

**APPENDIX C**  
**Formulas for Rates**

## Formulas for Rates

SEER\*Stat software was used to calculate rates. The following are the rate algorithms.

### Crude Rate

A crude rate is the number of new cases (or deaths) occurring in a specified population per year, usually expressed as the number of cases per 100,000 population at risk.

$$cruderate = \frac{count}{population} \times 100,000$$

SEER\*Stat allows you to display rates as cases per 1,000; 10,000; 100,000; or 1,000,000.

### Age-adjusted Rate

An age-adjusted rate is a weighted average of crude rates, where the crude rates are calculated for different age groups and the weights are the proportions of persons in the corresponding age groups of a standard population. Several sets of standard populations are included in SEER\*Stat. These include the total U.S. populations (1940, 1950, 1960, 1970, 1980, 1990, and 2000), 1991 and 1996 Canadian populations, European population, and the world population. The age-adjusted rate for an age group comprised of the ages x through y is calculated using the following formula:

$$aarate_{x-y} = \sum_{i=x}^y \left[ \left( \frac{count_i}{pop_i} \right) \times 100,000 \times \left( \frac{stdmili}{\sum_{j=x}^y stdmil_j} \right) \right]$$

where count is the number of cases for the  $i^{th}$  age group,  $pop_i$  is the relevant population for the same age group, and  $stdmili$  is the standard population for the same age group.

### Standard Error for a Crude Rate

This calculation assumes that the cancer counts have Poisson distributions.

$$SE_{crude} = \frac{\sqrt{count}}{population} \times 100,000$$

### Standard Error for an Age-adjusted Rate

This calculation assumes that the cancer counts have Poisson distributions. Suppose that the age-adjusted rate is comprised of age groups x through y.

$$SE_{AARate} = \left[ \sum_{i=x}^y \left( \frac{stdmil_i}{\sum_{j=x}^y stdmil_j} \right)^2 \times \left( \frac{count_i}{population_i^2} \right) \right]^{\frac{1}{2}} \times 100,000$$

### Crude Rate Confidence Intervals

The endpoints of a p x 100% confidence interval are calculated as:

$$CI_{low} = \frac{\left( \frac{1}{2} \left( ChiInv \left( \frac{p}{2}, 2 \times count \right) \right) \right)}{population} \times 100,000$$

$$CI_{high} = \frac{\left( \frac{1}{2} \left( ChiInv \left( 1 - \frac{p}{2}, 2 \times (count + 1) \right) \right) \right)}{population} \times 100,000$$

where  $ChiInv(p,n)$  is the inverse of the chi-squared distribution function evaluated at  $p$  and with  $n$  degrees of freedom, and we define  $ChiInv(p,0) = 0$ .

Although the normal approximation may be used with the standard errors to obtain confidence intervals when the count is large, we use the above exact method that holds even with small counts. When the count is large these two methods produce similar results.

#### Age-adjusted Rate Confidence Intervals

Suppose that the age-adjusted rate is comprised of age groups  $x$  through  $y$ , and let:

$$w_i = \frac{std\bar{r}_i}{\left( pop_i \times \sum_{j=x}^y std\bar{r}_j \right)}$$

$$v = \sum_{i=x}^y (w_i^2 \times count_i)$$

The endpoints of a  $p \times 100\%$  confidence interval are calculated as:

$$CI_{low} = \left( \frac{v}{2 \times rate} \right) \times \left( ChiInv \left( \frac{p}{2}, \frac{(2 \times rate^2)}{v} \right) \right) \times 100,000$$

$$CI_{high} = \left( \frac{v}{2 \times rate} \right) \times \left( ChiInv \left( 1 - \frac{p}{2}, \frac{(2 \times rate^2)}{v} \right) \right) \times 100,000$$

This method for calculating the confidence interval was initially developed by Fay and Feuer (1997) and further modified by Tiwari et al. (2006) on the algorithm for the upper confidence limit. The method produces similar confidence limits to the standard normal approximation when the counts are large and the population being studied is similar to the standard population. In other cases, the above method is more likely to ensure proper coverage.

Note: The rate used in the above formulas is not per 100,000 population.

#### References:

Johnson NL, Kotz S. Distributions in Statistics: Discrete Distributions. Johns Wiley, New York, 1969.

Fay MP, Feuer EJ. Confidence intervals for directly standardized rates: method based on the gamma distribution. *Statistics in Medicine* 1997 Apr 15;16(7):791-801.

Tiwari RC, Clegg LX, Zou Z. Efficient interval estimation for age-adjusted cancer rates. *Stat Methods Med Res* 2006 Dec;15(6):547-69.