Surveillance Research Program, NCI, Technical Report #2014-01

Variance of Delay-adjusted Age-adjusted Rates

Huann-Sheng Chen[1], Joe Zou[2], Doug Midthune[3], Martin Krapcho[2], Steve Scoppa[2], Eric J Feuer[1]

1. Statistical Methodologies and Applications Branch, Division of Cancer Control and Population Sciences, National Cancer Institute
2. Information Management Services Inc
3. Biometry Research Group, Division of Cancer Prevention, National Cancer Institute

Overview

The appropriate delay factor is matched to each record in SEER*Stat using the matching factor for each record. These factors potentially include age group, year of diagnosis, race, gender, and cancer site. The delay adjusted rates are computed by using weighted averages of the cases and this will produce proper delay adjusted rate estimates for cases across any grouping of registries. The variances of the delay-adjusted, age-adjusted rates can then be calculated. However, the variances are much more complicated and difficult to implement in SEER*Stat because of the covariances between the factors, which requires a large amount of storages of covariance matrices in SEER*Stat. To lessen the computing burden of the SEER*Stat, an approximation of the variance is used in the implementation. It has been shown empirically that the approximation tends to be very close to the exact variance.

Method

To illustrate the method, assume that we are interested in age adjusted rate for J age groups. For each cluster i $(i = 1, \dots, I)$ and an age group j $(j = 1, \dots, J)$, there is a delay adjustment factor, A_{ij}. J is usually equal to 19. Note that the delay adjustment factors A_{ij}'s are obtained from the delay model. The delay-adjusted age-adjusted rate (aarate) can be calculated as

$$Delay - adjusted\ aarate\ = \sum_{j=1}^{J} \frac{\sum_i c_{ij} A_{ij}}{\sum_i p_{ij}}\ s_j, \qquad (1)$$

where c_{ij} is the count, p_{ij} is the population, s_j is the standard population proportion and $\sum_{j=1}^{J} s_j = 1$. Note that A_{ij} is a function of $\beta = (\beta_1, \dots, \beta_k)'$, a parameter vector derived from the delay model. The count c_{ij} is assumed to follow a Poisson distribution. Also denote $c = \left(c_{11}, \dots, c_{IJ}\right)'$. The variance-covariance matrix of $(c, \beta)'$ is

$$Var\binom{c}{\beta} = \begin{pmatrix} Var(c) & 0 \\ 0 & Var(\beta) \end{pmatrix},$$

where

$$var(c) = \begin{pmatrix} c_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & c_{IJ} \end{pmatrix}.$$

Using the delta method, the variance of the delay-adjusted age-adjusted rate can be written as

$$var(delay\ adusted\ aarate)$$
$$= \sum_{j=1}^{J} \frac{\sum_i c_{ij} A_{ij}^2}{(\sum_i p_{ij})\hat{}2} s_j^2 + \left(\sum_{j=1}^{J} \frac{\sum_i c_{ij} \left(\frac{\partial A_{ij}}{\partial \beta}\right)'}{\sum_i p_{ij}} s_j \right) var(\beta) \left(\sum_{j=1}^{J} \frac{\sum_i c_{ij} \frac{\partial A_{ij}}{\partial \beta}}{\sum_i p_{ij}} s_j \right) \qquad (2)$$

To implement the variance of the delay-adjusted age adjusted rate, the second term in equation (2) requires a large amount of information and calculation. To simplify the calculation, we use the first term as an approximation:

$$var(delay\ adusted\ aarate) \approx \sum_{j=1}^{J} \frac{\sum_i c_{ij} A_{ij}^2}{\left(\sum_i p_{ij}\right)^2} s_j^2 \qquad (3)$$

A numerical study was conducted using the SEER*Stat to compare the variances (2) and (3). Across different cancer sites, registries, races, gender, and reporting year, for a total of 7894 models, the result show that the ratio of the standard error calculated from approximation (3) and from exact variance (2) ranges between (0.9834, 1.0066). It indicates that the approximation variance (3) tends to be very close to the exact variance (2).