

# A p p e n d i x A

## TECHNICAL NOTES

### Methods

#### Race and Ethnicity

This report restricts descriptions of race and ethnicity to short words and phrases. It is recognized that individual preference varies and that classification is not trivial. Considering the report's many text references, tables, and figures that make comparisons between races, readability and space require consistent and abbreviated usage. Thus, the report refers to African American, rather than Black or African. In tables and figures, African American is usually shortened to AfrAmer. Other standard terms are White; American Indian (sometimes shortened to AmerInd); Pacific Islander (sometimes shortened to PacIsl and sometimes referred to as Native Hawaiian or Other Pacific Islander); and Asian (sometimes combined with Pacific Islanders and shortened to API). Latino includes all those of Spanish- and Portuguese-speaking descent in the Americas, including people from Spain. Hispanic or Latino is considered by most data collectors such as the Census Bureau to be an ethnicity rather than a race. Thus, a Latino may be White or Asian or Black, but here all those persons are reported as Latino. Some data systems are allowing people to choose multiple races or simply a Multirace or Other category, so the report uses those designations when needed. Finally, race is often unreported, mis-reported, or unclassifiable in many data systems; the report often includes these for completeness, labeled as appropriate for the circumstance.

#### Rate Calculations

*Age-adjustment* All age-adjusted rates in this report are adjusted by the direct method to the 2000 U.S. Standard Population. In general, the number of deaths for specific causes of mortality in a community is affected by the size and age composition of the population. Because the risk of dying is primarily a function of age, simply calculating a crude rate for vital events such as death (number of deaths/population) can lead to misleading conclusions when comparing different subpopulations. This is because populations with a large component of elderly people tend to have a high death rate simply because the risk of dying is determined mostly by age. In order to nullify the effect of differences in the age composition of populations, disease rates can be age-adjusted. Age-adjusting methodology involves first calculating rates for each age category to determine age-specific rates. Each age-specific rate is multiplied by the proportion of the age category in a standard population. The sum of these weighted age-specific rates in a community is the age-adjusted rate for that community. Age-adjusted disease rates form a better basis for unbiased comparison across populations.

*Variability of rates* All vital statistics, including death rates, are subject to random variation. This variation is inversely related to the number of events (e.g. deaths) used to calculate the rate. The smaller the number of events, the greater the likelihood of random variation. In order to protect against providing misleading information based on statistically unreliable rates, the National Center for Health Statistics (NCHS) recommends presenting only rates based on 20 or more events.<sup>1</sup> For this

report, this standard has been relaxed to a requisite ten or more events for rates, a standard recently adopted by the Family Health Outcomes Project of the University of California, San Francisco.<sup>2</sup>

*Confidence interval* A good measure of the reliability of a rate is the confidence interval (CI) around the rate estimate. A confidence interval defines the range of rates that would be determined by repeated sampling of the same phenomenon. By statistical convention, a 95% confidence interval is considered a useful measure of the range of accuracy of an estimate. This means that with repeated sampling, one would obtain a rate within the confidence interval 95% of the time. These calculations normally use the binomial distribution. Based on recommendations of the National Center for Health Statistics (NCHS) regarding the calculation of rates and confidence intervals, the standard error of any rate based on fewer than 100 events is based on the Poisson distribution.<sup>1</sup> The Poisson distribution is similar to the binomial distribution but is characterized by very small numbers of events occurring in a large number of trials.<sup>3</sup>

### **Area-based Socioeconomic Measures and Health Inequities**

One way of looking at health inequities is by analyzing data according to socioeconomic status. Most datasets, however, do not carry these types of data. Absent appropriate data in the datasets themselves, socioeconomic status is best represented by poverty level according to census tracts. Nancy Krieger et al have shown in the Public Health Disparities Geocoding Project that the census tract is the best unit of analysis. They have also shown that the poverty level captures as much difference as more complex measures. This representation of economic status also takes into account neighborhood-level factors. The poverty rate is the rate for a tract, a proxy for neighborhood.

### **Years of Potential Life Lost**

Years of potential life lost is calculated based on YPLL-75, the number of years of life lost due to death before age 75. This method is used because average life expectancy in the United States is over 75 years. Years of potential life lost is derived by summing years of life lost over all age groups.

### **Life Expectancy**

Life expectancy at birth is calculated using abridged life tables by 10-year age intervals. The abridged method is used because 1) it is a short cut method, and 2) preparing a complete life table would not be suitable because data are sparsely distributed by single years of age.

### **Tests for Trends**

To test for significance of trends, Joinpoint Regression Program for Windows (v3.0) software was used.<sup>12</sup> Software is available through the National Cancer Institute. Method uses regression techniques to fit model with one or more line segments on a logarithmic scale. Statistics obtained from the models were average annual percent change (APC), and 95% confidence interval around APC, in rates associated with each line segment. Joinpoints (the years at which the slope of a line segment changes) are also provided.

## **Data Sources**

*Demographic and socioeconomic* U.S. Bureau of the Census, 1950-2000 Census; California Department of Education, Dataquest and Ed-Data; California Department of Finance; California Health Interview Survey; California Employment Development Department; National Economic Development

and Law Center; National Association of Homebuilders; California Department of Justice.

*Population estimates* The population estimate for Alameda County is from Census 2000 and California Department of Finance (DOF) estimates. Since most data are for 2001 to 2003, they have a midpoint of July 1, 2002. Age, sex, and race distributions are from Census 2000 (assuming that little had changed since the Census of April 1, 2000), but the total population is adjusted using the DOF benchmarks. For the years prior to 2000, the age and sex distributions are assumed to change linearly from Census 1990 to Census 2000, with the total number of persons taken from DOF estimates.<sup>4,5</sup>

*Births* Alameda County Public Health Department Vital Statistics Files obtained from the Alameda County Department of Public Health Automated Vital Statistics System (AVSS).

*Deaths* Alameda County Public Health Department Vital Statistics Files obtained from the Alameda County Department of Public Health Automated Vital Statistics System (AVSS) and the State of California Statistical Master Death file.

*Hospital discharge* Hospital inpatient discharge data collected by the California Office of Statewide Health Planning and Development (OSHPD).

*Cancer incidence* California Cancer Registry (CCR) data provided by the Northern California Cancer Center (NCCC) is the source of data on new cases of cancer. This data source has reports of cases diagnosed by site, as well as a range of clinical characteristics such as extent of disease and stage.

*Tuberculosis* Alameda County Department of Public Health Tuberculosis Information Management System (TIMS).

*HIV/AIDS* Alameda County Department of Public Health HIV/AIDS Reporting System (HARS).

*STDs* Alameda County STD surveillance system.

## Limitations of Data and Other Data Issues

*Hospital discharge data* Because persons with multiple hospitalizations during the year can be counted more than once, hospital discharge data produces the estimates for discharges, not persons. Changes in rates of hospitalization may be changes in hospital admission practices or the diagnostic coding of illnesses, or reflective of true changes in the patterns of disease. The data capture those illnesses or injuries serious enough to get people admitted to the hospital and are not the prevalence of a given disease or condition in the population since many who have the illness are not hospitalized. Race and ethnicity data are missing for many cases due to the omission of a race field on many hospital discharge forms. Consequently, race is not recorded in about 18% of records.<sup>6</sup> There are a large number of cases of ‘unknown’ and ‘other’ race which are not included in rate calculations, resulting in an overestimation of rates for some racial groups and an underestimation for others.

*Births* Information on the newborn is taken from the birth certificate. The race/ethnicity on the birth certificate is reported by self-identification according to the race and ethnicity of the mother.

*Deaths* The race and ethnicity of the decedent is from the death certificate as reported by family members to the funeral director. However, birth and census population data use the self-reported race of the respondent. Because of the combined effect of numerator and denominator biases, it has

been estimated that death rates are overestimated by about 1% among Whites and 5% among African Americans. They are underestimated by approximately 21% for American Indian or Alaska Natives, 11% for Asian and Pacific Islanders, and 2% for Latinos.<sup>7</sup>

*Change of International Classification of Disease* Mortality data for specific causes of death in this report are classified and coded according to the World Health Organization's (WHO) tenth revision of the International Classification of Diseases (ICD-10) implemented in the United States in 1999.<sup>8</sup> However, the mortality trend data for 1990 to 1998, and hospital discharge data are based on the ninth revision of the International Classification of Diseases (ICD-9). Since the beginning of the century, the International Classification of Disease for mortality has been modified about once every 10 years, except for the 20-year interval between the last two revisions. ICD-10 differs from ICD-9 in many respects: 1) ICD-10 is far more detailed than ICD-9, about 8,000 categories compared with 4,000 categories, mainly to provide more clinical detail for morbidity applications; 2) ICD-10 uses 4-digit alphanumeric codes compared with 4-digit numeric codes in ICD-9; 3) three additional chapters have been added, some chapters rearranged, cause of death titles have been changed, and conditions have been regrouped; 4) some coding rules have been changed.<sup>9</sup>

Introducing this tenth revision of International Classification of Disease creates discontinuities in time series and trends. This means the Healthy People 2010 objectives may not be strictly comparable with the tracking data for 1999 and subsequent years whose baseline data were 1997 and 1998.<sup>9</sup>

*Multiple race coding* The data on race in Census 2000 are not directly comparable to those collected in previous censuses. The October 1997 revised standards issued by the U.S. Office of Management and Budget (OMB) led to changes in the question on race for Census 2000. In Census 2000, respondents were allowed to select more than one category for race. Also, the "Asian and Pacific Islander" category was separated into two categories, "Asian" and "Native Hawaiian and Other Pacific Islander."

*Leading causes of death* Causes are ranked according to the number of deaths because it most accurately reflects the frequency of cause-specific mortality. In this report, leading causes of death were derived from the recommended list of 50 rankable causes from the 113 selected causes of death developed for use with ICD-10.<sup>10</sup> Leading causes of infant mortality were derived from a separate ranking procedure using the recommended list of 71 rankable causes from the 130 selected causes of infant death developed in accordance with ICD-10. Ranking leading causes of death is a tool for illustrating the relative burden of cause-specific mortality. However, the rankings do not necessarily indicate those causes of death of greatest public health importance. Some causes of death of public health importance, such as lung cancer and motor vehicle crashes are excluded from the ranking procedure and included in broader rankable categories, namely, all cancer and unintentional injuries, respectively. If they were included separately, both causes would rank among the ten leading causes of death.<sup>10</sup>

*Sexually transmitted diseases* The incidence of STDs depends on levels of screening. Since testing for STDs is not comprehensive or uniform throughout the jurisdiction, and since many STD infections are asymptomatic, the actual incidence of STDs is greater than that which is reported. In addition, STD data derive largely from laboratory reports which do not contain information on the race/ethnicity of the individual. Hence, the data is incomplete and conclusions about the distribution of STDs by race/ethnicity cannot be firmly drawn.

## Case Definitions

### Maternal and Child Health

*Infant mortality* Number of deaths to children less than one year old per 1,000 live births.

*Low birth weight* The percentage of live births weighing less than 2,500 grams.

*Early prenatal care* Care received during the first trimester (before 12 weeks) of pregnancy.

*Teen birth* Births to teenage mothers (15 to 19 years).

### Mortality

Cause of Death	ICD-10 Codes	Cause of Death	ICD-10 Codes
Diabetes	E10-E14	Motor Vehicle Crash	V02-V04, V09.0, V09.2, V12-V14, V19.0-V19.2, V19.4-V19.6, V20-V79, V80.3-V80.5, V81.0-V81.1, V82.0-V82.1, V83-V86, V87.0-V87.8, V88.0-V88.8, V89.0, V89.2
Coronary heart disease	I11; I20-I25		
Stroke	I60 - I69		
All cancer	C00-C97		
Lung cancer	C33-C34		
Colorectal cancer	C18-C21		
Female breast cancer	C50	Unintentional injury	V01-X59; Y85-Y86
Prostate cancer	C61	Suicide	X60-X84, Y87.0
Asthma	J45-J46	Homicide	X85-Y09; Y87.1

### Hospitalization

Diagnosis	ICD-9 Codes	Code Positions
Asthma	493.00-493.99	Primary Dx
Diabetes	250.00-250.99	Primary Dx-Dx4
Coronary heart disease	402.00-402.99; 410.00-414.99; 429.2	Primary Dx-Dx4
Stroke	430.00-438.99	Primary Dx-Dx4
Self-Inflicted Injury	E950-959	Primary E-code
Unintentional injury	E800-E949	Primary E-code
Assault	E960-E969	Primary E-code

### Cancer Incidence

Site	SEER Site Code
Breast	26000
Prostate	28010
Colorectal	21041-21049; 21051-21052
Lung	22030

A cancer case is defined in this report as a primary malignant tumor, that is, one originating in a particular organ or anatomic site rather than having spread from another location. Surveillance Epidemiology and End Results (SEER) site codes in the California Cancer Registry were used to define and select cases by site for this report.<sup>11</sup>

## Communicable Disease

*AIDS* The Centers for Disease Control and Prevention expanded the AIDS surveillance case definition in 1993 to include all HIV-infected persons with a CD4+ T-lymphocyte count of less than 200 cells/uL or with one of the AIDS-defining clinical conditions.

*Chlamydia* A case that is laboratory confirmed by isolation of *C. trachomatis* by culture.

*Gonorrhea* A case that is laboratory confirmed by isolation of *Neisseria gonorrhoea* by culture.

*Tuberculosis* Positive cultures for *M. tuberculosis* confirm the diagnosis of TB. However, TB may also be diagnosed by the medical provider on the basis of clinical signs and symptoms in the absence of positive cultures.

## References

1. National Center for Health Statistics. Technical Appendix from *Vital Statistics of United States, 1995*. Mortality. Hyattsville, MD. April 1999.
2. McCandless RR, Oliva G. *Guidelines for Statistical Analysis of Public Health Data with Attention to Small Numbers: A Collaborative Effort of the Family Health Outcomes Project Technical Advisory Group*. San Francisco, Calif. March 2002.
3. Kahn, HA, Sempos, CT. *Statistical Methods in Epidemiology*. New York. Oxford University Press. 1989.
4. State of California, Department of Finance. *E-5 City/County Population and Housing Estimates, 2004, Revised 2001-2003, with 2000 DRU Benchmark*. Sacramento, California. May 2004.
5. State of California, Department of Finance. *E-4 Revised Historical City, County and State Population Estimates, 1991-2000, with 1990 and 2000 Census Counts*. Sacramento, California. March 2002.
6. Kozak LJ. Underreporting of Race in the National Hospital Discharge Survey. *Advance Data from Vital and Health Statistics* No. 265. Hyattsville, Md: NCHS. 1995.
7. Rosenberg HM, Maurer JD, Sorlie PD, et al. Quality of Death Rates by Race and Hispanic Origin: A Summary of Current Research, 1999. National Center for Health Statistics (NCHS). *Vital and Health Statistics* 2(128). 1999.
8. World Health Organization. *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision*. Geneva: World Health Organization. 1992.
9. Anderson RN, Minino AM, Hoyert DL, Rosenberg HM. Comparability of Cause of Death Between ICD-9 and ICD-10: Preliminary Estimates. *National Vital Statistics Reports* 49(2). Hyattsville, Md: National Center for Health Statistics. 2001.
10. Anderson RN. Deaths: Leading Causes for 1999. *National Vital Statistics Reports* 49(11). Hyattsville, Md: National Center for Health Statistics. 2001.
11. Jackson NC, Valone T, Wormeli B. Cancer Reporting in California. Data standards for regional registries and California Cancer Registry. California Cancer Reporting Standards. Volume III. November 2002.
12. Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. *Statistics in Medicine*, vol 19: 335-351. 2000.

# Appendix B

## DATA TABLES

Table B.1: MCAH Indicators by Race/Ethnicity, 2001-2003, Alameda County, Total Counts, 3-Year Average Rates, and 95% Confidence Limits _____	160	Table B.15: Cancer Incidence by Gender and Race/Ethnicity, Alameda County, 2000-2002, Total Counts, 3-Year Average Rates, and 95% Confidence Limits _____	174
Table B.2: Historical Infant Mortality by Race/Ethnicity, 1990-2003, Annual Counts, Rates (per 1,000 Births), and 95% Confidence Limits _____	161	Table B.16: Historical Chronic Disease Hospitalization by Race/Ethnicity, Alameda County, 1991-2003, Annual Counts, Rates, and 95% Confidence Limits _____	175
Table B.3: Historical Low Birth Weight by Race/Ethnicity, Alameda County, 1990-2003, Annual Counts, Percentages, and 95% Confidence Limits _____	162	Table B.17: Historical Hospitalizations by Race/Ethnicity, Alameda County, 1991-2003, Annual Counts, Rates, and 95% Confidence Limits _____	176
Table B.4: Historical Early Prenatal Care by Race/Ethnicity, Alameda County, 1990-2003, Annual Counts, Percentages, and 95% Confidence Limits _____	163	Table B.18: Historical Injury Hospitalizations by Race/Ethnicity, Alameda County, 1991-2003, Annual Counts, Rates, and 95% Confidence Limits _____	177
Table B.5: Historical Teen Births by Race/Ethnicity, Alameda County, 1990-2003, Annual Counts, Rates (per 1,000 Females 15-19 years), and 95% Confidence Limits _____	164	Table B.19: Historical Cancer Incidence by Race/Ethnicity, Alameda County, 1990-2002, Annual Counts, Rates, and 95% Confidence Limits _____	178
Table B.6: Mortality by Gender and Race/Ethnicity, Alameda County, 2001-2003, Counts, 3-Year Average Rates, 95% Confidence Limits _____	165	Table B.20: Historical Cancer Incidence by Race/Ethnicity, Alameda County, 1990-2002, Annual Counts, Rates, and 95% Confidence Limits _____	179
Table B.7: Cancer Mortality by Gender and Race/Ethnicity, Alameda County, 2001-2003, Counts, 3-Year Average Rates, and 95% Confidence Limits _____	166	Table B.21: Tuberculosis Cases, Alameda County, 2002-2004, Total Counts, 3-Year Average Rates, and 95% Confidence Limits _____	180
Table B.8: Injury Mortality by Gender and Race/Ethnicity, Alameda County, 2001-2003, Counts, 3-Year Average Rates, and 95% Confidence Limits _____	167	Table B.22: AIDS Cases, Alameda County, 2002-2004, Total Counts, 3-Year Average Rates, and 95% Confidence Limits _____	181
Table B.9: Historical Mortality, Alameda County, 1990-2003, Annual Counts, Rates, and 95% Confidence Limits _____	168	Table B.23: Gonorrhea Cases, Alameda County, 2002-2004, Total Counts, 3-Year Average Rates, and 95% Confidence Limits _____	182
Table B.10: Historical Chronic Disease Mortality, Alameda County, 1990-2003, Annual Counts, Rates, and 95% Confidence Limits _____	169	Table B.24: Chlamydia Cases, Alameda County, 2002-2004, Total Counts, 3-Year Average Rates, and 95% Confidence Limits _____	183
Table B.11: Historical Cancer Mortality, Alameda County, 1990-2003, Annual Counts, Rates, and 95% Confidence Limits _____	170	Table B.25: Historical Communicable Disease, Alameda County, 1980-2004, Annual Counts, Rates, and 95% Confidence Limits _____	184
Table B.12: Historical Injury Mortality, Alameda County, 1990-2003, Annual Counts, Rates, and 95% Confidence Limits _____	171		
Table B.13: Chronic Disease Hospitalizations by Gender and Race/Ethnicity, Alameda County, 2001-2003, Counts, Rates, and 95% Confidence Limits _____	172		
Table B.14: Injury Hospitalization by Gender and Race/Ethnicity, Alameda County, 2001-2003, Counts, 3-Year Average Rates, and 95% Confidence Limits _____	173		