THE IMPACT OF A BIRTH CLERK TRAINING PROGRAM ON CALIFORNIA BIRTH DATA QUALITY 1998-2009

By

Linda L Remy, MSW PhD

Geraldine Oliva, MD MPH

Ted Clay, MS

Geraldine Oliva, MD MPH, 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

January 2013

Blank page intentionally inserted

TABLE OF CONTENTS

This work was carried out under Contract A107398, with the California Department of Public Health. The authors particularly appreciate the support of Gwyn Doebbert and Scott Fujimoto at California's Office of Health Informatics and Research, Alan Oppenheim at the California Center for Health Statistics, and Joyce Martin and her staff at the Centers for Disease Control, National Center for Health Statistics.

Suggested Citation:

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/

ACRONYMS

AVSS	Automated Vital Statistics System
BDQT	Birth Data Quality Training
BSMF	Birth Statistical Master File
CDC	Centers for Disease Control
CDPH	California Department of Public Health
CCHS	California Center for Health Statistics
CURROSH	OSHPDID after assigning consolidations and moves
DOB	Date of birth
ES	Effect Size
FHOP	Family Health Outcomes Project
HOSPCODE	Number assigned by Vital Statistics to facilities (licensed California hospitals, military hospitals, birthing centers, etc) submitting birth data for vital records
LHJ	Local health jurisdiction
LLB date	Date of last live birth
LMP date	Date of last menstrual period
MACRO	A special program for SAS code that is used over and over again in the same or different programs. Macros are data driven, in that the instructions let SAS decide what to do based on actual data values.
MCAR	Missing completely at random
MNAR	Missing not at random
NAPHSIS	National Association for Public Health Statistics and Information Systems
NCHS	National Center for Health Statistics
NCVS	National Center for Vital Statistics
OSHPD	Office of Statewide Health Planning and Development
OSHPDID	Hospital number assigned by OSHPD
OHIR	Office of Health Informatics Research
OOR	Out of range
OVR	Office of Vital Records
PEOPLE0	Number of people trained at first session to which hospital sent representatives
PEOPLEP	Number of people trained in later sessions to which hospital sent representatives
REMRESN	Reason for removing hospital from analysis
R-Sq	R-Squared or percent of variance explained
TRAINP	Number of trainings after first to which hospital sent representatives
TRNDATE0	First training date found for any hospital representative
ZIP	ZIP-code
UCSF	University of California, San Francisco
UNP	Unaffiliated Non-Profit Hospital

THE IMPACT OF A BIRTH CLERK TRAINING PROGRAM ON CALIFORNIA BIRTH DATA QUALITY 1998-2009

EXECUTIVE SUMMARY

The Birth Certificate Statistical Master File (BSMF) is the primary source of data elements used to calculate a variety of maternal and infant population health indicators. In addition, it is useful in documenting racial, ethnic, income, and geographic disparities. Many state and local health programs and agencies use these indicators to assess the health of women and children, compare their rates with national standards, and evaluate the effectiveness of various program and policy initiatives.

In 2003, the CDC/NCHS notified California Center for Health Statistics (CCHS) that rates of missing values for key data elements on the birth certificate did not meet Federal standards. The next year CCHS began to offer voluntary Birth Data Quality Trainings (BDQT) for birth clerks and others.

The Family Health Outcomes Project (FHOP) has been working with CCHS for many years on various aspects of data quality in the birth and death certificates. In 2009, CCHS asked FHOP to evaluate effectiveness of the BDQT. However, we did not receive all the needed data until late 2011.

METHODS: The study utilized an interrupted time series design over the period 1998-2009. The analysis is based on data from California's BSMF, structural information from the Annual Hospital Disclosure Report distributed by the Office of Statewide Health Planning and Development (OSHPD), and training attendance data provided by the BDQT staff. We worked closely with staff at the CDC/NCHS to develop SAS programs and macros that implemented their decision rules addressing data quality.

RESULTS: After controlling for time, hospital volume, and hospital auspice, we found little evidence that the BDQT had an impact on improving data quality. The extent to which results were impacted by data quality issues addressing BDQT attendance is unknown. However, trends are clear. Data quality deteriorated steadily or remained about the same over the study period for most measures evaluated. We did not have data to see if trends changed after 2009.

Most important, we confirmed that data were missing not at random (MNAR), a serious and statistically non-ignorable data problem [1-3]. Data quality varied by hospital volume and auspice, with considerable variation within and across years and geography (hospitals, counties). In a study with mainly moderate results for predicting birth certificate data quality as reflected by effect size (ES), hospital auspice was the most consistent predictor.

DISCUSSION: For many years, California has been well above the CDC/NCHS standard for missing data for key BSMF data fields. Results reported here show unequivocally that data quality worsened rather than improved between 1998-2009. We do not know why hospitals operating under some auspices have high data quality error rates while others under other auspices submit better quality data. We had intended to collect some key informant interview data from hospitals to explore possible causes of these variations. Due to budget constraints, this was not possible.

California has been calculating public health indicators assuming data were missing completely at random (MCAR) when in fact the data have been MNAR for many years. When data are MNAR, imputation is recommended [1-3].

Because they have known for many years that data often are MNAR, the CDC/NCHS developed detailed decision rules and some SAS programs to correct data before calculating Healthy People indicators

based on BSMF. We used their rules to do this analysis, converted several of their SAS programs to macros, and developed other new macros to clean data and calculate other indicators following CDC/NCHS rules. We are making these publicly available.

RECOMMENDATIONS

An important limitation of this study was missing and poor quality data on the BDQT attendance. We recommend that an evaluation plan be developed prior to initiating any future training to minimize the methodological problems we encountered with the training data provided to us.

We make a number of recommendations that might improve the BDQT effectiveness. We suggest exploring the possibility of moving to webinars with short-focused agendas that might be more palatable to both hospitals and birth clerks, can be archived for later viewing, and would more cost effective for both the hospitals and CCHS.

We also advise that there be some focused effort to involve hospital administrators. We should survey hospitals to assess staff qualifications and ratio of staff to births as well as QA activites and other systems issues that could impact data quality. Hospitals should be encouraged to make birth clerk attendance a required part of continuing education for staff. HISP should initiate a round of meetings with Chief Executive Officers of hospitals to review results of this research, and learn what might be done to gain their cooperation.

We very strongly recommend that CCHS adopt a policy to use available CDC/NCHS editing and imputation rules before calculating certain public health indicators. Until California finds an effective strategy to improve data quality, and until that strategy is followed for many years, public health indicators calculated as if data were MCAR will remain inaccurate and misleading. Imputation is the only solution for this very real and non-ignorable statistical problem. It affects every BSMF variable we studied, most importantly those used to calculate California’s Healthy People indicators for the Maternal, Child, and Adolescent Health (MCAH) population.

THE IMPACT OF A BIRTH CLERK TRAINING PROGRAM ON CALIFORNIA BIRTH DATA QUALITY 1998-2009

BACKGROUND

Birth certificate data are the primary source for calculating a variety of maternal and infant population health indicators. Among others, these include rates for fertility, inter-pregnancy interval, entry into and adequacy of prenatal care, delivery mode, preterm birth, low birth weight, and other important pregnancy outcomes. In addition, it is useful in documenting racial, ethnic, income, and geographic disparities in these rates.

Data quality is critical in accurate calculation of indicator rates. Many researchers have identified significant data quality issues with birth certificates and the implications for mis-estimating negative outcomes. Incorrect values can include missing, unknown, or unlikely. For example, how likely is it that a female age 15 just delivered her ninth live born child? Various elements of birth certificate data have been evaluated with respect to their reliability [4-7] and validity [8,9]. Another body of work has evaluated the impact of missing data [8-13].

In California, birth certificates are more likely to be incomplete for infants who later die [14]. Further, the higher a sub-population's risk of poor outcomes, the greater the likelihood that birth records will be incomplete [12,14-18]. Thus, when calculating indicator rates, excluding records with missing or unlikely values may lead to misleading rates and inaccurate estimates of progress toward Healthy People 2020 objectives and race/ethnic disparities.

The National Center for Health Statistics (NCHS) calculates the percent missing for each required birth certificate variable. NCHS issues an annual report comparing all states [19-21]. Since the early 2000s, California has not met NCHS data quality standards for a number of birth certificate variables. For example, it has been above the national standard for mother’s race since 2003. By 2005, California was above the national standard for: mother’s and father’s race, ethnicity, education, and others. Responding to these escalating problems with data quality, the California Center for Health Statistics (CCHS) started Birth Data Quality Trainings (BDQT) for birth clerks and others. This paper reports the results of an independent evaluation of BDQT effectiveness.

Previous work. For many years, the Family Health Outcomes Project (FHOP) at the University of California San Francisco has worked with the Maternal Child and Adolescent Health (MCAH) branch of the California Department of Health to calculate indicators for California local health jurisdictions (LHJ). MCAH requires counties to monitor these indicators as part of Federal Title V funded cooperative agreements between the state and the LHJs. Upon observing great variation in rates for missing or unlikely values that would impact the LHJs rates, FHOP initiated routine county-level data quality checks. The reports identify jurisdictions where poor quality data could lead to incorrect conclusions about MCAH population health status and/or the affect the evaluation of program interventions.

Data quality reports help counties assess the potential impact of data errors and give health department staff information to use to work with providers and hospitals to improve data quality. The data quality reports have identified significant variation among jurisdictions. For rural counties, even a few missing or unlikely values can result in misleading conclusions about the quality and adequacy of prenatal care or the effectiveness of doing outreach to bring women into care. This suggests that using the state average to gauge problems is not helpful for a state as large and diverse as California.

In 2003, in response to reports on missing values from NCHS, CCHS initiated an effort to improve the quality of information for the certificate of live births by providing training for hospital birth clerks and related staff. Initially, hospitals with the highest percent of missing data for maternal ethnicity, education, and month and year of last menstrual period were contacted and offered training for their birth clerks. Beginning in the last quarter of 2004, CCHS started to host regional BDQT throughout the state for birth clerks and related personnel.

CCHS also began to use the Automated Vital Statistics System (AVSS) to generate reports for every California delivery hospital that summarized counts for variables with unknown values. Note that AVSS has quality checks alerting the user to an unlikely value, but the user can bypass these.

In 2006, CCHS asked FHOP to examine the impact of data quality on birth-related indicators to provide evidence of the impact of errors on calculating health indicators. It was felt that presentation of this kind of data would be a useful addition to the BDQT and illustrate the importance of accurate data to the clerks.

After reviewing data quality between 1992 and 2003 for several perinatal indicators, FHOP decided to focus on date variables used to calculate preterm birth rates. This indicator is calculated from several date variables that can be either missing or unlikely. FHOP found that failing to impute missing or improbable values for variables used to calculate preterm birth rates can lead to statistically significant results and thus erroneous conclusions, particularly in jurisdictions with poor quality data [16]. The impact is similar both locally and at the state level. Findings were similar to those of other researchers [8,17,18].

While research has shown that birth certificate data quality problems exist, there have been no efforts to examine the impact of training birth clerks responsible for submitting the information. CCHS asked FHOP to do such a study. This is a report of our findings.

STUDY DESIGN AND METHODS

Study Question

Did the California training program for hospital birth clerks conducted between 2004-2009 improve the quality of birth certificate data?

Hypothesis

A targeted training program for birth clerks will result in significant improvement in the quality of birth certificate data.

Research Questions

- Are differences in hospital characteristics (e.g., volume, auspice) related to differences in quality and completeness?
- Did the California training program for birth clerks improve the quality of birth certificates? Specifically, when comparing data quality before and after training, is there a significant decrease in missing or unlikely values in fields targeted by the training?

- Is the number of trainees associated with greater improvement compared to hospitals with few or no attendees? That is, is there a dose effect?

- Are improvements sustained over time?

Study Design

The essence of the design is the presence of a periodic measurement process (routine collection of BSMF data through AVSS) and introduction of an experimental change (training) into this time series of measurements. This design is known as an interrupted time series [23].

Birth Statistical Master Files

We used confidential versions of California's BSMF to calculate data quality indicators over the 12-year period 1998 through 2009. These files have variables such as names, exact dates, addresses, and social security numbers that are not available in public release datasets. The BSMF that we use include reallocates, that is, out-of-state births to California resident women as well as home births. We previously prepared these data for longitudinal analysis using methods described elsewhere [24].

In the early 1960s, Vital Statistics began to assign "Hospital Codes" (HOSPCODE) to facilities delivering infants. Under state law, these facilities submit birth information to the California Department of Public Health (CDPH), which makes the official birth certificate. Hospital codes are assigned to stand-alone clinics, birthing centers, physician offices, and military hospitals, as well as hospitals licensed by the State of California [25]. We selected all births in California with a valid hospital code (N = 6,437,003) during the study period.

We consulted with Joyce Martin (CDC/NCHS) for guidance on methods to edit birth files. She provided information to define data quality problems [26-30]. Alan Oppenheim (CDPH-CHS) provided federally-generated reports comparing California BSMF data quality statistics to national standards [31].

After calculating quality indicators, we summarized the data by HOSPCODE, year, and birth month. The initial summary resulted in 41,344 records for 364 HOSPCODES.

Annual Hospital Disclosure Report

The Licensing and Certification Program (LCP) of CDPH is responsible for licensing, regulating, inspecting, and/or certifying various types of California hospitals and other healthcare facilities. LCP coordinates with the CDPH Office of Statewide Health Planning and Development (OSHPD). This agency is responsible for overseeing the building of California hospitals and has many data collection and dissemination responsibilities. Military hospitals are exempt from CDPH licensing and certification but are required to submit Vital Statistics information for births.

Before evaluating the impact of training on data quality, we wanted to control for temporally prior hospital characteristics. To do this, we used OSHPD's desk-audited Annual Hospital Disclosure Report (AHDR) [32]. OSHPD desk audits each submitted report for compliance with OSHPD's annual reporting system requirements, spending about 12 hours of professional review time on each submitted report, to assure data accuracy and reasonableness. The California Department of Health Services also conducts an on-site review of hospitals' accounting systems and further validates reported data [33].

All hospitals licensed by California submit these data. The information includes available staffing and services, ownership, types and number of beds, patient days and discharges, balance sheets and income statements, revenues by payer, and expenses by natural classification. For this study, our particular interest was hospital ownership (AHDR Pages 0 and 1) and volume (AHDR Page 4): number of total acute care and neonatal intensive care unit (NICU) discharges, and number of deliveries (from BSMF).

Because hospitals move and/or consolidate reporting, we had to implement methods to track such events longitudinally and modify data accordingly. The AHDR is based on the hospital's fiscal year. About half of hospitals use a calendar year and about half use a fiscal year. When hospitals change owners, move, close or reopen, the AHDR is supposed to close consistent with that date. Because of changes in California's hospital capacity, about 8% of AHDR reports have periods for less than a full year and about 2% for more than a year. A description of our methods to prepare the AHDR for longitudinal research is available elsewhere [34]. For this research, we updated our program tracking moves and consolidations, and changes in hospital ownership. .

OSHPD assigns facility identifiers (OSHPDID) which are unrelated to BSMF hospital codes. To do this study we had to crosswalk between OSHPDID and HOSPCODE. Appendix A describes the methods we followed to develop the crosswalk. Appendix B identifies hospitals we combined due to consolidations and moves.

Training Data

California began BDQT on 26-Oct-2004. Trainings were regional, with all facilities in a geographic area invited to send representatives. Trainers developed reservation sheets and attempted to have people sign in at arrival. However, an unknown number of attendees did not sign attendance sheets, and some attendance sheets were lost. Thus, we do not know for certain when hospitals first sent at least one person for training, exactly how many times they sent at least one person, the number of people a hospital sent over time, and for some number, whether they ever sent anyone. The extent to which we do not have this information may have a small but unknown impact on our results. See Appendix C for a full description of our methods to prepare the training data for analysis and identify hospitals to exclude.

Appendix D identifies excluded hospitals, the basis for exclusion, their training history, and number of births. We recommend that hospitals review Appendix D with an eye to verifying the accuracy of our exclusions. In particular, hospitals excluded for late training may have attended earlier but may not have been assigned correctly because of problems with the original attendance sheets. In addition, we may have missed some consolidations. To the extent that exclusions do not reflect what the hospital believes occurred, please notify us. We will make corrections and re-estimate the models.

Table 1: Hospitals and births by exclusion criteria

Category	Description	Hospitals N	Hospitals %	Births N	Births %
Total		324	100	6,411,944	100
Excluded	No Crosswalk	10	3.1	82,898	1.3
	Lt 100 any year	61	18.8	269,969	4.2
	Late training	15	4.6	387,981	6.1
	No training	12	3.7	283,971	4.4
	Cross-County	2	0.6	35,982	0.6
	Total Excluded	100	30.9	1,062,263	15.3
Included	Trained	224	69.1	5,350,529	83.4
	After Consolidation	210			

Table 1 summarizes results of steps to identify hospitals we could use to study the impact of training on birth certificate quality. From 1998-2009, 324 hospitals had 6,411,944 births. We excluded 100 hospitals (30.9%) with 15.3% of births and included 224 (69.1%) with 83.4% of births. After correcting for consolidations and moves, we had 210 hospitals.

Military hospitals were 6 of 10 facilities that did not crosswalk to the current OSHPDID (CURROSH). Birthing centers, the remainder of that group, accounted for only 2,288 births. Of 61 hospitals with less than 100 births any year, 24 closed before 2004, 8 closed after attending a first training, 5 opened in 2004 or later, and 14 had less than 100 cases every year. It is possible that some in this group had consolidations or moves that we were unable to link. It also is possible that hospitals with late or no training attended earlier as discussed above. Again, we urge hospitals to review the appendices and notify us of any changes they recommend.

Making the Analysis File

We assigned CURROSH to the month-by-month BSMF data quality files, setting aside records from hospitals lacking CURROSH, those with cross-county consolidations, and those with less than 100 records each year. We re-summarized the remaining records by CURROSH, again month-by-month. Even though we could not model the impact of training for 27 hospitals that we could not verify were trained or were trained after 2005, we retained their records for comparative purposes. For hospitals with unverified training, we assigned the first training date in their area.

Next, we pulled AHDR data for these hospitals. Because of short periods, some hospitals lacked AHDR data for some periods. In this situation, we imputed values using records from the preceding or succeeding record, depending on when short periods were temporally.

We merged these files, excluded short-period records, and calculated several sets of variables to use for analysis: time-related, volume, hospital ownership/auspice, and data quality rates expressed as percents. The analysis file had 2,520 records, or 12 annual periods for 210 hospitals as defined after accounting for consolidated reporting and moves, and 324 records for 27 hospitals excluded from analysis. Appendix E identifies hospitals used in the analysis. We recommend that hospitals review these appendices. If they find any errors, please contact us. We will make corrections and re-estimate the models.

Study Variables

Dependent Variables

The analysis focused on five sets of data quality variables: mother and father demographics, mother’s residence data, mother’s pregnancy history, prenatal care, and infant outcomes. Summary variables were created for these sets. In developing these measures, we were guided by data editing publications provided by staff at the CDC/NCHS [26-30]. See Appendix F for a full list of the dependent variables initially prepared for analysis.

Independent Variables

Time. For this analysis, we used three time variables. CALYEAR expresses time in calendar years and is used to place results in a framework familiar to most readers. REGYEAR expresses time relatively, with CALYEAR 1998 equal to zero, CALYEAR 1999 equal to 1, and so on. REGYEARSQ is the square of REGYEAR, used to model curvilinear trends. REGYEAR and REGYEARSQ were used in the regressions, CALYEAR for tables and graphs.

Volume. We tested several ways to rank number of NICU discharges, total births, and total acute care discharges. Best results were obtained by ranking these into quartiles where 0 was below the 25%ile and 3 was above the 75%ile. We summed the rank variables and re-ranked the results. The resulting variable VOL had values of 0 (Low or medium volume), 1 (Medium high), and 2 (High volume).

Auspice. We assigned ownership/auspice based on affiliation at end of the study period: unaffiliated non-profit (UNP), Kaiser (KAISER), Sutter (SUTTER), religious (CHURCH), district (DISTRICT), public (PUBLIC), and investor (INVESTOR). The variables UNP, CHURCH, SUTTER, and KAISER are sub-groups within the larger category of non-profit hospitals. We tested variations on these themes (ever one of these, one of these when the AHDR report was filed). As in earlier studies, the variable set that performed best was based on end of study period [35]. This also is consistent with decision rules for OSHPD's hospital outcome studies. In regression models, we used INVESTOR as the reference group.

Training variables. For this study we used number of people attending the first training (PEOPLE0) and number attending any training after that (PEOPLEP). They had the value of zero (0) for every year before training began. After examining results, we re-categorized them to the range of 0 to 3 (3 or more). A variable measuring the number of times the hospital sent people for training after the first was so highly correlated with PEOPLEP ($r = 0.85$, $p= <.0001$) that we did not use it. PEOPLE0 and PEOPLEP were not significantly correlated ($r = 0.09$, $p = 0.19$).

Statistical Analysis

The analysis dataset was constructed based on hospital-period [36], where each hospital has one record summarizing each measurement interval, in this case annually over the 12-year period 1998-2009. We developed multivariate regression models to predict quality rates using various independent variables. To minimize the influence of small numbers in the denominator, we used total births as a weight.

We developed regression models focused on explained variance. We calculated Type I sums of squares, based on the order variables enter into the model. For this study, the order was variable sets for time, volume, auspice (with INVESTOR used as the reference group), and training. Note that time, volume, and auspice temporally precede training. Time variables REGYEAR and REGYEARSQ were forced into all models to test for curvilinear trends.

With Type I sums of squares, each effect is adjusted for preceding effects in the model. The sum of Type I sums of squares equals the model explained variance (EV, also known as R-Square (R-Sq)). From this, one can estimate how much each set of variables contributes to EV, and the proportion of EV each set contributes.

All work was done in SAS, version 9.3, working in the Windows 7 64-bit environment. Output of data for tables and graphs was automated. We updated macros developed for earlier studies to save results of regression models. These were output to Excel files for review and to make tables for reporting. All programs and the ID crosswalk file are available on request.

RESULTS

Birth Data Quality Training

In late 2004, the Office of Vital Records (OVR) began providing Birth Data Quality Training (BDQT) for staff from general acute care hospitals, alternative birthing centers, and local health departments. OVR organized three-hour trainings regionally, inviting all hospitals in a given area. Training was in phases, with each phase having different agendas. Table 2 summarizes attendance focusing on hospitals and their staff who attended.

Table 2: Training attendees by phase

Description	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5
Start Date	26-Oct-04	02-Jun-05	02-Mar-06	13-Mar-08	21-Apr-09
End Date	10-May-05	15-Dec-05	17-Nov-06	08-Oct-08	28-Oct-09
Trainings	13	13	8	8	7
Hospitals	230	155	143	155	141
People	365	250	217	286	252
People per hospital	1.6	1.6	1.5	1.8	1.8

Because of budget cuts in 2006, OVR reduced the number of BDQT from 13 to 8, and offered no training in calendar year 2007. Training resumed in 2008 and 2009, with fewer sessions.

Over the five phases, we were able to verify that 1,370 hospital staff attended BDQT sessions. Given the number of people attending and the number of births over this interval, hospitals sent on average about one person per 2,000 births.

Counts of people are non-unique. Due to data limitations, we have no way to know if 1,370 trainees represents that many people trained once, or a smaller number trained several times. Similarly, some hospitals sent representatives to one training and never to another, while others sent staff to every session in their region. Some hospitals sent birth clerks, while others sent unit supervisors or other administrators. Information available to us did not allow us to reliably differentiate these roles.

Additionally, these numbers likely reflect an undercount of both hospitals and attendees. Some signup sheets were lost. Some hospital representatives registered for training and either did not attend or failed to sign attendance sheets. These issues affect the count of number of hospitals trained, when training first occurred, number of people attending the first training for a given hospital and people trained thereafter. In the analysis, we have no way to estimate the impact or compensate for these problems.

Table 3: Training topics by phase

Topic	Phase 1	2	3	4	5
Data Quality Reports	X	X	X	X	X
Regional Perinatal Program Coordinator (RPPC)	X	X	X		X
Improving Data Quality	X	X		X	X
Paternity Opportunity Program (POP)	X		X	X	X
Timeliness	X		X	X	X
Birth Certificate Revision		X	X	X	X
Best Practices	X	X		X	
Birth Certificate Issues				X	X
Race Data Collection	X				
Roundtables (Birth subjects only)	X				
Birth and Death Crossmatch	X				
Fetal Death Certificates		X			
Medical Data Supplemental Worksheet		X			
Occupation/Industry		X			
Special Issues in Birth Registration		X			
State Registered Domestic Partnerships		X			
What you need to know about your child's birth certificate		X			
Back to Basics: Review Phases 1-3			X		
Birth Cohort File			X		
Birth Certificate Amendment				X	

Table 3 shows BDQT topics, organized by phase and frequency. The agenda for the 3-hour sessions became shorter over time.

Data quality reports were prepared for each phase. In Phase 1, these focused on date of last menses, race of mother and father, mother's education, date prenatal care began, and number of prenatal care visits. These are core to the MCAH Healthy People indicators.

Trainings were regional, and the Regional Perinatal Program Coordinator usually spoke. Items addressing timeliness and best practices emphasized the birth clerk's role, and overcoming barriers to obtaining complete data by the required time.

The Paternity Opportunity Program (POP) is California's effort to encourage fathers who are not married to the mother to claim paternity [50]. If the mother is married or the father signs the POP form, then birth clerks are to obtain data on the father.

Between 2005 and 2007, California added eight new birth certificate variables that the CDC required. After Period 1, all trainings covered the newly introduced variables. Table 4 is extracted from a handout distributed at Period 5 trainings. It shows California's performance relative to the CDC/NCHS standard on six core variables available before and throughout the training period. We did not evaluate variables introduced after training began.

Table 4: State-level data quality report from Period 5 trainings, % Missing (Incomplete, Unknown, or Withheld) Values for Selected California Birth Certificate items available before training began

	NCHS 2009		Year				
Variable	Median	Standard	2004	2005	2006	2007	2008
Mother Hispanic (Unknown or withheld)	0.37	1.00	1.4	1.3	1.6	1.4	1.6
Mother's Race (Unknown or withheld)	0.35	1.00	1.4	1.3	1.5	1.5	1.6
Mother's Education (Unknown or withheld)	0.62	1.00	2.9	2.8	3	3.4	3.2
Date LMP Began (Incomplete or unknown)	7.87	11.81	14.7	12.2	11.2	10.5	9.3
Month Prenatal Care Began (Unknown)	2.33	3.49	1.6	0.8	0.8	2.1	2.1
Number of Prenatal Care Visits (Unknown)	2.18	3.27	2.7	1.2	1.1	2.5	3.3

The percent of missing values (error rate) stayed the same or increased for 5 of the 6 core variables, with 3 above the standard over the period and 2 above the median over the period. The LMP date error rate improved but remained above the NCHS median. Note that Table 4 only shows data after the BDQT started, and we cannot evaluate pre-training trends from this.

Hospital Characteristics

Table 5: Auspice and volume characteristics by number of hospitals and births, 1998-2009

		Hospitals		Births	
Measure	Description	Total	% Hosp	Total	% Births
Total		210	100	5,327,032	100
Auspice	Unaffiliated NP	34	16	1,073,843	20
	Kaiser	22	10	717,594	13
	Sutter	18	9	401,080	8
	Church	55	26	1,318,713	25
	District	22	10	449,670	8
	Public	20	10	513,892	10
	Investor	39	19	852,240	16
Volume	Low	54	26	367,253	7
	Medium Low	65	31	1,390,169	26
	Medium High	40	19	1,161,084	22
	High	51	24	2,408,526	45

Table 5 shows that the 55 hospitals organized under religious auspices (Church) comprised the largest auspice group. They accounted for 26% of facilities and 25% of births.

We grouped hospitals into categories depending on the number of deliveries over the 12-year study interval. Low-volume hospitals (N = 54, 26%) delivered 7% of infants over the interval, while high-volume hospitals (N = 51, 24%) delivered 45%. No Kaiser or Public hospitals were Low Volume, so we made three groups: Low-Medium, Medium High, and High.

Training

Table 6: Average number of people trained by auspice

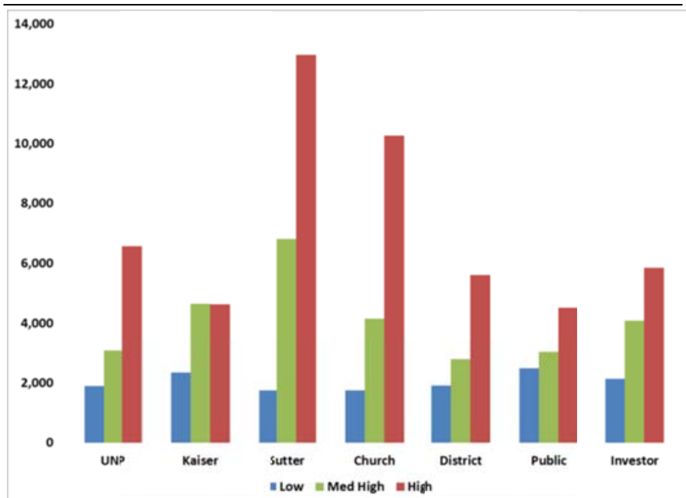
Description	Total	UNP	Kaiser	Sutter	Church	District	Public	Investor
Hospitals	210	34	22	18	55	22	20	39
People trained (First)	1.6	1.7	1.7	1.3	1.5	1.5	1.8	1.6
People trained (Later)	2.5	2.7	2.5	2.2	2.5	2.5	2.7	2.3
People trained (Total)	4.1	4.4	4.2	3.5	4.0	4.0	4.5	3.9
Trainings after first	2.2	2.5	2.1	1.9	2.1	2.1	2.4	2.0
Ratio births to trained people	3,626	3,671	4,105	4,177	4,055	2,780	3,408	3,049

Table 6 summarizes training participation by auspice. The Total column shows that hospitals sent an average of 1.6 people to the first training. Hospitals that sent anyone to later trainings sent an average of 2.5 people, for an average of 4.1 people per hospital over the entire study interval. If hospitals sent representatives after the first training, they sent people to an average of 2.2 more trainings. In terms of auspice, Sutter sent the fewest total people to trainings.

We calculated the ratio of births to trained people by dividing the total number of births (2004-2009) by the total number of people trained over the study interval. Sutter and Kaiser had the highest ratio of births to trained people, while District hospitals had the lowest ratio. We interpret a low ratio as a stronger institutional commitment to training than a higher ratio. On the hospital level, the ratio had a wide range from a low of 302 to a high of 40,807 births per trainee. Kaiser and Sutter may have felt less of a need to send people because of their computerized medical record systems.

Figure 1 shows the relationship between the ratio of births per trained by volume and auspice. At hospitals with the lowest volume, all auspices sent about the same number of trainees per births, plus or minus 2,000 births. The ratio increased at medium-high volume hospitals, with Sutter having the greatest increase among hospitals that size. For Kaiser, the ratio between medium high and high-volume was identical, about 1 person trained per 4,000 births.

Figure 1. Number of births (2004-2009) per trained by volume and auspice



The model to predict the ratio of births per trainee by auspice, volume, and the interaction of auspice by volume explained 39.8% of model variance (F = 6.25, P = <.0001). The interaction between auspice and volume accounted for 73% of EV, volume accounted for 22%, while training as a main effect accounted for 4% of EV, which was statistically non-significant. The greatest discrepancy for the ratio of births per trainee occurred among high-volume Sutter and Church-affiliated hospitals.

Recall that a high number on this ratio means that fewer people were trained given the number of births. Low volume hospitals sent about 1 trainee per 2,000 births. High-volume Church-affiliated hospitals sent about 1 per 10,000 births, and high-volume Sutter hospitals sent about 1 per 13,000 births.

Data Quality

Table 7 summarizes data quality statistics for studied variables using the 210 hospitals included in multivariate models. Data are displayed in groups: Father, mother, mother geography, pregnancy history, delivery, and infant. The 2009 NCHS median and standard is shown for variables that NCHS reports [21]. The NCHS median error is the middle value for all states on a given variable: 50% are below the value and 50% are above. The standard is 1.5 times the median.

Rates are calculated as percents, and are shown overall and by auspice. The grand mean is the average rate over the 12-year study period. Rates for father variables use the number of fathers who claimed paternity as the denominator. Rates based on last live birth use as the denominator records with previous live births, excluding records with birth order greater than one (that is, including only the first outcome of a pregnancy producing more than one infant). All other rates are based on total births.

The last rows of each group typically contain two summary variables: the percent of births with at least one problem with any variable in the group, and the average number of problems in the group. The latter give a sense of the additive nature of data quality problems. For example, in the group of variables examining mother's demographic characteristics, 3.5% of births had at least one error, and the average error rate per birth was 8.1%, with a range from 3.7% (Investor) to 14.9% (District). Note that these error rates are averaged over the 12-year period and do not account for underlying trends or training impact.

To evaluate the impact of training, we selected for multivariate modeling those variables with a grand mean of 1% or higher that were available over the entire study period (grayed). Because of uncertainty due to small numerators, we did not model indicators with rates below 1%. We adjusted for small denominators by using total births as a weight.

In the following sections, we present results of models for eight core variables used to calculate MCAH Healthy People indicators. Regression model details are available upon request.

Table 7: Percent of missing values by auspice, rates expressed as percent (%)

Group	Description	NCHS 2009 Median	NCHS 2009 Stand	Grand Mean	Auspice UNP	Kaiser	Sutter	Church	District	Public	Investor
Father	Paternity not established *	8.70	13.08	8.92	7.92	5.07	7.16	9.20	8.68	15.56	9.95
Demographics	Year of birth	1.20	1.79	0.12	0.09	0.06	0.17	0.09	0.05	0.38	0.13
Father	Month of birth	1.20	1.79	0.12	0.09	0.06	0.17	0.09	0.05	0.39	0.13
	Day of birth	1.20	1.79	0.13	0.09	0.06	0.17	0.09	0.06	0.40	0.13
	Any birth date problem			0.13	0.09	0.06	0.17	0.09	0.06	0.40	0.13
	Age			0.12	0.09	0.06	0.17	0.09	0.05	0.38	0.13
	Hispanic ethnicity	2.06	3.09	1.22	2.26	0.79	0.69	0.98	2.45	0.84	0.45
	Race	2.00	2.99	1.97	2.72	1.44	2.04	1.61	3.16	2.41	1.13
	Race/Ethnicity **	2.00	2.99	2.08	2.80	1.56	2.11	1.73	3.24	2.52	1.24
	Education	2.22	3.33	3.38	4.45	1.86	2.10	3.31	4.82	5.49	2.01
	Data problems (Any)			4.43	5.34	2.67	3.73	4.20	5.76	7.39	2.98
	Data problems (Sum)			9.70	14.53	6.09	7.07	8.39	15.64	12.46	5.15
Demographics	Year of birth	0.01	1.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01
Mother	Month of birth	0.01	1.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01
	Day of birth	0.01	1.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01
	Any birth date problem			0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.01
	Age			0.00	0.00	0.01	0.00	0.01	0.00	0.01	0.01
	Hispanic ethnicity	0.37	1.00	1.17	2.17	0.75	0.61	0.97	2.50	0.63	0.47
	Race	0.35	1.00	1.90	2.67	1.44	1.94	1.58	3.09	2.08	1.09
	Race/Ethnicity **	0.35	1.00	1.98	2.72	1.46	1.98	1.68	3.17	2.16	1.19
	Education	0.62	1.00	2.61	3.96	1.67	1.59	2.52	4.25	3.08	1.15
	Data problems (Any)			3.51	4.74	2.36	3.05	3.32	5.03	4.59	2.00
	Data problems (Sum)			8.10	13.28	5.45	5.46	7.06	14.86	7.17	3.66
Mother	ZIP/county mismatch			0.34	0.69	0.39	0.27	0.16	0.59	0.13	0.14
Geography	City/county mismatch			0.46	0.79	0.49	0.30	0.32	0.65	0.20	0.38
	Street address unknown			1.33	1.33	1.27	1.36	1.31	1.61	1.29	1.27
	Geography (Any) ***	0.70	1.05	1.81	2.12	1.81	1.68	1.64	2.26	1.50	1.66
	Geography (Sum)			2.13	2.81	2.15	1.93	1.79	2.85	1.62	1.79
Pregnancy	Birth history (Any)			0.17	0.15	0.16	0.12	0.22	0.12	0.20	0.15
History	Birth history (Sum)			0.24	0.20	0.25	0.15	0.34	0.15	0.31	0.21
	Last live birth month	3.35	5.02	1.35	0.94	0.64	1.20	2.17	0.23	3.26	0.69
	Last live birth year	1.31	1.96	0.12	0.08	0.13	0.05	0.20	0.03	0.16	0.10
	Last live birth (Any) **	1.31	1.96	1.50	1.06	0.81	1.29	2.39	0.28	3.48	0.81
	Last menstrual period year	4.65	6.97	4.36	5.03	5.92	1.17	4.38	3.68	5.55	3.30
	Last menstrual period month	4.96	7.44	4.55	5.49	5.98	1.29	4.54	3.72	5.84	3.38
	Last menstrual period day	7.87	11.81	12.09	15.16	10.74	5.70	14.95	11.02	11.53	8.85
	Last menstrual period (Any) **	7.87	11.81	12.11	15.18	10.80	5.71	14.97	11.04	11.54	8.86
	Last menstrual period (Sum)			21.00	25.68	22.63	8.16	23.87	18.42	22.92	15.53
Prenatal Care	Month prenatal care began	2.33	3.49	1.58	1.01	1.52	0.39	1.85	1.73	3.28	1.39
	Number of prenatal care visits	2.18	3.27	2.43	1.43	1.82	0.51	3.23	1.66	5.28	2.53
Delivery	Fetal presentation at birth	0.05	1.00	10.61	9.04	13.73	3.65	14.27	8.32	11.08	8.34
2005 forward	Forceps attempted	0.03	1.00	8.41	8.15	15.53	1.19	6.12	7.56	11.25	8.84
	Vacuum attempted	0.03	1.00	8.09	8.13	15.23	1.17	5.97	7.46	10.92	7.56
Infant	Infant birthweight	0.03	1.00	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01
	Gestational age *	2.27	3.40	5.04	5.88	6.72	1.50	5.04	4.25	6.45	3.79
Infant	Apgar 1 minute			2.00	0.48	2.01	0.27	4.64	2.54	1.31	0.50
2007 forward	Apgar 5 minutes	0.30	1.00	2.00	0.51	1.81	0.30	4.66	2.60	1.33	0.51
	Apgar 10 minutes	0.22	1.00	2.22	0.75	1.99	0.42	4.84	2.74	1.97	0.66
	Hearing screening			0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Infant test data problem (Any)			2.29	0.78	2.27	0.44	4.87	2.77	2.02	0.70
	Infant test data problem (Sum)			6.22	1.74	5.81	0.98	14.13	7.88	4.61	1.66

* Using California 2009 median and standard

** NCHS 2009 median and standard from preceding line

*** Using NCHS 2009 median and standard for mother's place of residence

Mother's Race/Ethnicity

Importance

The quality of race and ethnicity variables significantly affect the accuracy of comparisons of race-specific health indicator rates and thus of estimating race and ethnic disparities.

The CDC/NCHS considers race/ethnicity so vital that they publish decision rules and data products to impute race/ethnicity [30,37-41]. All CDC/NCHS vital statistics reports are based on race/ethnicity calculated using these methods and products.

Results

Table 8 shows results for mother's race/ethnicity model (F = 43.31, P = < 0.0001, explained variance (EV (also called R-square, or R-SQ), 7.93%, effect size (ES) = 19.28). Part A1 summarizes model statistics, specifically the contribution to EV and ES, summarized by variable sets. Part A2 presents variable statistics for significant predictors (P < 0.005) in the model, regression estimates (Est) and standard error (Std Err), and the standardized beta coefficient (Std Est) with its ES, calculated by squaring the Std Est. Part B visually displays trends. Missing data rose slowly through 2004 and then began to increase steeply. California was well above the CDC/NCHS Standard every year. Data quality was worse for father's race/ethnicity and trends almost exactly paralleled mother's.

Table 8: Results for mother's race/ethnicity

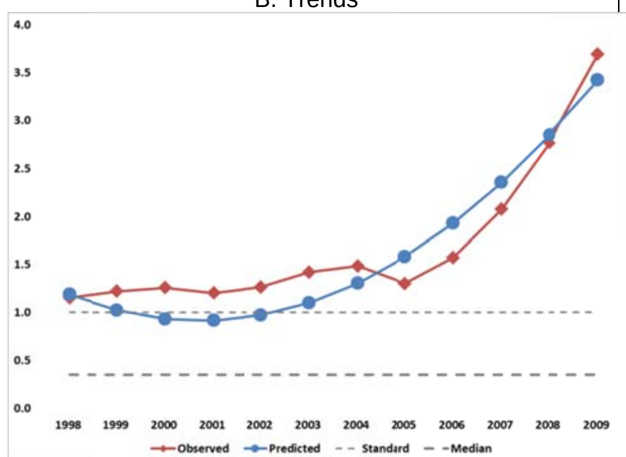
A1. Model Statistics

Contribution	EV	ES
Model	7.93	19.28
Time	3.86	15.29
Volume (after Time)	2.52	2.15
Auspice (after Time Volume)	1.55	1.84
Training (after Time Volume Auspice)	0.00	0.00

A2. Variable Statistics

Significant variables	Est	Std Err	Std Est	ES
Intercept	0.32	0.24	0.00	0.00
Year rel 1998	(0.20)	0.09	(0.17)	2.83
Year rel 1998 (Sq)	0.04	0.01	0.35	12.46
Hospital Volume	0.74	0.10	0.15	2.15
UNP	0.89	0.20	0.09	0.78
District	1.50	0.28	0.10	1.06

B. Trends



With hierarchical stepwise regression, which we did for this study, we are testing how much explained variance (EV) each step adds. Looking at A1, notice that time explained 3.9% of variance, and volume explained 2.5% after controlling for time. At 1.55%, after controlling for time and volume, auspice EV is statistically significant but small. Training had no effect for this variable. Notice also that the sum of step EV equals model EV. Variable statistics in A2 tell us that data quality tended to be poorer in larger hospitals, and those with UNP or District auspices.

From a policy perspective, regression results can be statistically significant but practically unimportant, particularly in population studies. In multivariate regression, EV generally decreases as the number of cases increases, although the results as here typically are highly significant. The ES is more helpful from a policy perspective because it is independent of sample size. Generally, an effect size of 2% is considered a small effect, 15% medium, and 35% high [42]. The overall model to explain data quality for mother's race/ethnicity had an EV of 7.9% and an ES of 19.3. The most important contribution comes from time (ES = 15.29, medium effect), with volume making a small but practically important contribution, auspices making a statistically significant but practically unimportant contribution, and training making no contribution.

Answers to Research Questions

Are differences in hospital characteristics (volume, auspice) related to differences in quality and completeness? For mother's race/ethnicity, differences in hospital volume were related to differences in quality and completeness. Larger hospital size was associated with higher error rates. Differences in auspice also were related to differences in quality and completeness in some instances. Specifically, UNP and District hospitals tended to have higher error rates. However, auspice was unimportant relative to time and volume.

Does a training program for birth clerks improve the quality of birth certificates? Specifically, when comparing data quality before and after training, is there a significant decrease in missing or unlikely values in fields targeted by the training? Training did not improve the error rate for the mother's race/ethnicity variable. This is demonstrated by the fact that neither number of people attending the first training nor number of people attending later trainings entered into the regression model.

Is the number of trainees associated with greater improvement compared to hospitals with few or no attendees? That is, is there a dose effect? There was no dose effect. The number of trainees attending first training or later trainings did not improve the error rate for the mother's race/ethnicity variable.

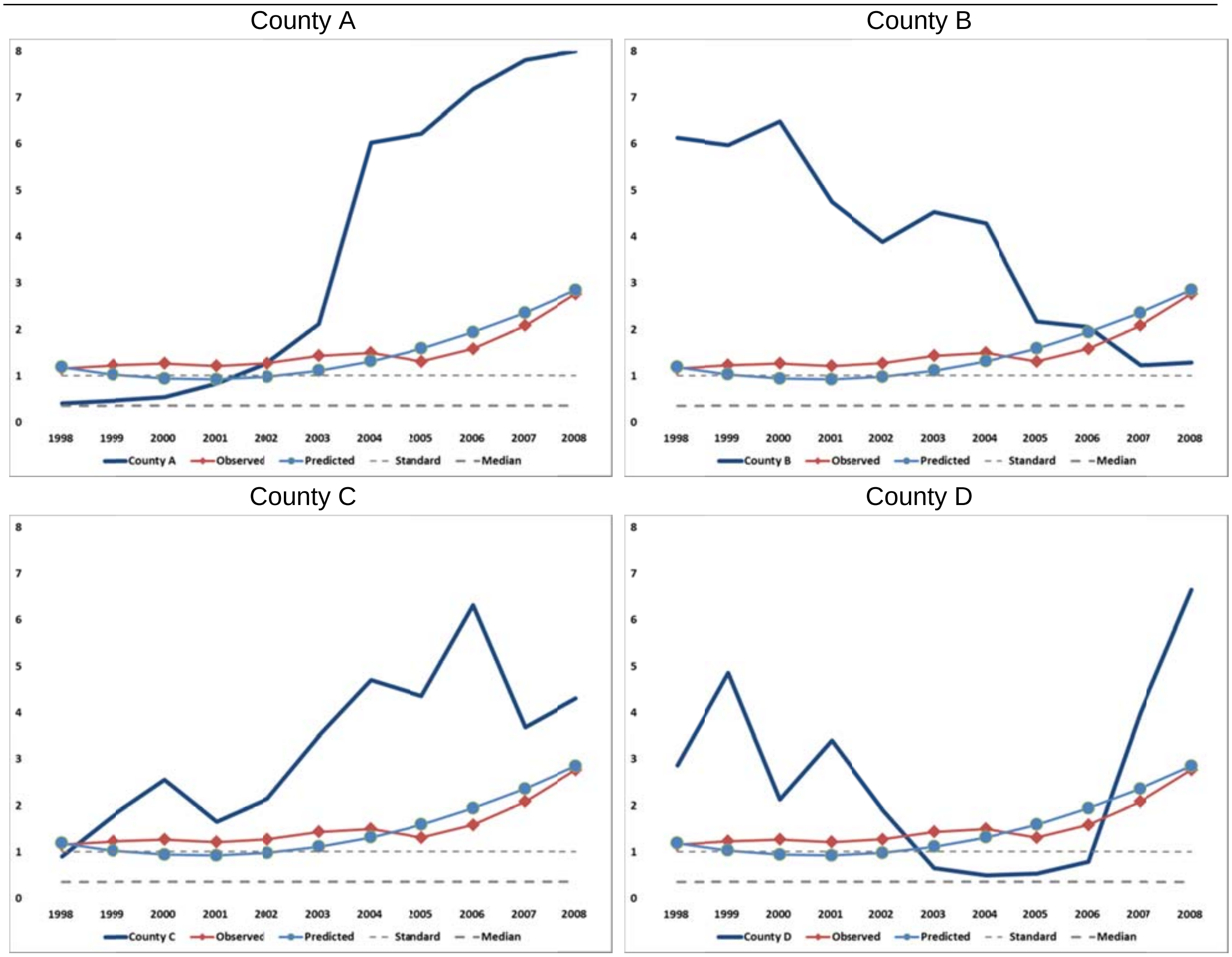
Are improvements sustained over time? Improvements were not sustained over time. In fact, the error rate increased over time.

Implications of County Variation in Data Quality

States that want to compare their rates with those published by CDC/NCHS are advised to implement CDC rules to impute missing or unknown values before calculating rates. This is particularly important when, as here, data quality is poor, deteriorating, and missing not at random (MNAR).

Although the state average in failing to report race/ethnicity has been rising overall, there has been widespread county-level variation in data quality trends. Figure 2 shows the nature of the problem focusing on four large counties. Data are shown by mother's county of residence, which is used to calculate HP 2020 indicators. The focus is on variation in county trends compared with the state trend.

Figure 2. County-level variation in maternal race/ethnicity missing %1998-2009



In an environment where data quality is in such flux within and across counties in addition to over time, coming to reliable conclusions about trends in race/ethnic disparities will be problematic. We strongly recommend that California add to the vital statistics files variables with bridged race and ethnicity to compensate for this very real problem. We also recommend that California require all Healthy People 2020 indicators using birth certificate data to be based on bridged race/ethnicity.

We will post on our website the macro we wrote incorporating the CDC/NCHS algorithm to impute race/ethnicity. Note that the CDC algorithm first uses father's race/ethnicity to impute mother's missing race/ethnicity. Because data quality is poor on both variables, imputation is a real presenting need.

Mother's Education

Importance

Education is an important underpinning to maternal and child health. Better-educated mothers (and fathers) tend to have better health, higher incomes, and longer lives. Educated women are more likely than uneducated women to ensure that their children eat well, finish school, and receive adequate health care [43].

Results

Table 9 shows results for mother's education model (F = 28.46, P = < 0.0005, EV = 7.35, ES = 8.94). Part A shows model statistics and variable statistics, and Part B shows trends. California was well above the CDC/NCHS Standard every year. Data quality was worse for father's education and those trends almost exactly paralleled mother's.

Table 9: Results for mother's education

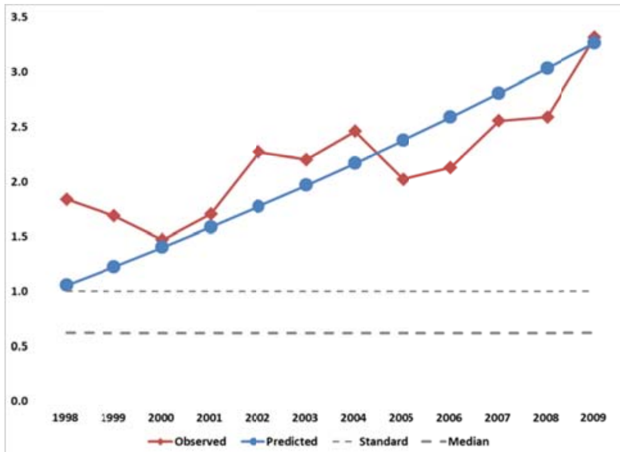
A1. Model Statistics

Contribution	EV	ES
Model	7.35	8.94
Time	1.70	1.17
Volume (after Time)	2.52	1.85
Auspice (after Time Volume)	3.13	5.91
Training (after Time Volume Auspice)	0.00	0.00

A2. Variable Statistics

Significant variables	Est	Std Err	Std Est	ES
Intercept	(0.61)	0.34	0.00	0.00
Year rel 1998	0.17	0.11	0.11	1.13
Year rel 1998 (Sq)	0.00	0.01	0.02	0.05
Hospital Volume	0.91	0.13	0.14	1.85
UNP	2.21	0.29	0.17	2.74
Church	1.03	0.27	0.08	0.68
District	2.69	0.39	0.14	1.94
Public	1.34	0.37	0.07	0.54

B. Trends



Time, volume, and auspice were the significant predictors, and no training variables entered the model. The overall ES of 8.94 is a medium effect, with the main contribution from auspice (ES = 5.91). Volume, while statistically significant, was practically unimportant.

Answers to Research Questions

Are differences in hospital characteristics (volume, auspice) related to differences in quality and completeness? For mother's education, differences in hospital volume and auspice were related to differences in quality and completeness. As size increased, data errors tended to increase. Differences in auspice also were related to differences in quality and completeness. Specifically, UNP, Church, District and Public hospitals tended to have higher error rates. Auspice was the most important contributor (ES = 5.91). While statistically significant, volume was practically unimportant (ES = 1.81).

Does a training program for birth clerks improve the quality of birth certificates? Specifically, when comparing data quality before and after training, is there a significant decrease in missing or unlikely values in fields targeted by the training? Training did not improve the error rate for mother's education.

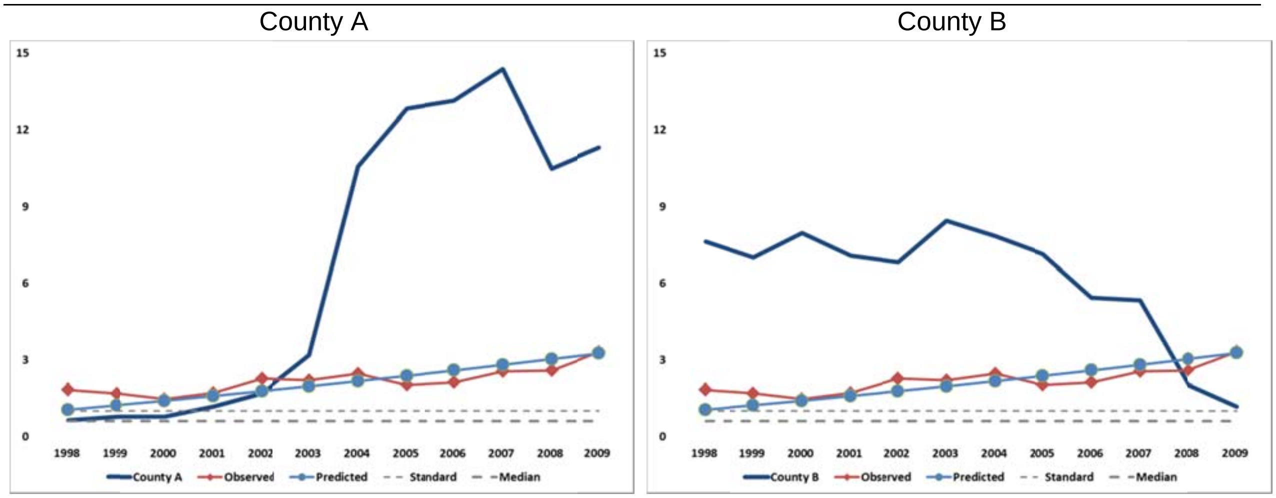
Is the number of trainees associated with greater improvement compared to hospitals with few or no attendees? That is, is there a dose effect? There was no dose effect. The number of trainees attending first or later trainings had no impact on the error rate for mother's education.

Are improvements sustained over time? Improvements were not sustained over time. In fact, similar to race/ethnicity, the error rate for mother's education rose over time.

Implications of County Variation in Data Quality

Although the state average in failing to report education has been rising overall, there has been widespread county-level variation in data quality trends. Figure 3 shows the nature of the problem focusing on two large counties. Data are shown by mother's county of residence, which is used to calculate HP 2020 indicators. The focus is on variation in county trends compared with the state trend.

Figure 3. County-level variation in maternal race/ethnicity missing %1998-2009



In an environment where data quality is in such flux within and across counties in addition to over time, coming to reliable conclusions about trends in education disparities will be problematic.

The CDC/NCHS does not have an algorithm to impute mother's education when it is missing. Jurisdictions attempting to stratify health indicators by education are advised to evaluate the impact of missing data on their conclusions. The variation in data quality over time also will impact conclusions that might be drawn from multivariate models using this variable.

Mother's Place of Residence

Importance

Health indicators are calculated based on place of residence. As uncertainty increases about where a mother lives, indicator rates will be increasingly inaccurate for both the jurisdiction where the mother truly lives and the jurisdiction where the mother does not live but was (incorrectly) recorded as living. Place of residence also is important to identify geographic disparities.

Results

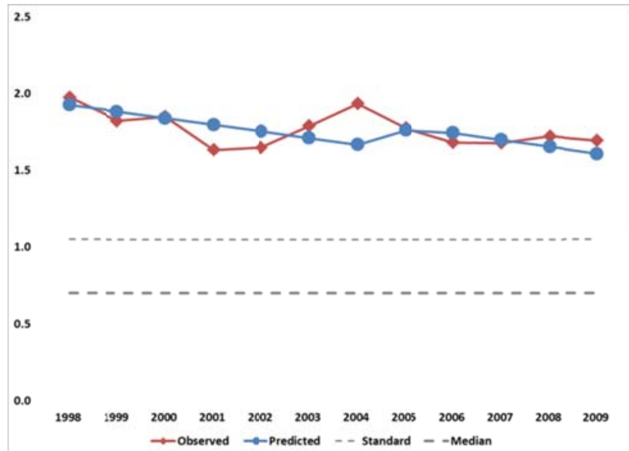
Table 10 shows results for mother's place of residence model (F = 13.99, P = < 0.0001, EV = 4.27, ES = 10.13). Part A presents model statistics and variable statistics, and Part B shows trends. California was steadily well above the CDC/NCHS Standard throughout the period. Unlike other indicators, the statewide error rate was relatively steady linearly, decreasing slightly over time.

Table 10: Results for mother's place of residence

A1. Model Statistics		
Contribution	EV	ES
Model	4.27	10.13
Time	0.27	0.92
Volume (after Time)	0.41	0.29
Auspice (after Time Volume)	2.27	2.40
Training (after Time Volume Auspice)	1.31	6.52

A2. Variable Statistics				
Significant variables	Est	Std Err	Std Est	ES
Intercept	1.75	0.09	0.00	0.00
Year rel 1998	(0.04)	0.03	(0.10)	0.92
Year rel 1998 (Sq)	(0.00)	0.00	(0.00)	0.00
Hospital Volume	0.10	0.04	0.05	0.29
UNP	0.40	0.08	0.11	1.11
District	0.57	0.11	0.11	1.10
Public	(0.22)	0.10	(0.04)	0.19
People trained (First)	0.34	0.06	0.21	4.51
People trained (Later)	(0.15)	0.04	(0.14)	2.01

B. Trends



Time, volume, auspice, and training were the significant predictors. Training appears to have had a significant and practical impact, but training never addressed variables related to mother's place of residence. Thus, it is difficult to conclude that this statistical result is due to training.

Answers to Research Questions

Are differences in hospital characteristics (volume, auspice) related to differences in quality and completeness? For mother's place of residence, differences in hospital volume and auspice were related significantly to differences in quality and completeness. Specifically, as volume increased, errors increased. UNP and District hospitals tended to have higher error rates, while public hospitals tended to have lower rates. The overall ES of 10.13 is moderate, with auspice making a small and practically important contribution (ES = 2.40), and volume making a statistically significant but practically unimportant contribution (ES = 0.29).

Does a training program for birth clerks improve the quality of birth certificates? Specifically, when comparing data quality before and after training, is there a significant decrease in missing or unlikely values in fields targeted by the training? As far as we have been able to identify, mother's place of residence never was a topic of training. Thus, we cannot establish that training improve the quality of this variable.

This result may reflect one of the sources of internal invalidity for the interrupted time series design. During the period when training was ongoing, CCHS also changed AVSS, implementing pull-down menus to guide clerks when entering geographic variables. This may explain how a training effect is found when we have no knowledge that it was addressed in training. That is, changes in the data collection system parallel with the training may underlie the seeming training effect, rather than training itself.

As Campbell and Stanley pointed out (1963, p. 41), "[This] design may frequently be employed to measure effects of a major change in administrative policy. Bearing this in mind, one would be wise to avoid shifting measuring instruments at the same time he shifts policy. In most instances, to preserve the interpretability of a time series, it would be better to continue to use a somewhat antiquated device rather than to shift to a new instrument."

Is the number of trainees associated with greater improvement compared to hospitals with few or no attendees? That is, is there a dose effect? Although training apparently never addressed mother's place of residence, the number of people attending the first training increased the error rate, while the number of people attending later decreased the error rate. This may be consistent with the time when the agency began to change AVSS.

Are improvements sustained over time? The effect of different training impacts (first to increase, then to decrease) was to maintain the trend over time.

Implications of County Variation in Data Quality

Although the state-level average error rate for mother's place of residence was relatively steady, variation in county-level quality was wide. Averaged over the 12-year period, only three counties had an average error rate below the NCHS Standard. Average error rates for half of the counties were higher than the state average, ranging as high as 12.4%. As with previous indicators, rates for some counties rose while rates for others decreased.

Unlike race/ethnicity and education, most jurisdictions with high error rates for mother's place of residence were small. These errors will not markedly influence state rates, but will significantly influence certainty of health indicator estimates for smaller areas. Small counties may seem to have high or low rates not because they truly do but because of significant uncertainty as to where people truly live.

The error rate trend for mother's place of residence had a small but statistically significant decline over the period. OVR has implemented pull-down screens for geographic variables. This may begin to speed improvement in geographic assignment.

Date of Last Live Birth

Importance

The date of last live birth is used to calculate the inter-pregnancy interval for women who previously had at least one live birth. The CDC/NCHS has published decision rules for calculating interpregnancy interval [30]. They do not impute date of last live birth when missing or invalid.

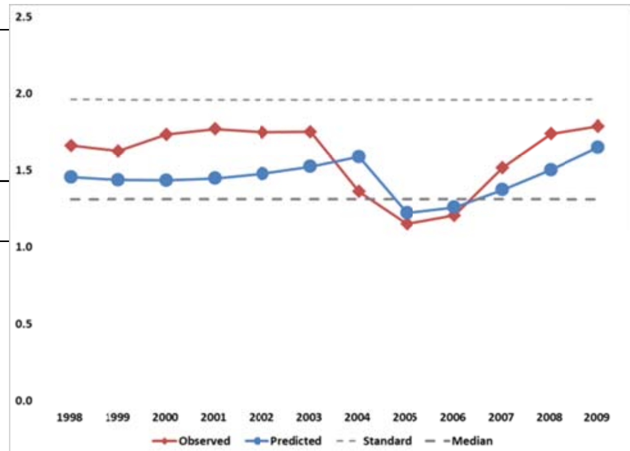
Results

Table 11 shows results for the model to explain the error rate for date of last live birth (F = 25.38, P = < 0.0001, EV = 5.71, ES = 7.46). Part A shows model and variable statistics, and Part B shows trends. Error rates were below the standard over the period, dipping briefly below the median in 2005 and 2006.

Table 11: Results for date of last live birth

A1. Model Statistics				
Contribution			EV	ES
Model			5.71	7.46
Time			0.01	0.74
Volume (after Time)			0.00	0.00
Auspice (after Time Volume)			5.52	6.21
Training (after Time Volume Auspice)			0.18	0.51
A2. Variable Statistics				
Significant variables	Est	Std Err	Std Est	ES
Intercept	0.99	0.22	0.00	0.00
Year rel 1998	(0.03)	0.09	(0.02)	0.06
Year rel 1998 (Sq)	0.01	0.01	0.08	0.68
Church	1.42	0.18	0.15	2.39
District	(0.66)	0.28	(0.05)	0.22
Public	2.54	0.27	0.19	3.60
People trained (Later)	(0.20)	0.09	(0.07)	0.51

B. Trends



Time, volume, and training were significant predictors, but the ES indicates that all these factors were practically unimportant. Auspice was the most important predictor for the date of last live birth error rate.

Answers to Research Questions

Are differences in hospital characteristics (volume, auspice) related to differences in quality and completeness? For date of last live birth, differences in hospital volume were not related to differences in quality and completeness, while auspice was related. Specifically, Church and Public hospitals tended to have higher rates while District hospitals tended to have lower rates. Auspice is the main contributor (ES = 6.21) to overall ES (7.46).

Does a training program for birth clerks improve the quality of birth certificates? Specifically, when comparing data quality before and after training, is there a significant decrease in missing or unlikely values in fields targeted by the training? The number of people attending later trainings temporarily reduced the error rate for date of last live birth. However the training ES of 0.51 indicates the impact, while statistically significant, was practically unimportant.

Is the number of trainees associated with greater improvement compared to hospitals with few or no attendees? That is, is there a dose effect? The number of trainees attending first training did not affect the error rate for date of last live birth. However, the number of trainees attending later trainings did temporarily reduce the error rate for this variable.

Are improvements sustained over time? Improvements were not sustained over time. After a brief decline, rates returned to previous ranges. With an ES of 0.51, the training effect was practically unimportant.

Implications of County Variation in Data Quality

Averaged over the 12-year period, twelve counties (21%) had an error rate above the NCHS Standard. Average error rates for those counties ranged as high as 9.2% over the interval and as high as 26% in some years. As with other indicators, rates for some counties rose while others fell.

In an environment where data quality is in such flux within and across counties in addition to over time, coming to reliable conclusions about trends for inter-pregnancy interval will be problematic.

Similar to race/ethnicity and education, a number of large counties had high error rates for date of last live birth. These are large enough to influence the accuracy of state rates for interpregnancy interval. Areas with high or unstable error rates face considerable uncertainty about interpregnancy interval estimates. They may seem to have high or low rates not because they truly do but because of uncertainty as to the true rate. The variation in data quality over time also will limit conclusions that might be drawn from multivariate models incorporating date of last live birth or interpregnancy interval.

Date of Last Menstrual Period

Importance

The algorithm to calculate infant's gestational age uses date of last menstrual period (LMP) and from gestational age, the preterm birth rate. The CDC/NCHS does not impute LMP when it is missing or invalid. However, in this circumstance, another algorithm imputes gestational age based on mother's race/ethnicity, infant's birth weight and parity.

Results

Table 12 shows results for the model to explain the error rate for LMP (F = 41.21, P = < 0.0001, EV = 10.30, ES = 8.72). Part A shows model and variable statistics, and Part B shows trends. Error rates were above the CDC/NCHS standard at start of period, below the standard before training began, and continued to decrease steadily. By 2009, the error rate was below the 2009 CDC/NCHS median.

Table 12: Results for date of last menstrual period

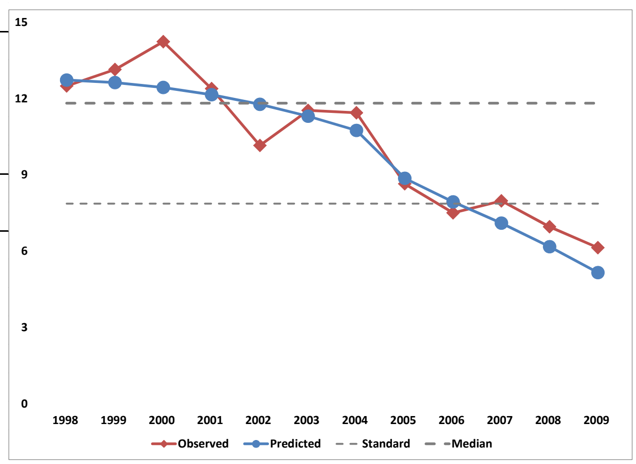
A1. Model Statistics

Contribution	EV	ES
Model	10.30	8.72
Time	3.64	1.74
Volume (after Time)	2.18	1.98
Auspice (after Time Volume)	4.34	4.66
Training (after Time Volume Auspice)	0.14	0.35

A2. Variable Statistics

Significant variables	Est	Std Err	Std Est	ES
Intercept	10.57	0.83	0.00	0.00
Year rel 1998	(0.07)	0.29	(0.02)	0.03
Year rel 1998 (Sq)	(0.05)	0.03	(0.13)	1.71
Hospital Volume	2.39	0.32	0.14	1.98
UNP	4.26	0.69	0.13	1.57
Sutter	(4.72)	1.01	(0.09)	0.83
Church	4.74	0.64	0.15	2.25
People trained (First)	(0.86)	0.43	(0.06)	0.35

B. Trends



Time, volume, and training were significant predictors, but the ES indicates that these factors were practically unimportant. Auspice was the most important predictor for the LMP error rate.

Answers to Research Questions

Are differences in hospital characteristics (volume, auspice) related to differences in quality and completeness? For LMP, differences in hospital volume and auspice were related to differences in quality and completeness. As volume increased, error rates tended to increase, but ES statistics indicate that volume was practically unimportant (ES = 1.98). As to auspice, UNP and Church hospitals tended to have higher rates while Sutter hospitals tended to have lower rates. Auspice made the largest and most practically important contribution (ES = 4.66) to understanding the LMP error rate.

Does a training program for birth clerks improve the quality of birth certificates? Specifically, when comparing data quality before and after training, is there a significant decrease in missing or unlikely values in fields targeted by the training? Although the general trend was for the LMP error rate to decrease, the number of people attending the first training further reduced the error rate. However, while statistically significant, the ES indicates training had no practical impact (ES = 0.35). This is because the strong trend was well under way before training began.

Is the number of trainees associated with greater improvement compared to hospitals with few or no attendees? That is, is there a dose effect? The number of trainees attending first training did slightly reduce the LMP error rate, albeit with no practical impact because of the strong trend already under way. The number of trainees attending later trainings did not further reduce the error rate for this variable.

Are improvements sustained over time? Improvements sustained over time. However, although the effect of the first training was statistically significant, it was practically unimportant to the overall trend. The strong trend was well under way before training began.

Implications of County Variation in Data Quality

Unlike the other variables, the LMP error rate tended to decrease over most jurisdictions. Nonetheless, averaged over the 12-year period, 20 counties (including Los Angeles) had an average error rate above the 2009 NCHS Standard ranging as high as 34% over the period and as high as 89% in one year for one county.

Similar to race/ethnicity and education, this problem is large enough to call into serious question the accuracy of state rates for indicators such as preterm birth that depend on this variable. Areas with high or unstable LMP error rates face considerable uncertainty about true rates for other indicators such as preterm birth that depend on this variable. Jurisdictions may seem to have high or low preterm birth rates not because they truly do but because of uncertainty as to the true rate related to the underlying LMP variable that is the basis for preterm birth and other indicators.

In an environment where data quality is in such flux within and across counties in addition to over time, coming to reliable conclusions about trends for indicators such as preterm birth that are based on LMP will be problematic.

The variation in data quality over time will severely limit conclusions one might otherwise draw from multivariate models incorporating variables relying on LMP. The CDC/NCHS does not have an algorithm to impute LMP date.

Month of First Prenatal Care Visit

Importance

Month of first prenatal care visit is a key indicator of health care access for pregnant women. It also is a core variable to calculate Kotelchuck's Adequacy of Prenatal Care Visits [56].

In 2007, California added two related variables, date of first and last prenatal care visits.

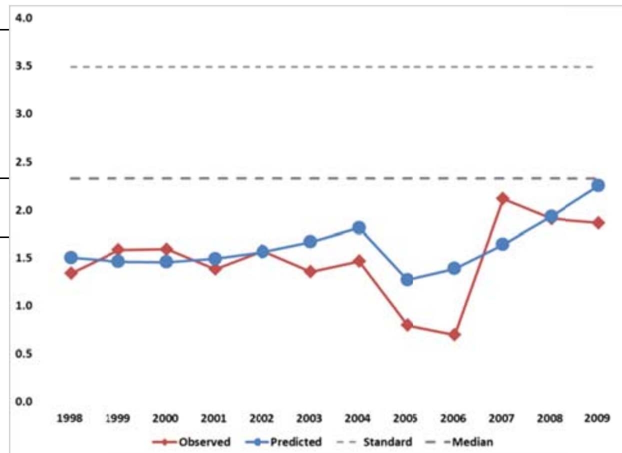
Results

Table 13 shows results for the model to explain the error rate for the month of first prenatal care visit (F = 13.14, P = < 0.0001, EV = 3.04, ES = 6.43). Part A shows model and variable statistics, and Part B shows trends. This is the only variable with error rates below the CDC/NCHS median over the period. It was missing steadily at about 1.5% through 2004, dropped briefly, and bounced to a higher rate by 2009.

Table 13: Results for month of first prenatal care visit

A1. Model Statistics				
Contribution			EV	ES
Model			3.04	6.43
Time			0.16	2.97
Volume (after Time)			0.00	0.00
Auspice (after Time Volume)			2.47	2.31
Training (after Time Volume Auspice)			0.41	1.15
A2. Variable Statistics				
Significant variables	Est	Std Err	Std Est	ES
Intercept	1.53	0.24	0.00	0.00
Year rel 1998	(0.06)	0.09	(0.05)	0.21
Year rel 1998 (Sq)	0.02	0.01	0.17	2.76
UNP	(0.60)	0.22	(0.06)	0.31
Sutter	(1.32)	0.32	(0.08)	0.67
Public	1.67	0.29	0.12	1.33
People trained (Later)	(0.33)	0.10	(0.11)	1.15

B. Trends



Time, auspice, and training were significant predictors, but training was practically unimportant. Auspice again was the most important predictor of the error rate for month of first prenatal care visit.

Answers to Research Questions

Are differences in hospital characteristics (volume, auspice) related to differences in quality and completeness? Differences in volume were not related to differences in quality and completeness. As to auspice, Public hospitals tended to have higher rates while UNP and Sutter hospitals tended to have lower rates, overall, a small but practically important effect (ES = 2.31).

Does a training program for birth clerks improve the quality of birth certificates? Specifically, when comparing data quality before and after training, is there a significant decrease in missing or unlikely values in fields targeted by the training? Although the general trend for month of first prenatal care visit was for the error rate to be flat and low, the number of people attending later trainings temporarily reduced the rate. However the training effect, while statistically significant, was not practically important (ES = 1.15).

Is the number of trainees associated with greater improvement compared to hospitals with few or no attendees? That is, is there a dose effect? The number of trainees attending training did affect the error rate for month of first prenatal care visit. Specifically, the rate decreased as a function of later training attendance. However, while the effect of the later trainings was statistically significant, it had only a minor impact on the overall results.

Are improvements sustained over time? Improvements were not sustained over time. We are of the opinion that the circumstance here is similar to the circumstance for mother's place of residence. Two new data variables were introduced that were relevant to month of first prenatal care visit. Here, however, the event affecting data quality appears to be instrumentation change (another type of threat to internal invalidity) [22]. Attention likely turned to the newly introduced variables and away from the core variable. When month of first prenatal care visit is missing, the CDC/NCHS has an algorithm to calculate this based on dates of last menstrual period and first prenatal visit.

Implications of County Variation in Data Quality

Similar to mother's place of residence, the error rate for month of first prenatal care visit tended to be relatively flat and later training temporarily reduced the error rate. Nonetheless, averaged over the 12-year period, eight counties had an error rate above the 2009 CDC/NCHS Median, including five above the CDC/NCHS Standard, ranging as high as 8.5% over the period and as high as 15.2% in some years.

Counties with the highest error rates were mixed in terms of size, and will not have a serious impact on reliability of the state rate. However, areas with high or unstable error rates face considerable uncertainty about true rates for indicators depending on this variable. They may seem to have high or low rates not because they truly do but because of uncertainty as to the true rate.

In an environment where data quality is in such flux within and across counties in addition to over time, coming to reliable conclusions about trends for indicators based on month prenatal care began will be problematic.

The variation in data quality over time will severely limit conclusions one might otherwise draw from multivariate models incorporating variables relying on month prenatal care began.

Number of Prenatal Care Visits

Importance

Number of prenatal care visits is a key indicator of health care access for pregnant women. It also is a core variable used to calculate Kotelchuck's Adequacy of Prenatal Care Visits [56].

In 2007, California added two related variables, date of first and last prenatal care visits.

Results

Table 14 shows results for the model to explain the rate for number of prenatal care visits to be missing (F = 17.80, P = < 0.0001, EV = 4.08, ES = 14.98). Part A shows model and variable statistics, and Part B shows trends. Error rates were below the CDC/NCHS Standard over the period. The error rate was steadily decreasing from 2000 through 2004, dropped briefly, then returned to the earlier highs.

Table 14: Results for number of prenatal care visits

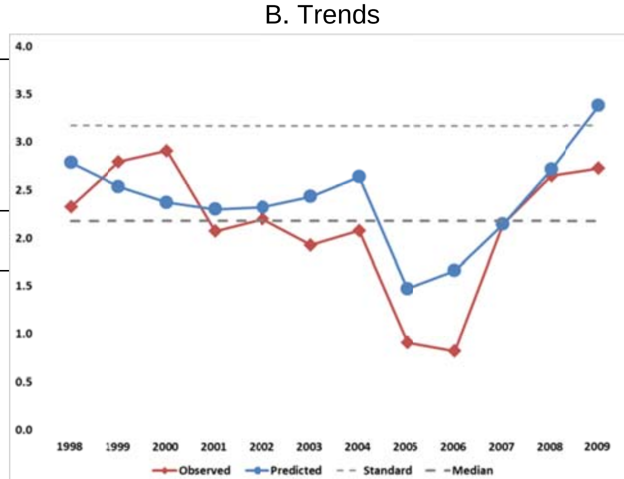
A1. Model Statistics

Contribution	EV	ES
Model	4.08	14.98
Time	0.22	9.60
Volume (after Time)	0.00	0.00
Auspice (after Time Volume)	3.16	3.45
Training (after Time Volume Auspice)	0.69	1.94

A2. Variable Statistics

Significant variables	Est	Std Err	Std Est	ES
Intercept	2.32	0.37	0.00	0.00
Year rel 1998	(0.30)	0.14	(0.15)	2.36
Year rel 1998 (Sq)	0.05	0.01	0.27	7.24
Sutter	(1.49)	0.51	(0.06)	0.35
Church	1.32	0.31	0.09	0.73
Public	3.48	0.45	0.15	2.37
People trained (Later)	(0.67)	0.16	(0.14)	1.94

B. Trends



Time, auspice, and training were significant predictors, but training was practically unimportant. Auspice again was the most important predictor of the error rate for number of prenatal care visits.

Answers to Research Questions

Are differences in hospital characteristics (volume, auspice) related to differences in quality and completeness? Differences in hospital volume were not related to differences in quality and completeness. As to auspice, Church and Public hospitals tended to have higher error rates while Sutter hospitals tended to have lower rates. Overall, auspice was a statistically significant predictor for number of prenatal care visits, with a small but practically important effect (ES = 3.45).

Does a training program for birth clerks improve the quality of birth certificates? Specifically, when comparing data quality before and after training, is there a significant decrease in missing or unlikely values in fields targeted by the training? Similar to mother's place of residence, we have no information to indicate that training specifically targeted number of prenatal care visits. While the general trend was for this error rate to drop, the number of people attending later trainings temporarily further reduced the rate.

Is the number of trainees associated with greater improvement compared to hospitals with few or no attendees? That is, is there a dose effect? The number of trainees attending training did affect the quality of this variable. Specifically, the error rate decreased as a function of later training.

Are improvements sustained over time? Improvements were not sustained over time. Although the effect of attending later trainings was statistically significant, it was not practically important (ES = 1.94) in predicting the error rate for number of prenatal care visits.

We are of the opinion that the circumstance here is similar to the circumstance for mother's place of residence. Two new date variables were introduced that were relevant to number of prenatal care visits. Here, however, the event affecting data quality appears to be instrumentation change (another type of threat to internal invalidity) [22]. Attention turned to the newly introduced variables.

Implications of County Variation in Data Quality

Averaged over the 12-year period, 13 counties had an error rate above the 2009 CDC/NCHS Median, including nine above the CDC/NCHS Standard, ranging as high as 15.5% over the period and as high as 25% in some years.

Counties with the highest error rates tended to be among the largest in the state (including Los Angeles). These counties will have a serious impact on reliability of the state rate. In addition, areas with high or unstable error rates face considerable uncertainty about true rates for indicators depending on this variable. They may seem to have high or low rates not because they truly do but because of uncertainty as to the true rate.

In an environment where data quality is in such flux within and across counties in addition to over time, coming to reliable conclusions about trends for indicators based on number of prenatal care visits will be problematic. The variation in data quality over time will severely limit conclusions one might otherwise draw from multivariate models incorporating variables relying on this variable.

Gestational Age in Weeks

Importance

Gestational age is a computed variable based initially on infant's birth date and date of mother's last menstrual period (LMP). If the computed variable is outside the range 17 to 47 completed weeks, the CDC/NCHS imputes gestational age using edited values for infant's birth date and LMP, mother's race/ethnicity (unknown/other values imputed), infant's birth weight (improbable values set to missing), and edited plurality (values outside the range 1-5 set to 1). All published vital statistics reports for preterm birth rates are based on complete gestational age estimates in the range 17 to 47 completed weeks edited as summarized here [30,44,45,46].

The obstetric estimate of gestation (in weeks) has been available in California since 2007.

Results

Table 15 shows results for the model to explain the error rate for gestational age in weeks (F = 36.71, P = < 0.0001, EV = 11.63, ES = 28.48). Part A shows variable and model statistics, and Part B shows trends. This trends graph is based on California's 2009 median rate, with the standard equal to 1.5 times the median (lower line). The error rate increased steadily through 2004, dropped precipitously with the start of training, and remained low thereafter.

Table 15: Results for gestational age in weeks

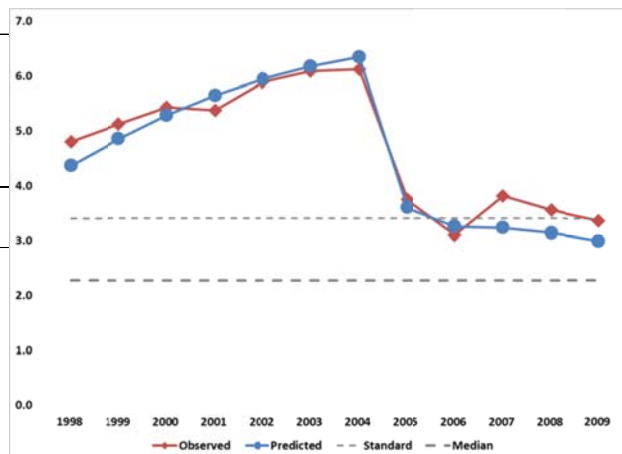
A1. Model Statistics

Contribution	EV	ES
Model	11.63	28.48
Time	3.09	12.52
Volume (after Time)	0.75	0.48
Auspice (after Time Volume)	4.71	6.82
Training (after Time Volume Auspice)	3.08	8.67

A2. Variable Statistics

Significant variables	Est	Std Err	Std Est	ES
Intercept	3.29	0.40	0.00	0.00
Year rel 1998	0.52	0.13	0.29	8.32
Year rel 1998 (Sq)	(0.03)	0.01	(0.20)	4.20
Hospital Volume	0.53	0.15	0.07	0.48
UNP	1.84	0.35	0.12	1.44
Kaiser	2.58	0.40	0.14	2.05
Sutter	(2.75)	0.48	(0.12)	1.39
Church	1.02	0.33	0.07	0.51
Public	2.50	0.44	0.12	1.44
People trained (Later)	(1.31)	0.14	(0.29)	8.67

B. Trends



Time, volume, auspice, and training were significant predictors of the gestational age error rate, with both auspice and training having practically important ES. However, we are uncertain as how much of the training ES truly is due to training and how much is due to the 2007 introduction of the clinical estimate of gestational age. That is, we may have another threat to internal validity.

Answers to Research Questions

Are differences in hospital characteristics (volume, auspice) related to differences in quality and completeness? For gestational age in weeks, differences in hospital volume were related to differences in quality and completeness. As volume increased, error rates tended to increase. While statistically significant, volume is practically unimportant (ES = 0.48). As to auspice, UNP, Kaiser, Church, and Public hospitals tended to have higher rates while Sutter hospitals tended lower. This result is both statistically significant and practically important (ES = 6.82).

Does a training program for birth clerks improve the quality of birth certificates? Specifically, when comparing data quality before and after training, is there a significant decrease in missing or unlikely values in fields targeted by the training? Although the general trend was for this error rate to increase, the number of people attending later trainings significantly reduced the rate. This result is both statistically significant and practically important (ES = 8.67).

Is the number of trainees associated with greater improvement compared to hospitals with few or no attendees? That is, is there a dose effect? The number of trainees attending training did affect the quality of this variable. Specifically, the rate decreased as a function of later training.

Are improvements sustained over time? Improvements were sustained over time. Further, like mother's place of residence, this is the only other core variable where the training ES was greater than 2%.

How much of this improvement is due to training and how much is due to the 2007 introduction of the clinical estimate of gestational age is subject to debate. Martin reported that gestational age estimates improved after the introduction of clinical estimates [45]. This suggests the initial drop was due to training, but the 2007 introduction of clinical estimates is somehow involved in the rates continuing to drop. That is, another instrument change likely caused another threat to internal validity [22], and that may be producing artificially stronger results in favor of a training effect.

Implications of County Variation in Data Quality

Despite significant improvement, 16 counties had error rates above 5% in 2009, and rates ranged as high as 23% in some counties in some years. Three of the 16 were large counties in southern California, while small counties in northern California and the central valley comprised the remainder. These counties will impact reliability of the state rate. In addition, areas with high or unstable error rates face considerable uncertainty about true rates. They may seem to have high or low rates not because they truly do but because of uncertainty as to the true rate.

In an environment where data quality is in such flux within and across counties in addition to over time, coming to reliable conclusions about trends for indicators based on gestational age in weeks will be problematic. The variation in data quality over time will severely limit conclusions one might otherwise draw from multivariate models incorporating variables relying on this variable.

Two national workshops of perinatal experts examined whether to change the definition of gestational age based upon LMP or clinical estimate. After reviewing their results, NAPHSIS concluded there was insufficient evidence to recommend using clinical estimate [46]. Births with missing LMP are known to be higher risk births associated with preterm delivery. Not addressing this issue may artificially lower the preterm delivery rate.

Although the CDC/NCHS recommends imputation based on birth weight, sex, and race/ethnicity, California has not done this in calculating preterm birth rates. As a result, there are discrepancies in preterm birth rates between those computed by NCHS and those computed by California. We strongly recommend that California use the CDC/NCHS algorithm to compute gestational age before reporting local jurisdiction rates. We modified the NCHS software to run as a macro, and we will post it on our website.

DISCUSSION

Methodology

California has not met NCHS quality standards for birth certificate data for many years. Responding to NCHS concerns, CCHS began the BDQT to improve data quality in 2004, with no initial intent to formally evaluate its efficacy. We identified several methodological issues that may have affected the results.

For example, trainers used different structures to record information about training attendees. CCHS staff reported that some signup sheets were lost and some people who attended may not have signed the sheets.

None of the sign-up sheets indicated the CCHS hospital code, and different attendees for what we took to be the same hospitals gave different but similar hospital names and different addresses, perhaps for offices across the campuses. We faced significant challenges in assigning CCHS hospital codes. For these reasons, we included several appendices to this report with the request that hospitals review them to see if we made correct decisions.

In the end, we made variables identifying the first date we thought a hospital sent at least one person to a training session, the number of people sent to that first training, the number of trainings at least one person attended later, and the total number of people sent to those later trainings. The good news is that most hospitals sent at least one person by the end of 2005, or about 15 months after the effort began.

We also do not know how many birth clerks hospitals employ, or the number of births they are responsible for recording. The ratio of births per hospital to trainees (2004-2009) had a wide range from a low of 302 to a high of 40,807 births per trainee over the interval. We have no information providing guidance as to the number of births one clerk should be responsible for recording.

The training agenda changed with each wave. As CCHS added new variables to the BSMF, they were added to the agenda, with a first set in 2005, and a second in 2007. As time went on, the agenda shortened, and the focus on core variables appears to have been lost. Sessions were about four hours long, regardless of the number of topics covered.

We used an interrupted time series design over the period 1998-2009 to evaluate the impact of the BDQT [22]. Often, as here, circumstances do not permit meeting all the conditions of a true experimental design. However, among quasi-experimental designs, the interrupted time-series rivals the true experiment. It has become the standard method of causal analysis in applied behavioral research.

The addition of a second time series for a comparison group can provide a check on some threats to validity. We initially planned a control group of hospitals that never attended training, but found that most attended within the first 15 months. Thus, we were unable to construct a meaningful control group. We identified several possible threats to internal validity that may have affected study results.

Known threats to internal validity of an interrupted time series design [22] provide some guidance on what might have happened with the BDQT. For example, adding new sets of variables two times during the study period (instrumentation) and changing the agenda with each new wave may have directed attention away from the core variables (curriculum). Four hours once a year may be too little training to have an impact (dose response). The method of measuring attendance changed frequently (instrumentation). Attendance was voluntary, some hospitals sent representatives only to one training, some sent supervisors rather than clerks, and no attention was paid to training enough clerks at each hospital (dropouts).

Study Factors

For this evaluation, we did trend analysis with the time series interrupted after the first hospital representative attended training. After controlling for variables that other research has associated with data quality (time, hospital volume, auspice) and that temporally preceding training, the error rate increased substantially for most core variables, and the BDQT failed to have any practically important, substantive impact. It is possible that instrumentation changes created threats to internal validity for four variables.

Volume was a statistically significant predictor for four of the eight core variables (mother's race/ethnicity, education, place of residence, and LMP date). However, ES was practically important only for mother's race/ethnicity (ES = 2.15). Thus while volume is important in many ways, in terms of California's data quality, it appears to be a small factor in predicting data errors.

Auspice was a consistent predictor of data errors in all core variables, with ES in the range 2.31 to 6.82. Kaiser hospitals were associated with increased error rates on only one of the core variables, and Sutter hospitals were associated with lower rates on four. UNP, Church, and Public hospitals were associated with higher rates on five core variables. Hospitals operating under these latter auspices may lack the resources to collect BSMF data systematically, in contrast to resources that Kaiser and Sutter hospitals may have. This may suggest that automating hospital data collection could have a more significant impact on error rates than training per se.

Training was statistically significant for six core variables (mother's place of residence, LLB date, LMP date, month of first prenatal care visit, number of prenatal care visits, and gestational age) and did not enter the models for mother's race/ethnicity or education. While statistically significant, ES was below 2 (indicating training had a very weak effect but was practically unimportant) for month of first prenatal care visit, number of prenatal care visits, and gestational age. The effect was somewhat mixed in that training sometimes increased rates, or reduced them temporarily, with no sustained impact.

Possible threats to internal validity, through changes in the way data were entered and/or introduction of related variables, were associated with four of these measures. Training effect size was moderate for mother's place of residence (ES = 6.52), with no evidence that the variable was addressed in training and with a likely threat to internal validity through the introduction of pull-down menus in AVSS. ES was larger for gestational age in weeks (ES = 8.67), again with no evidence the variable had been addressed in training and with a threat to internal validity that the CDC documented with introduction of clinical estimates of gestational age. Similarly, month prenatal care began and number of prenatal care visits may have been impacted by the introduction of dates of first and last visits.

In the end, we were unable to document any sustained, positive, substantive impact on data error rates that we clearly could attribute to the BDQT.

Non-random Missing Data

When data are missing completely at random (MCAR), missingness does not depend on observed or unobserved data [1-3]. In this context, analyses yield unbiased parameter estimates (i.e., estimates close to population values), albeit with some loss of statistical power. If BSMF data was MCAR, losing statistical power would not be a major issue in calculating rates for population indicators, because California numbers are more likely to be large enough even in most small counties.

The methodology literature strongly recommends that analysts test their data to be sure they can proceed with the simpler assumption that data are missing MCAR. The analyst should not proceed as if data was MCAR without first testing that assumption [1-3].

Our research design explicitly tested the assumption that data were MCAR. We found that volume and particularly auspice usually predicted missingness. That is, data are missing not at random (MNAR), a serious and non-ignorable problem, especially in the longitudinal data context [1-3]. An important wrinkle with MNAR is that it applies jointly to the data and the analysis: the design must address the fact of missingness.

We presume that other factors we did not measure also affect randomness assumptions, because most of our models had modest ES. Several possibilities that come to mind include birth clerk staffing, training, supervision, full-time equivalents, and turnover, as well as other factors that only could be identified by surveying hospitals. That is, we need to understand the local hospital culture to understand data quality [47]. We hoped to do such surveying but were unable to obtain funding for this.

We also did not have data enabling us to test if training might have become more effective after 2009, after the “noise” went down from the new variables added in 2005 and 2007.

California has been calculating public health indicators assuming data were MCAR when in fact data have been MNAR for many years, a statistically non-ignorable situation. Data quality has varied across time, and within and across health jurisdictions and hospitals. When data are MNAR, imputation is recommended. Different contexts require different statistical strategies.

The CDC/NCHS has known for many years that BSMF data are MNAR. They developed detailed decision rules and some SAS programs to correct data before calculating Healthy People indicators based on BSMF. We used their rules to do this analysis, converted several of their SAS programs to macros, and developed other new macros to clean data and calculate other indicators following CDC/NCHS rules.

In 2006, we completed a study for CCHS explicitly examining the impact of applying CDC/NCHS rules to correct variables underlying indicators for low birth weight and preterm birth over the period 1992-2003 [48]. We found that data quality had begun to deteriorate perhaps as early as 1992, that data quality problems occurred most frequently in high-risk infants leading to under-estimated rates (another example of MNAR), and that following CDC/NCHS rules produced more robust results.

At that time, we recommended that CCHS scientifically monitor the effectiveness of its then recently initiated BDQT. CCHS followed up that recommendation by asking us to evaluate the training impact, the results of which we reported here. We discuss our other recommendations below.

RECOMMENDATIONS

As implemented, the BDQT did not have a meaningful impact on birth certificate data quality. The results surprised us. We did not begin this study with a null hypothesis: we expected the training would be effective. It took a while for our staff to come to terms with the findings. However, after all the technical discussions, the state-level data quality report distributed during Training Period 5 (Table 4) demonstrated in the simplest way that the BDQT indeed was ineffective in improving data quality for the core variables.

In this section, we make our recommendations focusing first on how to modify the training if it is to continue, then on the need to examine the hospital culture, and finally, how to move forward methodologically given that California’s BSMF data have been longitudinally MNAR for many years, creating a scientifically non-ignorable data problem.

Training

1. If resources permit, training should be 6-8 hours. Four hours is not enough time to teach (or reinforce) what needs to be learned. If this is not possible, trainings should be broken up into shorter segments and video taped for web-posting.

2. Another possibility is to develop a standardized webinar-based training curriculum, with each segment focused on one topic. The webinars would be recorded and posted on the web, with pre- and post-training tests automated to capture what was learned. These would be available for reference any time clerks needed to know about the topic. Because of increasing resource constraints, FHOP has been producing webinars for several years through our contract with Maternal, Child, and Adolescent Health. Local health jurisdictions have responded very positively. We suspect that hospital birth clerks would positively respond to a similar training strategy by CCHS. We also suspect hospitals would find this a valuable resource given staff turnover. Any such effort should be monitored for effectiveness from the beginning.

3. The core agenda should be stabilized, and sufficient time allocated to each topic to ensure that learning can occur.

4. If face-to-face training or webinars are offered, more consistent and rigorous collection of registration and attendance data is needed to evaluate the impact of the training. We would be pleased to make recommendations to OVR if appropriate.

5. If training continues in any form, we recommend that evaluation materials focus on measurable objectives related to knowledge, and less on how people "liked" the training.

Some pessimists think training will not be effective until some mechanism forces hospitals to provide good quality data. For example, hospitals batch submit patient data to OSHPD, which returns the entire batch if the error rate is above some number. Another strategy might be to develop a relationship with accreditation agencies, share data quality information with them, and have them add this to their reviews. It might be possible to obtain legislation to fine hospitals submitting poor data quality, as California now is able to do for hospitals experiencing severe adverse patient outcomes. Another possibility might be to work with unions for legislation to establish staffing ratios and training standards for birth clerks.

In making these suggestions, we are acutely aware that each may engender strong reactions from hospitals and perhaps motivate them to implement their own programs to improve birth certificate data quality.

Learn More about the Local Hospital Culture

The data strongly suggest that the ratio between the number of births and people trained likely was too thin for training to be effective. In this context, we recommend that CCHS survey hospitals to gain some sense about how many people work as birth clerks, the full-time equivalent dedicated to this work, and staff turnover. Hospitals should be encouraged to make birth clerk attendance a required part of continuing education for staff.

We also recommend that CCHS initiate a round of meetings with hospital Chief Executive Officers, nursing supervisors, and others who are responsible for overseeing the completing of birth certificates. The purpose would be to review results of this research, and learn what might be done to gain their cooperation in providing better quality data. If requested, we would be able to generate hospital-level longitudinal trend data similar to what we do for Healthy People indicators.

We continue to believe that it is important to gain a better understanding of how hospital culture influences data quality. Every year, half of the states submit data to CDC/NCHS that is below federal medians. Another quarter of hospitals are above the median but below the standard. California has been well above the standard for many years and results reported here show unequivocally that data quality is worsening rather than improving. When three-quarters of our nation's states submit data of sufficient quality and California does not, then something about California's hospital culture is in play.

Having done this study, we have a much better understanding than we had before of how to design a hospital-based survey, both with respect to hospitals to sample and questions to ask. We would very much like to find funding to carry out such work.

How to Address Non-Ignorable Longitudinal Data Quality Problems

Ten years ago, the CDC/NCHS notified CCHS that data quality problems were no longer ignorable. The training program as implemented has been ineffective, and data quality problems in 2009 mainly were the same or worse. This study clearly has shown data are MNAR. In this context, continuing to use MCAR methods to calculate public health indicators is questionable statistically. The problem is particularly acute in the longitudinal context. In our 2006 report [48], we made the following recommendations that address this issue:

1. While we understand that California legislation requires the state to provide a file with unedited data that mirrors the birth certificate data as collected, we advise that the state add new variables to the BSMF implementing Federal data cleaning algorithms, with flags indicating corrected variables.
2. Require state contractors to use cleaned variables to calculate population-based outcomes, in order to more accurately report progress toward Healthy People objectives.
3. Continue to report data quality to counties to engage them as ongoing partners in the quest for more accurate population-based health outcomes monitoring.

These recommendations will solve the problem prospectively but will do nothing to correct data quality for earlier years. Toward this end, we are posting on our website the programs we developed with close cooperation from CDC/NCHS staff. Following CDC/NCHS rules to correct race/ethnicity is the most pressing, as that is the foundation upon which all indicators stand to examine progress in eliminating race/ethnic disparities.

The results of this study showing that California now has many more years of increasingly poor quality data lends a sense of urgency to these recommendations.

Page intentionally left blank

APPENDIX A: HOSPITAL CROSSWALK

OSHPD assigns facility identifiers (OSHPDID) which are unrelated to BSMF hospital codes. To do this study we had to crosswalk between OSHPDID and HOSPCODE. Appendix A describes the methods we followed to develop the crosswalk.

Herrchen et al developed a crosswalk methodology to make a linked birth, hospital, and death file to study newborns [49]. She graciously provided crosswalk files for the years 1995-2006. These crosswalk within a given year but do not account for longitudinal changes like moves or consolidated reporting. We prepared her files for our purposes. The resulting file had 398 records with 383 HOSPCODES associated with 395 OSHPDID (the Herrchen files).

We read into SAS the 2009 Vital Statistics file with historic HOSPCODE, address, city, and ZIP (N = 781) from 1960 forward (the historic file) [25], removing military hospitals (N = 32) and doctor's offices (N = 12). This left 737 possible facilities.

The problem with the historic BSMF file is that it has original names and addresses when first assigned in 1960 or later. 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. This makes the crosswalk difficult when, as here, we are developing a methodology based on geography. We began with a simple cleaning of available names, addresses, and cities to make linkage more possible.

In preparing BSMF and OSHPD files for longitudinal analysis, a standard final task we do is summarize cases across all facilities and years from 1989 forward (N = 546) for the BSMF and 1983 forward for OSHPD files, which include free-standing ambulatory facilities from 2005 forward (N = 1,274) [24] (the YEARS files). These files are a resource to identify abrupt changes in number of cases, or when facilities appear, which might indicate closures, consolidations, or moves. We merged the historic BMSF file with the BSMF years file to identify the first and last year with data for the facility and the number of births from 1989 to 2009 (N = 11,486,483).

AHDR Page 0 has information about the current and previous hospital name and owner, and the physical address, mailing address, and address of the person completing the report. Having this longitudinal resource of address and name data from OSHPD has always been valuable. For example, this provides a first step to identify possible ownership changes in moves. We also have a series of other files from OSHPD identifying hospital changes back to the early 1980s.

Page 0 variables are full of typographic errors. We applied the same minimal code to clean address data that we used with the BSMF historic file, then output one record for each combination of OSHPDID, name, and street address (N (Reports) = 15,649; N (OSHPDID) = 752; N (Combo) = 48,130).

Next, we applied some macros we developed for other studies to link HOSPCODE and OSHPDID. We did not permit cross-county linkages, for example, hospitals in one county that move to another county, or that consolidate reporting with a hospital in another county. Applying 14 different criteria, we made one or more potential linkages for 532 HOSPCODE. We assigned the current OSHPDID (CURROSH) to handle moves and consolidations. Of 206 unlinked HOSPCODE, 151 had no births from 1989 forward, while the remainder was birthing centers that OSHPD does not track. Unlinked HOSPCODE accounted for 10,351 (0.09%) of 11,486,483 births over the interval 1989-2009.

The next step had two goals. First, we had to resolve linkages where OSHPDID and HOSPCODE pointed to different facilities. Second, we wanted to check linkage reliability for the years that coincided with the

Herrchen files. Of 532 hospitals we linked, 381 were in the Herrchen files, including 50 (13.1%) that we linked to her OSHPDID and one other. This gave an overall reliability of 86.9% between her link and ours to the same OSHPDID. We broke the 50 ties by selecting her OSHPDID. One of her links was to an OSHPDID that OSHPD's ALIRTS system did not recognize. She linked one OSHPDID to two HOSPCODE, but this was a cross-county link that we did not permit.

We found 151 HOSPCODE with no deliveries during the Herrchen file years. We linked 132 to one OSHPDID (91.7%). We linked 14 HOSPCODE to 28 OSHPDID. We broke these cross-links in favor of the one with the strongest results. We manually over-rode five linkages. In the end, we linked 532 HOSPCODE to 441 CURROSH.

The crosswalk between OSHPDID and HOSPCODE will be posted on our website, together with the SAS programs, so others may review and comment on our methodology. If hospitals notify us that we crosswalked incorrectly, we will make the changes and re-estimate our models.

APPENDIX B: HOSPITALS COMBINED DUE TO CONSOLIDATIONS OR MOVES

We merged the training file with the Vital Statistics YEARS file and the file with the ID crosswalk, keeping records that were in the YEARS and training files. This resulted in a list of 324 hospitals. We then assigned CURROSH, identified HOSPCODE with more than one CURROSH. A total of 28 HOSPCODE combined to 14 CURROSH because of moves or consolidated reporting to OSHPD. Table 16 identifies the hospitals combined due to consolidations or moves. We suggest that hospitals review Table 16 to verify that these correctly reflect moves or consolidated reporting to OSHPD. To the extent that they do not, please notify us and we will make corrections and re-estimate the models.

Table 16: Hospitals combined due to consolidations or moves

County	Hospital Code	OSHPD Identifier	Elig	BSMF	Train	Births	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
10 Fresno	0058 Community Regional M	100717 Community Regional	1	1	1	75,579	4,792	4,813	5,145	5,259	5,511	6,738	7,484	7,586	7,903	6,951	6,849	6,548
10 Fresno	0067 University Medical Cent	100717 Community Regional	0	1	0	9,236	2,093	1,706	1,604	1,567	1,608	649	7	2				
12 Humboldt	0070 General Hospital	121080 St. Joseph Hospital -	0	1	0	1,083	313	319	375	76								
12 Humboldt	0075 St. Joseph Hospital	121080 St. Joseph Hospital -	1	1	1	7,081	393	433	349	518	639	623	665	675	679	692	710	705
19 Los Angeles	0151 Centinela Hospital Med	190148 Centinela Hospital M	1	1	1	28,591	2,128	1,856	2,324	2,341	2,233	2,485	2,304	2,236	3,775	3,090	2,056	1,763
19 Los Angeles	0162 Daniel Freeman Memor	190148 Centinela Hospital M	0	1	0	16,337	3,087	2,829	1,879	1,849	1,490	1,771	1,707	1,716	9			
19 Los Angeles	0200 Citrus Valley Medical C	190636 Citrus Valley Medica	0	1	0	1,111	746	354	2	1		2	1	1			2	2
19 Los Angeles	0265 Citrus Valley Medical C	190636 Citrus Valley Medica	1	1	1	58,360	3,588	3,582	4,506	4,396	4,529	4,871	5,392	5,540	5,638	5,588	5,549	5,181
19 Los Angeles	0134 Bay Harbor Hospital	190680 Providence Little Cor	0	1	0	1,031	573	458										
19 Los Angeles	0275 Providence Lcm San Pe	190680 Providence Little Cor	1	1	1	9,070	454	583	828	767	799	816	765	795	889	831	712	831
30 Orange	0374 Anaheim Memorial Mec	301098 Anaheim Memorial M	1	1	1	21,588	772	1,093	1,668	2,253	2,366	2,404	1,845	1,768	1,993	1,898	1,802	1,726
30 Orange	0392 Anaheim Memorial Mec	301098 Anaheim Memorial M	0	1	0	3,670	1,692	1,411	567									
30 Orange	0397 San Clemente Hospital	301317 Saddleback Memoria	0	1	0	785	383	281	84	34	1					1	1	
30 Orange	0736 Saddleback Memorial M	301317 Saddleback Memoria	1	1	1	35,027	2,416	2,501	2,736	3,181	3,171	3,212	3,270	3,056	2,881	2,804	2,877	2,922
33 Riverside	0744 Inland Valley Regional I	334068 Southwest Healthcar	1	1	1	17,577	1,046	1,230	1,405	1,369	1,550	1,534	1,546	1,644	1,648	1,595	1,472	1,538
33 Riverside	0759 Rancho Springs Medica	334068 Southwest Healthcar	1	1	0	20,572	1,211	1,110	1,184	1,222	1,392	1,857	1,883	1,941	1,976	2,122	2,130	2,544
36 San Bernardir	0455 Montclair Hospital Medi	361166 Montclair Hospital Mo	1	1	1	13,077	1,228	1,318	743	763	1,139	1,334	1,065	1,248	1,229	954	956	1,100
	0459 Montclair Community H	361166 Montclair Hospital Mo	0	0	1													
37 San Diego	0489 Sharp Memorial Hospita	370694 Sharp Memorial Hos	0	1	0	3,842	3,802	7	2	4	4	2		3	5		3	10
37 San Diego	0778 Sharp Mary Birch Hosp	370694 Sharp Memorial Hos	1	1	1	86,511	3,705	6,527	6,982	6,518	6,373	7,237	7,650	8,034	8,211	8,374	8,595	8,305
37 San Diego	0480 Scripps Mercy Hospital	370744 Scripps Mercy Hospi	1	1	1	22,806	1,586	1,611	1,720	1,892	1,984	2,069	2,122	1,909	1,870	2,025	2,083	1,935
37 San Diego	0498 Scripps Mercy Hospital	370744 Scripps Mercy Hospi	1	1	1	28,865	2,622	2,864	2,616	2,479	2,278	2,268	2,257	2,348	2,280	2,417	2,340	2,096
43 Santa Clara	0568 Comm Hosp Los Gatos	430763 El Camino Hospital	1	1	1	9,908	934	865	814	862	1,011	1,190	971	758	771	751	650	331
43 Santa Clara	0569 El Camino Hospital	430763 El Camino Hospital	1	1	1	52,763	4,269	4,193	4,465	4,188	4,639	4,521	4,505	4,335	4,413	4,509	4,439	4,287
43 Santa Clara	0746 St. Louise Regional Ho	434138 St. Louise Regional I	1	1	0	8,118	540	680	852	773	690	715	666	639	695	550	645	673
43 Santa Clara	0747 St. Louise Hospital	434138 St. Louise Regional I	0	1	0	848	496	352										
56 Ventura	0645 Ventura County Medica	560481 Ventura County Med	1	1	1	27,539	1,760	1,612	1,739	1,797	1,863	2,027	2,097	2,708	3,145	2,967	2,992	2,832
56 Ventura	0654 Santa Paula Hospital	560481 Ventura County Med	0	1	1	3,010	269	260	231	274	287	273			129	340	441	506

APPENDIX C: TRAININGS

California began BDQT on 26-Oct-2004. Trainings were regional, with all facilities in a geographic area invited to send representatives. Trainers maintained reservation sheets and attempted to have people sign in at arrival. However, an unknown number of attendees did not sign attendance sheets, and some attendance sheets were lost. Thus, we do not know for certain when hospitals first sent at least one person for training, exactly how many times they sent at least one person, the number of people a hospital sent over time, and for some number, whether they ever sent anyone. The extent to which we do not have this information may have some unknown impact on our results. Table 17 summarizes what we know about the trainings.

We entered available attendance data into an Excel spreadsheet, one record per hospital per training, counting the number of people from the facility who signed attendance sheets on that date. Because attendance sheets only had hospital name and address, we had to identify the HOSPCODE to assign. As described above, the historic HOSPCODE list did not have current names of many hospitals. Using a combination of county and city where the hospital was located enabled us to identify most hospitals. We also searched AHDR Page 0 to find others. Non-hospital attendees (local or state departments of public health, etc) were assigned the value 9999. The extent to which we were unable to assign HOSPCODE correctly may have some unknown impact on our results.

We initially had 1,091 records summarizing attendance between October 2004 and late 2009. We removed records where hospitals registered in advance but for whom we found no attendance verification (N = 126), and records for non-hospital attendees (N = 141)), leaving a remainder of 824 records. Table 17 identifies the training dates, locations, number of hospitals sending representatives, and number of people who signed in.

We summarized this file to one record per training to obtain the verified number of hospitals and hospital staff that attended. For each hospital, we counted the documented number of people attending the first training (PEOPLE0), number of trainings attended later that we could document through 2009 (TRAINP), and total number of people attending those later trainings (PEOPLEP). Data limitations did not permit us to use other information such as the number of uniquely trained people, the role of trained people, or evaluations of training quality. We identified 262 facilities that sent at least one person to at least one training.

Table 17: Training dates, hospitals, and staff

Date	City	Hospitals	Staff	Average
	Total	824	1369	1.7
26-Oct-04	Fresno	20	26	1.3
3-Nov-04	Bay Area	44	71	1.6
10-Nov-04	Riverside	23	42	1.8
18-Nov-04	Ventura	11	16	1.5
1-Dec-04	Long Beach	39	67	1.7
8-Dec-04	Sacramento	18	30	1.7
16-Dec-04	Santa Barbara	6	11	1.8
12-Jan-05	Shasta	6	8	1.3
26-Jan-05	Oakland Kaiser	8	12	1.5
2-Feb-05	San Diego	14	32	2.3
28-Feb-05	Santa Ana	19	19	1.0
2-Mar-05	Humboldt	6	6	1.0
10-May-05	Fresno	16	25	1.6
2-Jun-05	Alameda	29	46	1.6
23-Aug-05	Humboldt	2	4	2.0
31-Aug-05	Santa Cruz	7	10	1.4
14-Sep-05	Stockton	17	25	1.5
22-Sep-05	Oxnard	9	12	1.3
28-Sep-05	Shasta	3	5	1.7
4-Oct-05	Orange	13	22	1.7
12-Oct-05	San Diego	11	21	1.9
19-Oct-05	Santa Rosa	10	16	1.6
15-Nov-05	Pasadena	15	23	1.5
29-Nov-05	Long Beach	17	30	1.8
8-Dec-05	San Bernardino	16	27	1.7
15-Dec-05	Santa Maria	6	9	1.5
2-Mar-06	Fresno	20	20	1.0
12-Apr-06	Santa Rosa	16	24	1.5
25-Apr-06	Orange County	17	28	1.6
11-May-06	Roseville	20	26	1.3
28-Jun-06	San Diego	12	28	2.3
18-Jul-06	Oxnard	10	18	1.8
25-Jul-06	Pasadena	17	28	1.6
17-Nov-06	Sacramento	31	45	1.5
18-Mar-08	Alameda	23	33	1.4
9-Apr-08	San Diego	14	37	2.6
15-May-08	Fresno	15	30	2.0
4-Jun-08	Sacramento	25	42	1.7
10-Jul-08	Long Beach	27	60	2.2
20-Aug-08	Pasadena	22	34	1.5
18-Sep-08	San Bernardino	15	25	1.7
8-Oct-08	Ventura	14	25	1.8
21-Apr-09	Riverside	17	40	2.4
20-May-09	Fresno	19	30	1.6
23-Jun-09	Long Beach	23	29	1.3
14-Jul-09	Pasadena	32	58	1.8
18-Aug-09	Sacramento	26	39	1.5
8-Sep-09	San Francisco	18	40	2.2
28-Oct-09	Redding	6	15	2.5

APPENDIX D: EXCLUDED HOSPITALS

We next removed 100 hospitals with the following characteristics: no OSHPD linkage, cross-county consolidations, less than 100 births in any year of the study period, no documented training, or first training after 2005. We removed hospitals with cross-county consolidations because AHDR data are not available after consolidation. See Table 18.

Of 61 hospitals with less than 100 births any year, 24 closed before 2004, 8 closed after attending a first training, 5 opened in 2004 or later, and 14 had less than 100 cases every year. It is possible that some in this group had consolidations or moves that we were unable to link. It also is possible that hospitals with late or no training attended earlier as discussed above. Again, we urge hospitals to review the exclusions and notify us of any changes they recommend.

Table 18: Excluded hospitals

County	Hospital Code	OSHPD Facility ID	REMRESN	SRCBC	SRCTRN	ELIG	Births
01 Alameda	0001 Alameda Hospital	010735 Alameda Hospital	LT12YR	1	0	0	1,071
01 Alameda	0015 Kaiser Hospital: Oakland	010856 Kaiser Fnd Hosp - Oakland Campus	LT12YR	1	1	0	13,176
01 Alameda	0016 Summit Medical Center - Hawthorne	010937 Alta Bates Summit Med Ctr-Summit Campus-Hawtl	LT12YR	1	0	0	21,949
05 Calaveras	0037 Mark Twain St. Joseph's Hospital	050932 Mark Twain St. Joseph's Hospital	LT12YR	1	1	0	918
07 Contra Costa	0040 Doctors Medical Center	070904 Doctors Medical Center - San Pablo	LT12YR	1	1	0	5,174
07 Contra Costa	0048 Mt. Diablo Medical Center	071018 John Muir Medical Center-Concord Campus	LT12YR	1	0	0	2,737
07 Contra Costa	0785 Kaiser Foundation Hospital Antioch	074097 Kaiser Found Hsp-Antioch	LT12YR	1	1	0	2,679
10 Fresno	0056 Coalinga Regional Medical Center	100697 Coalinga Regional Medical Center	LT12YR	1	1	0	124
10 Fresno	0062 Sanger General Hospital	100791 Central Valley Orthopedic And Spine Institute	LT12YR	1	0	0	1,313
11 Glenn		110889 Glenn Medical Center	LT12YR	0	1		0
16 Kings	0109 Hanford Community Med Cntr	160725 Hanford Community Medical Center	LT12YR	1	0	0	1,101
17 Lake	0115 St Helena Clearlake	171049 St. Helena Hospital - Clearlake	LT12YR	1	0	0	1,906
19 Los Angeles	0136 Bellwood General Hospital	190069 Bellwood General Hospital	LT12YR	1	0	0	3,534
19 Los Angeles	0237 Mission Hospital	190197 Community And Mission Hsp Of Hntg Pk - Slauson	LT12YR	1	1	0	18,877
19 Los Angeles	0164 Lakewood Regional Medical Center	190240 Lakewood Regional Medical Center	LT12YR	1	0	0	6,827
19 Los Angeles	0182 Granada Hills Community Hospital	190348 Granada Hills Community Hosp	LT12YR	1	0	0	5,274
19 Los Angeles	0186 Robert F. Kennedy Medical Center	190366 Robert F. Kennedy Medical Center	LT12YR	1	0	0	5,400
19 Los Angeles	0254 Suburban Medical Center	190468 Promise Hospital Of East Los Angeles-East L.A. C	LT12YR	1	1	0	22,650
19 Los Angeles	0214 Long Beach Community Medical Center	190475 Community Hospital Of Long Beach	LT12YR	1	0	0	4,875
19 Los Angeles	0267 North Hollywood Medical Center	190654 North Hollywood Medical Center	LT12YR	1	0	0	519
19 Los Angeles	0277 Santa Marta Hospital And Clinic	190685 Elastar Community Hospital	LT12YR	1	0	0	5,165
19 Los Angeles	0279 Santa Teresita Hospital	190691 Santa Teresita Hospital	LT12YR	1	0	0	1,468
19 Los Angeles	0283 South Bay Medical Center	190734 South Bay Hospital	LT12YR	1	0	0	116
19 Los Angeles	0291 St. Luke Hospital	190759 St. Luke Med. Ctr - Pasadena	LT12YR	1	0	0	2,531
19 Los Angeles	0302 Nhmc-Sherman Way Campus	190810 Northridge Hospital Medical Center - Sherman Way	LT12YR	1	0	0	16,482
19 Los Angeles	0309 Doctors Hospital Of West Covina	190857 Doctors Hospital Of West Covina, Inc	LT12YR	1	0	0	567
19 Los Angeles	0223 Martin Luther King Jr Harbor Hospital	191230 Martin Luther King Jr.-Harbor Hospital	LT12YR	1	1	0	8,774
24 Merced	0349 Mercy Medical Center - Dominican	240948 Mercy Medical Center Merced-Dominican Campus	LT12YR	1	0	0	6,189
26 Mono	0674 Mammoth Hospital	260011 Mammoth Hospital	LT12YR	1	1	0	1,227
30 Orange	0373 Anaheim General Hospital	301097 Anaheim General Hospital	LT12YR	1	1	0	6,205
30 Orange	0377 Chapman Medical Center	301140 Chapman Medical Center	LT12YR	1	0	0	707
30 Orange	0716 Placentia Linda Hospital	301297 Placentia Linda Hospital	LT12YR	1	1	0	6,273
30 Orange	0396 Santa Ana Hospital Med Center	301314 Santa Ana Hospital Med Center	LT12YR	1	0	0	10,753
30 Orange	0399 South Coast Medical Center	301337 South Coast Medical Center	LT12YR	1	1	0	5,597
30 Orange	0404 West Anaheim Medical Center	301379 West Anaheim Medical Center	LT12YR	1	0	0	890
30 Orange	0748 Irvine Regional Hospital & Medical Center	304045 Hoag Hospital Irvine	LT12YR	1	1	0	19,702
30 Orange	0786 Kaiser Foundation Hospital Irvine	304306 Kaiser Foundation Hospital Irvine	LT12YR	1	1	0	3,422
31 Placer	0788 Kaiser Foundation Hospital Roseville	314024 Kaiser Fnd Hosp - Sacramento/Roseville-Eureka	LT12YR	1	1	0	4,815
32 Plumas	0414 Plumas District Hospital	320986 Plumas District Hospital	LT12YR	1	0	0	1,098
33 Riverside	0768 Eisenhower Memorial Hospital	331168 Eisenhower Memorial Hospital	LT12YR	1	0	0	9,118
33 Riverside	0424 Palo Verde Hospital	331288 Palo Verde Hospital	LT12YR	1	1	0	1,900
36 San Bernardir	0453 Chino Valley Medical Center	361144 Chino Valley Medical Center	LT12YR	1	1	0	7,587
36 San Bernardir	0460 Mountains Community Hospital	361266 Mountains Community Hospital	LT12YR	1	1	0	1,459
36 San Bernardir	0732 Hi Desert Medical Center	362041 Hi-Desert Medical Center	LT12YR	1	1	0	2,670
37 San Diego	0479 Alvarado Hospital Med. Cntr.	370652 Alvarado Hospital	LT12YR	1	0	0	4,452
37 San Diego	0488 Sharp Coronado Hospital	370689 Sharp Coronado Hospital And Healthcare Center	LT12YR	1	0	0	1,980
39 San Joaquin	0755 Kaiser Permanente Hospital	394009 Kaiser Fnd Hosp-Manteca	LT12YR	1	0	0	2,962
40 San Luis Obis	0544 San Luis Obispo General Hospital	400511 San Luis Obispo General Hsp	LT12YR	1	0	0	1,916
42 Santa Barbar	0557 Goleta Valley Cottage Hospital	420483 Goleta Valley Cottage Hospital	LT12YR	1	0	0	3,271
42 Santa Barbar	0563 St. Francis Med Cnt Of Santa Barbara	420528 St. Francis M/C-Santa Barbara	LT12YR	1	0	0	965
42 Santa Barbar	0565 Valley Community Hospital	420535 Valley Community Hospital	LT12YR	1	0	0	1,004
43 Santa Clara	0575 Columbia San Jose Medical Center	430879 San Jose Medical Center	LT12YR	1	0	0	2,068
45 Shasta	0587 Mayers Memorial Hospital	450936 Mayers Memorial Hospital	LT12YR	1	1	0	1,086
45 Shasta	0588 Redding Medical Center	450940 Shasta Regional Medical Center	LT12YR	1	0	0	803
49 Sonoma	0606 Healdsburg General Hospital	490964 Healdsburg District Hospital	LT12YR	1	0	0	559
50 Stanislaus	0787 Kaiser Hospital Of Modesto		LT12YR	1	1	0	2,563
54 Tulare	0632 Lindsay District Hospital	540746 Lindsay District Hospital	LT12YR	1	0	0	1,174
56 Ventura	0649 Ojai Valley Community Hospital	560501 Ojai Valley Community Hospital	LT12YR	1	0	0	347
16 Kings	0112 Naval Hospital		NOOSH	1	1	1	4,497
36 San Bernardir	0461 Naval Hospital 29 Palms		NOOSH	1	0	1	5,149
37 San Diego	0501 Naval Hospital		NOOSH	1	1	1	18,545
37 San Diego	0502 Naval Medical Center (Balboa)		NOOSH	1	0	1	41,652
48 Solano	0597 60th Medical Group		NOOSH	1	1	1	6,688
36 San Bernardir	0693 Weed Army Community Hospital		NOOSH	1	0	1	4,079
37 San Diego	0711 Best Start Birth Center		NOOSH	1	0	0	1,284
36 San Bernardir	0740 Inland Midwife Services		NOOSH	0	1		411
49 Sonoma	0770 Women's Health & Birth Center		NOOSH	1	1	0	1,004
34 Sacramento	0782 The Birth Center		NOOSH	0	1		177
15 Kern	0100 Ridgecrest Regional Hospital	150782 Ridgecrest Regional Hospital	TRNLATE	1	1	1	5,793
19 Los Angeles	0191 Hollywood Presbyterian Medical Center	190382 Hollywood Presbyterian Medical Center	TRNLATE	1	1	1	57,470
19 Los Angeles	0226 Providence Tarzana Medical Center	190517 Providence Tarzana Medical Center	TRNLATE	1	1	1	35,207
19 Los Angeles	0229 Glendale Memorial Hospital	190522 Glendale Memorial Hospital And Health Center	TRNLATE	1	1	1	21,298
19 Los Angeles	0247 Northridge Hospital Medical Center	190568 Northridge Hospital Medical Center	TRNLATE	1	1	1	31,411
19 Los Angeles	0263 Presbyterian Intercommunity Hospital	190631 Presbyterian Intercommunity Hospital	TRNLATE	1	1	1	42,859
19 Los Angeles	0272 San Dimas Community Hospital	190673 San Dimas Community Hospital	TRNLATE	1	1	1	8,845
19 Los Angeles	0280 Pacifica Hospital Of The Valley	190696 Pacifica Hospital Of The Valley	TRNLATE	1	1	1	10,602
19 Los Angeles	0310 West Hills Hospital And Medical Center	190859 West Hills Hospital And Medical Center	TRNLATE	1	1	1	15,607
19 Los Angeles	0315 White Memorial Medical Center	190878 White Memorial Medical Center	TRNLATE	1	1	1	39,656
33 Riverside	0425 Parkview Community Hospital Med. Cntr.	331293 Parkview Community Hospital Medical Center	TRNLATE	1	1	1	24,119
37 San Diego	0492 Grossmont Hospital	370714 Grossmont Hospital	TRNLATE	1	1	1	37,141
39 San Joaquin	0537 St. Joseph's Medical Center	391042 St. Joseph's Medical Center Of Stockton	TRNLATE	1	1	1	25,153
40 San Luis Obis	0545 Sierra Vista Regional Medical Center	400524 Sierra Vista Regional Medical Center	TRNLATE	1	1	1	13,521
44 Santa Cruz	0586 Watsonville Community Hospital	444013 Watsonville Community Hospital	TRNLATE	1	1	1	19,299
04 Butte	0032 Oroville Hospital	040937 Oroville Hospital	TRNNONE	1	0	1	7,021
18 Lassen	0119 Banner Lassen Medical Center	180919 Lassen Community Hospital	TRNNONE	1	0	1	2,996
19 Los Angeles	0140 Beverly Hospital	190081 Beverly Hospital	TRNNONE	1	0	1	23,330
19 Los Angeles	0146 California Hospital Medical Center	190125 California Hospital Medical Center - Los Angeles	TRNNONE	1	0	1	53,045
19 Los Angeles	0262 Pomona Valley Hospital Medical Center	190630 Pomona Valley Hospital Medical Center	TRNNONE	1	0	1	77,559
19 Los Angeles	0729 Kaiser Hospital: Woodland Hills	191450 Kaiser Fnd Hosp - Woodland Hills	TRNNONE	1	0	1	21,935
30 Orange	0761 Western Medical Ctr - Anaheim	301188 Western Medical Center Hospital - Anaheim	TRNNONE	1	0	1	28,077
36 San Bernardir	0477 Victor Valley Community Hospital	361370 Victor Valley Community Hospital	TRNNONE	1	0	1	14,389
38 San Franciscc	0525 St. Luke's Hospital	380964 St. Luke's Hospital	TRNNONE	1	0	1	12,425
41 San Mateo	0550 Seton Medical Center	410817 Seton Medical Center	TRNNONE	1	0	1	8,937
43 Santa Clara	0746 St. Louise Regional Hospital	434138 St. Louise Regional Hospital	TRNNONE	1	0	1	8,118
50 Stanislaus	0617 Memorial Medical Center	500939 Memorial Hospital Medical Center - Modesto	TRNNONE	1	0	1	26,139
10 Fresno	0063 Selma Community Hospital	100793 Selma Community Hospital	XCNTYCON	1	1	1	10,551
51 Sutter	0623 Fremont Medical Center	510882 Fremont Medical Center	XCNTYCON	1	1	1	25,431

APPENDIX E: HOSPITALS CARRIED FORWARD

County	OSHPD Identifier	REMRESN	Auspice	Volume	Births	TrnDate0	City	People0	PeopleP	TrainP	Ratio
Alameda	010739 Alta Bates Summit Med Ctr-Alta		3 Sutter	2 High	7,502	11/03/2004	Bay Area	2	3	2	1,500
Alameda	010805 Eden Medical Center		3 Sutter	0 Low	965	11/03/2004	Bay Area	3	3	2	161
Alameda	010846 Alameda Co Med Ctr - Highland		6 Public	1 Med High	1,341	11/03/2004	Bay Area	2	3	3	268
Alameda	010858 Kaiser Fnd Hosp - Hayward		2 Kaiser	2 High	2,845	11/03/2004	Bay Area	1	3	3	711
Alameda	010967 St. Rose Hospital		4 Church	0 Low	1,344	11/03/2004	Bay Area	2	3	3	269
Alameda	010987 Washington Hospital - Fremont		5 District	0 Low	2,878	11/03/2004	Bay Area	1	3	2	720
Alameda	014050 Valleycare Medical Center		1 UNP	0 Low	1,402	11/03/2004	Bay Area	3	1	1	351
Amador	034002 Sutter Amador Hospital		3 Sutter	0 Low	138	09/14/2005	Stockton	1	2	2	46
Butte	040875 Feather River Hospital		4 Church	0 Low	654	01/12/2005	Shasta	2	3	3	131
Butte	040937 Oroville Hospital	TRNNONE	1 UNP	0 Low	618	01/12/2005	Shasta	0	0	0	618
Butte	040962 Enloe Medical Center- Esplanad		1 UNP	0 Low	1,572	09/28/2005	Shasta	2	2	1	393
Colusa	060870 Colusa Regional Medical Center		1 UNP	0 Low	205	12/08/2004	Sacramento	1	3	3	51
Contra Costa	070924 Contra Costa Regional Medical (		6 Public	1 Med High	2,025	11/03/2004	Bay Area	2	3	2	405
Contra Costa	070934 Sutter Delta Medical Center		3 Sutter	0 Low	1,025	11/03/2004	Bay Area	1	3	2	256
Contra Costa	070988 John Muir Medical Center-Walnu		1 UNP	2 High	3,055	11/03/2004	Bay Area	1	3	2	764
Contra Costa	070990 Kaiser Fnd Hosp - Walnut Creek		2 Kaiser	2 High	4,055	11/03/2004	Bay Area	3	3	3	676
Contra Costa	074017 San Ramon Regional Medical C		7 Investor	0 Low	843	11/03/2004	Bay Area	1	3	3	211
Del Norte	084001 Sutter Coast Hospital		3 Sutter	0 Low	294	03/02/2005	Humboldt	1	1	1	147
El Dorado	090793 Barton Memorial Hospital		1 UNP	0 Low	562	12/08/2004	Sacramento	2	3	3	112
El Dorado	090933 Marshall Medical Center (1-Rh)		1 UNP	0 Low	600	12/08/2004	Sacramento	1	1	1	300
Fresno	100005 Clovis Community Medical Cente		1 UNP	0 Low	3,105	10/26/2004	Fresno	2	3	3	621
Fresno	100717 Community Regional Medical Ce		1 UNP	2 High	7,449	10/26/2004	Fresno	1	3	3	1,862
Fresno	100797 Sierra Kings District Hospital		5 District	0 Low	1,456	10/26/2004	Fresno	1	3	3	364
Fresno	100899 St. Agnes Medical Center		4 Church	1 Med High	2,759	05/10/2005	Fresno	2	3	2	552
Fresno	104062 Kaiser Fnd Hosp - Fresno		2 Kaiser	0 Low	1,267	10/26/2004	Fresno	1	2	2	422
Humboldt	121002 Mad River Community Hospital		7 Investor	0 Low	459	03/02/2005	Humboldt	1	3	2	115
Humboldt	121051 Redwood Memorial Hospital		4 Church	0 Low	364	03/02/2005	Humboldt	1	0	0	364
Humboldt	121080 St. Joseph Hospital - Eureka		4 Church	0 Low	692	03/02/2005	Humboldt	1	3	3	173
Imperial	130699 El Centro Regional Medical Cen		6 Public	0 Low	1,242	02/02/2005	San Diego	2	3	3	248
Imperial	130760 Pioneers Memorial Hospital		5 District	0 Low	1,535	02/02/2005	San Diego	3	3	2	256
Inyo	141273 Northern Inyo Hospital		5 District	0 Low	191	10/26/2004	Fresno	1	2	2	64
Kern	150706 Delano Regional Medical Center		7 Investor	0 Low	820	10/26/2004	Fresno	3	1	1	205
Kern	150722 Bakersfield Memorial Hospital- 3		4 Church	1 Med High	2,380	10/26/2004	Fresno	1	3	3	595
Kern	150736 Kern Medical Center		6 Public	2 High	4,180	10/26/2004	Fresno	1	3	3	1,045
Kern	150761 Mercy Hospital - Bakersfield		4 Church	2 High	3,139	10/26/2004	Fresno	1	1	1	1,570
Kern	150782 Ridgecrest Regional Hospital		1 UNP	0 Low	452	09/18/2008	San Bernardino	3	2	1	90
Kern	150788 San Joaquin Community Hospital		4 Church	0 Low	1,952	05/10/2005	Fresno	2	2	1	488
Kings	160787 Central Valley General Hospital		4 Church	0 Low	1,830	10/26/2004	Fresno	1	3	3	458
Lake	171395 Sutter Lakeside Hospital		3 Sutter	0 Low	299	10/19/2005	Santa Rosa	2	2	2	75
Lassen	180919 Lassen Community Hospital	TRNNONE	1 UNP	0 Low	260	01/12/2005	Shasta	0	0	0	260
Los Angeles	190034 Antelope Valley Hospital		5 District	2 High	5,082	11/18/2004	Ventura	2	3	2	1,016
Los Angeles	190053 St. Mary Medical Center		4 Church	1 Med High	2,473	12/01/2004	Long Beach	1	3	3	618
Los Angeles	190066 Bellflower Medical Center		7 Investor	0 Low	1,047	12/01/2004	Long Beach	2	3	3	209
Los Angeles	190081 Beverly Hospital	TRNNONE	1 UNP	0 Low	1,850	11/10/2004	Riverside	0	0	0	1,850
Los Angeles	190125 California Hospital Medical Cent	TRNNONE	4 Church	2 High	4,863	11/10/2004	Riverside	0	0	0	4,863
Los Angeles	190148 Centinela Hospital Medical Cent		7 Investor	1 Med High	4,013	12/01/2004	Long Beach	2	3	2	803
Los Angeles	190198 Los Angeles Community Hospita		7 Investor	0 Low	380	12/01/2004	Long Beach	1	0	0	380
Los Angeles	190200 San Gabriel Valley Medical Cent		4 Church	1 Med High	2,051	11/15/2005	Pasadena	1	3	2	513
Los Angeles	190243 Downey Regional Medical Cente		1 UNP	1 Med High	1,455	12/01/2004	Long Beach	2	3	2	291
Los Angeles	190256 East Los Angeles Doctors Hosp		7 Investor	0 Low	642	12/01/2004	Long Beach	1	2	2	214
Los Angeles	190298 Foothill Presbyterian Hospital-Jo		7 Investor	0 Low	778	12/01/2004	Long Beach	1	3	3	195
Los Angeles	190307 Pacific Alliance Medical Center,		7 Investor	0 Low	2,265	11/15/2005	Pasadena	2	3	2	453
Los Angeles	190315 Garfield Medical Center		7 Investor	2 High	3,808	11/10/2004	Riverside	1	3	3	952
Los Angeles	190323 Glendale Adventist Medical Cent		4 Church	1 Med High	2,296	11/18/2004	Ventura	1	2	1	765
Los Angeles	190328 East Valley Hospital Medical Ce		7 Investor	0 Low	449	12/01/2004	Long Beach	2	3	2	90
Los Angeles	190352 Greater El Monte Community Ho		7 Investor	0 Low	1,085	12/01/2004	Long Beach	1	3	3	271
Los Angeles	190382 Hollywood Presbyterian Medical	TRNLATE	4 Church	2 High	4,428	07/14/2009	Pasadena	2	0	0	2,214
Los Angeles	190385 Providence Holy Cross Medical (		4 Church	1 Med High	2,312	11/15/2005	Pasadena	3	3	2	385
Los Angeles	190392 Good Samaritan Hospital-Los A		4 Church	2 High	3,283	11/15/2005	Pasadena	1	3	3	821

County	OSHPD	Identifier	REMRESN	Auspice	Volume	Births	TrnDate0	City	People0	PeopleP	TrainP	Ratio
Los Angeles	190400	Huntington Memorial Hospital		1 UNP	2 High	3,505	12/01/2004	Long Beach	1	3	3	876
Los Angeles	190422	Torrance Memorial Medical Cen		1 UNP	2 High	3,903	12/01/2004	Long Beach	2	3	3	781
Los Angeles	190429	Kaiser Fnd Hosp - Sunset		2 Kaiser	2 High	2,242	12/01/2004	Long Beach	1	2	2	747
Los Angeles	190430	Kaiser Fnd Hosp - Bellflower		2 Kaiser	2 High	3,334	12/01/2004	Long Beach	2	2	1	834
Los Angeles	190431	Kaiser Fnd Hosp - Harbor City		2 Kaiser	1 Med High	1,774	12/01/2004	Long Beach	2	2	1	444
Los Angeles	190432	Kaiser Fnd Hosp - Panorama City		2 Kaiser	1 Med High	1,629	11/15/2005	Pasadena	1	3	2	407
Los Angeles	190434	Kaiser Fnd Hosp - West La		2 Kaiser	0 Low	1,429	12/01/2004	Long Beach	3	3	3	238
Los Angeles	190470	Providence Little Company Of M		4 Church	1 Med High	2,832	12/01/2004	Long Beach	2	3	2	566
Los Angeles	190517	Providence Tarzana Medical Ce	TRNLATE	4 Church	1 Med High	3,531	08/20/2008	Pasadena	1	0	0	3,531
Los Angeles	190521	Memorial Hospital Of Gardena		7 Investor	0 Low	848	12/01/2004	Long Beach	1	3	3	212
Los Angeles	190522	Glendale Memorial Hospital And	TRNLATE	4 Church	1 Med High	1,992	07/14/2009	Pasadena	1	0	0	1,992
Los Angeles	190525	Long Beach Memorial Medical C		1 UNP	2 High	6,059	12/01/2004	Long Beach	3	3	3	1,010
Los Angeles	190529	Methodist Hospital Of Southern		4 Church	1 Med High	2,117	11/15/2005	Pasadena	2	3	2	423
Los Angeles	190547	Monterey Park Hospital		7 Investor	0 Low	2,099	11/18/2004	Ventura	2	2	2	525
Los Angeles	190555	Cedars Sinai Medical Center		1 UNP	2 High	7,285	12/01/2004	Long Beach	2	2	1	1,821
Los Angeles	190568	Northridge Hospital Medical Cen	TRNLATE	4 Church	2 High	2,511	07/14/2009	Pasadena	3	0	0	837
Los Angeles	190587	Pacific Hospital Of Long Beach		7 Investor	0 Low	1,259	12/01/2004	Long Beach	3	3	3	210
Los Angeles	190630	Pomona Valley Hospital Medical	TRNNONE	1 UNP	2 High	6,697	11/10/2004	Riverside	0	0	0	6,697
Los Angeles	190631	Presbyterian Intercommunity Ho	TRNLATE	1 UNP	2 High	3,549	07/25/2006	Pasadena	2	0	0	1,775
Los Angeles	190636	Citrus Valley Medical Center - Q		1 UNP	2 High	5,404	11/15/2005	Pasadena	2	3	3	1,081
Los Angeles	190673	San Dimas Community Hospital	TRNLATE	7 Investor	0 Low	742	07/14/2009	Pasadena	1	0	0	742
Los Angeles	190680	Providence Little Company Of M		4 Church	0 Low	765	12/01/2004	Long Beach	3	3	1	128
Los Angeles	190687	Santa Monica - Ucla Medical Ce		6 Public	0 Low	1,960	12/01/2004	Long Beach	1	2	2	653
Los Angeles	190696	Pacifica Hospital Of The Valley	TRNLATE	7 Investor	0 Low	1,088	08/20/2008	Pasadena	1	0	0	1,088
Los Angeles	190754	St. Francis Medical Center		4 Church	2 High	6,446	12/01/2004	Long Beach	1	0	0	6,446
Los Angeles	190756	St. John's Health Center		4 Church	0 Low	1,396	12/01/2004	Long Beach	2	3	3	279
Los Angeles	190758	Providence Saint Joseph Medica		4 Church	1 Med High	2,320	12/01/2004	Long Beach	1	3	3	580
Los Angeles	190796	Ronald Reagan Ucla Medical Ce		6 Public	1 Med High	1,951	12/01/2004	Long Beach	2	3	3	390
Los Angeles	190812	Valley Presbyterian Hospital		4 Church	2 High	3,664	12/01/2004	Long Beach	1	3	3	916
Los Angeles	190818	Verdugo Hills Hospital		1 UNP	0 Low	961	11/15/2005	Pasadena	1	3	3	240
Los Angeles	190854	Los Angeles Metropolitan Medic		7 Investor	0 Low	1,166	12/01/2004	Long Beach	1	0	0	1,166
Los Angeles	190859	West Hills Hospital And Medical	TRNLATE	7 Investor	0 Low	1,297	07/10/2008	Long Beach	1	2	1	432
Los Angeles	190878	White Memorial Medical Center	TRNLATE	4 Church	2 High	3,476	08/20/2008	Pasadena	1	3	2	869
Los Angeles	190883	Whittier Hospital Medical Center		7 Investor	0 Low	2,151	12/01/2004	Long Beach	2	1	1	717
Los Angeles	190949	Henry Mayo Newhall Memorial H		1 UNP	0 Low	1,301	11/18/2004	Ventura	1	3	3	325
Los Angeles	191227	Lac/Harbor-Ucla Medical Center		6 Public	1 Med High	1,008	12/01/2004	Long Beach	2	3	3	202
Los Angeles	191228	Lac+usc Medical Center		6 Public	1 Med High	1,438	12/01/2004	Long Beach	1	0	0	1,438
Los Angeles	191231	Los Angeles County Olive View-		6 Public	1 Med High	856	12/01/2004	Long Beach	1	3	3	214
Los Angeles	191450	Kaiser Fnd Hosp - Woodland Hil	TRNNONE	2 Kaiser	1 Med High	1,718	11/10/2004	Riverside	0	0	0	1,718
Los Angeles	196035	Kaiser Fnd Hosp - Baldwin Park		2 Kaiser	1 Med High	3,121	12/01/2004	Long Beach	1	3	3	780
Madera	201281	Madera Community Hospital		1 UNP	0 Low	1,626	10/26/2004	Fresno	1	3	3	407
Marin	211006	Marin General Hospital		3 Sutter	1 Med High	1,811	11/03/2004	Bay Area	1	0	0	1,811
Mendocino	231013	Mendocino Coast District Hospit		5 District	0 Low	183	03/02/2005	Humboldt	1	2	1	61
Mendocino	231396	Ukiah Valley Medical Center/Hos		4 Church	0 Low	821	03/02/2005	Humboldt	1	0	0	821
Merced	240924	Memorial Hospital Los Banos		3 Sutter	0 Low	697	10/26/2004	Fresno	1	3	3	174
Merced	240942	Mercy Medical Center Merced-(		4 Church	0 Low	2,663	05/10/2005	Fresno	1	3	3	666
Monterey	270744	Community Hospital Monterey P		1 UNP	1 Med High	1,226	12/16/2004	Santa Barbara	1	3	3	307
Monterey	270777	George L Mee Memorial Hospita		1 UNP	0 Low	582	08/31/2005	Santa Cruz	2	1	1	194
Monterey	270875	Salinas Valley Memorial Hospita		5 District	1 Med High	1,975	08/31/2005	Santa Cruz	2	3	3	395
Monterey	274043	Natividad Medical Center		6 Public	1 Med High	3,008	08/31/2005	Santa Cruz	1	3	1	752
Napa	281047	Queen Of The Valley Hospital - I		4 Church	0 Low	914	11/03/2004	Bay Area	3	3	3	152
Napa	281078	St. Helena Hospital		4 Church	0 Low	338	10/19/2005	Santa Rosa	2	1	1	113
Nevada	291023	Sierra Nevada Memorial Hospita		4 Church	0 Low	469	12/08/2004	Sacramento	1	3	3	117
Nevada	291053	Tahoe Forest Hospital		5 District	0 Low	449	09/14/2005	Stockton	1	3	3	112
Orange	300225	Orange Coast Memorial Medica		1 UNP	0 Low	1,501	02/28/2005	Santa Ana	1	3	3	375
Orange	301098	Anaheim Memorial Medical Cent		7 Investor	1 Med High	1,798	02/28/2005	Santa Ana	1	1	1	899
Orange	301132	Kaiser Fnd Hosp - Anaheim		2 Kaiser	1 Med High	3,129	02/28/2005	Santa Ana	1	3	2	782
Orange	301175	Fountain Valley Rgnl Hosp And M		7 Investor	2 High	3,608	02/28/2005	Santa Ana	1	1	1	1,804
Orange	301188	Western Medical Center Hospita	TRNNONE	7 Investor	0 Low	2,921	11/18/2004	Ventura	0	0	0	2,921
Orange	301205	Hoag Memorial Hospital Presbyt		4 Church	2 High	4,665	02/28/2005	Santa Ana	1	3	3	1,166
Orange	301234	La Palma Intercommunity Hospit		7 Investor	0 Low	679	02/28/2005	Santa Ana	1	3	3	170
Orange	301248	Los Alamitos Medical Center		7 Investor	0 Low	1,609	02/28/2005	Santa Ana	1	3	3	402
Orange	301258	Coastal Communities Hospital		7 Investor	0 Low	2,742	02/28/2005	Santa Ana	1	3	3	686
Orange	301262	Mission Hospital Regional Medic		4 Church	1 Med High	3,138	02/28/2005	Santa Ana	1	3	3	785
Orange	301279	University Of California Irvine Me		6 Public	1 Med High	1,505	02/28/2005	Santa Ana	1	3	3	376
Orange	301283	Garden Grove Hospital And Mec		7 Investor	1 Med High	3,139	02/28/2005	Santa Ana	1	3	3	785
Orange	301317	Saddleback Memorial Medical C		1 UNP	2 High	3,278	12/01/2004	Long Beach	1	3	3	820
Orange	301340	St. Joseph Hospital - Orange		4 Church	2 High	4,803	02/28/2005	Santa Ana	1	3	2	1,201
Orange	301342	St. Jude Medical Center		4 Church	1 Med High	2,273	11/18/2004	Ventura	2	3	3	455
Orange	301566	Western Medical Center - Santa		7 Investor	1 Med High	3,559	02/28/2005	Santa Ana	1	3	3	890

County	OSHPD	Identifier	REMRESN	Auspice	Volume	Births	TrnDate0	City	People0	PeopleP	TrainP	Ratio
Placer	310791	Sutter Auburn Faith Hospital		3 Sutter	0 Low	467	09/28/2005	Shasta	1	2	2	156
Placer	311000	Sutter Roseville Medical Center		3 Sutter	1 Med High	2,679	12/08/2004	Sacramento	2	3	3	536
Riverside	331152	Corona Regional Medical Center		7 Investor	0 Low	1,920	11/10/2004	Riverside	1	3	1	480
Riverside	331164	Desert Regional Medical Center		7 Investor	2 High	3,300	11/10/2004	Riverside	3	3	1	550
Riverside	331194	Hemet Valley Medical Center		5 District	0 Low	1,062	11/10/2004	Riverside	1	0	0	1,062
Riverside	331216	John F Kennedy Memorial Hospi		7 Investor	0 Low	3,123	11/10/2004	Riverside	2	1	1	1,041
Riverside	331293	Parkview Community Hospital M	TRNLATE	1 UNP	1 Med High	1,773	04/21/2009	Riverside	2	0	0	887
Riverside	331312	Riverside Community Hospital		7 Investor	2 High	3,311	11/10/2004	Riverside	3	3	2	552
Riverside	331326	San Gorgonio Memorial Hospital		5 District	0 Low	374	11/10/2004	Riverside	2	3	3	75
Riverside	334025	Kaiser Fnd Hosp - Riverside		2 Kaiser	1 Med High	3,196	11/10/2004	Riverside	2	0	0	1,598
Riverside	334048	Kaiser Fnd Hospital - Moreno Va		2 Kaiser	0 Low	1,342	11/10/2004	Riverside	1	3	2	336
Riverside	334068	Southwest Healthcare System-M		7 Investor	1 Med High	3,446	11/10/2004	Riverside	3	2	2	689
Riverside	334487	Riverside County Regional Medi		6 Public	2 High	2,809	11/10/2004	Riverside	1	3	3	702
Sacramento	340913	Kaiser Fnd Hosp - Sacramento/I		2 Kaiser	2 High	3,812	12/08/2004	Sacramento	1	3	2	953
Sacramento	340947	Mercy General Hospital		4 Church	1 Med High	2,783	12/08/2004	Sacramento	2	3	3	557
Sacramento	340950	Mercy San Juan Hospital		4 Church	2 High	3,032	12/08/2004	Sacramento	2	3	3	606
Sacramento	340951	Methodist Hospital Of Sacramer		4 Church	0 Low	1,049	12/08/2004	Sacramento	2	3	3	210
Sacramento	341006	University Of California Davis Me		6 Public	2 High	2,648	12/08/2004	Sacramento	2	3	2	530
Sacramento	341051	Sutter General Hospital		3 Sutter	2 High	5,295	12/08/2004	Sacramento	1	1	1	2,648
Sacramento	342344	Kaiser Fnd Hosp - South Sacran		2 Kaiser	1 Med High	3,499	12/08/2004	Sacramento	2	3	3	700
Sacramento	344029	Mercy Hospital - Folsom		4 Church	0 Low	1,187	12/08/2004	Sacramento	2	3	3	237
San Benito	350784	Hazel Hawkins Memorial Hospita		5 District	0 Low	584	12/16/2004	Santa Barbara	1	3	3	146
San Bernardino	361105	Barstow Community Hospital		7 Investor	0 Low	290	11/10/2004	Riverside	2	0	0	145
San Bernardino	361166	Montclair Hospital Medical Cente		7 Investor	0 Low	1,053	11/10/2004	Riverside	3	1	1	263
San Bernardino	361223	Kaiser Fnd Hosp - Fontana		2 Kaiser	1 Med High	3,967	11/10/2004	Riverside	1	3	3	992
San Bernardino	361246	Loma Linda University Medical C		4 Church	2 High	2,551	12/08/2005	San Bernardino	2	3	3	510
San Bernardino	361308	Redlands Community Hospital		1 UNP	1 Med High	2,282	11/10/2004	Riverside	1	3	2	571
San Bernardino	361318	San Antonio Community Hospita		1 UNP	1 Med High	2,273	12/08/2005	San Bernardino	2	3	2	455
San Bernardino	361323	Community Hospital Of San Ber		4 Church	1 Med High	2,541	11/10/2004	Riverside	2	3	3	508
San Bernardino	361339	St. Bernardine Medical Center		4 Church	1 Med High	2,624	12/08/2005	San Bernardino	2	3	1	525
San Bernardino	361343	St. Mary Regional Medical Cent		4 Church	1 Med High	2,111	11/10/2004	Riverside	1	0	0	2,111
San Bernardino	361370	Victor Valley Community Hospita	TRNNONE	1 UNP	0 Low	1,255	11/10/2004	Riverside	0	0	0	1,255
San Bernardino	364144	Desert Valley Hospital		7 Investor	0 Low	447	11/10/2004	Riverside	2	3	3	89
San Bernardino	364231	Arrowhead Regional Medical Ce		6 Public	2 High	3,307	11/10/2004	Riverside	3	3	2	551
San Diego	370694	Sharp Memorial Hospital		1 UNP	2 High	7,755	02/02/2005	San Diego	2	3	3	1,551
San Diego	370705	Fallbrook Hospital District		5 District	0 Low	525	10/12/2005	San Diego	1	0	0	525
San Diego	370714	Grossmont Hospital	TRNLATE	5 District	2 High	3,487	04/09/2008	San Diego	2	0	0	1,744
San Diego	370730	Kaiser Fnd Hosp - San Diego		2 Kaiser	2 High	3,951	02/02/2005	San Diego	3	2	1	790
San Diego	370744	Scripps Mercy Hospital		4 Church	2 High	4,391	02/02/2005	San Diego	2	3	3	878
San Diego	370755	Palomar Medical Center		5 District	2 High	4,279	02/02/2005	San Diego	1	3	3	1,070
San Diego	370759	Paradise Valley Hospital		7 Investor	0 Low	1,946	02/02/2005	San Diego	1	3	2	487
San Diego	370771	Scripps Memorial Hospital - La J		1 UNP	1 Med High	3,993	02/02/2005	San Diego	1	3	3	998
San Diego	370780	Tri-City Medical Center		5 District	2 High	3,352	02/02/2005	San Diego	2	3	3	670
San Diego	370782	University Of Calif-San Diego Me		6 Public	2 High	2,865	02/02/2005	San Diego	3	3	3	478
San Diego	370875	Sharp Chula Vista Medical Cente		1 UNP	1 Med High	2,724	02/02/2005	San Diego	3	3	3	454
San Diego	370977	Pomerado Hospital		5 District	0 Low	1,212	02/02/2005	San Diego	2	1	1	404
San Diego	371394	Scripps Memorial Hospital - Enc		1 UNP	0 Low	1,593	10/12/2005	San Diego	2	1	1	531
San Francisco	380857	Kaiser Fnd Hosp - Geary S F		2 Kaiser	1 Med High	2,443	01/26/2005	Oakland Kaiser	1	2	1	814
San Francisco	380929	California Pacific Med Ctr-Pacifi		3 Sutter	2 High	6,026	06/02/2005	Alameda	1	2	2	2,009
San Francisco	380939	San Francisco General Hospital		6 Public	1 Med High	1,237	11/03/2004	Bay Area	2	3	3	247
San Francisco	380964	St. Luke's Hospital	TRNNONE	3 Sutter	0 Low	1,058	11/03/2004	Bay Area	0	0	0	1,058
San Francisco	381154	Ucsf Medical Center		6 Public	1 Med High	1,856	10/19/2005	Santa Rosa	2	1	1	619
San Joaquin	390846	Dameron Hospital		1 UNP	1 Med High	2,547	09/14/2005	Stockton	3	3	3	425
San Joaquin	390923	Lodi Memorial Hospital		1 UNP	0 Low	1,312	11/03/2004	Bay Area	2	3	3	262
San Joaquin	391010	San Joaquin General Hospital		6 Public	1 Med High	2,496	11/03/2004	Bay Area	2	3	3	499
San Joaquin	391042	St. Joseph's Medical Center Of	TRNLATE	4 Church	2 High	2,064	11/17/2006	Sacramento	1	1	1	1,032
San Joaquin	391056	Sutter Tracy Community Hospital		3 Sutter	0 Low	704	11/03/2004	Bay Area	1	3	3	176
San Joaquin	392287	Doctors Hospital Of Manteca		7 Investor	0 Low	671	11/03/2004	Bay Area	1	3	3	168
San Luis Obispo	400480	French Hospital Medical Center		4 Church	0 Low	769	12/15/2005	Santa Maria	1	2	2	256
San Luis Obispo	400524	Sierra Vista Regional Medical C	TRNLATE	7 Investor	0 Low	1,195	10/08/2008	Ventura	1	0	0	1,195
San Luis Obispo	400548	Twin Cities Community Hospital		7 Investor	0 Low	711	12/16/2004	Santa Barbara	3	3	2	119

County	OSHPD	Identifier	REMRESN	Auspice	Volume	Births	TrnDate0	City	People0	PeopleP	TrainP	Ratio
Santa Clara	430705	Regional Medical Of San Jose		7 Investor	1 Med High	2,307	11/03/2004	Bay Area	1	1	1	1,154
Santa Clara	430763	El Camino Hospital		5 District	2 High	5,548	11/03/2004	Bay Area	2	3	3	1,110
Santa Clara	430779	Good Samaritan Hospital-San Jo		7 Investor	2 High	3,122	11/03/2004	Bay Area	1	3	2	781
Santa Clara	430837	O'connor Hospital - San Jose		4 Church	1 Med High	2,923	11/03/2004	Bay Area	2	2	1	731
Santa Clara	430883	Santa Clara Valley Medical Cent		6 Public	2 High	5,129	11/03/2004	Bay Area	2	3	2	1,026
Santa Clara	431506	Kaiser Fnd Hosp - San Jose		2 Kaiser	1 Med High	2,497	11/03/2004	Bay Area	2	3	3	499
Santa Clara	434040	Lucile Salter Packard Children's		1 UNP	2 High	5,077	11/03/2004	Bay Area	3	3	3	846
Santa Clara	434138	St. Louise Regional Hospital	TRNNONE	4 Church	0 Low	679	11/03/2004	Bay Area	0	0	0	679
Santa Clara	434153	Kaiser Fnd Hosp - Santa Clara		2 Kaiser	2 High	3,357	01/26/2005	Oakland Kaiser	2	3	2	671
Santa Cruz	440755	Dominican Hospital-Santa Cruz/S		4 Church	1 Med High	1,123	08/31/2005	Santa Cruz	1	3	1	281
Santa Cruz	444012	Sutter Maternity And Surgery Ce		3 Sutter	0 Low	835	08/31/2005	Santa Cruz	1	0	0	835
Santa Cruz	444013	Watsonville Community Hospital	TRNLATE	7 Investor	0 Low	1,759	05/15/2008	Fresno	2	0	0	880
Shasta	450949	Mercy Medical Center		4 Church	1 Med High	2,001	01/12/2005	Shasta	2	0	0	1,001
Siskiyou	470871	Mercy Medical Center Mt. Shast		4 Church	0 Low	125	01/12/2005	Shasta	1	2	1	42
Siskiyou	474007	Fairchild Medical Center		1 UNP	0 Low	211	01/12/2005	Shasta	1	3	3	53
Solano	480989	Kaiser Fnd Hosp - Rehabilitation		2 Kaiser	1 Med High	2,255	11/03/2004	Bay Area	3	3	3	376
Solano	481094	Sutter Solano Medical Center		3 Sutter	0 Low	930	11/03/2004	Bay Area	1	3	3	233
Solano	481357	North Bay Medical Center		1 UNP	0 Low	1,552	11/03/2004	Bay Area	1	3	3	388
Sonoma	490919	Sutter Medical Center Of Santa		3 Sutter	0 Low	1,937	11/03/2004	Bay Area	1	2	1	646
Sonoma	491001	Petaluma Valley Hospital		4 Church	0 Low	523	11/03/2004	Bay Area	1	3	2	131
Sonoma	491064	Santa Rosa Memorial Hospital-M		4 Church	1 Med High	1,391	11/03/2004	Bay Area	1	2	2	464
Sonoma	491076	Sonoma Valley Hospital		5 District	0 Low	241	11/03/2004	Bay Area	1	3	2	60
Sonoma	494019	Kaiser Fnd Hosp - Santa Rosa		2 Kaiser	0 Low	1,770	11/03/2004	Bay Area	2	1	1	590
Stanislaus	500852	Doctors Medical Center		7 Investor	2 High	4,557	10/26/2004	Fresno	1	3	2	1,139
Stanislaus	500867	Emanuel Medical Center, Inc		4 Church	1 Med High	2,371	10/26/2004	Fresno	2	3	3	474
Stanislaus	500939	Memorial Hospital Medical Cent(	TRNNONE	3 Sutter	1 Med High	1,779	10/26/2004	Fresno	0	0	0	1,779
Stanislaus	500967	Oak Valley District Hospital (2-R		5 District	0 Low	359	10/26/2004	Fresno	1	3	1	90
Tehama	521041	St. Elizabeth Community Hospital		4 Church	0 Low	653	01/12/2005	Shasta	1	3	2	163
Tulare	540734	Kaweah Delta Medical Center		5 District	2 High	4,118	10/26/2004	Fresno	2	3	3	824
Tulare	540798	Sierra View District Hospital		5 District	0 Low	1,976	10/26/2004	Fresno	1	3	3	494
Tulare	540816	Tulare District Hospital		5 District	0 Low	1,010	10/26/2004	Fresno	2	1	1	337
Tuolumne	554011	Sonora Regional Medical Center		4 Church	0 Low	511	12/08/2004	Sacramento	1	2	2	170
Ventura	560473	Community Memorial Hospital-S		1 UNP	2 High	3,386	11/18/2004	Ventura	1	3	3	847
Ventura	560481	Ventura County Medical Center		6 Public	1 Med High	2,130	11/18/2004	Ventura	2	3	3	426
Ventura	560492	Los Robles Hospital & Medical C		7 Investor	1 Med High	1,968	11/18/2004	Ventura	2	3	3	394
Ventura	560508	St. John's Pleasant Valley Hospi		4 Church	0 Low	522	11/18/2004	Ventura	1	3	3	131
Ventura	560525	Simi Valley Hospital And Health		4 Church	0 Low	620	11/18/2004	Ventura	1	3	3	155
Ventura	560529	St. John's Regional Medical Cen		4 Church	1 Med High	2,377	11/18/2004	Ventura	2	3	3	475
Yolo	571086	Woodland Memorial Hospital		4 Church	0 Low	673	12/08/2004	Sacramento	3	1	1	168
Yolo	574010	Sutter Davis Hospital		3 Sutter	0 Low	1,205	12/08/2004	Sacramento	2	3	3	241

APPENDIX F: DEPENDENT VARIABLES PREPARED FOR ANALYSIS

Demographic Set. This set of variables is core to calculating public health statistics related to birth outcomes. The CDC/NCHS has published detailed step-by-step instructions for editing demographic variables for data quality, preparatory to imputing values to improve the reliability of public health statistics derived from demographics [26,30]. These variables are fundamental to assessing and tracking longitudinal changes in disparities based on age and/or race/ethnicity. Variable names in the demographic dataset were prefixed to distinguish between mother and father, respectively with M or MOM or F or DAD. Table 2 identifies the demographic variables.

Table 19: Demographic data quality variables for mother and father

Variable Name	Definition
FATHERQ	Paternity not established as evidenced by the field for father's last name fully missing or coded Unknown, withheld, none, XX, YY, ZZ. Rates for father characteristics are based on cases with established paternity.
RACEQ	Race coded other/unknown/missing
HISPQ	Ethnicity coded unknown/missing
RACETHQ	Race and/or Hispanic ethnicity coded other/unknown/missing
BTHDATECQY	Year of birth missing or out-of-range
BTHDATECQM	Month of birth missing or out-of-range
BTHDATECQD	Day of birth missing our out-of-range
DOBQ	Any data quality problem with Mother/Father birthdate
AGEQ	Mismatch between age calculated from birthdate and reported age
MOM(DAD)S	Sum of the above data quality problems for Mother/Father
MOM(DAD)A	At least one of the above data quality problems for Mother/Father

NCHS calculates data quality and various birth statistics based on father characteristics where the mother is married [31]. California asks mothers if they are married but does not provide the variable in the BSMF. However, if the father's name is in the file, the father claimed paternity, either through marriage or by signing a Declaration of Paternity (POP form) [50].

Using presence of father's last name to indicate paternity acknowledgement [51], we calculated father's statistics. For the set of father variables, rates were calculated using a variable FATHER as the denominator. This was calculated by subtracting FATHERQ from TOTALBC.

Rates for the set of maternal demographic variables and the variable FATHERQ were calculated using total births (TOTALBC) as the denominator. Rates are expressed as percents, for example, MOMAR = MOMA/TOTALBC * 100.

The BDQT repeatedly addressed the importance of obtaining the POP form. They did not address data quality for father's demographics other than race, and never addressed parents' birthdates or age.

Mother's geographic location. Public health indicators are based on geography, for example, where the mother lives. Uncertainty in underlying geography variables increases the error rate when calculating public health indicators for geographic regions.

The FIPS place variable NCHS uses for the annual data quality report has not been covered in the training program. California uses non-standard FIPS place codes that are not on the NCHS FIPS code list [51]. Thus we did not look at FIPS place of residence for this analysis. Instead, we looked at mismatches between city, ZIP, and county, and the presence of a street address for mother's residence. We focused on these variables because of their importance to geocoding, and because of problems identified for these variables in our analysis of geographic data quality in death certificates [52].

ZIP was classified as an error when it was missing, OOR, or not in the correct county. For example, ZIP 94904 is in Marin County, not Mendocino County.

We had ZIPs for other states when the mother was reported to be a California resident. We also found California ZIPS and cities for mothers recorded as residents of other counties and states. For example, Berkeley and Cleveland are not in Los Angeles County, and San Francisco is not in Utah. Finally, we also flagged as an error when mother's street address was missing or unknown. We did not address other typographic errors in street address that make geocoding so challenging.

In identifying city, ZIP, and county errors, we were assisted by a library of SAS formats derived from a master geography file that we have maintained for many years [53]. It is built upon summaries of geographic data from all Vital Statistics and OSHPD files, augmented by three types of commercial files: (1) tracking ZIPs that are newly created, discontinued, or split from larger zips [54], (2) assignment of existing ZIPs to cities and counties [55], with both preferred and allowed cities identified, and (3) ArcGIS data provided with that software package.

Table 20: Data quality variables for mother's residence

Variable Name	Definition
ZIPQ	Mismatch between county and ZIP of residence
CITYQ	Mismatch between county and city of residence
MADDREQ	Address fully missing or coded unknown, withheld, none, XX, YY, ZZ
GEOGS	Sum of the above variables
GEOGA	At least one of the above variables

Rates for the set of geographic variables were calculated using total births (TOTALBC) as the denominator. Rates are expressed as percents, for example, GEOGSR = GEOGS/TOTALBC * 100. Note than when we are calculating rates for sum variables (e.g., GEOGS), the result is not a rate. It is more appropriately described as the number of errors per 100 records. This contrasts with GEOGAR which is the percent of records with at least one error.

Mother's pregnancy history. This set of variables includes the total children born alive, total children born, previous pregnancy terminations and last live birth. Training did not cover any variables in this set, probably because few public health indicators are reported directly for these variables. However, this set is so fundamental to calculating other indicators (such as inter-pregnancy interval) that CDC/NCHS developed decision rules to identify and correct errors, in order to publish better statistics for indicators that rely on them. For example, the interval between the last live birth and the current birth is not calculated if the total number of children born is unknown (she must have had at least one prior birth).

Table 21: Data quality variables for mother's pregnancy history

Variable Name	Definition
LBNOWLQ	Live born children now living not in range 0-30
LBNOWDQ	Live born children now dead not in range 0-30
TERMLT20Q	Terminations less than 20 weeks gestation not in range 0-30
TERMGE20Q	Terminations 20 weeks or more gestation not in range 0-30
TCHILDBAQ	Total children born alive not in range 0-39 or unrealistic given mom's age
TCHILDBQ	Total children ever born not in range 0-40 or unrealistic given mom's age
BHISS	Sum of the above variables
BHISA	At least one of the above variables
LLBNOMO	Month of last live birth missing
LLBNOYR	Year of last live birth missing
LLBNEGINT	Negative interval between date of last live birth and current birth
LLBSHORT	Singleton and interval since last live birth shorter than 4 months,
LLBIMPAGE	Biologically implausible age of mother at previous live birth
LLBINTA	Any of the above LLB variables. Note that these are mutually exclusive and no sum variable is calculated because ANY equals SUM.

Rates for the set of birth history variables LBNOWLQ through BHISA were calculated using total births (TOTALBC) as the denominator. The denominator for LLB* quality rates (LLBDEN) is total births excluding records with no previous live births and records with birth order greater than 1. That is, LLB error rates are calculated only for records with a singleton birth and women who had a previous live birth, after correcting the birth history variables.

Prenatal care variables. We evaluated two prenatal care variables that are critical in and of themselves, and together with gestational age in weeks are used to calculate the Kotelchuck Index for Adequacy of Prenatal Care Utilization [56]. These variables are the month prenatal care began and the number of prenatal care visits, in Table 5.

Table 22: Data quality variables for prenatal care

Variable Name	Definition
MOINCQ	Month prenatal care missing
NPREXAMQ	Number of prenatal care visits missing

Infant variables. Over the entire study period, birth order and birth type had only two errors. Infant's birthdate was always present. Table 6 identifies the infant outcome variables we evaluated.

Table 23: Data quality variables for infant outcomes

Variable Name	Definition
GAGEQ	Gestational age outside the range 17-47 weeks
BWTQ	Birth weight outside the range 227-8165 grams
APGAR01Q	Apgar 1 minute test not in range 0-10 (2007 forward)
APGAR05Q	Apgar 5 minute test not in range 0-10 (2007 forward)
APGAR10Q	Apgar 10 minute test not in range 0-10 (2007 forward)
HEARSCRNQ	Hearing screening is missing or unknown (2007 forward)
TESTS	Sum of data quality problems for newborn tests (2007 forward)
TESTA	Any data quality problems for newborn tests (2007 forward)

CDC/NCHS processes birth weight and gestational in other ways than the stated range criteria. In some situations, inconsistent values may be changed to unknown. The CDC/NCHS algorithm then imputes values to both variables to improve calculation of rates based on birth weight and/or gestational age. We did not extend this analysis to these additional errors types.

ENDNOTES

1 Graham JW (2009) Missing data analysis: Making it work in the real world. Annu. Rev. Psychol. 2009. 60:549–576.

2 Milliken GA, Johnson DE (1984). Analysis of Messy Data. Volume 1: Designed Experiments. New York: Von Nostrand Reinhold Company.

3 Milliken GA, Johnson DE (1989). Analysis of Messy Data. Volume 2: Nonreplicated Experiments. New York: Von Nostrand Reinhold Company.

4 Germain JM, Czernichow P, Josset V, Torre JP, Marpeau L, Le Meur H, et al. [Quality of data acceptable for perinatal epidemiology surveillance: assessment of the health certificate at birth and the national obstetrics medical file. Study in three Seine-Maritime maternal wards]. J Gynecol Obstet Biol Reprod (Paris) 1998;27(4):384-8.

5 Adegboye AR, Heitmann B. Accuracy and correlates of maternal recall of birthweight and gestational age. BJOG 2008;115(7):886-93.

6 DiGiuseppe DL, Aron DC, Ranbom L, Harper DL, Rosenthal GE. Reliability of birth certificate data: a multi-hospital comparison to medical records information. Matern Child Health J 2002;6(3):169-79.

7 Northam S, Knapp TR. The reliability and validity of birth certificates. J Obstet Gynecol Neonatal Nurs 2006;35(1):3-12.

8 Brender JD, Suarez L, Langlois PH. Validity of parental work information on the birth certificate. BMC Public Health 2008;8:95.

9 Roohan PJ, Josberger RE, Acar J, Dabir P, Feder HM, Gagliano PJ. Validation of birth certificate data in New York State. J Community Health 2003;28(5):335-46.

10 Headley AJ, Fulcomer MC, Bastardi MM, Im W, Sass MM, Chung K. The use of missing birth record data as a marker for adverse reproductive outcomes: a geocoded analysis of birth record data. J Natl Med Assoc 2006;98(7):1078-88.

11 Martin JA. United States vital statistics and the measurement of gestational age. Paediatr Perinat Epidemiol 2007;21 Suppl 2:13-21.

12 Pearl M, Wier ML, Kharrazi M. Assessing the quality of last menstrual period date on California birth records. Paediatr Perinat Epidemiol 2007;21 Suppl 2:50-61.

13 Wier ML, Pearl M, Kharrazi M. Gestational age estimation on United States livebirth certificates: a historical overview. Paediatr Perinat Epidemiol 2007;21 Suppl 2:4-12.

14 Gould JB, Chavez G, Marks AR, Liu H. Incomplete birth certificates: a risk marker for infant mortality. Am J Public Health 2002;92(1):79-81.

15 Anthony S, van der Pal-de Bruin KM, Graafmans WC, Dorrepaal CA, Borkent-Polet M, van Hemel OJ, et al. The reliability of perinatal and neonatal mortality rates: differential under-reporting in linked professional registers vs. Dutch civil registers. Paediatr Perinat Epidemiol 2001;15(3):306-14.

16 Pedrosa LD, Sarinho SW, Ximenes RA, Ordonha MR. [Quality of data on early neonatal deaths]. Rev Assoc Med Bras 2007;53(5):389-94.

17 Romero DE, Cunha CB. [Evaluation of quality of epidemiological and demographic variables in the Live Births Information System, 2002]. Cad Saude Publica 2007;23(3):701-14.

18 Suthutvoravut S, Hogue CJ, Guyer B, Anderka M, Oberle MW. Are preterm black infants larger than preterm white infants, or are they more misclassified? J Biosoc Sci 1989;21(4):443-51.

19 Natality Not Classifiable and Unknown Data for 2003-2004, Revised States Only. National Center for Health Statistics. Dated 27-Jan-2005.

20 Natality Not Classifiable and Unknown Data for 2007-2009, Revised States Only. National Center for Health Statistics. Dated 19-Jan-2010.

21 Natality Not Classifiable and Unknown Data for 2008-2010, Revised States Only. National Center for Health Statistics. Dated 10-Nov-2011.

23 Campbell DT, Stanley JC. (1963) Experimental and Quasi-Experimental Designs for Research. Houghton-Mifflin Company.

24 Remy L, Clay T. (2012) 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/

25 California Department of Public Health (2010) Hospital Codes 1960-2009. Provided as part of the documentation for the 2009 Birth Statistical Master File.

26 Division of Vital Statistics. Instruction manual part 12: Computer edits for natality data, effective 1993.Vital statistics, data preparation. Hyattsville, Maryland: National Center for Health Statistics. 1995. Last accessed 14-Oct-2011 at: www.cdc.gov/nchs/data/dvs/instr12.pdf.

27 National Center for Health Statistics. Technical appendix from vital statistics of the United States: 1999, Natality. 2001. Hyattsville, Maryland: National Center for Health Statistics. Last accessed 14-Oct-2011 at: www.cdc.gov/nchs/data/techap99.pdf.

28 National Center for Health Statistics. Technical appendix from vital statistics of the United States: 2003, Natality. 2005. Last accessed 14-Oct-2011 at: http://www.cdc.gov/nchs/data/techapp03_1-09.pdf.

29 Emails between Linda Remy and Joyce Martin (CDC/NCHS), Various, September 2011. See CDC_MOM_AGE.PDF.

30 Martin J. (2011) National Vital Statistics System (NVSS) Computer Edits for Natality Data (Revised States). September 23, 2008 (Updated 2/11/2011).

31 Email from Alan Oppenheim (CDPH-CHS) to Linda Remy dated 12-Dec-2011.

32 Healthcare Information Division - Annual Financial Data. Last accessed 24-Jul-012 at: http://www.oshpd.ca.gov/HID/Products/Hospitals/AnnFinanData/CmplteDataSet/index.asp

33 See http://www.usc.edu/schools/price/research/healthresearch/OSHPD.html. Last accessed 16-Jan-2013.

34 Remy L, Clay T, Oliva G (2004) Creating Longitudinal Hospital-Level Datasets. UCSF Family Health Outcomes Project. Last accessed 24-Jul-2012 at: http://fhop.ucsf.edu/fhop/docs/pdf/pubs/METH_HOSP.pdf

35 Remy L, Oliva G, Clay T (2008) Maternal morbidity and outcomes including mortality, California2001-2006. Family Health Outcomes Project, University of California San Francisco.

36 Singer JD, Willett JB (2003) Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. New York: Oxford.

37 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.

38 Office of Management and Budget (2000). 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_guidance2000update.pdf.

39 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.

40 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 19-May-2011 at: http://www.cdc.gov/nchs/data/dvs/Multiple_race_docu_5-10-04.pdf.

41 Johnson DP (2008). Using the Multiple Race Bridging Matrix. Documentation and NCHS race bridging file emailed to Linda Remy on 27-May-2011 from Dave Johnson, Survey Statistician, DHHS/PHS/CDC/CCHIS/NCHS/DVS/SPSRB. CDC/NCHS uses this file to bridge multi-race for all Federal vital statistics publications. Per Dave Johnson, CDC revisited whether to update the file in 2010, determined it did not need to be updated.

42 Cohen J. (1988) Statistical Power Analysis for the Behavioral Sciences. Second edtion. Hillsdale NJ: Lawrence Erlbaum Associates.

43 State of the World's Mothers. (2012). Save the Children, Publisher. Last accessed 24-Sep-2012 at: http://www.savethechildren.org/atf/cf/%7B9def2ebe-10ae-432c-9bd0-df91d2eba74a%7D/STATE-OF-THE-WORLDS-MOTHERS-REPORT-2012-FINAL.PDF.

44 Xu J. (2005) Comp_gest_new_fmt.sas. Effective 2003 for Revising States, Written by JIAQUAN XU, 08-Feb-2005 DVS/NCHS/CDC. Last accessed 17-Jan-2012 at http://www.naphsis.org/index.asp?bid=1312

45 Martin JA (2007) United States vital statistics and measurement of gestational age. Paediatric and Perinatal Epidemiology, 21 (Suppl. 2), 13–21. Last accessed 25-Sep-2012 at: http://www.naphsis.org/NAPHSIS/files/ccLibraryFiles/Filename/000000000968/Martin%20paper.pdf.

46 NAPHSIS (2009) Recommendations on Measuring and Improving Quality of Gestational Age Data. Last accessed 25-Sep-2012 at: http://www.naphsis.org/NAPHSIS/files/ccLibraryFiles/Filename/000000000967/Gestational%20Age%20NAPHSIS%20report.doc.

47 Skinner J, Staiger D, Fisher ES (2010). Looking back, moving forward. NEJM, 362:7, 569:574.

48 Remy LL, Oliva G (2006) Impact of data quality on birth-related health indicators. Last accessed 30-Nov-2012 at: http://fhop.ucsf.edu/fhop/docs/pdf/pubs/dATAQUAL7.pdf.

49 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.

50 Paternity Opportunity Program. Last accessed 09-Apr-2012 at: http://www.childsup.ca.gov/Resources/EstablishPaternity/POPFAQ.aspx

51 Email from Alan Oppenheim (CDPH-CHS) to Linda Remy dated 21-Feb-2012.

52 Remy L. (2009) A geocoding case study: California death certificates 2005 and 2007. UCSF Family Health Outcomes Project.

53 Remy L, Clay T. (2012) 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/

54 Western Economic Research (2011) California Zip-code Change. Last ordered 07-Nov-2011 at: www.wermaps.com.

55 CD Light (2011) ZIPLIST5. Last ordered 07-Nov-2011 at: www.zipinfo.com.

56 Alexander GR, Kotelchuck M (1996) Quantifying the adequacy of prenatal care: a comparison of indices. Public Health Rep. 1996 Sep-Oct; 111(5): 408–419.