

TECHNICAL NOTES

DATA SOURCES

The primary data source used for this report was the DATA2010 system, developed and maintained by the U.S. Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS), Division of Health Promotion Statistics. A complete listing of the data sources used by DATA2010 for monitoring HP2010 objectives can be found at <http://wonder.cdc.gov/data2010/source.htm>. Where California data were unavailable in DATA2010, secondary data sources were used for extracting numerator data. These included, but were not limited to: birth and death records maintained by the California Department of Public Health (CDPH), Center for Health Statistics (CHS); hospital discharge records maintained by the California Office of Statewide Health Planning and Development (OSHPD); case registries maintained by the CDPH Center for Infectious Disease and the CDPH Center for Chronic Disease Prevention and Health Promotion; and prevalence data developed by the California Health Interview Survey (CHIS) and the California Behavioral Risk Factor Surveillance Survey (CA-BRFSS).¹⁻⁷

Population denominator data by age, sex, and race-ethnicity used in the calculation of population-based mortality and morbidity rates were supplied by the NCHS in cooperation with the U.S. Census Bureau, and by the California Department of Finance (CDOF), Demographic Research Unit (DRU).^{8,9} Denominator data used in the calculation of natality rates were extracted from the CDPH/CHS Birth Statistical Master Files and from the Birth Cohort Files.

DATA DEFINITIONS

Race and Ethnicity

The Population Estimates Program of the U.S. Census Bureau, in collaboration with the NCHS, releases bridged-race population estimates of the resident population of the United States for calculating vital rates in DATA2010.⁸ These estimates result from bridging the 31 race categories used in Census 2000, as specified in the 1997 Office of Management and Budget (OMB) standards for the collection of data on race and ethnicity, to the four race categories specified under the 1977 standards (i.e., White, Black or African American, American Indian or Alaska Native, Asian or Pacific Islander), and two Hispanic origin categories (i.e., Hispanic or Latino, not Hispanic or Latino). The bridging methodology is described in the 2003 NCHS report, "Census 2000 Populations with Bridged Race Categories", which is available from http://www.cdc.gov/nchs/data/series/sr_02/sr02_135.pdf. More detailed information on these data files is available at <http://www.cdc.gov/nchs/about/major/dvs/popbridge/popbridge.htm>.

The population data supplied by the CDOF/DRU for use in calculating vital statistics rates includes seven mutually exclusive racial and ethnic categories (American Indian, Asian, Black, Hispanic, Multirace, Pacific Islander, and White), by age, and by sex.⁹

Numerator data (i.e., the number of hospitalizations used in the calculation of morbidity rates) were extracted using the "Ethnicity" and "Race" variables available in the OSHPD Patient Discharge Data files, which provide these mutually exclusive categories:

- Hispanic/Latino
- African American/Black, non-Hispanic
- American Indian/Alaska Native, non-Hispanic
- Asian/Pacific Islander, non-Hispanic
- White, non-Hispanic
- Other/Unknown, non-Hispanic

Mortality Data

Beginning in 1999, deaths are coded using the International Classification of Diseases, Tenth Revision (ICD-10).¹⁰ Readers and users of these data are cautioned that prior year's mortality tables based on ICD-9 are not comparable and should not be used for direct comparisons with ICD-10 data.

Following is a list of the ICD-10 codes that were used to operationally define each Healthy People 2010 objective that use mortality data for monitoring:¹¹

Obj. 3-1	All cancers.....	C00-C97
Obj. 3-2	Lung cancer.....	C33-C34
Obj. 3-3	Female breast cancer.....	C50 (female only)
Obj. 3-4	Cervical cancer.....	C53
Obj. 3-5	Colorectal cancer.....	C18-C21
Obj. 3-6	Oropharyngeal cancer.....	C00-C14
Obj. 3-7	Prostate cancer.....	C61 (male only)
Obj. 3-8	Malignant melanoma.....	C43
Obj. 5-5	Diabetes-related.....	E10-E14
Obj. 10-4a	Food allergy.....	T78.0, T78.1
Obj. 12-1	Coronary heart disease.....	I11, I20-I25
Obj. 12-7	Stroke.....	I60-I69
Obj. 13-14	HIV.....	B20-B24
Obj. 15-3	Firearm-related.....	U01.4, W32-W34, X72-X74, X93-X95, Y22-Y24, Y35.0
Obj. 15-8	Poisoning.....	X40-X49, X60-X69, X85-X90, Y10-Y19, Y35.2
Obj. 15-9	Suffocation.....	W75-W84, X70, X91, Y20
Obj. 15-13	Unintentional injury.....	V01-X59, Y85-Y86
Obj. 15-15	Motor vehicle crash.....	V02-V04 (.1, .9), V09.2, V12-V14 (.3-.9), V19 (.4-.6), V20-V28 (.3-.9), V29-V79 (.4-.9), V80 (.3-.5), V81.1, V82.1, V83-V86 (.0-.3), V87 (.0-.8), V89.2
Obj. 15-16	Pedestrian.....	V02-V04 (.1, .9), V09.2
Obj. 15-25	Residential fire.....	X00-X09
Obj. 15-27	Fall-related.....	W00-W19
Obj. 15-29	Drowning.....	V90, V92, W65-W74
Obj. 15-32	Homicide.....	U01-U02, X85-Y09, Y87.1
Obj. 16-4	Maternal deaths due to obstetric causes.....	A34, O00-O95, O98-O99
Obj. 18-1	Suicide.....	U03, X60-X84, Y87.0
Obj. 24-1	Asthma.....	J45-J46
Obj. 24-10	Chronic obstructive pulmonary disease.....	J40-J44
Obj. 26-2	Cirrhosis.....	K70, K73, K74
Obj. 26-3	Drug-induced.....	D52.1, D59.0, D59.2, D61.1, D64.2, E06.4, E16.0, E23.1, E24.2, E27.3, E66.1, F11-F16 (.0-.5, .7-.9), F17 (.0, .3-.5, .7-.9), F18-F19 (.0-.5, .7-.9), G21.1, G24.0, G25.1, G25.4, G25.6, G44.4, G62.0, G72.0, I95.2, J70.2-J70.4, L10.5, L27.0, L27.1, M10.2, M32.0, M80.4, M81.4, M83.5, M87.1, R78.1-R78.5, X40-X44, X60-X64, X85, Y10-Y14

Infant Mortality

The infant death rate is the number of deaths among infants under one year of age per 1,000 live births, and is a universally accepted and easily understood indicator that is useful in pointing to problems with the health status of infants and mothers, and possible problems in the delivery of health care and related services to these groups in a community. The following HP2010 objectives use data from the linked birth and infant death records available in the California Birth Cohort Files:

Obj. 16-1f	Birth Defect Deaths	Q00-Q99
Obj. 16-1g	Congenital Heart and Vascular Defect Deaths	Q20-Q28
Obj. 16-1h	Sudden Infant Death Syndrome Deaths.....	R95
Obj. 16-2	Infant Deaths	A00-Y89

Nativity

The following HP2010 objectives use natality data obtained from the Birth Statistical Master Files for 2000- 2003:

Obj. 16-6a The prenatal care indicator, Month Prenatal Care Began, has been associated with access to care. Early prenatal care is defined as the percentage of mothers who began prenatal care in their first trimester (the first three months of pregnancy). However, as a health indicator, the percentage of births in which the mother's prenatal care began in the first trimester fails to document whether the prenatal care was adequate, nor whether it actually continued for the course of the pregnancy.

Obj. 16-9a Primary Cesarean section deliveries among low-risk females are defined as singleton, full-term, vertex presentation births among women giving birth for the first time.

Obj. 16-9b Repeat Cesarean section deliveries among low-risk females are defined as full-term, singleton, vertex presentation births among women who have had a previous Cesarean section delivery.

Obj. 16-10a,b Low birth weight (LBW) and very low birth weight (VLBW) have been associated with negative birth outcomes and as an indicator of access problems and/or need for prenatal care services. LBW is defined as the percentage of live births weighing less than 2,500 grams; VLBW is defined as the percentage of live births weighing less than 1,500 grams.

Obj. 16-11a-c Preterm births are defined as: a) less than 37 completed weeks of gestation; and b) between 32 and 36 completed weeks of gestation. Very preterm births are defined as c) less than 32 completed weeks of gestation.

Morbidity

Following is a list of the ICD-9-CM codes that were used to operationally define each Healthy People 2010 objective that uses cause-specific hospitalization data for monitoring:¹²

Obj. 1-9a	Pediatric asthma.....	493
Obj. 1-9b	Uncontrolled diabetes.....	250.02-250.03, 250.10-250.13, 250.20-250.23, 250.30-250.33
Obj. 1-9c	Pneumonia and influenza.....	481, 487
Obj. 2-10	Vertebral fractures.....	805.0, 805.2, 805.4, 805.8
Obj. 5-10	Lower extremity amputations.....	84.1 (amputation procedure code)
	In persons with diabetes	250 (diabetes)
Obj. 12-6	Congestive heart failure	428.0
Obj. 14-17	Peptic ulcer.....	531-534

Obj. 15-1	Nonfatal head injuries.....	800-801, 803-804, 850-854, 870-873, 925
Obj. 15-2	Nonfatal spinal cord injuries.....	806, 952
Obj. 15-5	Nonfatal firearm injuries	E922.0-922.9, E955.0-955.4, E965.0-965.4, E970, E985.0-985.4
Obj. 15-7	Nonfatal poisonings.....	E850-E869, E950-E952, E962, E972, E980- E982
Obj. 15-12	ER visits for injuries.....	E800-E869, E880-E929, E950-E999
Obj. 15-28	Hip fractures.....	820
Obj. 15-30	Nonfatal dog bite injuries.....	E906.0
Obj. 16-5a	Maternal morbidity.....	641, 642.0-642.7, 642.9, 643.2, 646.2, 646.6- 646.7, 648.0, 648.5-648.6, 648.8, 658.4, 659.2, 659.3, 664.2-664.3, 664.5, 665, 666, 668, 669.0- 669.4, 670, 671.3-671.5, 672, 673, 674.0-674.3, 674.8-674.9
	Delivery	V27
Obj. 16-15	Spina bifida and other NTD's.....	740-742.0
Obj. 16-21	Sickle Cell Disease.....	282.60-282.69
Obj. 24-2	Asthma	493
Obj. 24-3	Asthma.....	493
Obj. 28-12	Otitis media.....	381.0-381.4, 382

California data on end-stage renal disease cases used for monitoring Objective 4-1 were extracted from the USRDS Renal Data Extraction & Reference (RenDER) query system available at http://www.usrds.org/odr/xrender_home.asp.

AGE-ADJUSTED RATES

The numerator data used to compute rates and percentages were extracted by place of residence of the decedent or patient for objectives using mortality or morbidity data, and by place of residence of the mother for objectives using natality data.

An unstandardized rate (usually referred to as a "crude rate") is obtained by dividing the total number of vital events (e.g. deaths) by the total population at risk, then multiplying by some standard basis (e.g. per 100,000). Crude death rates, which include the effects of age, are the rates that should be applied when measuring the actual risk of dying in a specific population. Subpopulations with varying age compositions can have highly disparate death rates, since the risk of dying is primarily a function of age. Therefore, subpopulations with a large component of elderly tend to have higher death rates simply because the risk of dying is determined mostly by age. Any effect of different age compositions among population subgroups can be removed by the process known as "age-adjustment."

Age-adjusted death rates are hypothetical rates obtained by calculating age-specific rates for each subpopulation and multiplying these rates by proportions of the same age categories in a standard population, then summing the apportioned specific rates to a total. The "standard population" used in the age-adjusted death rates in this report is the 2000 United States Standard Population. The age-adjusted rates put all subgroups on an equal basis with respect to the effect of age and permit direct comparisons among subgroups and with HP 2010 mortality objectives. It is important to understand that age-adjusted death rates should be viewed as hypothetical constructs or index numbers rather than as actual measures of the risk of mortality. For further information on age-adjusted rates, see the NCHS report by Klein and Schoenborn.¹³

Natality data, prenatal care and low birth weight, were not age-adjusted and represent percentages based on the number of live births. Comparisons of infant mortality among subpopulations reflect the actual risk of dying within one year of birth and are unaffected by

confounding of different age compositions because the decedents are all of the same age (under one year). Age-adjusting is not applicable to these data.

RELIABILITY OF RATES

All vital statistics rates are subject to random variation. This variation is inversely related to the number of events (e.g. death) used to calculate the rate. The smaller the frequency of occurrence of an event, then the greater the likelihood of random fluctuations within a specified time period or a certain subpopulation. The more rare an event, the less stable its occurrence is from observation to observation. As a consequence, subpopulations with only a few deaths or a few events can have highly unstable rates from year to year.

The "standard error" of a death rate and "coefficient of variation" (or Relative Standard Error, RSE) provide a rational basis for determining which rates may be considered "unreliable." In this report, subpopulations with a relative standard error of greater than or equal to 23 percent of the rate or percent are marked with the notation "DSU" (Data Statistically Unreliable). This criterion conforms with the standard used by the NCHS in determining the reliability cut-off for rates and percents.¹⁴ In addition, for rates based on zero events the standard error cannot be calculated and is indeterminate. Whenever the standard error is indeterminate, confidence limits are not calculated and zeroes are used to denote the confidence intervals.

TESTS OF STATISTICAL SIGNIFICANCE

A confidence interval (C.I.) is a range of values that is normally used to describe the uncertainty around a point estimate of a rate. Confidence intervals are a measure of the variability in the data and describe how much different the point estimate could have been if the underlying conditions stayed the same but chance had led to a different set of data. Confidence intervals are calculated with a stated probability (i.e., 95%), such that there is a 95% chance that the C.I. covers the true value. The true population value is a constant, even though its value is unknown, but a C.I. is a random quantity whose value depends on the random sample or data from which it is calculated. Therefore we describe a 95% C.I. as having a 95% probability of covering the true value, rather than saying that there is a 95% probability that the true value falls within the confidence interval.

When the number of events is small in relation to the population at risk (i.e., the event is rare), calculation of 95% CIs based on the Poisson probability distribution is recommended. (95% CIs correspond to a p-value of 0.05, so if you are making many comparisons approximately 5% of the comparisons may be statistically significant due to chance alone.) In general, if CIs for two separate rates overlap, there is no statistically significant difference between the two rates. In a one sample case, as for example if one is comparing the age-adjusted rate for a particular racial or ethnic group to a standard value, confidence intervals are equivalent to statistical tests. That is, if a 95% confidence interval around a particular age-adjusted rate excludes the comparison value, then a statistical test for the difference between the two values would be significant at the 0.05 level.¹⁵

A least squares linear regression model was used to perform trend analyses for this report.¹⁶⁻¹⁸ For HP2010 objectives having a minimum of five data points available for trend analysis, if the 95% C.I. of the slope of a linear equation fitted to the data by the method of least squares did not include the value zero then the null hypothesis was rejected at the .05 level of significance and the trend was considered to be statistically significant. As more data points become available, parameter estimates derived from the regression model will be used to calculate predicted values for each objective to the year 2010.

FORMULAS USED IN THIS REPORT

$$\text{IDR} = \left(\frac{nD}{nB} \right) \times 1000$$

$$\text{AADR} = \sum W_a \left(\frac{nD_a}{N\text{pop}_a} \right) \times B$$

$$\text{ASDR} = \left(\frac{nD_a}{N\text{pop}_a} \right) \times B$$

$$\text{SE}_x = \left(\frac{\text{CDR}}{\sqrt{nD}} \right)$$

$$\text{SE}_y = \sqrt{\sum \frac{(W_a \times \text{ASDR})^2}{nD_a}}$$

$$\text{RSE}_x = \left(\frac{\text{SE}_x}{\text{CDR}} \right) \times 100$$

$$\text{RSE}_y = \left(\frac{\text{SE}_y}{\text{AADR}} \right) \times 100$$

$$\text{Lower 95\% CI} = \text{AADR} - (1.96 \times \text{SE}_y) \quad \text{Upper 95\% CI} = \text{AADR} + (1.96 \times \text{SE}_y)$$

- Where:
- IDR = Infant Death Rate
 - AADR = Age-Adjusted Death Rate
 - ASDR = Age-Specific Death Rate
 - n^B = Number of Live Births
 - nD = Number of Deaths

 - Npop = Population Size
 - nD_a = Number of Deaths in an Age Group
 - $N\text{pop}_a$ = Population Size in Same Age Group
 - B = Base (100,000)
 - W_a = Age-Specific Weight (Standard Population Proportion)

 - SE_x = Standard Error of a Crude Rate
 - RSE_x = Relative Standard Error of a Crude Rate
 - SE_y = Standard Error of an Age-Adjusted Death Rate
 - RSE_y = Relative Standard Error of an Age-Adjusted Death Rate

REFERENCES

1. California Department of Public Health, Center for Health Statistics. Death Statistical Master Tape File Documentation.
2. California Department of Public Health, Center for Health Statistics. Documentation for the Underlying and Multiple Cause-of-Death Public Use Data Files.
3. California Department of Public Health, Center for Health Statistics. Birth Statistical Master Tape File Documentation.
4. California Department of Public Health, Center for Health Statistics. California Birth Cohort File Documentation.
5. California Office of Statewide Health Planning and Development, Healthcare Information Resource Center. Patient Discharge Data File Documentation.
6. University of California – Los Angeles, Center for Health Policy Research. California Health Interview Survey Public Use Data Dictionary.
7. California Department of Public Health, Cancer Surveillance and Research Branch, Survey Research Group. California Behavioral Risk Factor Surveillance Survey.
8. U.S. Centers for Disease Control and Prevention, National Center for Health Statistics. United States Census 2000 population with bridged race categories. *Vital and Health Statistics*, Series 2, No. 135, DHHS Publication No. (PHS) 2003-1335, September 2003.
9. California Department of Finance, Demographic Research Unit. Estimated race/ethnic population with age and sex detail, 2000-2050, May 2004.
10. World Health Organization. International Statistical Classification of Diseases and Related Health Problems, Tenth Revision. Geneva, Switzerland, WHO, 1999.
11. U.S. Department of Health and Human Services. Healthy People 2010, 2nd Edition. Washington, D.C.: U.S. Government Printing Office, November 2000.
12. U.S. Public Health Service/Health Care Financing Administration. International Classification of Diseases, 9th Revision, Clinical Modification. 4th ed. Washington, DC: US Department of Health and Human Services, Public Health Service, 1991. DHHS Publication No. (PHS) 91-1260.
13. Klein, R.J. and Schoenborn, C.A. Age adjustment using the 2000 projected U.S. population. Hyattsville, MD: *Statistical Notes*, No. 20. Hyattsville, MD: National Center for Health Statistics, January 2001.
14. Klein, R.J., et al. Healthy People 2010 criteria for data suppression. Hyattsville, MD: National Center for Health Statistics, *Statistical Notes*, No. 24, July 2002.
15. Armitage, P. and Berry, G. *Statistical Methods in Medical Research*. Boston, MA: Blackwell Scientific Publications, 1987.
16. Fleiss, J.L. *Statistical methods for rates and proportions*. New York: John Wiley and Sons, 1981.
17. Kleinman, J.C. State trends in infant mortality. *American Journal of Public Health*, 1986;76(6), 681-687.
18. Oreglia, A. Public Health Rate Program (V.1..0). California Department of Health Services, Health Data and Statistics Branch, July 1990.

End Note

Summary data for each HP2010 objective currently being monitored by the California Department of Public Health are provided in the Executive Summary to this report (<http://cdphinternet/data/indicators/goals/Documents/summary.pdf>). Included in the summary is a table showing the HP2010 target, the latest statewide measure, an indicator of whether the objective was being achieved for the total California population, for each gender, and for each racial and ethnic population for which data were available, with notations on the significance and direction of trends found in the data and the data source.