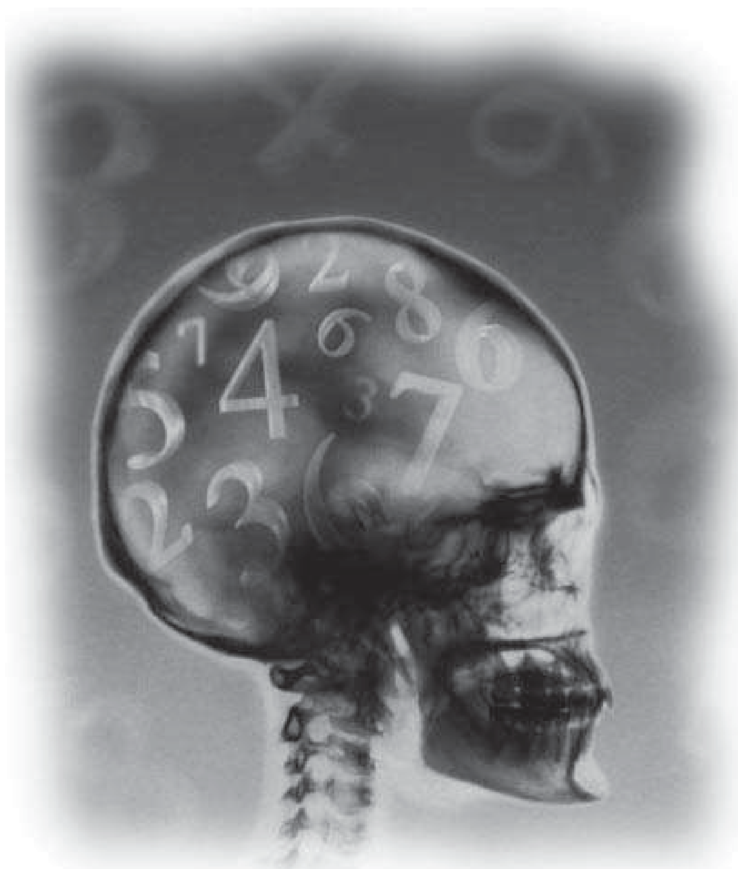
Statistical Methods



Why Do We Use Rates?

A rate provides a meaningful way to compare deaths between population groups of different sizes.

How to calculate a rate

A death rate is calculated by dividing the number of deaths by the total population, and then multiplying the result by a standard population size such as 100,000.

$$\text{Rate} = \frac{\text{Number of Deaths} \times 100{,}000}{\text{Total Population}}$$

How to make comparisons

A higher death rate for a population means that people in that group have an increased risk of dying from a particular disease.

A lower death rate for a population means that people in that group have a lower risk of dying from a particular disease.

A local example:

If we want to compare unintentional injury deaths between Contra Costa and California, it is important that we use rates.
California's population is much larger than that of Contra Costa - we would expect California to have many more unintentional injury deaths. Rates allow us to see if Contra Costa County has proportionally more (or less) of its "fair share" of unintentional injury deaths.

Table 70. Calculating rates: Contra Costa & California Unintentional Injury deaths (2000-2002)

	3 Years of Unintentional Injury Deaths	Population (2000-2002)
Contra Costa	666	2,921,403
California	27,970	104,111,745

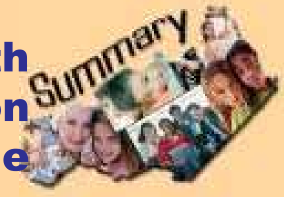
To calculate death rates, we divide the number of deaths in each group by its total population, and then multiply the results by 100,000.

(NOTE: Multiplying our rate by 100,000 does not really change its size. This is simply a statistical tradition, which allows our local rates to be compared to other rates around the world.)

These calculations gives us a rate of 22.8 unintentional injury deaths per 100,000 in Contra Costa and 26.9 unintentional injury deaths per 100,000 in California.
This means that the unintentional injury death rate is higher in California - people living in California have a higher risk of dying from unintentional injury than residents of Contra Costa.

Why are Age-adjusted Rates Important?

Age-adjusted rates allow you to compare health statistics (like death rates) between population groups, even though the size of the groups or the age of group members might be very different.



An age-adjusted rate is the best summary statistic for comparing the impact of diseases like heart disease, cancer, stroke and diabetes that are heavily influenced by age.

Age-adjusted rates are useful for identifying differences that are due to environmental or behavioral risk factors instead of age.

How to calculate an age-adjusted rate

Step 1:
Break up the population into age groups and then track down the number of cases and the corresponding population total for each age group.

Step 2:
Calculate an age-specific rate for each age group.

An age-specific rate per 100,000 is calculated by dividing the number of cases by each group's total population and then multiplying that number by 100,000. For more information about how to calculate rates, please see page 155.

Step 3:
Choose a standard population and find the percentage of the standard population that is found in each age group. This report uses the 2000 US Census standard population to determine the percentage for each age group (shown below).
You can download US standard population files for 1940, 1950, 1960, 1970, 1980, 1990 and 2000 (at no cost) from the National Cancer Institute's website (found at http://seer.cancer.gov/stdpopulations/).

Step 4:
Calculate the weighted age-specific rate for each age group by multiplying each age-specific rate by the percent of the standard population that is found in that particular age group.

Step 5:
The sum of these weighted age-specific rates in a community or race/ethnic group is the age-adjusted rate for that particular community or race/ethnic group. See table below.

Example:

Heart disease deaths among all Contra Costa residents, 2000-2002

Age Group	Number of deaths (2000- 2002)	Population total (2000- 2002)	Age-specific rate*	2000 US Census (%) distribution United States	Weighted age-specific rate*
0-4	2	196,720	1.0	6.8%	0.1
5-9	2	219,369	0.9	7.3%	0.1
10-14	3	226,333	1.3	7.3%	0.1
15-19	3	203,912	1.5	7.2%	0.1
20-24	4	155,259	2.6	6.7%	0.2
25-29	10	166,444	6.0	6.9%	0.4
30-34	9	212,070	4.2	7.3%	0.3
35-39	31	246,185	12.6	8.1%	1.0
40-44	64	253,983	25.2	8.0%	2.0
45-49	120	232,038	51.7	7.1%	3.7
50-54	173	210,307	82.3	6.2%	5.1
55-59	230	157,862	145.7	4.8%	7.0
60-64	292	112,024	260.7	3.8%	9.9
65-69	304	86,837	350.1	3.4%	11.9
70-74	494	77,517	637.3	3.1%	19.8
75-79	785	70,758	1109.4	2.6%	28.8
80-84	974	49,865	1953.3	1.8%	35.2
>85	2123	43,920	4833.8	1.5%	72.5
Total:	5623	2,921,403		99.9%	198.1

Age-Adjusted Rate (w/ 95% confidence interval): *198.1 (192.9 to 203.3)

* Per 100,000 persons.

What is Statistical Significance?

In statistics, "significant" does not mean important. It means "probably true."

Throughout this report, when we say that a difference between two rates or percentages is significant, we mean that we are at least 95% certain this difference is not due to chance alone.

What is a confidence interval?

The confidence interval (CI) shows the lowest and highest boundaries between which we would expect the true rate to fall. A rate is considered a better estimate when it is based on a large number of cases and the confidence interval is narrow. It is common practice to put a 95% confidence interval around a rate.

Important things to remember

Just because two rates appear different, it doesn't mean that the difference is worth talking about.

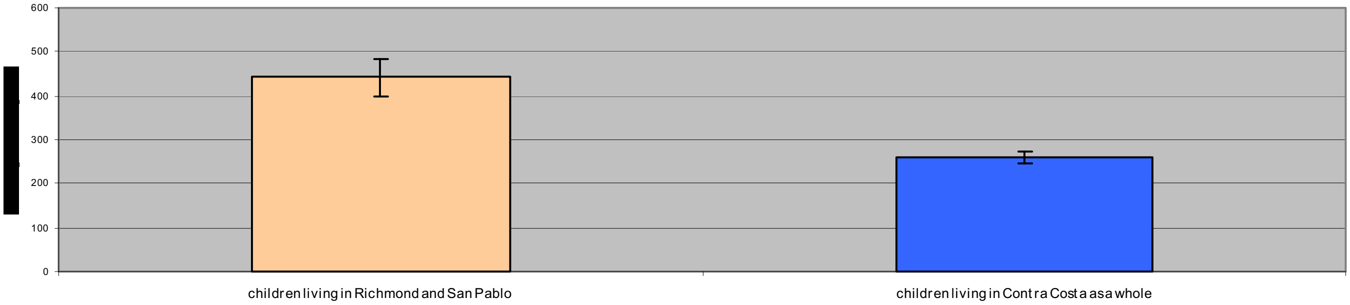
- A difference of 5 deaths per 100,000 may look impressive, but we do not expect this difference to be exactly the same every year. We expect small ups and downs.
- A statistically significant difference is a difference that is greater than would be expected through normal fluctuations.
- A difference is considered statistically significant when the confidence intervals for two rates do not overlap.

Local example:

When looking at our local data, it appears that the crude rate of hospital treatment for asthma is higher among children living in **Richmond** and **San Pablo** (441.6 per 100,000) than among children living in the county as a whole (259.8 per 100,000).

Comparing the confidence intervals
We compare the confidence intervals (Figure 21, page 160) to see if the upper and lower boundaries for children living in **Richmond** and **San Pablo** overlap with the boundaries for children living in Contra Costa as a whole.

Figure 21.



As shown in Fugure 21, the lower and upper boundaries of the 95% confidence interval are shown by "I-Bars." The "I-Bars" show that the lower boundary for children living in Richmond and San Pablo (399.6) is higher than the upper boundary for children living in Contra Costa as a whole (272.3).

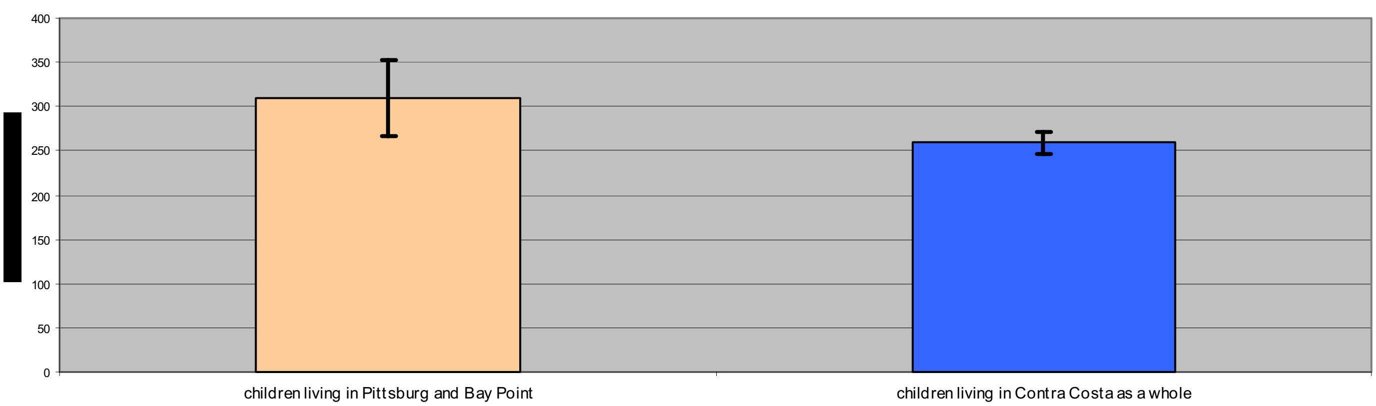
The confidence intervals do not overlap.

This is a statistically significant difference.

Therefore, we can say that the crude rate of hospital treatment for asthma is significantly higher among children living in Richmond and San Pablo compared to children living in Contra Costa as a whole

Another local example:

Figure 22.



It appears that the crude rate of hospital treatments for asthma is higher among children living in **Pittsburg** and **Bay Point** (310.2 per 100,000) than among children living in the county as a whole (259.8 per 100,000).

Comparing the confidence intervals

Again, we look at the confidence intervals to see if this is a statistically significant difference.

When looking at Figure 22, we notice that the lower boundary for children living in Pittsburg and Bay Point (267.3) is lower than the upper boundary for children living in the county as a wholde (272.3). The confidence intervals overlap. **This is NOT a statistically significant difference.**

How to Calculate Relative Risk (RR)

The relative risk (sometimes called "risk ratio") compares the likelihood of dying for a certain population group against the risk of death for all other population groups.

Calculating relative risk draws our attention to unfair racial differences in disease deaths and diagnosed cases.
Relative risk is calculated by dividing the death or disease risk in a specific population group (Group A) by the risk of people from all other groups.

$$RR = \frac{\text{Risk in One Group (Group A)}}{\text{Risk in All Other Groups}}$$

What relative risk tells us

A **relative risk that is greater than 1.0** shows that there is an increased risk among the people in Group A.

- This means if the relative risk was 1.5, people in Group A would be 50% more likely than people in all other groups to die from a cause.

- Or if the relative risk were 3.0, people in Group A would be three times as likely as people from other groups to die from a cause.

A **relative risk that is less than 1.0** indicates that there is a lower risk among the people in Group A.

- If the relative risk were 0.8, people in Group A would be 20% less likely than people in all other groups to die from a cause.

A local example:

Through tests of statistical significance, it was found that African Americans had the highest death rate from AIDS, compared to the Contra Costa County population as a whole.

For this example, African Americans will be our "Group A," the group with the highest death rate from AIDS. Our comparison group will be made up of people from all other race/ethnic groups.

Table 71. Calculating Risk.
Contra Costa County: AIDS Deaths1999-2001

	People who died from AIDS	Population
African Americans (Group A)	50	260,553
All Others	51	2,585,895
Total	101	2,846,448

To calculate the risk in each group, we divide the number of people who died of AIDS by the population totals in each group.

This gives us a risk of 0.0001919 among African Americans (50 divided by 260,553), and a risk of 0.0000198 among people from other race/ethnic groups (51 divided by 2,585,448).

To calculate the relative risk, we divide the risk among African Americans by the risk among people from other race/ethnic groups. This gives us a relative risk of 9.7 (0.0001919 divided by 0.0000198).

From this example, we can say that African Americans are ten times as likely to die from AIDS compared to people from other race/ethnic groups (RR=9.7).

Why Don’t Our Rates Match the Others?

We have better population estimates

Rates in this report may differ from those in other reports because previously reported rates may have been calculated using older population estimates or data from another source.

We sought the most up-to-date population estimates for use in this report. In May 2004, the California Department of Finance (DOF) released a series of population estimates and projections for California cities and counties. These new population estimates document the population growth that has occurred since the 2000 US Census. These estimates are now considered the best data to use when calculating 2000-2002 health statistics for California communities.

The "multi-race" category is new

For the first time, the 2000 US Census allowed county residents to record themselves and their family members as belonging to "one or more races." The California DOF 2000-2002 population data also includes this new category for multi-race persons.

In most cases, the multi-race category is not included in the statistics or data tables. This is for a number of reasons that vary by dataset:

- Death dataset. There was a small number of deaths among multi-race persons leading to unstable rates.
- California Health interview survey. Residents choosing the multi-race option were lumped together with 'other' into a general and ill-defined category: 'other single/two or more race groups.'
- Birth (AVSS) dataset. Residents choosing more than one race group were left uncoded and lumped into a general and ill-defined category: 'other/not coded.'
- Hospitalization dataset. There was no multi-race reporting category for the 2000-2002 data.

Some residents who had previously been categorized into a single-race group are now categorized as multi-racial. This may especially affect the statistics for Asian/ Pacific Islanders, American Indian/Alaska Natives and African Americans.

Hospitalization rates required extra work

Unlike the California DOF population data, the hospital information does not include the race/ethnic category of two or more race groups.

In order to calculate rates using the hospital data, we had to reallocate all multiracial residents found in the DOF population data. These multiracial residents were moved into the other single race groups.

This reallocation was done in accordance with the DOF guidelines that specify that 54% of multiracial residents in Contra Costa should be reallocated as Asian/Pacific Islander, 25% should be reallocated as American Indian/Alaska Native and 21% should be reallocated as African American. (More information about the DOF guidelines for Contra Costa and other California counties is available online at http://www.dof.ca.gov/HTML/DEMOGRAP/MultiraceAllctns2000-2050.htm.)

In this report, the reallocated population totals were used in calculating the rates of hospital treatments for asthma, mental disorders and substance abuse. (See table below.)

For all other health statistics, we used the unmodified race/ethnic population totals from the DOF. (See table ## below.)

Table 72. **Reallocated** mid-year estimates race/ethnicity groups Contra Costa 2000-2002 (3-year total)

	Population Size
White	1,656,936
Hispanic/Latino	549,215
Asian/Pacific Islander	392,148
African American	288,556
American Indian/Alaska Native	34,548
Total:	2,921,403

Table 73. California DOF mid-year estimates for race/ethnic groups and gender
Contra Costa for 2000-2002 (3-year total)

	Men	Women	Population Size
White	805,802	851,134	1,656,936
Hispanic/Latino	281,272	267,943	549,215
Asian	160,949	178,073	339,022
African American	126,981	145,184	272,165
American Indian/ Alaska Native	7,412	7,623	15,035
Native Hawaiian/ Other Pacific Islander	5,270	5,708	10,978
Two or more race groups	**38,480**	**39,572**	**78,052**
Total:	1,426,166	1,495,237	2,921,403

City-level population estimates

Unfortunately, the *California Department of Finance (DOF)* does not provide the mid-year population estimates for cities that are needed to calculate city-level rates.

Based on the DOF city-level population estimates for January 1st 2001 and January 1st 2002, we calculated a 2001 mid-year population estimate for residents in each city. The 2001 mid-year population estimates were then multiplied by three for calculating 2000-2002 rates.

Table 74. Calculated mid-year city-level estimates by gender
Contra Costa 2000-2002 (3-year total)

	Men	Women	Population Size
Antioch	139,336	145,214	284,550
Bay Point	33,577	33,393	66,970
Brentwood	40,448	41,602	82,050
Concord	183,518	188,052	371,569
Martinez	54,326	55,174	109,500
Oakley	39,520	38,705	78,225
Pinole	27,883	30,167	58,050
Pittsburg	86,728	89,822	176,550
Richmond	146,790	155,160	301,950
San Pablo	44,986	46,614	91,601
Walnut Creek	90,924	106,026	196,950
Total:	1,426,166	1,495,237	2,921,403

City-level estimates for men and women

The California DOF does not provide the mid-year population estimates for cities by gender that are needed to calculate city-level rates for men and women.

We assumed that the rate of population growth is the same among men and women, and calculated 2001 mid-year city-level population estimates for male and female residents. These calculations were based on statistics from the 2000 US census and the average rate of population growth in each city between 04/01/2000, the date of the 2000 US Census statistics, and 06/30/2001. The 2001 mid-year population estimates were then multiplied by three for calculating 2000-2002 rates.

City-level estimates by age group

The California DOF does not provide the mid-year population estimates for cities by age group that are needed to calculate age-specific and age-adjusted rates.

We assumed that the rate of population growth is the same in each age group, and calculated 2001 mid-year city-level population estimates for each age group in each city, for all residents, male residents and females residents. These calculations were based on statistics from the 2000 US Census and the average rate of population growth in each city between 04/01/2000, the date of the 2000 US Census statistics, to 06/30/2001. The 2001 mid-year population estimates were then multiplied by three for calculating 2000-2002 rates.

These city-level estimates, by age group and gender, were used for the age-adjusted rates and for the analyses on specific topics such as teen births and asthma hospitalizations among children. These additional denominators are shown in detailed tables with 95% confidence intervals, available online at: http://cchealth.org/health_data/hospital_council/.