GIS and Vital Statistics in California

Linette Scott, MD, MPH

Abstract
Data about California's registered births and deaths are collected and maintained by the Center for Health Statistics, California Department of Health Services (CHS). Come and learn how GIS has been used to enhance the quality, analysis, and presentation of these datasets. With statewide implementation of an Electronic Death Registration System in progress and recent improvements to the statewide Automated Vital Statistics System, the CHS is in a position to begin routinely geocoding California's vital statistics. The addition of longitude and latitude to birth and death records facilitates their use for a multitude of public and preventive health policy and business functions. Participants will leave with a clear understanding of how GIS enhances the quality and utility of California's birth and death datasets, important data privacy considerations in the geographical display of vital record data, and benefits from specific public health uses of geocoded vital record data.

California Center for Health Statistics

The Office of State Registrar was originally created in California in 1858 to collect information regarding marriages, births, divorces and deaths. While original information collected was incomplete, after changes to the Political Code in 1905 the records became very complete. (Coy, 1919) The original Office of State Registrar is now within the Center for Health Statistics. "The mission of the Center for Health Statistics (CHS) is to facilitate the collection, validation, statistical analysis, and dissemination of health data in support of the mission of the California Department of Health Services (CDHS)." (CHS, 2005)

In 1981, the Automated Vital Statistics System (AVSS) was first implemented in California. (Williams, 1996) This system improved timeliness, emphasized accuracy, minimized redundant data entry, and provided feedback to providers of birth data. Since its inception, AVSS has been augmented to register deaths and reportable communicable disease data at local health departments. In 2005, the California Electronic Death Registration System (CA-EDRS) was piloted in two counties. (Biomedical Informatics Research and Consulting Service, 2006) CA-EDRS is a web-based system that allows funeral directors, coroners, physicians, hospitals, and local registration districts to enter and interact with information on death certificates 24 hours per day, 7 days per week. This system leverages new technologies and models to provide improved registration of death

certificates. During 2005, funeral homes, coroners, and local registration districts registered death certificates with CA-EDRS in the pilot counties. In non-pilot counties, death certificates were key data entered at the CHS after the registration process was complete. In 2006, funeral homes, coroners, and local registration districts in additional counties have been implementing CA-EDRS to register death certificates. When death certificates are registered through CA-EDRS within local registration districts, the data undergo more than 1,000 validations or "edits."

AVSS and CA-EDRS create electronic records of births and deaths in database formats, thus the data from these systems are well suited for increased use in geographic information systems (GIS). As the collector/ steward of birth and death data, the CHS is uniquely poised to fulfill its mission to CDHS by leading efforts to improve and increase uses of GIS, and in particular geocoding (with x,y coordinate) records of California's vital events.

Geocoding California's Vital Records

The Center for Health Statistics has long had an interest in GIS and geocoding. When geocoding to street level accuracy was explored in the past, it was found that the key data entry needed to prepare the addresses for geocoding was cost prohibitive. The CHS has invested time and personnel to explore and advocate for public health uses of GIS, thus helping CDHS prepare the way for routine street level geocoding and geoanalysis of vital record and other public health data. While the statewide files of California births and deaths have included geographic codes for nation, state, county, and zip codes, not until recently has the CHS had the resources to geocode street addresses with a latitude and longitude (x,y coordinate).

The CHS recently was able to have several years of birth and death data geocoded to street level accuracy in batch processes. The death records are limited as the address of residence on the death certificate has not been routinely entered electronically prior to CA-EDRS implementation. A CHS partner key entered the addresses for death certificates for 1999-2001, which we have been able to batch geocode. This data has also been used and geocoded by other programs and partners, resulting in redundancy, inconsistent use of geocoding methods, and varying quality of geocode assignments over the years. The body of geographic work that has occurred clearly demonstrates the need for data geocoded to street level.

The CDHS has been working to obtain a centralized geocoding resource as this is a crucial first step toward integrating the use of GIS throughout CDHS programs. The CHS has been supporting these efforts with the objective of routinely geocoding address data on birth and death certificates at the point of data capture into our electronic systems.

GIS Enhances Quality and Utility of California's Vital Statistics

California is a large state, consisting of 58 counties, 61 local health departments, 163,696 square miles, and an estimated 2006 population of over 37 million people. Our counties range in size from San Francisco County with 45 square miles to San Bernardino County with 20,117 square miles and range in population from Alpine County with an estimated 2006 population of 1,241 to Los Angeles County with an estimated 2006 population of 10,245,572. (CA DOF, 2006) Within this context, California has a strong need for subcounty information and analysis to better understand the issues in the local jurisdictions. In particular, there is increased need and interest for neighborhood and community analysis for public health.

California's vital records play an integral role in analysis and policy decisions throughout a wide range of public and private organizations. Birth and death records form a particularly valuable data set in that they represent every person in the population. Also, these records are essential legal documents required by Californians to obtain a wide range of services and benefits. As such, these data sets have a wide constituency of users with many different objectives. The ability to have these records geocoded to street addresses will significantly increase their utility and enhance community and neighborhood analyses. In particular, birth and death data are essential to the development of demographic estimates and projections for regions and communities.

Birth and death records are high quality data sets. Our initial batch geocoding achieved approximately 94% match rate to address or zip+4 accuracy. By moving toward geocoding at point of entry with an address validation process, we anticipate further improvement in the accuracy of addresses on birth and death records. These data sets are also very complete, with more than 500,000 births per year and more than 250,000 deaths per year.

The greatest benefit to the geocoding of birth and death records comes from the increased utility of vital record data. There are many additional benefits of GIS to be realized when GIS is considered as a decision support system that uses geospatially oriented information for problem solving and planning.

Confidentiality Considerations for Geocoded Vital Records

Birth and death files contain large amounts of data for each event, including personal identifiers, general demographics, and personal health data. Many parts of these records are considered confidential by state law. In 2003, there was an update to the Health and Safety Code applying restrictions to the release of certified copies of birth and death certificates. (California Law, 2006) A continual evaluation process must balance essential public health

needs for the data while ensuring the privacy and reliability of personal health data.

In the administration of public health programs, we are stewards of large volumes of personalized data that are to be used for public health practice. While this use is legally excluded from compliance with the Health Insurance Portability and Accountability Act of 1996 (HIPAA), we have a responsibility to our stakeholders to meet community data management standards, taking into account the restrictions placed on hospitals, clinics and health care providers by HIPAA. (U.S. Department of Health & Human Services, 2006)

As we consider personal data in the birth and death records, the address is one of a set of unique personal identifiers that in combination with other descriptives may easily identify an individual. (U.S. Department of Health & Human Services, 1998) This address can be converted to an x,y coordinate through the geocoding process, thus creating the ability to associate a set of data with a geographic point, described as point data. The x,y coordinate enables mapping data both for legitimate and illegitimate purposes. Powerful home computers can now easily combine information from a variety of readily available sources to identify individuals and more easily breach confidentiality protections that have been sufficient in the past. (Sweeney, 2001) With this in mind, the x,y coordinate creates the need for more care with the distribution, and in particular display, of point data than was needed with the address alone. (Duncan, 2002)

For many public health programs that involve identifying and tracking specific cases, individuals, or exposures, point data is required. Public health practitioners at local, state and federal levels with legitimate needs and legal rights to birth and death point data would be eligible to receive the x,y coordinates as well. Several local health departments have invested in GIS and geocoding. Those that have not would benefit significantly from state support of geocoding and those that have may be able to eliminate some redundancy of efforts if the state provides geocoding as a routine part of data collection.

In order for vital record and other public health point data to be used in research, specific guidelines for approval by the Committee for the Protection of Human Subjects (CPHS) must be followed. Current guidelines from CPHS at the California Health and Human Services Agency regard table cells with fewer than 16 research subjects to be small cells requiring special care not to reveal the identity of individuals. (CPHS, 2005) A wide range of other data users, including companies interested in using birth and death data, etc. are not legally eligible to receive these point data. As the quantity and quality of public data has changed in recent years, rules regarding the

amount of deidentification required to protect an individual's data are being reconsidered.

Public Health Uses of Geocoded Vital Records

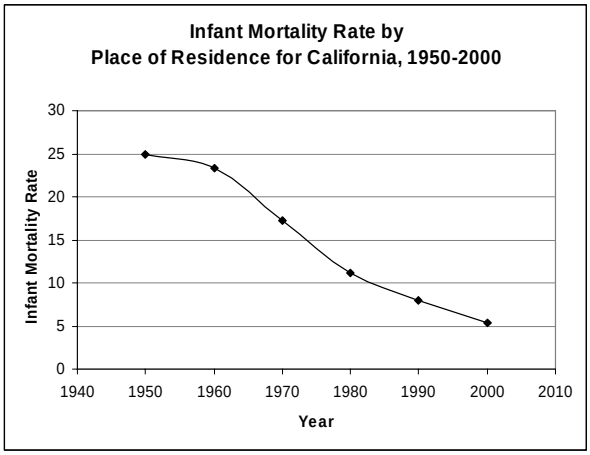
The mission of Public Health is to "fulfill society's interest in assuring conditions in which people can be healthy." (Institute of Medicine, 1998) Geographic data is integral to understanding many aspects of public health; including access to health services, disease tracking and outbreak investigation, emergency response and preparedness, environmental hazards, integrated service delivery, health status and outcomes, and resource allocation and management.

We now consider the following three examples of how we have used geocoded birth and death data and GIS within the Center for Health Statistics this past year: (1) Health Indicators – using GIS to look at subcounty populations; (2) Leading causes of death – using GIS to provide graphic presentation; and (3) Training to improve quality – using GIS to show improvement for trainees.

Health Indicators – using GIS to look at subcounty populations

In public health we consider health indicators as measures for the health of a community. Many of these health indicators are part of the national Healthy People 2010 goals (http://www.healthypeople.gov).

Infant mortality is one key indicator calculated with vital record data, which is the number of deaths in children under the age of one year old per 1,000 live births. It is a widely used indicator of the health of a community and has a Healthy People 2010 goal of 4.5 deaths per 1,000 live births. (Ficenec, 2005)



In California the infant mortality rate has dropped from 24.9 in 1950 to 5.4 in 2000. However, it has not improved uniformly across California.

By using geocoded data, the variation in this indicator throughout the state is apparent. Figure 1 shows the improvement of the infant mortality rate over time from 1994 to 2003 by county.

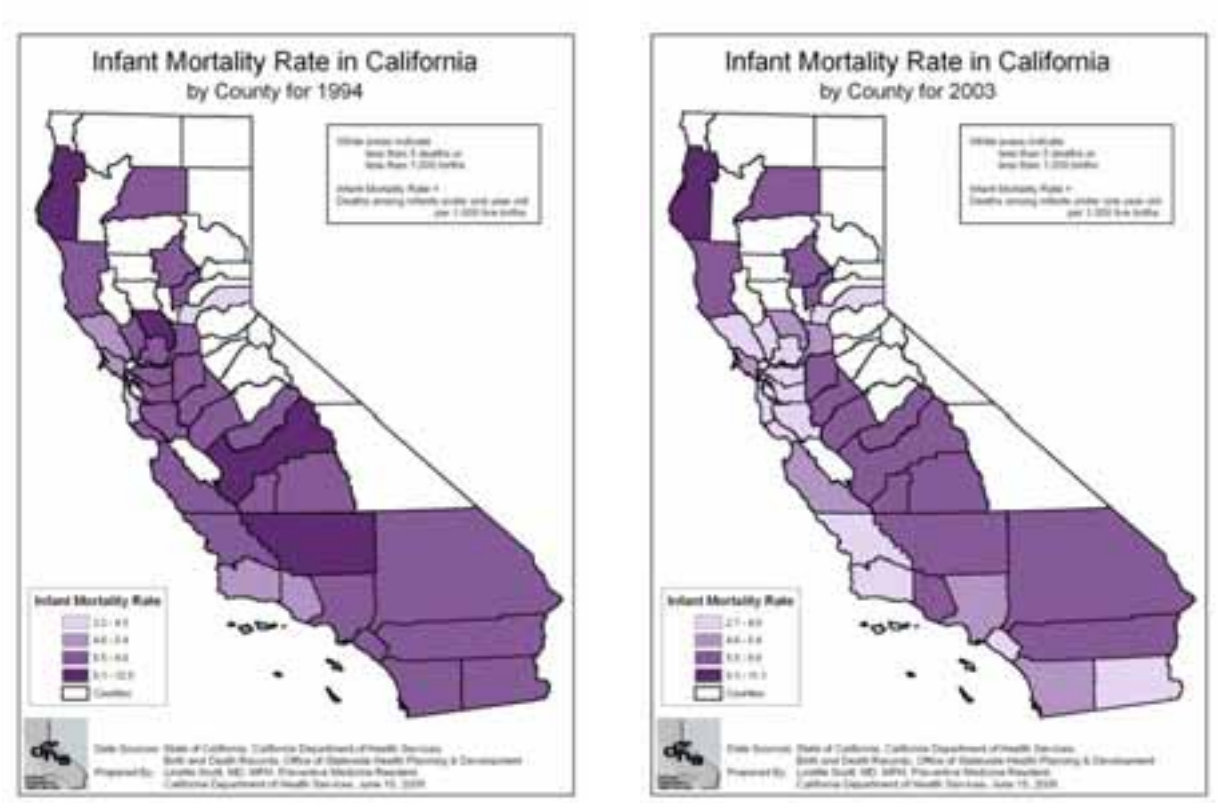


Figure 1: Infant Mortality Rate in California by County comparing 1994 and 2003.

The Medical Service Study Areas (MSSA) were redefined following the 2000 census through community stakeholder meetings throughout the state. These boundaries are accepted as Rational Service Areas by the Health Resources Services Administration and are used for allocation of health care resources to medically underserved areas. The MSSA boundaries are maintained by the California Office of Statewide Health Planning and Development (OSHPD) and are being considered by several programs within CDHS for subcounty analysis. There are 541 MSSAs in California which follow census tract boundaries allowing for calculation of demographic denominator data. (Byrne, 2005)

In Figure 2, the infant mortality rate was calculated for each MSSA aggregated for 1999 - 2001. White areas indicate insufficient births or deaths to calculate a stable rate. Each map demonstrates a different administrative boundary overlying this health indicator. The first is county boundaries, the second is California Senate District Boundaries and the third is California Assembly District Boundaries.

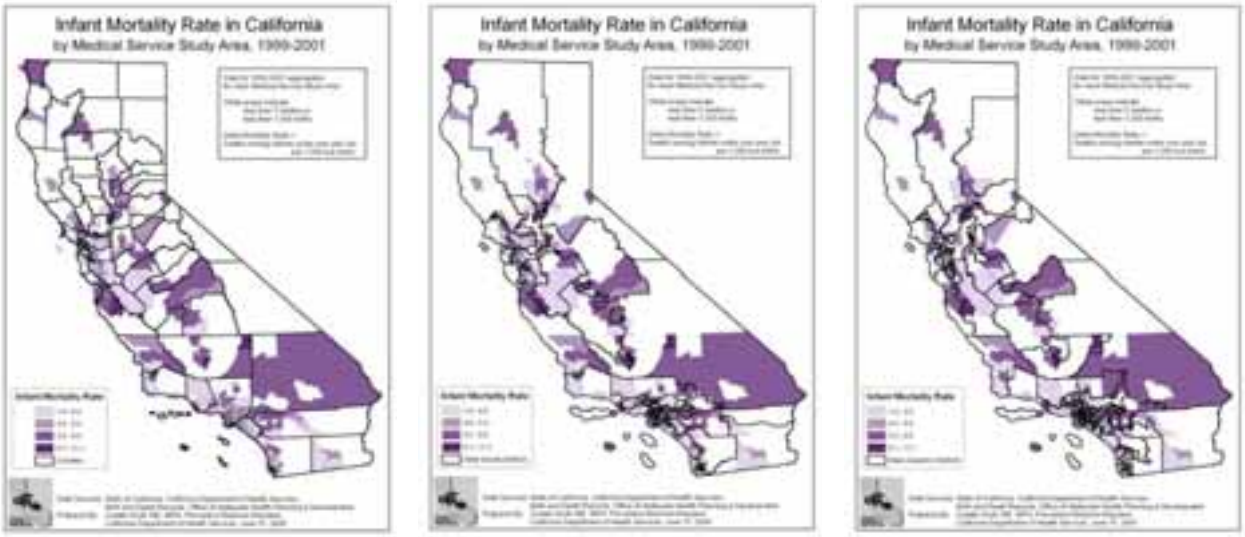


Figure 2: Infant Mortality Rate in California by Medical Service Study Area, 1999-2001 with overlay of different administrative boundaries.

<u>Leading causes of death – using GIS to provide graphic presentation</u>
Within the CHS, the Office of Health Information and Research publishes Data Summaries, which are a series of leading cause of death reports. This past year GIS has been used to create graphics supporting the information provided in these reports.

One example is from the report, *Influenza and Pneumonia Deaths, California 2004*. In Figure 3 we see some of the variation throughout the state in Age-Adjusted Death Rates for 2002 through 2004 due to Influenza and Pneumonia. (Jew-Lochman, 2006) This demonstrates the impact of the visual representation of information in contrast with tabular representations of data.

Figure 3: Example of map being used to show county variation from *Influenza and Pneumonia Deaths, California 2004*

Training to improve quality – using GIS to show improvement for trainees

Vital records in California are very high quality documents, but we continually seek to improve the data collected, since these are essential data sets for so many purposes. Over the past several years, the CHS has designed and implemented a series of trainings for birth clerks, supervisors, nurse managers and county vital records staff throughout the state. This training has provided direct feedback to the individuals who collect and enter birth data to demonstrate the importance of their work and how they are doing in comparison with their peers. Most data items collected on the birth certificate in California are reported to the National Center for Health Statistics (NCHS) for inclusion in national statistics. In this process, the percent of unknown responses for each data item is monitored in order to assess compliance with national (NCHS) and international (World Health Organization) standards for the percent unknown per data item. Progress toward meeting, or maintaining, these standards is communicated during CHS trainings and workshops. In the workshops, local birth clerks and hospital administrators receive tables with their status for percent unknown of selected data fields with an anonymous comparison to other facilities in their county.

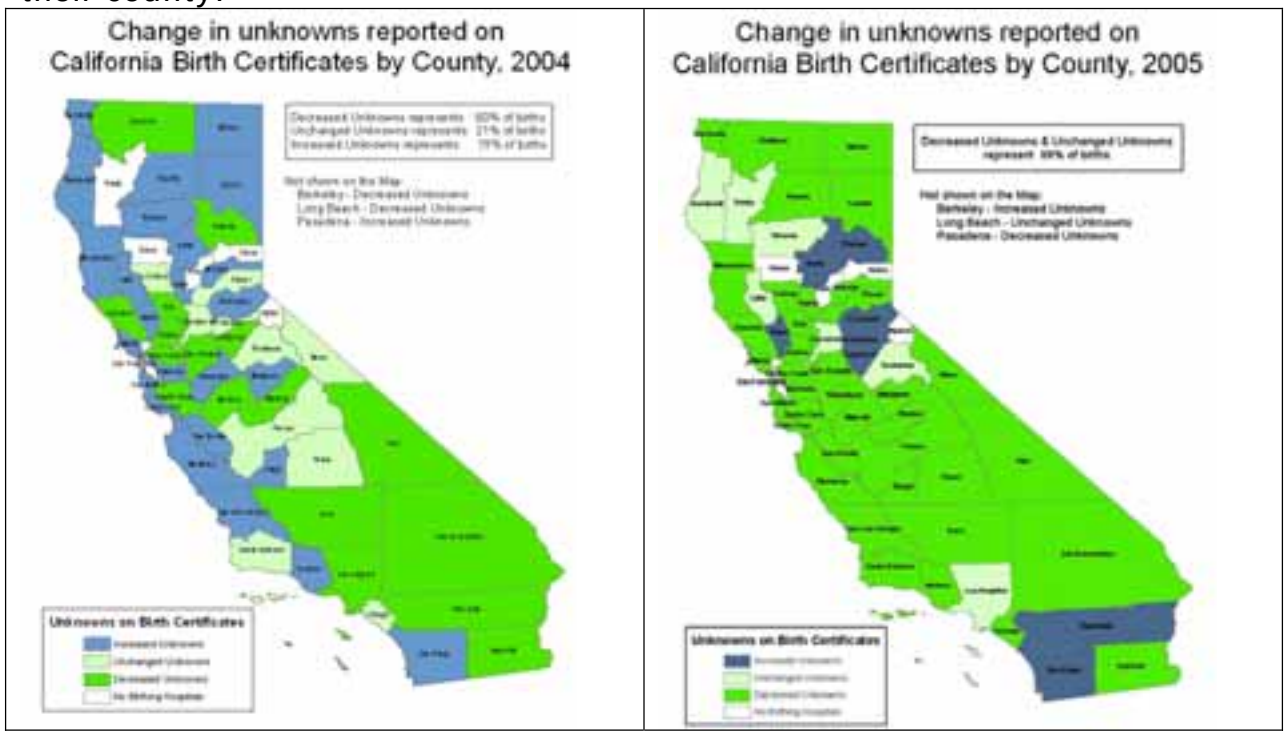


Figure 4: Change in percent unknown over one year for all data items on the California Birth Certificate. The 2004 map represents changes occurring in 2004, at the beginning of the Regional Workshops. The 2005 map represents changes occurring in 2005, after two cycles of the Regional Workshops.

Figure 4 shows two maps. The map on the left shows the change in percent unknown for all data items from 2003 to 2004, when the Regional

Workshops first began. The map on the right shows the change in percent unknown for the data items from 2004 to 2005, after two full cycles of Regional Workshops throughout the state had been completed. This graphic demonstration of the improvement has a substantial impact on the audience.

Conclusions and Future Goals

GIS is a powerful tool that brings the speed and power of new technologies to many public health tasks that have been performed for decades, even centuries. A key component of GIS is collaboration. It inherently involves associating information from varied sources to draw new conclusions and coordinate response efforts. Thus, it provides momentum for increased collaboration within CDHS, with local health departments, and with other public and private agencies.

The Center for Health Statistics is the most logical place for geocoding birth and death data to provide consistency and quality. The long tradition of careful and detailed management of Californians' vital record data will continue as the CHS considers policy issues related to maintaining the confidentiality of this geocoded data while preserving the geographic accuracy of the data.

Acknowledgements

The author would like to acknowledge the many colleagues that have been involved directly and indirectly with different aspects of this work at the California Department of Health Services, the Office of Statewide Health Planning and Development, the University of California Davis, and many other organizations. GIS is truly a collaborative effort.

In particular, the author would like to thank the following individuals for their support for and input into this project:

- Gregory Franklin, MHA, Deputy Director, Health Information and Strategic Planning
- Gwendolyn Doebbert, MA, MS, Health Information and Strategic Planning
- Matthew Stone, MPH, Health Information and Strategic Planning
- Teresita (Terry) Trinidad, MS, Chief, Center for Health Statistics and her staff
- Michael Quinn, Chief, Office of Health Information and Research and his staff
- Melissa Reyes, RN, California Office of Vital Records
- Michael Byrne, MA, Office of Statewide Health Planning and Development
- Scott Christman, MPDS, Office of Statewide Health Planning and Development

References

Biomedical Informatics Research and Consulting Service, University of California Davis. 2006. Retrieved June 15, 2006 from http://www.edrs.us/edrs/index.jsp

Byrne, M, Christman, S, McAdara, S, California's experience with implementing GIS to support equitable healthcare accessibility, California Office of Statewide Health Planning and Development, ESRI Health Users Group Conference, Chicago, 2005. Retrieved June 23, 2006 from http://gis.esri.com/library/userconf/health05/docs/pap1056.pdf

California Department of Finance (CDOF), Retrieved June 18, 2006 from http://www.dof.ca.gov/HTML/DEMOGRAP/E-1table.xls

California Department of Health Services, Center for Health Statistics (CHS), Retrieved May 17, 2005 from http://www.dhs.ca.gov/hisp/chs/

Committee for the Protection of Human Subjects (CPHS), CPHS Bulletin & Update #3, January, 2005. Retrieved June 15, 2006 from http://oshpd.ca.gov/cphs/Bulletins/3CPHS.pdf

California Law, Health and Safety Code 103526. Retrieved June 20, 2006 from http://www.leginfo.ca.gov/calaw.html

Coy, OC, Guide To The County Archives of California. California Historical Survey Commission, 1919. Retrieved May 17, 2005 from http://www.mariposaresearch.net/VITALS.html

Duncan, GT, Statistical Data Stewardship in the 21st Century: An Academic Perspective, Joint Statistical Meetings, New York City, 2002.

Ficenec, S, California' Infant Mortality Rate, 2003, California Department of Health Services, Center for Health Statistics, April, 2006. Retrieved June 20, 2006 from http://www.dhs.ca.gov/hisp/chs/OHIR/reports/others/infantdeath2003.pdf

Institute of Medicine, Committee for the Study of the Future of Public Health, Division of Health Care Services. The Future of Public Health. National Academy Press, Washington, D.C., 1988.

Jew-Lochman, S, Influenza and Pneumonia Deaths, California, 2004, California Department of Health Services, Center for Health Statistics, April, 2006. Retrieved June 20, 2006 from

http://www.dhs.ca.gov/chs/OHIR/reports/leadingcause/Influenza2004.pdf

Sweeney, L, Information Explosion, *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies*, L Zayatz, P Doyle, J Theeuwes and J Lane (eds), Urban Institute, Washington, DC, 2001.

United States Department of Health and Human Services, Unique Health Identifier for Individuals: A White Paper, 1998.

United States Department of Health and Human Services, Office for Civil Rights – HIPAA, Retrieved June 15, 2006 from http://www.hhs.gov/ocr/hipaa/

Williams, RL, AVSS History, University of California, Santa Barbara, Retrieved May 18, 2005 from http://www.avss.ucsb.edu/history.htm

Author's Information

Linette Scott, MD, MPH
California Department of Health Services
Center for Health Statistics
1616 Capitol Avenue-MS 5101
PO Box 997410
Sacramento, CA 95899-7410
Phone: 916-440-7849
Fax: 916-650-6889
lscott@dhs.ca.gov