specified in the 1977 OMB standards (White, Black, API, AIAN) by Hispanic and non-Hispanic [24].
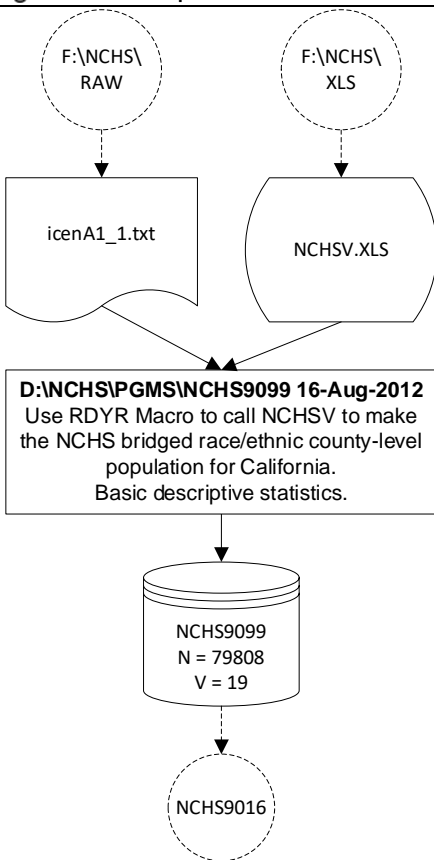
Figure 5. Prepare NCHS 1990-1999 files

Figure 5 summarizes steps to prepare 1990-1999 NCHS data. NCHS distributes these years as zipped text files by sets of states. Thereafter, all years are available as single SAS files with all states included. On our Working Drive, in this case D:\NCHS, we pull and store California data from the Non-Confidential (F) Drive.

To prepare the 1990-2000 data, we set up the RDYR macro to give control to the NCHSV spreadsheet. We describe this process elsewhere [3].

Figure 6. Prepare NCHS 2000-2019 files
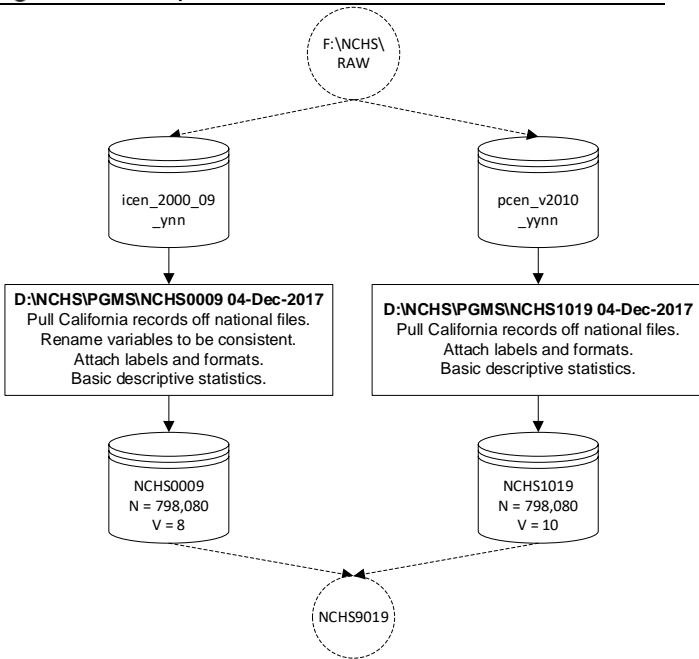
NCHS distributes these years of data as text or SAS files. We downloaded the annual SAS files. The variables are slightly different in the two sets. Note that the outgoing file NCHS1019 becomes obsolete annually. As NCHS updates the post-censal projections, a new program is run for each annual revision of the series.

Figure 7. The NCHS population files
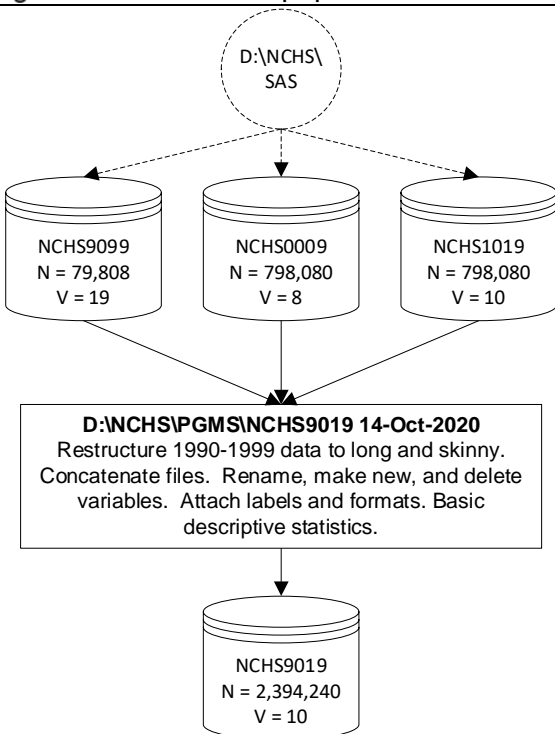
At this point, we put the sets of intercensal estimates together. Note that the 1990-1999 file is "fat", with annual population having a variable name "POPyyyy". To be consistent with our methods for the California DOF population files, we make this file long and skinny before concatenating the other files. The resulting product has one record per year, county, single-year age up to 85, race, and Hispanic ethnicity.

To distribute undefined (multi-race, other, unknown) race groups, as found in our population health datasets, NCHS assigns single race randomly, using local distributions for Hispanic, sex, and age. NCHS provided FHOP with the file it uses to bridge multiple-race data into the single race format for vital statistics reporting [25]. This file and the SAS macro to carry out the reassignment is available upon request.

Intercensal Population Estimates

About mid-way through each census period, we obtain commercial intercensal population estimates to enable us to estimate sub-county population changes. For the current intercensal period, we turned to Geolytics. We purchased two national files, one structured like the standard SF1, and the other like the P5 table.
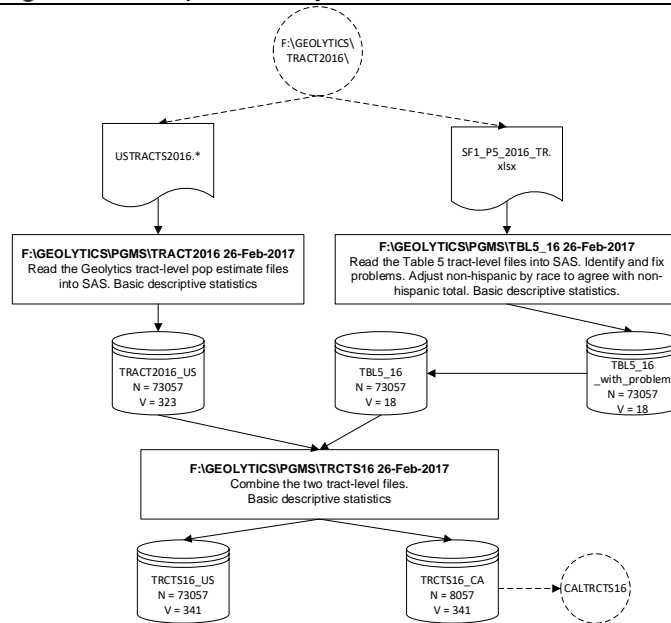

Figure 8.   Prepare Geolytics 2016 estimates

Figure 8 summarizes steps to prepare Geolytics intercensal estimates data, starting with SF1 file, USTRACTS2016. The large number of fields (N = 323) reflects that race by sex by ethnicity by age are arrayed across the file rather than vertically. Geolytics also has a set of fields to identify other geographic levels (County, ZIP, Place, etc) for data summary.

The SF1 P5 file is a summary of ethnicity and race over all age groups. In addition to the geography variables, it has counts for Total, Not Hispanic by race, and Hispanic by race.

The program TRCTS16 puts the two files together, saves the national file then pulls out the California records.

Figure 9. Bring back lost ZIPS
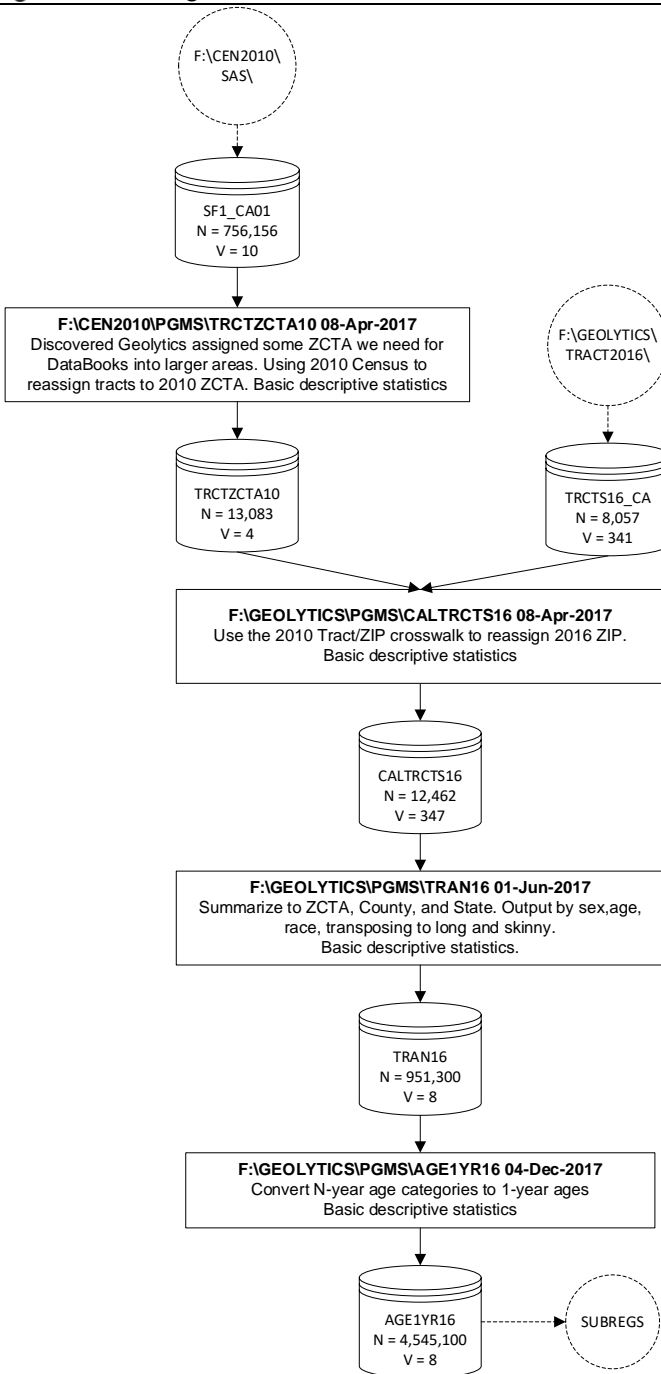
When we first ran programs to estimate or project small area 2016 intercensal population, we discovered that Geolytics "disappeared" some ZIPs in the 2010 census, assigning them to larger areas.

We returned to the multi-level 2010 SF1 file, and pulled off records with both tract and ZIPs, keeping only the geography variables.

We merged this with the Geolytics file to reassign ZIPs, enabling us to proceed with making the files we needed to calculate small area estimates and projections.

The next steps involved our standard of transforming the file into long and skinny (TRAN16) and then the one-year age file (AGE1YR16).

INTERCENSAL SMALL AREA POPULATION ESTIMATES

To estimate intercensal small area population, FHOP uses census and intercensal files, summarized to the geographic levels of interest, and either DOF or NCHS annual county-level population files. Note that we have previously standardized race/ethnicity in all sets of incoming files, with undefined race groups randomly assigned to defined groups. For our example in this

document, we use DOF data for county-level population, and the 1990, 2000, 2010 census and the 2016 Geolytics intercensal files to describe how we prepare small area files.

The Geography Master

We maintain an excel file (GEOGMAS.XLSX) with all ZIPs found in any year of every dataset we process. This includes a count of the number of times the ZIP appeared within and over all the datasets, and the earliest and most recent year the ZIP occurred. We call this our Geography Master [11].

Various commercial ZIP-code products help us classify ZIPS as to their type (unique, military, Post Office, or standard), whether the USPS discontinued the ZIP, if the ZIP split into two or more parts, or if it is new. Following a rule from OSHPD for its hospital datasets, we do not allow ZIPs to cross county boundaries. When in doubt, we assign the cross-county ZIP portion to the ZIP in the county with the largest portion of the population rather than the largest area.

We use this file to identify areas below the county level but higher than the ZIP, such as Supervisor Districts, Service Planning Areas (SPA), Special Jurisdictions (JURIS), or Health Facility Planning Areas (HFPA). Then we make a series of formats to classify data into these smaller areas before summarizing to make products such as our DataBooks. Here, we use those formats to classify ZIPs into SPA and JURIS before summarizing the files to get populations.

We work with locals to identify which ZIPs to assign to what small areas. We treat Supervisor Districts as if they were SPA. Counties can have both SPA and JURIS. Los Angeles is an example, with eight SPA and two JURIS (Long Beach and Pasadena).

Before running the programs to generate small area statistics, it is **critical** to confirm in the geography master that ZIP-code assignments to small areas are correct and to re-build the geography format library as needed. If assignments are incorrect, small area numerators and population denominators will be much larger or smaller than expected.

Prepare Small-Area Files
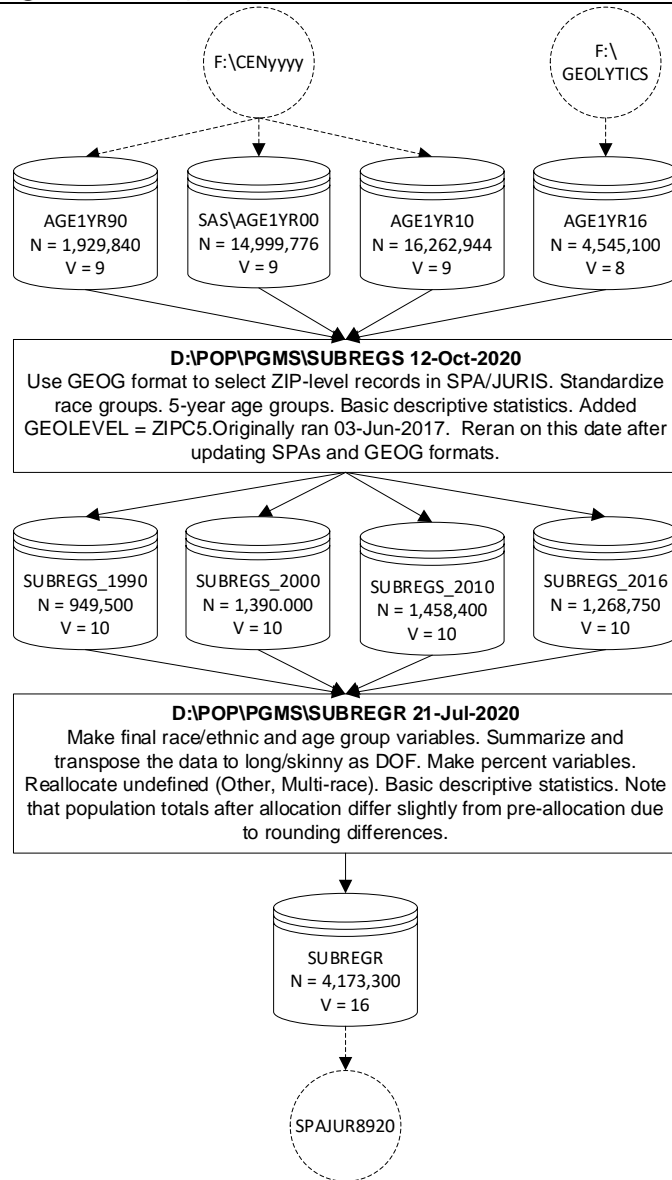
Figure 10. Prepare small-area files

Figure 10 summarizes steps to prepare Census data to interpolate small-area intercensal population estimates using the 1-year age files. Note that census year populations tie exactly to the California DOF population files.

In the program SUBREGS, we use Geography Master formats to classify small areas such as SPA, JURIS, or others. Because different periods had different race/ethnicity definitions, we standardize to common groups within time. We output data for counties with sub-regions.

The program SUBREGR redefines race/ethnicity to prepare for bridging, and summarizes the data. The macro RACE_ALLOC then follows Federal decision rules to allocate undefined race groups (other, multi-race) to the final race/ethnicity. A new variable POPB (Population, bridged) is the sum of the original POP variable plus the number assigned to it by bridging.

To provide longitudinal consistency, all years of SUBREGR have the 5-category race/ethnic variable (White, Black, Hispanic, API (Asian/Pacific Island), and AIAN (American Indian/American Native) available before 2000. We also carry along a 7-category variable (White, Black, Hispanic, Asian, Pacific Island, AIAN, and Multi-race) available from 2000 forward. We only do small-area statistics using the 5-category race/ethnicity variable.

Interpolate Population between Censuses
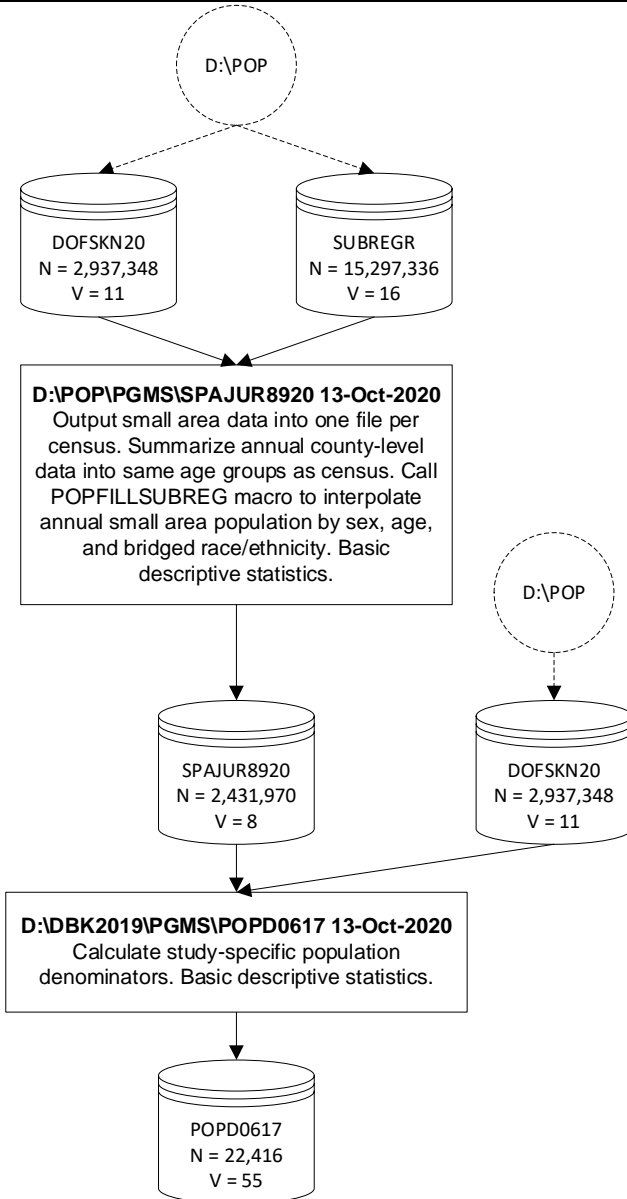
Figure 11. Interpolate population

Figure 9 summarizes the last steps to interpolate small-area intercensal estimates and prepare study-specific population denominator files. In this example, we used DOF data, which we have to use for our DataBooks. But we could as well have used NCHS data, as both files are similarly structured.

Recall that the file SUBREGR has small-area census data spaced a decade apart, while DOFSKN20 has annual data in one-year age groups. We begin the program SPAJUR8920 by outputting data for each census into temporary separate files and then further splitting those by sub-regions. From DOFSKN20, we select counties with sub-regions and summarize county-level data into the same 5-year age groups as the census data.

Then SPAJUR8920 calls the macro POPFILLSUBREG to interpolate annual small area population. It does the following steps for each stratum of sex, age, and race/ethnicity. First, for a year with data for both the county and sub-county regions ("census" years), the macro divides the two populations to obtain the percent of county contained in each sub-region. Note as here, that "census" data can be from a census year or an inter-censal year.

Second, the macro fills in a sub-region percent of county for all years, This step uses linear interpolation if the year is between two censuses, or uses the sub-region percent from the nearest census if not. Finally, for all years, including census years, we multiply the percent of county times the DOF county population to obtain the sub-region population.

To make a study-specific population, we bring together the interpolated small area and county-level populations, calculate denominator variables of interest, and store the resulting file in the study-specific directory. In the example POPD0617, we are preparing 12-year annual population numbers needed to make FHOP's DataBook products for California counties and special jurisdictions over the interval 2006-2017. Notice that we have gone from a skinny file to a wide file, with one record per geography level (county, SPA, jurisdiction) and race/ethnicity per year, and population for each age group needed (female 15-44, children 0 to 14, etc).

Validating the Small Area Allocations

Before developing our current method of estimating population for county sub-regions, Los Angeles County provided us with their SPA estimates, which we used when making DataBooks. They also provided lists of ZIPs assigned to each SPA.

RESOURCES

We have focused on preparing non-confidential population data to use for denominators, including methods when small area data are not available consistently over long periods. Be aware of the limitations of small area geographic data. Do not push the design, analysis, or interpretation beyond what is supportable given known methodologic problems.

All programs are available upon request. FHOP has only two people who can provide a limited amount of handholding to learn how to use these resources. Users will have to contract for more than one hour of support.

ENDNOTES

1   Remy L, Clay T. (2016) Managing Longitudinal Research Studies: The Basic Computing Environment. San Francisco, CA: University of California, San Francisco, Family Health Outcomes Project. Available at: http://fhop.ucsf.edu/data-management-methods.

2   Remy L, Clay T. (2016) Managing Longitudinal Research Studies: Standardizing Variables Over Time. San Francisco, CA: University of California, San Francisco, Family Health Outcomes Project. Available at: http://fhop.ucsf.edu/data-management-methods.

3   Remy L, Clay T. (2017) Managing Longitudinal Research Studies: Preparing Master Files. San Francisco, CA: University of California, San Francisco, Family Health Outcomes Project. Available at: http://fhop.ucsf.edu/data-management-methods

4   Geolytics. http://geolytics.com/USCensus,Neighborhood-Change-Database-1970-2000,Products.asp

5   State of California, Department of Finance, Race/Ethnic Population with Age and Sex Detail, 1970–1989. Sacramento, CA, December 1998. Last accessed 03-Dec-2017 at: http://www.dof.ca.gov/Forecasting/Demographics/Estimates/Race-Ethnic/1970-89/.

6   State of California, Department of Finance, Race/Ethnic Population with Age and Sex Detail, 1990–1999. Sacramento, CA, Revised May 2009. Last accessed 03-Dec-2017 at: http://www.dof.ca.gov/Forecasting/Demographics/Estimates/Race-Ethnic/1990-99/index.html.

7   State of California, Department of Finance, State of California, Department of Finance, Race/Hispanics Population with Age and Gender Detail, 2000–2010. Sacramento, California, September 2012. September 6, 2012 (Original Release - integer data only), September 14, 2012 (Revised to include decimal detail), March 19, 2013 (Removed data for July 2010) Last accessed 03-Dec-2017 at: http://www.dof.ca.gov/Forecasting/Demographics/Estimates/Race-Ethnic/2000-2010/

8   State of California, Department of Finance, State and County Population Projections by Race/Ethnicity, Sex, and Age 2010-2060, Sacramento, California, December 2014. Note that we use the full P-3 file. Last accessed 03-Dec-2017 at: http://www.dof.ca.gov/Forecasting/Demographics/Projections/documents/P3_Complete.csv.zip. The internal file date is 20-Jun-2017.

9   Remy LL, Clay T, Oliva G. (2011). Issues and Decisions to be made on Collecting, Coding and Reporting Race and Ethnicity for Public Health Indicators. Family Health Outcomes Project, University of California, San Francisco. Updated January 2017. Available at: http://fhop.ucsf.edu/data-management-methods

10   Ingram DD, Parker JD, Schenker N, Weed JA, Hamilton B, Arias E, Madans JH (2003) United States Census 2000 population with bridged race categories. National Center for Health Statistics. Vital Health Stat(2)135. 2003. Last accessed 14-Apr-2016 at: http://www.cdc.gov/nchs/data/series/sr_02/sr02_135.pdf

11   Remy L, Clay T. (2017) Managing Longitudinal Research Studies: Methods to Prepare the Geography Master. San Francisco, CA: University of California, San Francisco, Family Health Outcomes Project. Available at: http://fhop.ucsf.edu/data-management-methods.

12   National Vital Statistics System. Bridged-Race Population Estimates - Data Files and Documentation. Last access 14-Oct-2020 at: https://www.cdc.gov/nchs/nvss/bridged_race/data_documentation.htm

13   1990 Census Summary File 1 ASCII files and CD-ROM discs. Last accessed 04-Dec-2017 at: http://www2.census.gov/census_1990/1990STF1.html

14   1990 Census Summary File 3 CD-ROM discs. Last accessed 04-Dec -2017 at: http://www2.census.gov/census_1990/1990STF3.html

15   2000 Census Summary File 1 ASCII files. Last accessed 04-Dec-2017 at: http://www2.census.gov/census_2000/datasets/Summary_File_1/California/

16   2000 Census Summary File 3 ASCII files. Last accessed 04-Dec-2017 at: http://www2.census.gov/census_2000/datasets/Summary_File_3/California/

17   Last accessed 04-Dec-2017 at: http://www.census.gov/support/2000/SF1/SF1SAS.zip

18   Last accessed 04-Dec-2017 at: https://www.census.gov/support/2000/SF3/SF3SAS.zip

19   Last accessed 04-Dec-2017 at: http://www2.census.gov/census_2010/04-Summary_File_1/California/

20   Last accessed 04-Dec -2017 at: http://www.sascommunity.org/wiki/Converting_2010_Census_Summary_File_1_%28SF1%29_Data_into_SAS_Data_Sets#SAS_CODE

21 Parker JD, Schenker N, Ingram DJ, Weed JA, Heck KE, Madans JH (2004) Bridging between two standards for collecting information on race and ethnicity: An application to Census 2000 and vital rates. Public Health Reports, 119:Mar-Apr, 192-205.

22 Office of Management and Budget (2000). Appendix C: The Bridge Report: Tabulation Options for Trend Analysis. Provisional guidance on the implementation of the 1997 standards for federal data on race and ethnicity. Last accessed 12-Apr-2011 at: http://www.whitehouse.gov/sites/default/files/omb/assets/information_and_regulatory_affairs/re_app-ctables.pdf.

23 Office of Management and Budget. Revisions to the standards for the classification of Federal data on race and ethnicity, Federal Register 62FR58781–58790, October 30, 1997. Last accessed 19-Apr-2016 at: http://www.whitehouse.gov/omb/ fedreg/ombdir15.html.

24 Office of Management and Budget. Race and ethnic standards for Federal statistics and administrative reporting. Statistical Policy Directive 15, 1977. Last accessed 19-Apr-2016 at: http://wonder.cdc.gov/wonder/help/populations/bridged-race/Directive15.html

25 National Center for Health Statistics (2004). NCHS Procedures for Multiple-Race and Hispanic Origin Data: Collection, Coding, Editing, and Transmitting. Division of Vital Statistics, National Center for Health Statistics, Centers for Disease Control and Prevention. May 7, 2004. Last accessed 05-Jun-2017 at: http://www.cdc.gov/nchs/data/dvs/Multiple_race_docu_5-10-04.pdf.