MANAGING LONGITUDINAL RESEARCH STUDIES:

BIRTH AND FETAL DEATH STATISTICAL MASTER FILES

By

Linda L Remy, MSW PhD

Ted Clay, MS

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: http://www.ucsf.edu/fhop

November 2018

TABLE OF CONTENTS

Overview **1**

Steps to Create Master Files **2**

Document Incoming Data 2
Define Data Structure 3
Make the Master File 5

Checking Longitudinal Consistency **6**

Contents Report 6
Frequency Reports 7

Geographic Classification **8**

VSMF Nation, State, and County Variables 8
City Variables 9
Hospital Codes 10

Resources **11**

Endnotes **11**

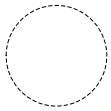
Table of Figures

Figure 1. Steps to create Vital Statistics Master Files 5

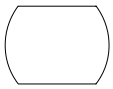
Table of Tables

Table 1. Sample of documentation for incoming files 2
Table 2. Sample of file defining structure 3
Table 3. Sample contents report of VSMF confidential variables 7
Table 4. Sample frequencies report for mother's education variables 8

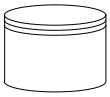
Table of Legends



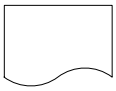
From program creating incoming file or name of next program using file



Excel file input or output



SAS file input or output



Flat or text file input or output

SAS program with brief description of steps.

Suggested Citation

Remy L, Clay T. (2018) Managing Longitudinal Research Studies: Birth and Fetal Death Statistical Master Files. San Francisco, CA: University of California, San Francisco, Family Health Outcomes Project. Available at: http://fhop.ucsf.edu/data-management-methods

ACRONYMS

ANSI	American National Standards Institute
BSMF	Birth Statistical Master File
CADPH	California Department of Public Health
CPHS	Committee for the Protection of Human Subjects
FDSMF	Fetal Death Statistical Master File
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
OSHPD	Office of Statewide Health Planning and Development
SSN	Social Security Number
UCSF	University of California, San Francisco
VSMF	Vital Statistics Master Files
VSAC	Vital Statistics Advisory Committee

BIRTH AND FETAL DEATH MASTER FILES

This document describes methods the UCSF Family Health Outcomes Project (FHOP) uses to prepare confidential versions of Vital Statistics Master Files (VSMF), specifically the Birth Statistical Master File (BSMF) and Fetal Death Statistical Master File (FDSMF) distributed by the California Department of Public Health (CADPH). When discussing characteristics that apply to both the BSMF and FDSMF, we use the acronym VSMF.

We assume that the user of this document has read other documents describing the foundation of our methodology:

Volume One: The Basic Operating Environment

Volume Two: Standardizing Variables Over Time

Volume Three: Preparing Master Files

These and related documents are available on our website:
https://fhop.ucsf.edu/data-management-methods

OVERVIEW

Confidential versions of VSMF include: detailed demographic information related to the infant, mother, and father; medical data related to the vital event; and confidential personal identifiers such as names, addresses, and other fields that could identify an individual. To have these files users must obtain approval from the California Health and Human Services Agency's Committee for the Protection of Human Subjects (CPHS) and the CADPH Vital Statistics Advisory Committee (VSAC).

This document summarizes steps to import confidential VSMF into SAS, check variables longitudinally, and prepare formats. We discuss issues associated with certain types of variables. At the writing of this document, we have VSMF files from 1989 through 2013. All work is in SAS, assisted by Microsoft Excel and Visio.

Although FHOP uses confidential VSMF, processes and resources described here will help users of other versions. We are making this basic methodology and its associated software public to help population health 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 studies using VSMF.

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 internal documentation as possible. Because public funding supports development of these programs, they are in the public domain. This is why we are making them available.

STEPS TO CREATE MASTER FILES

Document Incoming Data

Incoming VSMF are stored on the Confidential Drive in password-protected ZIP files, documented in Excel files on the Master Drive, and read into SAS on the Working Drive using a macro program. We described these steps in Volume 3, Preparing Master Files. Volume 1, The Basic Operating Environment, defines the Confidential, Master, and Working Environments.

Table 1 shows a few lines of the initial BSMF documentation file (BC_DOC.XLS). VSMF arrive as flat text files with the DAT extension. Some years they arrive as two flat files. Beyond the year designation, notice that names for incoming flat files and files documenting their contents are not standardized. Other information we document about incoming VSMF files is the number of bytes per record, whether the file includes names, addresses, and Social Security Number (SSN) for various family members (I = infant, M = mother, F = father, N = None). We do not show the notes column. It has sensitive information such as the name of the person sending the file, location of passwords, etc.

Table 1. Sample of documentation for incoming files

		FILE NAME		OTHER INFORMATION			
YEAR	ZIP FILE	FILE NAME	RECORD LAYOUT SOURCE	BYTES	NAMES	ADDR	SSN
2005	BC2005.ZIP	BSMF05-980.DAT	2005BC980.DOC	980	I M F	M	I M F
2006	BC2006.ZIP	B06SEAL.DAT	2006BC980.DOC	980	I M F	M	I M F
2007	BC2007.ZIP	B08SEALA.DAT B08SEALB.DAT	2007BD1300.DOC	1300	I M F	M	M F
2008	BC2008.ZIP	B08SEALA.DAT B08SEALB.DAT	2008BC1300.doc	1300	I M F	M	M F
2009	BC2009.ZIP	B09SEAL.DAT	BSMF09 1300-Byte-Layout.doc	1300	I M F	M	N

The FDSMF never had SSN, while the BSMF had at least the mother's SSN from 1997 through 2008. In 2010, CADPH announced that the Social Security Act, Title 42 of the Code of Federal Regulations does not permit the release of SSN in the BSMF. Beginning with the 2009 BSMF, SSN are no longer available. This is a significant blow to longitudinal research.

Define Data Structure

Table 2 shows a few lines of the Excel file defining BSMF contents (BCV.XLS) to import the flat files into SAS. The example focuses on a few variables in the INFANT group for the 2009 file. We do not show columns that are not relevant for this example.

Table 2. Sample of file defining structure

VARNAME	LABEL	LENGTH	TYPE	FORMAT	OUT1	STATS	SASCODE	_2009
IBTHDATEO	Date of birth (Infant) (CCYYMMDD)	8	C		Y	M	IBTHDATEO=left(trim(IBTHDATEO));	159-166
IBTHDATEC	Date of birth (Infant) (CCYYMMDD)	8	C		Y	M	if 1989 le &yyyy le 1990 then do;%fixdate(var=IBTHDATEO,out=IBTHDATEC,pattern=C9YYMMDD,year=&yyyy);end;else if 1991 le &yyyy le 2003 then do; %fixdate(var=IBTHDATEO,out=IBTHDATEC,pattern=C9YYMMDD);end;else IBTHDATEC=IBTHDATEO;	CALC
IBTHDATE	Date of birth (Infant)	4	N	date9.	Y	MI	%datevar(IBTHDATEC,IBTHDATE,0,0);	CALC
IBTHHR	Hour of birth (HHMM)	4	C		Y	M	IBTHHR = left(trim(IBTHHR));if length(IBTHHR)=3 then IBTHHR='0'||IBTHHR;else if length(IBTHHR)=2 then IBTHHR='00'||IBTHHR;	167-170
SEX	Sex of child	3	N	sex.	Y	F		175
IRACE	Race (Infant)	3	N	race.	Y	F		
BTYPE	Type of birth	3	N	btype.	Y	F		176
BORDER	Birth order	3	N	border.	Y	F		177
BW	Birthweight (Grams)	4	N	bw.	Y	FU	if BW in (9998, 9999) then BW = .;	171-174

Many longitudinal date variables arrive in different lengths, not structured to the full 8-byte length CCYYMMDD, where CC is century, YY is year, MM is month, and DD is day. In VSMF, dates had one structure for 1989-1990, another for 1991-2003, and yet another thereafter.

We show how we standardize dates by focusing on steps to make the infant's birth date variable IBTHDATE. We first read in IBTHDATEO, the original date. Column _2009 shows the incoming flat file has this in columns 159-166. RDYR, the controlling macro that reads the data into SAS, looks to the SASCODE column and carries out the instruction, which is to import this field trimmed and left aligned. The LENGTH column shows we import it as 8 bytes long, regardless of its incoming length. The TYPE column shows we import it as a character variable (C).

Next, we calculate IBTHDATEC. Column _2009 contains CALC. The controlling macro to read the data into SAS looks to the SASCODE column and calls the FIXDATE macro to structure the variable appropriately. FIXDATE converts IBTHDATEO to IBTHDATEC with the structure CCYYMMDD.

To make IBTHDATE, column _2009 again contains CALC. The controlling macro looks to the SASCODE column and calls the DATEVAR macro to structure IBTHDATE as a length 4 numeric variable, formatted date9 (DD-MMM-CCYY). The OUT1 column indicates that all date variables are output. The STATS column indicates the listing file (LST) will show IBTHDATEO

and IBTHDATEC frequencies based on our special format for missing values (M), and IBTHDATE will have M frequencies plus a listing of invalid dates (I) that the DATEVAR macro corrected.

The infant's birthdate is always present in the VSMF, but mother and father birthdates are sometimes missing. In 2009, 136 mothers had no birthdate, and two more were invalid. Here is an example of the MI listing for the sequence of date variables:

```
MBTHDATEC_                              Cumulative    Cumulative
Missing       Frequency     Percent     Frequency      Percent
-----------------------------------------------------------------
Missing             136        0.03           136         0.03
Present          528489       99.97        528625       100.00


MBTHDATE_                               Cumulative    Cumulative
Missing      Frequency     Percent     Frequency      Percent
-----------------------------------------------------------------
Missing            138        0.03           138         0.03
Present         528487       99.97        528625       100.00


Invalid Date Conversions for MBTHDATE

   Obs     MBTHDATEO     MBTHDATEC      MBTHDATE

234164     19400320      19400320
289737     19341228      19341228
```

Two dates of mother's birth were set to missing. They were outside the upper range for a woman to bear a child, which we set at 60. In 2009, the women whose birthdates we set aside would be respectively 69 and 75 years old.

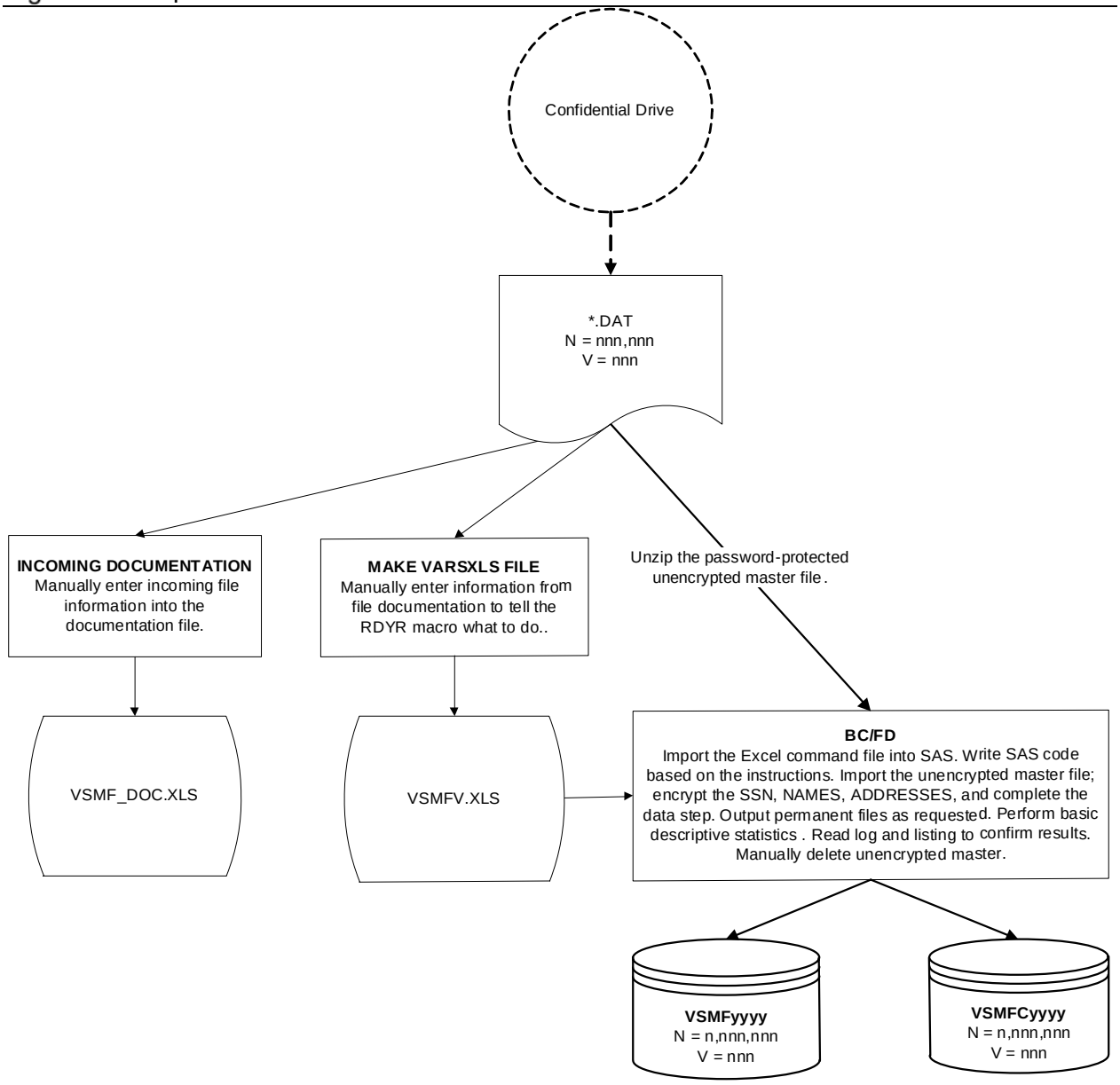
The time of the infant's birth (IBTHHR) is a length 4 variable based on a 24-hour clock, where 0000 is midnight and 2359 is just before midnight. Time occasionally is not entered as defined. Does the number 11 mean 11 minutes after midnight (0011) or 11 in the morning (1100)? We read the variable for time of the infant's birth into SAS as a character variable with length 4 from columns 167-170. The SASCODE column shows that when we find an observation with a shorter length, we add zeros to align hours and minutes.

Notice that the other variables have defined formats. Our policy is to give the format the same name as the variable. Formats for the variables SEX, RACE, BTYPE, and BORDER are based on definitions in the VSMF documentation file. VS discontinued calculating infant's race (based on mother and father) in 1996. We defined external formats for BW (500-gram ranges). The final LST will show standard frequency counts (PROC FREQ) for these variables, and BW also will have a PROC UNIVARIATE (U).

Make the Master File

Figure 1 shows a Visio diagram summarizing steps to convert confidential VSMF into a longitudinally consistent set of files structured per FHOP standards. For details on setting up the VARSXLS file, refer to Volume Three: Preparing Master Files.

Figure 1. Steps to create Vital Statistics Master Files



CHECKING LONGITUDINAL CONSISTENCY

Contents Report

After making all years of all files in a given set, we run CONTSRPT.SAS to verify that the internal structure is consistent longitudinally. This program does a PROC CONTENTS over all available years, and outputs a file with the variable names, labels, type (character or numeric), length, and format for each variable. These are merged by year, then merged with a subset of variables (STORDER, GROUP, and VARNAME) from the VARSXLS file that created the masters. The final step outputs the information to an excel file for review.

Our CONTSRPT review focuses on whether the same variable has the same label, type, length, and format in every year. Over time, some variables appear or disappear, and we check for such discontinuities. Early experience with these reports led us to develop the system we now use, so data can be consistent longitudinally. This report looked pretty sad “back in the day.”

Table 3 shows a cross-section of the confidential data tab (BCC) in the CONTSRPT.XLS, for the BSMF decade 1991-2000. Bear in mind that we have these files from 1989 forward. Variables are grouped by STORDER, as specified in the VARSXLS file that was the source for the master files, in this case BCV.XLS. Each yearly column shows the type ((N)umeric, (C)haracter, (D)ate) and format associated with a given variable that year. Confidential VSMF variables do not have formats

Notice that mother’s SSN was available through 2008, father’s from 2003-2008, and infant’s from 2004-2006. The suffix C indicates that SSN are encrypted. The absence of a consistent variable to link health events over the life course significantly limits longitudinal research options.

VSMF names have two versions. One is truncated, and the other is the full name. Mother’s mailing address became available in 2007. In this file, variables without a C or E suffix are not encrypted.

Table 3. Sample contents report of VSMF confidential variables

GROUP	NAME	LABEL	_2000	_2001	_2002	_2003	_2004	_2005	_2006	_2007	_2008	_2009
LINKAGE	YR_OBS		n7	n7	n7	n7	n7	n7	n7	n7	n7	n7
LINKAGE	LFN	LOCAL FILE NUMBER	c13	c13	c13	c13	c13	c13	c13	c13	c13	c13
LINKAGE	SFN	STATE FILE NUMBER	c13	c13	c13	c13	c13	c13	c13	c13	c13	c13
LINKAGE	SFNP	STATE FILE NUMBER (PAPER CERTIFICATE)					c6	c6	c6	c6	c6	c6
LINKAGE	MSSNC	SOCIAL SECURITY NUMBER (MOTHER) (ENCR)	c9	c9	c9	c9	c9	c9	c9	c9	c9	
LINKAGE	FSSNC	SOCIAL SECURITY NUMBER (FATHER) (ENCR)	c9			c9	c9	c9	c9	c9	c9	
LINKAGE	ISSNC	SOCIAL SECURITY NUMBER (INFANT) (ENCR)					c9	c9	c9			
INFANT	IDTHSFN	DEATH STATE FILE NUMBER (INFANT)					c19	c19	c19	c19	c19	c19
INFANT	IFIRSTAE	FIRST NAME (INFANT) (AVSS) (ENCR)					c30	c30	c30	c30	c30	c30
INFANT	IMIDDAE	MIDDLE NAME (INFANT) (AVSS) (ENCR)					c24	c24	c24	c24	c24	c24
INFANT	ILASTAE	LAST NAME (INFANT) (AVSS) (ENCR)					c34	c34	c34	c34	c34	c34
INFANT	IFIRSTE	FIRST NAME (INFANT) (ENCR)	c12	c12	c12	c12	c12	c12	c12	c12	c12	c12
INFANT	IMIDDE	MIDDLE NAME (INFANT) (ENCR)	c12	c12	c12	c12	c12	c12	c12	c12	c12	c12
INFANT	ILASTE	LAST NAME (INFANT) (ENCR)	c20	c20	c20	c20	c20	c20	c20	c20	c20	c20
MOTHER	MFIRSTAE	FIRST NAME (MOTHER) (AVSS) (ENCR)					c25	c25	c25	c25	c25	c25
MOTHER	MMIDDAE	MIDDLE NAME (MOTHER) (AVSS) (ENCR)					c18	c18	c18	c18	c18	c18
MOTHER	MLASTAE	LAST NAME (MOTHER) (AVSS) (ENCR)					c33	c33	c33	c33	c33	c33
MOTHER	MFIRSTE	FIRST NAME (MOTHER) (ENCR)	c12	c12	c12	c12	c12	c12	c12	c12	c12	c12
MOTHER	MMIDDE	MIDDLE NAME (MOTHER) (ENCR)					c12	c12	c12	c12	c12	c12
MOTHER	MLASTE	LAST NAME (MOTHER) (ENCR)	c20	c20	c20	c20	c20	c20	c20	c20	c20	c20
FATHER	FFIRSTAE	FIRST NAME (FATHER) (AVSS) (ENCR)					c25	c25	c25	c25	c25	c25
FATHER	FMIDDAE	MIDDLE NAME (FATHER) (AVSS) (ENCR)					c18	c18	c18	c18	c18	c18
FATHER	FLASTAE	LAST NAME (FATHER) (AVSS) (ENCR)					c33	c33	c33	c33	c33	c33
FATHER	FFIRSTE	FIRST NAME (FATHER) (ENCR)					c25	c25	c25	c25	c25	c25
FATHER	FMIDDE	MIDDLE NAME (FATHER) (ENCR)					c18	c18	c18	c18	c18	c18
FATHER	FLASTE	LAST NAME (FATHER) (ENCR)	c33	c33	c33	c33	c33	c33	c33	c33	c33	c33
GEOG	MADDRE	STREET ADDRESS (MOTHER) (ENCR)	c50	c50	c50	c50	c50	c50	c50	c50	c50	c50
GEOG	MADDRME	MAILING ADDRESS (MOTHER) (ENCR)								c50	c50	c50

Frequency Reports

The program FRQYRS.SAS outputs an excel file of the same name, with tabs for categorical variables in a given group. This includes formatted continuous variables, for example, age at admission. Here, we are looking for unformatted values, or sharp changes in distributions that might indicate variables read into SAS incorrectly.

FRQYRS.SAS outputs both the number of cases and the percent of cases in that year. When definitions change (for example, mother's education (MEDU to MEDUN)), we look for changes in the number of cases in a given category that might indicate definitional issues. Table 4 shows the number of cases for mother's education variables, again focusing on the 2000-2009 cross-section. Notice that formatted values (FMTVALUE) show the underlying number. This allows programmers to make next generation variables without having to refer to the original codebook.

It also highlights how variables measuring the same construct have different values and definitions. This is an example of a variable we will have to bridge for consistency if we use it in a longitudinal study that includes the changeover years. Bridging is a method to reduce the number of categories for time trends. The purpose is to make groups comparable, across

indicators and time, between numerators and denominators. Bridging continues until numerators and denominators are available over the period of interest in the current format.

Another problem is that the number of mothers whose education was unknown almost tripled over the 10-year period, when the number of births remained relatively stable. This is an example of a data quality problem.

Table 4. Sample frequencies report for mother's education variables

VARIABLE	LABEL	FMTVALUE	_2000	_2001	_2002	_2003	_2004	_2005	_2006	_2007	_2008	_2009
MEDU	Years of education (Mother)	0 No Education	2,249	2,143	2,060	2,158	1,998	2,447				
MEDU	Years of education (Mother)	1-12 Elementary or Secondary	304,995	301,081	296,059	296,811	296,886	299,226				
MEDU	Years of education (Mother)	13 One Yr College	33,188	31,790	31,064	31,670	30,821	31,099				
MEDU	Years of education (Mother)	14 Two Yr College	52,330	51,210	51,635	52,291	52,075	52,475				
MEDU	Years of education (Mother)	15 3 Yr College	19,312	18,562	18,333	18,632	18,415	18,748				
MEDU	Years of education (Mother)	16 4 Yr College	63,423	63,926	65,728	69,167	69,532	69,676				
MEDU	Years of education (Mother)	17 5+ Yr of College	49,757	51,618	54,590	59,050	61,185	61,954				
MEDU	Years of education (Mother)	99 Unknown or Not Available	7,591	8,655	11,457	12,753	15,598	15,303				
MEDUN	Years of education (Mother)	1 8th or less							56,847	53,171	48,292	42,150
MEDUN	Years of education (Mother)	2 9th-12th, no dipl, at least age 9							148,668	99,305	94,617	85,979
MEDUN	Years of education (Mother)	3 High school grad/GED/ at least age 16							99,956	149,751	144,762	135,666
MEDUN	Years of education (Mother)	4 Some college, no degree, at least age 17							88,258	90,645	92,856	91,677
MEDUN	Years of education (Mother)	5 AA, at least age 18							29,498	28,958	28,412	27,951
MEDUN	Years of education (Mother)	6 Bachelors, at least age 20							82,963	84,863	84,031	82,128
MEDUN	Years of education (Mother)	7 Masters, at least age 21							30,032	30,647	31,243	32,051
MEDUN	Years of education (Mother)	8 Doctorate/Professional, at least age 23							11,453	11,661	11,506	11,702
MEDUN	Years of education (Mother)	9 Unknown							16,652	19,339	17,730	19,321

GEOGRAPHIC CLASSIFICATION

VSMF have a number of geography-associated variables. Nation, state, and county-level examples include where the infant was born, mother's and father's place of birth, mother's place of residence, and health jurisdiction filing the record of the vital event. Sub-county variables include such things as facility where the child was born, mother's city of residence, addresses, ZIP-codes, and Census Place and Tract codes.

We introduce some longitudinal issues for VSMF nation, state, county, and city variables, and facilities where children are born. Detailed information related to these issues and other geography-related variables are in our Geography Master document.

VSMF Nation, State, and County Variables

The fundamental problem with this group of variables is that different codes describe the same geographic entity across several variables at the same time and over time. For example, the code for Hawaii is variously HI, 112, and 12. The code for Humboldt County is 12 and 012.

Before the national 2003 revision of the VSMF, the National Center for Health Statistics (NCHS) used geographic codes unique to it [1]. With the 2003 revision, they adopted Federal FIPS codes [2]. Further complicating matters, the US Census converted from FIPS to ANSI codes for the 2010 Census [3]. We addressed these various longitudinal changes by creating a crosswalk from older geographic variables to 2010 Census standards. This is useful for assigning geographic variables consistently over time.

Families from certain regions of the world have different kinds of health problems. For example, women born in Central America or Mexico often have better birth outcomes than Hispanic women born in the United States. Mothers born in certain Middle East or African nations are more likely to have experienced genital mutilation, often associated with truly severe birth complications. We developed a format to crosswalk nation of birth to larger regions.

We discuss development of the VSMF region, nation, and state crosswalks in the document about our Geography Master. These formats enable merging VSMF data with census data.

Another issue is that the VSMF have two sets of nativity variables for mothers and fathers. The set with the suffix A (AVSS) is a more complete list than the set without the suffix.

City Variables

VSMF city variables are released as written because they are shown on the vital records certificate. Thousands of typographic errors exist, which increases geocoding and mapping problems. For example, we found 19 incorrect spellings for Los Angeles and 12 for San Francisco. As part of the ongoing maintenance of our geography master series, we summarize all city names found in various files including commercial vendors. We manually correct misspellings and convert them to a format.

Format $cityc. replaces the incorrectly spelled city with the correct name. If we lack a city name but have a ZIP, we call the format $zipcity. to impute city. Note our format naming convention: left side of the format name is the variable we have, right side of the format name is the variable we are imputing. The following is code we use to make the BSMF:

```
if 1989 le &yyyy le 1996 then MCITY = put(ZIPC5,$zipcity.);
else if &yyyy in (2001, 2002) then MCITY = put(ZIPC5,$zipcity.);
* clean city;
MCITY = upcase(MCITY);
if MCITY gt ' ' then MCITY = put(MCITY,$cityc.);
```

We do not use variables for Census Place (FIPCITY) or Tract. The VSMF collected the 4-byte Census Place code from 1985-1997, the 3-byte Census Place code from 1998-2002, and the 5-byte FIPS code from 2003. Regardless of the version, the same FIPS city/place code refers to multiple cities. Tract is an optional field with many missing. We have done a number of analyses on relationships between ZIP, City, County, Place, and Tract. Geography assignment errors are pervasive in the VSMF, which was part of the motive for developing our geography master series. VS knows of the problems and has undertaken an agency-wide initiative to improve geocoding.

In the meantime, ZIP-code has been available since 1989, and while still problematic, returns a better city name.

Hospital Codes

Babies are born in more places than hospitals. Some mothers choose to have their baby at home or in birthing clinics. Other babies are born by the side of the road, trapped in commute traffic on their way to the hospital. Thus, codes for where babies are born differ in important respects from facility codes the Office of Statewide Health Planning and Development (OSHPD) issues.

Further, hospitals and clinics open, close, and change names. They may disappear in the data just when another facility in their area shows huge increases in patients. This may reflect that the first facility indeed closed and another picked up the extra cases, or that it moved and VS gave it a new identifier.

Another issue with VSMF facility codes is that they differ completely from OSHPD's codes. This has a basis in history. In 1960, VS began to assign facility codes for VSMF events. OSHPD did not come into existence until 1973. Thus, VSMF codes developed completely independently of OSHPD.

Code numbers for hospitals open at that time reflect the result of sorting by county and hospital name, then assigning numbers sequentially. Thus, Alameda County had 25 facilities delivering infants numbered sequentially from 0001 to 0025. Some years later, another facility opened that VS assigned code 0737, and final facility given code 0787. In 1989, only 20 Alameda County facilities were still delivering infants. By 2009, only 8 facilities remained.

The parallel development of the VSMF and OSHPD facility codes has hampered longitudinal research into the role of structural characteristics on patient outcomes. The best understanding of the vital event comes from linking VSMF with OSHPD files. The latter have separate records for infant and mother, with a more complete listing of diagnoses, procedures, and outcomes.

We developed a crosswalk between the VSMF and OSHPD facility codes [4]. Related to this is a crosswalk showing which hospitals consolidated or moved and when, and another for ownership and auspices changes, and other important structural changes. We now can link VSMF records to hospital records, and also to OSHPD's Hospital Annual Disclosure Report, which provides detailed information on structural characteristics such as 24-hour availability of an emergency room, beds (newborn nursery, neonatal intensive care unit), ownership, auspices (non-profit, public), and other characteristics that research has shown impact outcomes.

These various resources are available upon request. Separate documents describe their development.

RESOURCES

We have focused on the making of VSMF per FHOP standards. 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 NCHS Geographic Coding 2003. Division of Vital Statistics, CDC/NCHS. Last accessed 19-Nov-2018 at: http://www.cdc.gov/nchs/data/dvs/Geocode_pres.pdf. .

2 Specifications for collecting and editing the United States standard certificates of birth and death and the report of fetal death -- 2003 Revision. Dated Apr-2004. Updated 18-Mar-2005. Last accessed 19-Nov-2018 at: https://www.cdc.gov/nchs/data/dvs/Guidelinesbirthspecs1101acc.pdf. See also: http://www.cdc.gov/nchs/nvss/vital_certificate_revisions.htm.

3 American National Standards Institute (ANSI) Codes (2010). US Census Bureau. Last accessed 19-Nov-2018 at: http://www.census.gov/geo/www/ansi/ansi.html.

4 Remy L, Clay T. (2016) Managing Longitudinal Research Studies: Crosswalking Hospital Identifiers. San Francisco, CA: University of California, San Francisco, Family Health Outcomes Project. Available at: http://familymedicine.medschool.ucsf.edu/fhop/