

# MANAGING LONGITUDINAL RESEARCH STUDIES:

---

## CROSSWALKING HOSPITAL IDENTIFIERS

---

By

Linda L Remy, MSW PhD

Ted Clay, MS

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>

September 2016

## TABLE OF CONTENTS

<b>Background</b> .....	<b>1</b>
<b>Analysis Files</b> .....	<b>4</b>
Linkage files .....	4
Validation files .....	5
<b>Data Linkage</b> .....	<b>6</b>
<b>Validation</b> .....	<b>7</b>
Compare births by source and hospital.....	7
Compare FHOP and Herrchen linkage .....	10
<b>An Example</b> .....	<b>11</b>
<b>Summary</b> .....	<b>13</b>
<b>Endnotes</b> .....	<b>14</b>

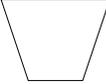
## Table of Figures

Figure 1.	Consolidated Reporting Example .....	3
Figure 2.	Link OSHPDID and HOSPCODE using geography variables.....	6
Figure 3.	Examine linkage results by source .....	7
Figure 4.	Examine FHOP and Herrchen linkage results .....	11
Figure 5.	OUM data quality by source and level .....	12

## Table of Tables

Table 1.	Summary of linkage results.....	7
Table 2.	Comparison of number of births by hospital, data source, and time.....	9

## Table of Legends

	From source of incoming file or out to next program
	SAS file input or output
	Excel, flat or text file, input or output including manual operation
	SAS program with brief summary of steps.

### Suggested Citation

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://fhop.ucsf.edu/data-management-methods>

## ACRONYMS

ALIRTS	Automated Licensing Information and Report Tracking System
ASC	Ambulatory Surgery Center data distributed by OSHPD
BSMF	Birth Statistical Master File
CDPH	California Department of Public Health
CHOA	Change of auspice, for example from publicly-owned to investor-owned
CHON	Change of name
CHOW	Change of owner
ED	Emergency Department data distributed by OSHPD
FACNO	Facility identifier identified by Licensing and Certification
FDSMF	Fetal Death Statistical Master File
GACH	General Acute Care Hospitals
HAFD	Hospital Annual Financial Disclosure Report
HOSPCODE	3-digit identifier assigned by California Vital Statistics to facilities delivering babies. Note that the ID is assigned to hospitals, military hospitals, birthing centers, and other types of facilities where infants are born.
HIRS	Health Information and Research Section
LCP	Licensing and Certification Program, CDPH
MOVE	Facility physical location moved from Point A to Point B
OSHPD	Office of Statewide Health Planning and Development
OSHPDID	6-digit identifier OSHPD assigns to facilities that report to it. Military hospitals and birthing centers are exempt from reporting. The first 2-digits indicate the county where the hospital is located. The other is the identifier within the county.
OUM	Other, unknown, missing
PARENT	Owns or controls another facility and includes OSHPD subsidiary data in its report. This can include the HAFD or patient data or both.
PDD	Patient discharge data distributed by OSHPD
PHF	Psychiatric Health Facility
SUBSIDIARY	Owned or controlled by another facility, and data reported to OSHPD combined with the parent organization. This can include the HAFD or patient data or both.



# CROSSWALKING HOSPITAL IDENTIFIERS

---

Using information about hospital structural characteristics (e.g., ownership, auspices, licensed capacity, physician specialties) or financial performance can strengthen population health research. Unfortunately, such information is not available in Vital Statistics birth or fetal death statistical master files (respectively BSMF and FDSMF) distributed by a branch of the California Department of Public Health (CDPH) [1]. Nor is such information included in the various patient abstract files distributed by the Office of Statewide Health Planning and Development (OSHPD), specifically patient-level discharge data (PDD), emergency department encounters (ED), and ambulatory surgery encounters (ASC) [2].

Detailed structural information is available in OSHPD's Hospital Annual Financial Data (HAFD) [3]. The HAFD has desk-audited data collected from all acute care hospitals licensed by the State of California. Among other things, it includes type of ownership, type and number of beds, physicians by specialty, patients and days of care by bed types, balance sheets and income statements, revenues by payer, and expenses by natural classification.

Many researchers trying to relate outcomes in BSMF or FDSMF data to hospital data (whether inpatient or structural) are limited because they do not know how to link the datasets. A primary limitation is that OSHPD and CDPH use different identifiers for the same facility. Another major obstacle over time is that the agencies reassign identifiers for various reasons, and in OSHPD files, facilities can disappear even though they continue to exist because of consolidated reporting [4].

In this complex environment, the UCSF Family Health Outcomes Project (FHOP) developed a crosswalk between the OSHPD identifier (OSHPDID) and the Vital Statistics identifier (HOSPCODE). Note that the Vital Statistics to Care Delivery Linkage that Health Information Solutions developed for OSHPD looks within a given year [5-7], while we additionally address longitudinal changes such as consolidations or moves.

We are making the crosswalk publicly available, in the hope that researchers will use it as a tool to help understand California's longitudinal health outcome trends. This document describes our methods to develop the crosswalk. We close with an example of its use.

## BACKGROUND

---

CDPH registers all California births, deaths, and fetal deaths. It compiles this information into statistical master files used to do research and to calculate various public health indicators. In the early 1960s, CDPH began to assign HOSPCODE to facilities delivering infants. These

include stand-alone clinics, birthing centers, physician offices, and military hospitals, as well as hospitals licensed by the State of California. These facilities submit birth information to CDPH, which makes the official birth certificate and related data files (BSMF, FDSMF). Initially, Vital Statistics assigned HOSPCODE within county alphabetically based on facility name, then assigned HOSPCODE sequentially as facilities entered the health services delivery stream [8].

Under state law, facilities submit birth information to CDPH, which makes the official birth certificate. Both the BSMF and FDSMF include HOSPCODE. Since the early 1960s, CDPH sometimes did and sometimes did not assign a new HOSPCODE when ownership changed or the hospital moved and sometimes did and sometimes did not change the name when it changed. Through 2013, the historic hospital code list identifies 786 entities.

The CDPH Licensing and Certification Program (LCP) is responsible for licensing, regulating, inspecting, and/or certifying various types of California hospitals and other healthcare facilities. Military hospitals, stand-alone birthing centers, and certain physician-owned surgery centers are exempt from CDPH oversight. LCP assigns its facility identifier (FACNO) based on geographic location. If the owner of a health facility at a given location ceases to provide licensable services at the site and moves to another geographic location, LCP assigns a new facility identifier. OSHPD includes the FACNO in some of its files.

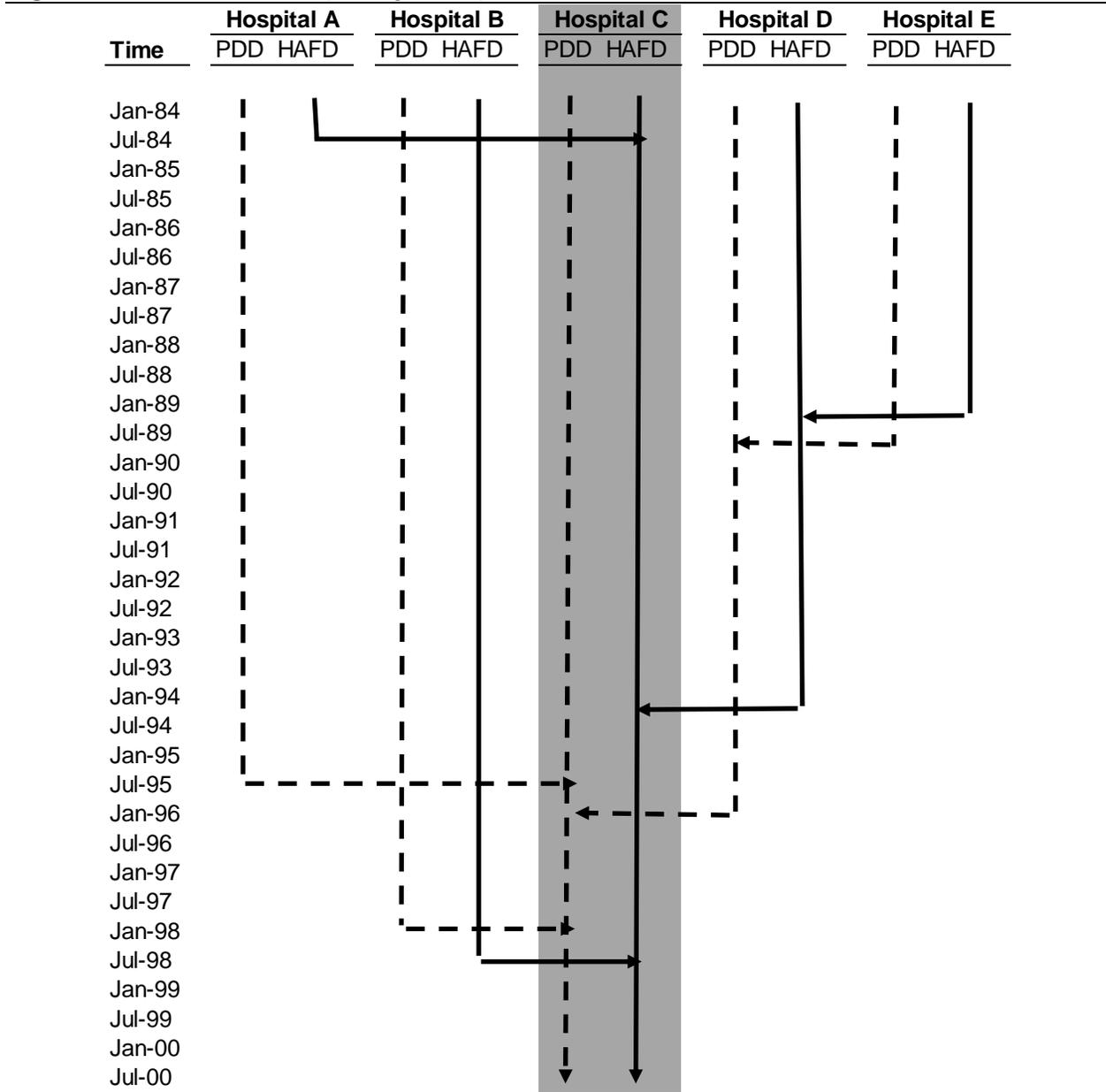
Hospital uniform accounting and reporting began with the passage of the California Hospital Disclosure Act in 1971 [9]. The Legislature tasked the Health and Welfare Agency with creating and maintaining uniform accounting standards for health providers and mandated public oversight of the health care industry to force it to account for public resources. OSHPD became a state agency in 1979 [10]. It emerged from reorganization of the Health and Welfare Agency brought about by SB363 (Gregorio, Chapter 1252, Statutes of 1977) that also created CDPH. OSHPD released the HAFD in 1981 and the PDD in 1983. In 2005, OSHPD added ED and ASC data. They also distribute long term care data, which we do not address here.

OSHPD assigns identifiers only to California licensed hospitals and related facilities. Military and other Federal hospitals in California are exempt from their oversight. OSHPD originally assigned its 6-digit identifier based loosely on geographic location. The first two digits indicate the county in which the facility is located. The last four digits are unique within each county. Sometimes in the past, OSHPD allowed facilities to move without changing the OSHPDID. Now they change the OSHPDID when facilities move. On some occasions, OSHPD changed the OSHPDID when ownership changed. Like LCP, the OSHPDID now is based on geography.

For facilities with common ownership, OSHPD allows consolidated reporting for the HAFD and/or patient data, and they allow it both within and across county lines. To illustrate the complexity, Figure 1. shows consolidated reporting for the HAFD and PDD for one group of San Francisco hospitals. In 2000, all facilities were open, but we had HAFD and PDD data only for Hospital C and birthing data for two. Thus, after 2000, only combined structural information is

available for these hospitals. A sixth San Francisco hospital with the same owner still reports HAFD, PDD, and birth data separately.

Figure 1. Consolidated Reporting Example



To handle the needs of longitudinal studies, we assign the facility identifier associated with the subsidiary (OSHPDID, HOSPCODE) to the parent identifier. Over a given study period, we treat all data for consolidated facilities as if they were one. We developed decision rules based on variable type (categorical or continuous) to role the HAFD data into one record. A description of our methods to prepare the HAFD for longitudinal research is available elsewhere [11].

One problem with the historic BSMF file of names, addresses and HOSPCODES is that CDPH inconsistently updates it. Over time, hospitals change names, or use an address on another side of the hospital as their primary address, or move and keep the same or a different name. By contrast, OSHPD maintains detailed annual information on these important characteristics in various files. FHOP follows OSHPD's policy, to use current rather than historic name in reports.

We describe our methodology to link OSHPD and Vital Statistics identifiers based on geography (county, city, ZIP-code, address) and name. Then we report results of validating the linkage. Note that we previously established temporal sequencing of consolidations and moves [11].

## ANALYSIS FILES

---

### Linkage files

---

As part of routine HAFD processing [11], we make a file of street addresses from Page 00 (P00) over the period, currently encompassing 1981 through 2013. We correct city spelling and minimally edit street addresses (Street to St, Avenue to Ave, etc). The resulting file currently has 1,818 unique addresses and variables identifying the first and last year the address appeared for 759 hospitals that ever submitted HAFD data (the ADDR file). We use this to identify moves [11], to make our geography master file [12], and here to link HOSPCODE to OSHPDID.

We also make a file summarizing all names found in P00, which we use to identify possible changes in ownership, name, or auspice [11]. In addition to the P00 names, OSHPD provides other files identifying various structural or ownership changes that occur from year to year. We imported these files into SAS, added the result to the P00 list, and deleted duplicate names. Through 2013, the file prepared for linkage had 3,029 unique names for 759 hospitals.

As part of routine processing to prepare BSMF files for longitudinal research, we read into SAS the most recent Vital Statistics excel file with historic HOSPCODE, address, city, and ZIP from 1960 forward (N = 785) [8]. Our standard routine corrects the spelling of city names and does minimal editing to standardize addresses. In addition to using this file to identify linkages, we use this to make HOSPCODE formats (the HOSPCODE file) [13].

In preparing BSMF and OSHPD files for longitudinal analysis, another standard task is to summarize cases across all facilities and years from 1989 forward for the BSMF/FDSMF and from 1983 forward for OSHPD files, which include ED and AS data from 2005 forward. These are a resource to identify abrupt changes in number of cases or when facilities disappear or appear, which might indicate closure, consolidation, unconsolidation, or a move. The summaries identify the first and last year with data for the facility and the annual number of records. In the case of births, this presently covers the period 1989 to 2013 (N = 13,463,128 births) for 547 HOSPCODEs (the YRS\_HOS\_BC file).

We manually maintain an excel file with three tabs addressing linkage (LINK\_MANUAL). One tab (Unlinkable) is a list of HOSPCODES that we previously established will not link with the OSHPD files. These include facilities such as Federal hospitals, stand-alone birthing clinics, and physician offices where births occurred. After completing the linkage described below, the next version of the Unlinkable tab will have more HOSPCODES.

A second tab (Force) contains a list of HOSPCODES we are forcing to link with OSHPDID. We identify forced links in a variety of ways, including on-line searches.

The last tab (Reject) identifies links we reject temporarily or permanently. Temporary rejections typically involve stopping an ID associated with a later event from linking with an ID associated with an earlier event. For example, consider a move where the hospital had the same name before and after the move, Vital Statistics did not issue a new HOSPCODE, but OSHPD did. We take care of these situations later but temporarily set them aside.

A permanent reject involves situations such the following. The County Board of Supervisors and the Directors of a Healthcare District agreed to locate a county-managed Psychiatric Health Facility (PHF) on the District hospital grounds. The facilities were legally and administratively separate but shared the same physical address. No infants were born at the PHF. Vital Statistics assigned one HOSPCODE for the District Hospital and (appropriately) never assigned one to the PHF, while OSHPD correctly assigned two OSHPDIDs. In this case, because of different ownership, we do NOT want to link the PHF OSHPDID with the HOSPCODE.

## Validation files

---

Identifying structural changes in hospitals is an important part of our longitudinal research. Toward this end, we manually maintain what we call the ID report (IDRPyyyy), which chronicles structural changes over time. Among the various uses for this file, the most relevant for the current task is making the IDMAP file. This tracks OSHPDID changes due to moves, consolidation, and unconsolidation. IDMAP currently identifies 211 such changes. We use this file to begin the process of summarizing data to the surviving OSHPDID.

For validation purposes, we made a file counting the number of PDD records with admission source "Newborn" over the interval 1983-2014 by OSHPDID. The resulting file identified 445 OSHPDIDs, with 16,550,130 newborn records (the CKPDDBT file). Note that CKPDDBT does NOT account for consolidations or moves.

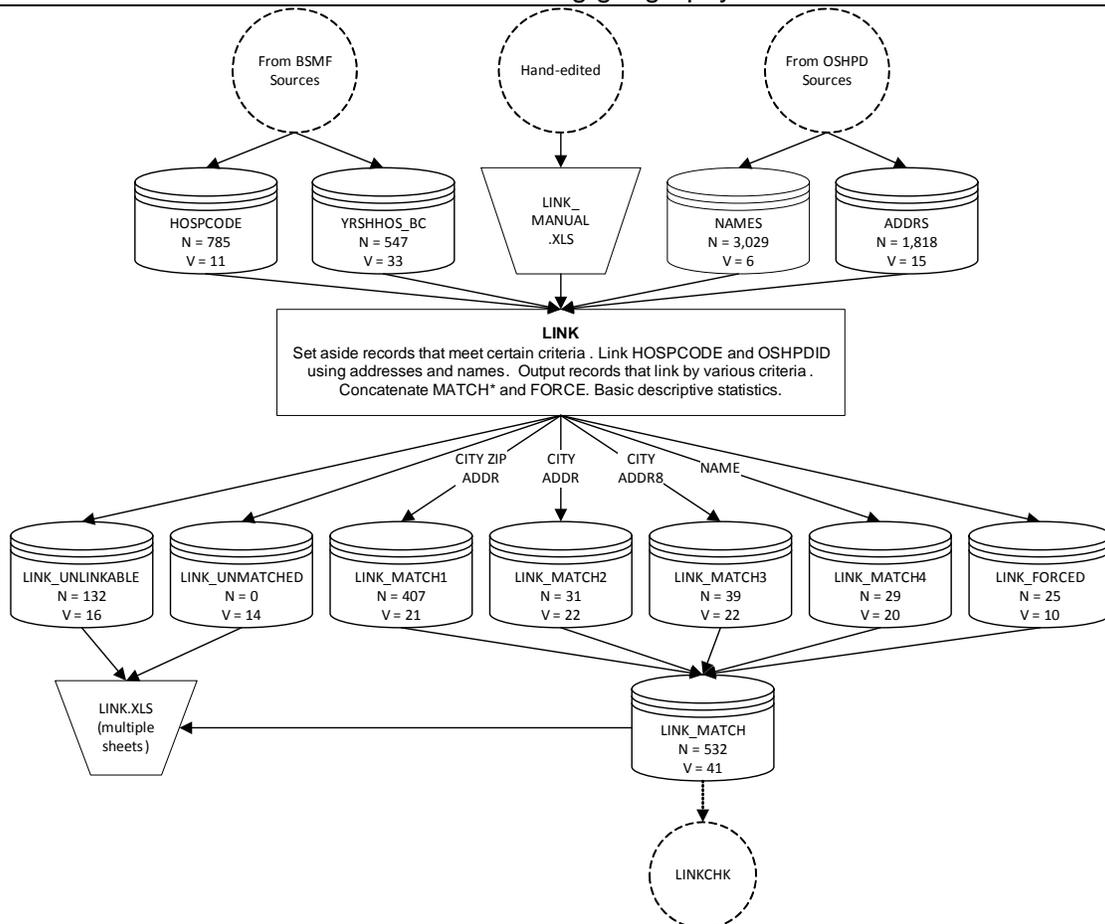
Beate Herrchen developed a methodology to link birth, death and hospital files [5]. OSHPD contracts with her company, Health Information Solutions, to make the linked file it distributes. Her method crosswalks identifiers within a year but does not address longitudinal changes due to moves or consolidated reporting. To validate our work, she generously provided her linkage

files (1995-2011, the most recent available). We transposed her files to one record per HOSPCODE, with flags indicating each year linkage occurred (BEATEXWKS).

## DATA LINKAGE

We first merged the HOSPCODE and YRSHOS\_BC files, and incorporated data in the Unlinkable tab from LINK\_MANUAL.XLS, separating facilities we knew could not link (N = 132, 212,965 births) from those potentially linkable (HOSP0, N = 653). We also prepared the ADDR0 (ADDRS0 N = 1,818) and NAMES (NAMES0 N = 3,029) files for matching. Finally, we prepared the LINK\_MANUAL tabs Forced and Rejected. Calling our MATCHMRG macro, we merged ADDR0 with HOSP0 by city, ZIP-code, and address outputting links that were not rejected (MATCH1 N = 407). Using residuals, the second merge was by city and address (MATCH2 N = 31). The third merge was by city and the first eight digits of the address (MATCH3 N = 39), and the last match was by hospital name (MATCH4 N = 29), with 25 records forced. The final file LINK\_MATCH had 532 records linked by HOSPCODE and OSHPDID. The unmatched HOSPCODE residual (LINK\_unmatched, N = 147) had 70 births (range 0 to 11). Figure 2. summarizes major steps in data linkage.

Figure 2. Link OSHPDID and HOSPCODE using geography variables



**Table 1. Summary of linkage results**

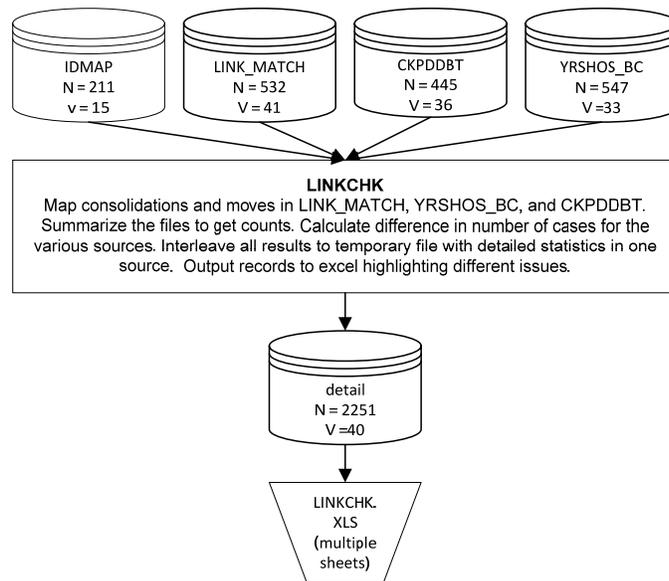
Match ID	OSHPDID matches		HOSPCODE matches		Multiple OSHPDID
	N	Col %	N	Col %	
Total	532	100	507	100	
1 CITY ZIPC5 ADDR	407	77	406	80	1
2 CITY ADDR	31	6	25	5	6
3 CITY ADDR8	39	7	21	4	18
4 NAME	29	5	29	6	0
9 FORCED	26	5	26	5	0

Table 1 shows linkage results. We linked 532 OSHPDID to 507 HOSPCODE with most in the first step, exact matches on city, ZIP-code and address.

## VALIDATION

### Compare births by source and hospital

**Figure 3. Examine linkage results by source**



Our next task was to validate linkages by comparing number of births in the Vital Records and OSHPD files. Figure 3 summarizes this work. First, we mapped consolidations and moves (IDMAP) for accepted linkages (LINK\_MATCH), Vital Statistics (YRSHOS\_BC), and PDD (CKPDDBT) births. Then we summarized the resulting files to reflect the current OSHPDID (SUMPDD N = 358; SUMBC N = 402). We merged the SUM files and calculated differences in number of births. Finally, we interleaved the files and output various results to excel for review.

The linkages we made over the interval 1983-2013 involved 13,250,093 BC births and 13,313,583 PDD births, an overall difference of 63,490 (0.48%).

We found 22 OSHPDID/HOSPCODE linkages with 575,229 PDD births (4.3% of PDD births) but no HOSPCODE births. Of these, 12 had their last birth (N = 24,984) before 1989 when Vital Statistics data became available. Another 10 involved consolidations or moves with the Subsidiary having no births and the Parent having births (N = 550,245).

We identified 78 linkages with HOSPCODE births but no OSHPDID births. Of these, 77 had 136 HOSPCODE births over the interval 1989-1998. One HOSPCODE had 11,892 births. This facility opened in 2008. It involved a cross-county consolidation, which our rules do not permit.

Table 2. shows results for the State and a few hospitals selected to highlight the complexity of how structural changes affect longitudinal reporting. The first column identifies the example number. The next identifies the county where the facility is located. The column “OSHPD Label” is the number OSHPD assigned the facility and the name OSHPD currently uses. The column “Vital Statistics Label” identifies the HOSPCODE and name Vital Statistics uses. The column SOURCE shows the source for the interleaved records discussed earlier. The next column shows total births over the interval 1989-2013 followed by annual counts for the most recent 12-year interval. FHOP typically uses 12 years for our DataBook products [14].

We begin with a simple case, **Example 1**. Kaiser Foundation opened a hospital in the County of Fresno and City of Fresno in 1995. The column “OSHPD Label (Original)” identifies that OSHPD assigned OSHPDID 104062 to this facility. The column “Vital Statistics Label” gives the HOSPCODE (772) and label that Vital Statistics assigned. The HOSPCODE links to the OSHPDID. Note that the Vital Statistics label differs from the OSHPD label. The column “Source” identifies the source of the data, with BC from YRSHOS\_BC and PDD from CKPDDBT. The first row in the column “Births” identifies Vital Statistics birth counts (BC = 23,650). The second row reports OSHPD birth counts (PDD = 23,671). In the simple case as here, with no consolidations or moves, SUMBC = BC, and SUMPDD = PDD. Over the interval, the difference in births reported by these sources (X\_DIFF) is 21 records (0.6%). Looking over the row, some years BC had a few more records, other years PDD had a few more.

**Example 2** involves consolidated reporting. Again in Fresno County and City, Community Health System owns “Community Regional Medical Center – Fresno” (OSHPDID 100717, HOSPCODE 058, Set 2, Row 1). We show this involves a consolidation, because the OSHPDID column has that code, indicating it is the Parent, while the OSHPD Label shows two OSHPDID. Specifically, in 1996, Fresno County sold to Community Health System the previously named “Valley Medical Center of Fresno” (OSHPDID 100822, HOSPCODE 067, Set Row 2). From the HAFD, this change of ownership (CHOW) involved a change of auspice (CHOA) from public to non-profit, and a change of name (CHON) to the current “University Medical Center”. Under column SOURCE, two lines report BC births and two report PDD births, with SUMBC and SUMPDD the sums of those lines. Although larger in absolute number of records than Example 1, the difference of -1112 between data sources reflects just -0.6% over the interval.

In 1999, Community Health System consolidated HAFD reporting with 100717 becoming the Parent and 100822 becoming the Subsidiary. In 2003, they consolidated patient reporting and, from the HAFD, closed delivery units in 100822 in favor of 100717. However, we have no way from this to know if Community Health System closed just the delivery unit, or indeed, if OSHPDID 100822 continues to survive. From OSHPD’s Automated Licensing Information and Report Tracking System (ALIRTS), 100822 closed in 2009 [15]. Consolidated reporting masked the loss of one hospital in Fresno County.

**Table 2. Comparison of number of births by hospital, data source, and time**

Example	County	OSHPDID	OSHPD Label (Original)	Vital Statistics Label	Source	Births	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	
State Total						BC	13,250,093	517235	529204	533624	536499	550299	554084	540298	515915	497159	488867	486461	477461
						PDD	13,313,583	519250	530527	534474	538384	550795	554906	539274	514671	497823	489883	491835	479184
						X_DIFF	63,490	2015	1323	850	1885	496	822	-1024	-1244	664	1016	5374	1723
1	10	Fresno	104062 104062 Kaiser Foundation Hospital - Fresno	0772 Kaiser Permanente Medical Center	BC	23,650	1311	1296	1255	1225	1317	1355	1210	1112	1012	1123	1112	1125	
						PDD	23,671	1311	1298	1258	1222	1318	1366	1211	1107	1009	1124	1114	1125
						SUMBC	23,650	1311	1296	1255	1225	1317	1355	1210	1112	1012	1123	1112	1125
						SUMPDD	23,671	1311	1298	1258	1222	1318	1366	1211	1107	1009	1124	1114	1125
						X_DIFF	21	0	2	3	-3	1	11	1	-5	-3	1	2	0
2	10	Fresno	100717 100717 Community Regional Medical Center - Fresno	0058 Community Regional Medical Center	BC	149,916	5511	6738	7484	7586	7903	6951	6849	6548	6617	6477	6375	5976	
						BC	37,520	1608	649	7	2								
						PDD	148,505	5543	6736	7486	7625	7921	7000	6848	6542	6629	6529	6436	6040
						PDD	37,819	1628	678		2	1							
						SUMBC	187,436	7119	7387	7491	7588	7903	6951	6849	6548	6617	6477	6375	5976
						SUMPDD	186,324	7171	7414	7486	7627	7922	7000	6848	6542	6629	6529	6436	6040
						X_DIFF	-1,112	52	27	-5	39	19	49	-1	-6	12	52	61	64
3	16	Kings	160787 160787 Central Valley General Hospital	0111 Central Valley General Hosp	BC	39,859	2056	1803	1895	1880	2077	2236	2232	2044	1877	1884	1968	1970	
						PDD	40,136	2061	1806	1901	1900	2088	2255	2226	2062	1893	1912	1993	1999
						SUMBC	39,859	2056	1803	1895	1880	2077	2236	2232	2044	1877	1884	1968	1970
						SUMPDD	40,136	2061	1806	1901	1900	2088	2255	2226	2062	1893	1912	1993	1999
						X_DIFF	277	5	3	6	20	11	19	-6	18	16	28	25	29
						BC	12,324	3	2	5	2	2	3	2	1				
						BC													
						PDD	17,665				133	763	1049	1014	1227	1073			
						PDD	2,961								86	1151	955	769	
						SUMBC	12,324	3	2	5	2	2	3	2	1				
						SUMPDD	20,626				133	763	1049	1014	1227	1159	1151	955	769
						X_DIFF	8,302	-3	-2	-5	131	761	1046	1012	1226	1159	1151	955	769
10	Fresno	100793 100793 Selma Community Hospital	0063 Selma Community Hospital	BC	20,414	893	1150	1230	891	763	1049	1012	1225	1159	1132	948	769		
						PDD	12,121	896	1150	1226	767								
						SUMBC	20,414	893	1150	1230	891	763	1049	1012	1225	1159	1132	948	769
						SUMPDD	12,121	896	1150	1226	767								
						X_DIFF	-8,293	3	0	-4	-124	-763	-1049	-1012	-1225	-1159	-1132	-948	-769
10	Fresno	100797 100797 Adventist Medical Center - Reedley	0065 Sierra Kings District Hospital	BC	28,149	1344	1392	1488	1569	1735	1790	1601	1596	1256	1194	1192	1187		
						PDD	28,174	1340	1395	1485	1575	1740	1788	1588	1598	1262	1197	1201	1171
						SUMBC	28,149	1344	1392	1488	1569	1735	1790	1601	1596	1256	1194	1192	1187
						SUMPDD	28,174	1340	1395	1485	1575	1740	1788	1588	1598	1262	1197	1201	1171
						X_DIFF	25	-4	3	-3	6	5	-2	-13	2	6	3	9	-16

**Example 3** involves five hospitals now owned by Adventist Healthcare [16]. In 1965, Adventist built Hanford Community Medical Center (OSHPDID 160725, HOSPCODE 109) in the City of Hanford, Kings County. In 1998, they purchased the then named “Sacred Heart Hospital” (OSHPDID 160787, HOSPCODE 111) also in Hanford. This CHOW did not involve a CHOA as both systems are religiously affiliated, but did involve a CHON to “Central Valley General Hospital”. Note that both the OSHPD and Vital Statistics labels reflect this CHON.

One year later, Adventist purchased Selma Community Hospital (OSHPDID 100793, HOSPCODE 063). This CHOW involved a CHOA from public (Selma Healthcare District) to religious and a CHON reflected in both the OSHPD and Vital Statistics labels. Only 15 miles from Hanford, Selma is located in Kings County.

In 2005, Adventist consolidated OSHPD reporting for Hanford (Parent 160725) and Selma (Subsidiary 100793). Notice that births continue to be reported to Vital Statistics from Selma (HOSPCODE = 63) in Fresno County, while in the PDD, births are reported from Hanford (OSHPDID 164029) in Kings County. We do not link IDs across counties. Reflecting this complex history, we see a large difference longitudinally in the number of birth records at these facilities: 8,302 vs -8,293. Looking across the X\_DIFF row, we see that total differences are small until 2005, then differ markedly thereafter.

Further confusing the situation, in 2010, Adventist moved Hanford (160725) to a new location (OSHPDID 164029) and changed the name to “Adventist Medical Center - Hanford”. Vital Statistics assigned a new HOSPCODE (791) and name “Community General Medical Center” that does not reflect the CHON while the OSHPD label does. Note that OSHPDID 164029 groups these records; 160725 was the Parent, then moved to 164029, which becomes the renumbered Parent.

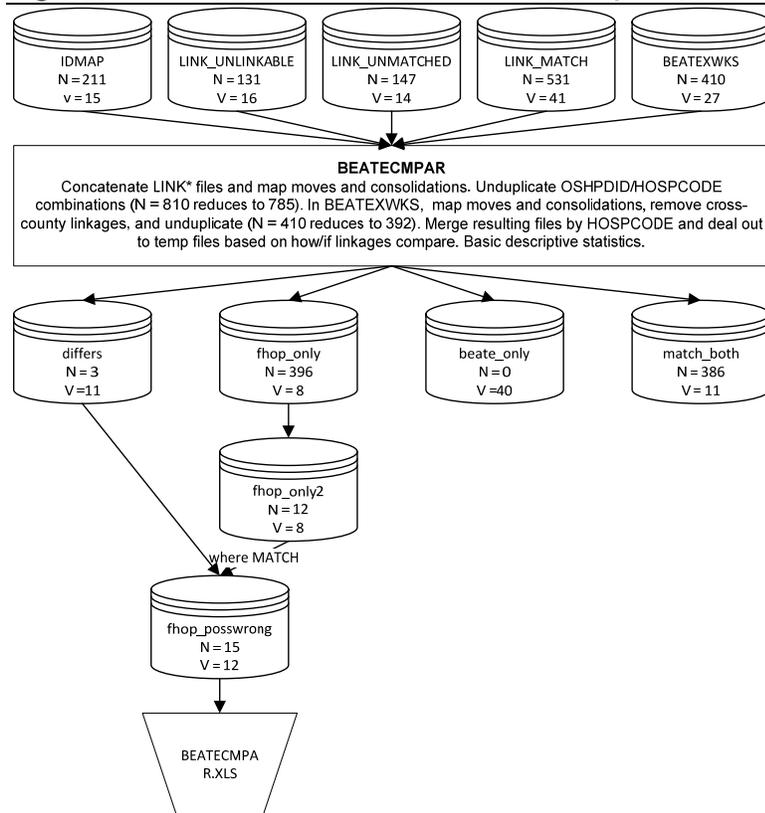
Finally, in 2011, Adventist purchased the Sierra Kings District Hospital in Fresno County. This CHOW again involved a CHOA from public to religious, and a CHON to “Adventist Medical Center – Reedley”. Again, the Vital Statistics label does not reflect the CHON.

## Compare FHOP and Herrchen linkage

---

The transposed Herrchen file (1995-2011) had 410 records with HOSPCODE linked to OSHPDID (BEATEXWLS). Because she links a shorter period than FHOP (1983-2013), Herrchen has fewer links than FHOP. Herrchen links within a given year and does not track longitudinal changes. In this context, she linked 6 OSHPDID to 11 HOSPCODE, with 3 intra-county and 3 cross-county, and the remaining 398 OSHPDID linked to a single HOSPCODE. She also linked 10 HOSPCODE to 18 OSHPDID with 2 cross-county, and 369 HOSPCODE linked to a single OSHPDID. Because of our interest in county structural capacity, our algorithm does not permit cross-county linkages. Figure 4. summarizes steps to assess the reliability of FHOP’s linkage algorithm relative to Herrchen’s linkage.

Figure 4. Examine FHOP and Herrchen linkage results



Of 785 unduplicated FHOP records, 386 linkages were in the Herrchen files (match\_both) and three differed. The 396-record FHOP-only file had 117 linkages that Herrchen did not make, with 12 having births in 1995 or later. We output 15 records (3+12, fhop\_posswrong) to Excel for manual review. All involved consolidations or moves and were determined to be correct given our slightly different rules.

Herrchen did not make any linkage that we declared unlinkable (N = 131) or that we had been unable to match (N = 147), and no records remained in her comparison file.

## AN EXAMPLE

We have been interested for many years in developing strategies to monitor data quality longitudinally [17,18], because of the impact on reliability of population health indicators. Since a given hospital submits all data to OSHPD (for PDD or ED) and Vital Statistics (BC), an interesting question is whether different hospital departments submit data of similar or different quality. This is in the context that the CDC sets a 1% standard for undefined (other, unknown, missing - OUM) race or ethnicity. For longitudinal statistical reports, the CDC bridges OUM and multi-race to a single 5-category race/ethnicity variable: Hispanic all-race, and non-Hispanic White, Black, Asian/Pacific Island, and American Indian/Alaska Native [19-22].

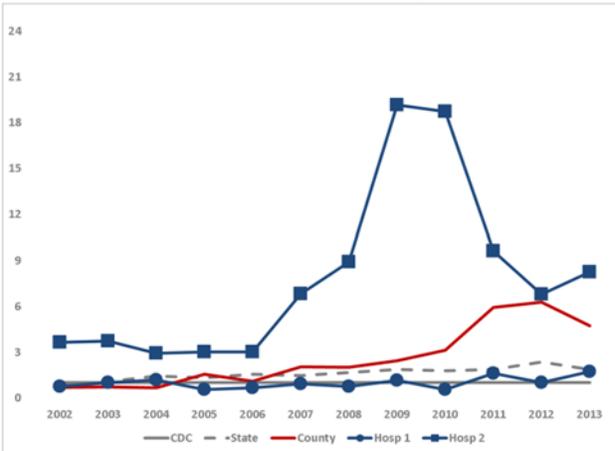
The issue is further complicated because OSHPD and Vital Statistics differently define race. OSHPD assigns North American natives to one code, and assigns Central and South American natives plus multi-racial people to "Other". Following CDC rules, Vital Statistics assigns natives of North, Central and South America to one group, and facilitates multi-race calculation. Developing comparable race/ethnic statistics across data sources is difficult in this situation.

As an example, focusing on OUM, we summarized data quality by year (2002-2013), county of occurrence and hospital, for females age 15 to 44 admitted to hospital (PDD) or emergency department (ED, introduced in 2005), and females delivering live-born infants as reported in the

BSMF. Given that OSHPD permits consolidated reporting and Vital Statistics does not, we had to crosswalk the OSHPD and Vital Statistics data to the current, respective codes. Figure 5 compares OUM data quality trends across data sources for the combined race/ethnic variable for the State, one County and two hospitals in that County, over the period 2002-2013.

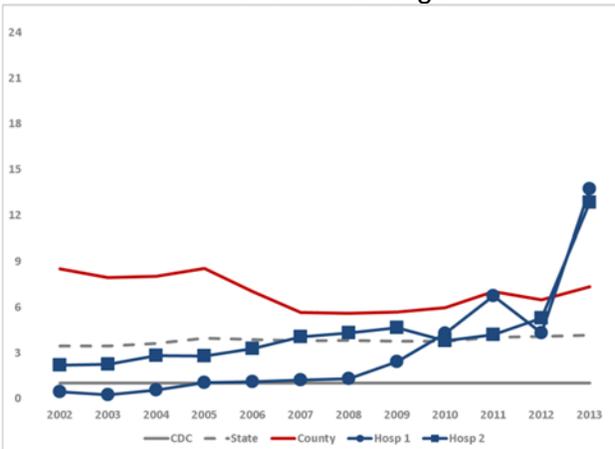
Figure 5. OUM data quality by source and level

A. Birth certificate



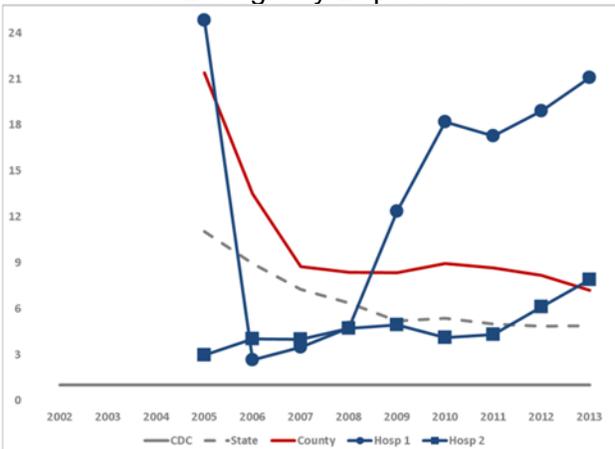
In BC data, the State had a steady small increase in OUM data quality problems from below the standard (0.8%) to 2% by period end. The County rose steadily from below standard (0.7%) to 6.3% in 2012 then dropped slightly. Hospital 1 had a steady small increase from below standard (0.8%) to 2% by period end. Hospital 2 began above standard (3.6%) peaked in 2009-2010 at 19% then dropped back to 8%.

B. Patient Discharge



In the PDD, the State rose steadily from 3.2 to 4.4%. The County began the period at 8.5% and dropped to 7.3% by period end. Hospital 1 began below standard (0.4%) and Hospital 2 began above standard at 2.2%. Both hospitals rose over the interval then spiked in 2013 to about 13%

C. Emergency Department



OSHPD introduced ED datasets in 2005. At State and County levels, OUM data quality rates improved steadily from 12% and 21% respectively but both remained above standard, 4.9% and 7.2% by period end. In 2005, 25% of Hospital 1 ED records were undefined, plummeting to 2.7% then rising markedly to 21% by period end. Hospital 2 rose throughout the period from 3% (well below State) to 7.9% (slightly above State).

From this simple example, we can begin to understand the strengths of linking facility identifiers. Our example shows that wild swings at local hospitals average out at the County level, and trend lines become ever smoother at the State level. Different departments in the same hospitals clearly provide data of different quality for the same population in the same or different years, and over time. Local health jurisdictions could use linked data such as this example illustrates to approach hospitals and encourage them to improve data quality.

## SUMMARY

---

We posted the OSHPDID/HOSPCODE crosswalk on our website, and the file to assign current IDs for longitudinal studies. If hospitals notify us that our crosswalk is incorrect, we will make the appropriate changes and re-estimate any affected work. SAS programs are available on request.

A central part of the FHOP research agenda is to examine relationships between changes in hospital infrastructure and changes in hospital outcomes. A fundamental infrastructure issue is whether hospitals are open and if so, whether they care for a given population. We initially thought it would be simple to answer the question, "How many hospitals are open each year and where are they?" Answering this turned out to be one of the most time-consuming tasks.

In theory, hospital license numbers identify a given physical location. When hospitals disappear from various files, the basis is not readily apparent. We must determine if the facility closed, merged, converted to consolidated reporting, or moved, resulting in a new license ID. Yet another possibility is that the licensing entity assigned a new ID to a facility at the same location. Some hospitals move without getting a new ID.

Consolidated reporting obscures the number of open hospitals and where they are located. The California legislature should mandate one uniform number to use across agencies (Licensing and Certification, OSHPD, Vital Statistics, etc.). The number should reflect the physical location with type of care provided as a sub-number. Our view is that the state should not permit consolidated reporting across county lines.

Only by interviewing people and engaging colleagues across the state and nation were we able to understand what we were seeing. The overall problem is one of transparency, the fact that it is so difficult to clarify and resolve (to the extent we were able) the data problems we encountered. Opaqueness benefits hospitals and thwarts legitimate public information needs.

We particularly want to acknowledge the generosity of Beate Herrchen. To validate our linkage, she provided the Health Information Solutions files used to make OSHPD's linked files. It was heartening to confirm that, working independently, we made the same linkages.

## ENDNOTES

---

- 1 See Data Files. Last accessed 21-Jan-2016 at: <http://www.cdph.ca.gov/data/dataresources/requests/Pages/VitalStatisticsBirthDeathFetalDeathMarriageData.aspx>
- 2 See Data Request Center. Last accessed 21-Jan-2016 at: [http://oshpd.ca.gov/HID/Data\\_Request\\_Center/](http://oshpd.ca.gov/HID/Data_Request_Center/)
- 3 See Annual Financial Data. Last accessed 21-Jan-2016 at: <http://www.oshpd.ca.gov/HID/Products/Hospitals/AnnFinanData/CmplteDataSet/index.asp>
- 4 See questions 6A and 6B, last accessed 21-Jan-2016 at: [http://www.oshpd.ca.gov/HID/Data\\_Request\\_Center/Data\\_Documentation.html](http://www.oshpd.ca.gov/HID/Data_Request_Center/Data_Documentation.html)
- 5 Health Information Solutions. California Vital Stats to Care Delivery Linkage See: [http://health-info-solutions.com/HIS\\_support.html](http://health-info-solutions.com/HIS_support.html)
- 6 OSHPD. Linked birth and death data. See: [http://www.oshpd.ca.gov/HID/Data\\_Request\\_Center/Types\\_of\\_Data.html](http://www.oshpd.ca.gov/HID/Data_Request_Center/Types_of_Data.html)
- 7 Herrchen B, Gould JB, Nesbitt TS (1997). Vital statistics linked birth/infant death and hospital discharge record linkage for epidemiological studies. *Comput Biomed Res.* 1997 Aug;30(4):290-305.
- 8 California Department of Public Health (2013) California Birth Statistical Master File Hospital Codes 1960-2013. Provided as part of the documentation for the 2013 Birth Statistical Master File.
- 9 OSHPD (2009) Accounting and Reporting Manual for California Hospitals, Second edition. Last accessed 19-Jul-2016 at: <http://www.oshpd.ca.gov/HID/Products/Hospitals/AnnFinanData/Manuals/index.html>
- 10 California State Archives (2009) Inventory of the Office of Statewide Health Planning and Development Records. Last accessed 24-Nov-2014 at: <http://pdf.oac.cdlib.org/pdf/csa/oshpd.pdf>.
- 11 Remy L, Clay T, Oliva G (2004) Creating Longitudinal Hospital-Level Datasets: Annual Hospital Disclosure Report. UCSF Family Health Outcomes Project. Last accessed 24-Jul-2012 at: [http://fhop.ucsf.edu/fhop/docs/pdf/pubs/METH\\_AHDR.pdf](http://fhop.ucsf.edu/fhop/docs/pdf/pubs/METH_AHDR.pdf)
- 12 Remy LL, Clay T (2015) Managing Longitudinal Research Studies: The Geography Master. San Francisco, CA: University of California, San Francisco, Family Health Outcomes Project. Available at: <http://familymedicine.medschool.ucsf.edu/fhop/>.
- 13 Remy L, Clay T. (2015) 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://familymedicine.medschool.ucsf.edu/fhop/>
- 14 Remy LL, Clay T, Oliva G. (2013) Title V 5-Year Needs Assessment Indicators: DataBook 2000- 2011. San Francisco, CA: University of California, San Francisco, Family Health Outcomes Project. Available at: <http://fhop.ucsf.edu/fhop/>.
- 15 Automated Licensing Information and Report Tracking System (ALIRTS). Last accessed 07-Jul-2016 at: <https://www.alirts.oshpd.ca.gov/UtilReports.aspx?FacId=57&FType=hosp&FName=UNIVERSITY%20MEDICAL%20CENTER&OID=106100822&isReg=1>

- 16 Adventist Health Central Valley Network History. Last accessed 24-Jan-2016 at: <https://www.adventisthealth.org/central-valley/pages/about-us/history.aspx>
- 17 Remy LL, Oliva G (2006) Impact of Data Quality on Birth Related Health Indicators. Family Health Outcomes Project, University of California, San Francisco. Last accessed 19-Jul-2016 at: <https://fhop.ucsf.edu/sites/fhop.ucsf.edu/files/wysiwyg/dATAQUAL7.pdf>
- 18 Remy LL, Oliva G, Clay T (2013). The impact of a birth clerk training program on California birth data quality 1998-2009. Family Health Outcomes Project, University of California, San Francisco. Available at: [fhop.ucsf.edu/](http://fhop.ucsf.edu/)
- 19 U.S. Census Populations With Bridged Race Categories. CDC, NCHS, NVSS. Last accessed 17-Jun-2016 at: [http://www.cdc.gov/nchs/nvss/bridged\\_race.htm](http://www.cdc.gov/nchs/nvss/bridged_race.htm)
- 20 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). Last accessed 17-Jun-2016 at: [http://www.cdc.gov/nchs/nvss/bridged\\_race.htm](http://www.cdc.gov/nchs/nvss/bridged_race.htm)
- 21 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 access 17-Jun-2016 at: [http://www.cdc.gov/nchs/data/dvs/Multiple\\_race\\_docu\\_5-10-04.pdf](http://www.cdc.gov/nchs/data/dvs/Multiple_race_docu_5-10-04.pdf)
- 22 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.