Istituto di Ricerche sulla Popolazione
e le Politiche Sociali - CNR

IRPPS Working Papers

ISSN 2240-7332

# Numerical estimation of the variance of the completeness index applied to cancer data

Anna Gigli

**What is IRPPS?**

**IRPPS** is an Interdisciplinary Research Institute that conducts studies on demographic and migration issues, welfare systems and social policies, on policies regarding science, technology and higher education, on the relations between science and society, as well as on the creation of, access to and dissemination of knowledge and information technology.

**www.irpps.cnr.it**

IRPPS WPs n. 13 (2007)

# Numerical estimation of the variance of the completeness index applied to cancer data

*Anna Gigli*

**Abstract**

Cancer prevalence is the proportion of people in a population diagnosed with cancer in the past and still alive. One way to estimate prevalence is via population-based registries, where data on diagnosis and life status of all incident cases occurring in the covered population are collected. In this report a numerical method for the estimation of the variance of the completeness index (NUMCOMP) is developed, and comparisons are made with a previous analytical method (VARCOMP) proposed by Gigli et al.(2006).

The paper is organized as follows: section 1 introduces the problem; section 2 illustrates the new method; section 3 describes the algorithm in details; finally in section 4 the new and old methods are applied to the following cancer sites: all sites, anus, brain, colorectal for males and females, and to breast cancer for females (all races/ethnicities) and results are compared and commented.

**Keywords:** Complete prevalence, Cancer registries, Incidence, Survival, Numerical derivatives, SEER*Stat software

**Citazione consigliata**:

Gigli, Anna. Numerical estimation of the variance of the completeness index applied to cancer data. *IRPPS Working Papers*, n. 13, 2007.

**Anna Gigli** è ricercatrice e presso presso l'Istituto Superiore di Sanità (e-mail: anna.gigli@irpps.cnr.it).

Istituto di Ricerche sulla Popolazione e le Politiche Sociali - CNR
Via Palestro, 32 - 00185 Roma
http://www.irpps.cnr.it/it

# 1 Introduction

The Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute annually reports the number of persons alive following a diagnosis of cancer, or *complete prevalence* of cancer. This statistic is estimated in 2 steps:

1. The number of alive cancer cases reported to the SEER data after 1975 are counted and survivors among people lost to follow-up are estimated and added. This is denoted *limited duration prevalence* and can be calculated using the SEER*Stat software.

2. The proportion of prevalence that is unobserved, i.e. prevalence of cases diagnosed prior to 1975, is estimated using the *completeness index* method. Limited duration prevalence is adjusted to represent complete prevalence or lifetime prevalence. The completeness index method (Capocaccia and De Angelis, 1997) is implemented into COMPREV, a new software that calculates complete prevalence by adjusting limited duration prevalence imported from SEER*Stat with the completeness index.

Variance of limited duration prevalence is given in Clegg *et al.*(2002), variance of the completeness index (VARCOMP) and consequently of complete prevalence is given in Gigli *et al.*(2006). VARCOMP is based on analytical approximation of the variance and is implemented in the COMPREV software. From a practical point of view VARCOMP has a limitation: since the completeness index method, used to estimate complete prevalence, is based on the parametric modelling of the incidence and survival functions, VARCOMP is based on analytical derivatives of the completeness index with respect to the incidence and survival parameters. Thus it depends on the analytical models used to estimate incidence and survival. Experience has shown that a single parametric form of survival and/or incidence functions is not able to fit all different cancer sites; for example, for acute lymphocytic leukemia we need a survival function by at least two age groups, to describe the childhood and adult acute lymphocytic leukemia survival.

Therefore it is necessary to develop an approximation to VARCOMP which is independent of the parametric assumptions.

In this report a numerical method for the estimation of the variance of the completeness index (NUMCOMP) is developed, and comparisons are made with the old method.

The paper is structured as follows: section 2 illustrates the new method; section 3 describes the algorithm in details; finally in section 4 the new and old methods are applied to the following cancer sites: all sites, anus, brain, colorectal for males and females, and to breast cancer for females (all races/ethnicities); results are compared.

# 2 Numerical estimation of the variance of the completeness index

We briefly recall some notation and the previous method (VARCOMP); further details can be found in Gigli *et al.*(2006).

For a fixed birth cohort $c$ and a fixed age at prevalence $x$

- *complete prevalence* $N_x(0,x)$ is defined as the portion of people aged $x$ alive on a certain date who had been diagnosed of the disease between ages 0 and $x$;
- *limited duration prevalence* $\tilde{N}_x(x-L,x)$ is the prevalence at age $x$ estimated by population-based cancer registries and is based on a limited observational period $L$ (Gail *et al.*, 1999);
- *modelled prevalence* $\hat{N}_x(0,x;\underline{\hat{\psi}})$ is a parametric estimate of the prevalence at age $x$, based on a complex function (convolution) of incidence and survival parametric models; $\underline{\hat{\psi}}$ is the maximum likelihood estimate of the vector of the incidence and survival parameters (Verdecchia *et al.*, 1989);
- *completeness index* $R_x(L;\underline{\hat{\psi}}) = \dfrac{\hat{N}_x(x-L,x;\underline{\hat{\psi}})}{\hat{N}_x(0,x;\underline{\hat{\psi}})}$ is the proportion of modelled prevalence at age $x$ that is observed (Capocaccia and De Angelis, 1997);
- an estimate of the complete prevalence at age $x$ is obtained by combining the limited duration prevalence and the completeness index

$$N_x^*(0,x;\underline{\hat{\psi}}) = \frac{\tilde{N}_x(x-L,x)}{R_x(L;\underline{\hat{\psi}})}$$

- the analytical approximation to the variance of the completeness index is

$$\hat{\text{var}}_x = \text{var}[R_x(L;\hat{\underline{\psi}})] \approx \left(\frac{\partial R_x}{\partial \underline{\psi}}\Big|_{\underline{\psi}=\hat{\underline{\psi}}}\right)^T \hat{\boldsymbol{V}} \left(\frac{\partial R_x}{\partial \underline{\psi}}\Big|_{\underline{\psi}=\hat{\underline{\psi}}}\right), \qquad (1)$$

  where $\hat{\boldsymbol{V}}$ is the covariance matrix of the mle vector $\hat{\psi}$ and $\partial R_x/\partial\underline{\psi}$ is the vector of partial derivatives of $R_x$ with respect to the components of the parameter vector $\underline{\psi}$ (Gigli *et al.*, 2006);

Equation (1) corresponds to the diagonal of the covariance matrix of $R_x(L;\hat{\underline{\psi}})$. In this report we will estimate the whole covariance matrix of the completeness indices computed at different ages

$$\text{cov}(R_{x_i}, R_{x_j}) \approx \left(\frac{\partial R_{x_i}}{\partial \underline{\psi}}\Big|_{\underline{\psi}=\hat{\underline{\psi}}}\right)^T \hat{\boldsymbol{V}} \left(\frac{\partial R_{x_j}}{\partial \underline{\psi}}\Big|_{\underline{\psi}=\hat{\underline{\psi}}}\right). \qquad (2)$$

The computation of $\partial R_{x_i}/\partial\underline{\psi}$ for each component of $\underline{\psi}$ and for each age group $x_i$ is to be solved by means of successive iterations.

Let $\hat{\underline{\psi}} = (\hat{\psi}_1, \ldots, \hat{\psi}_p) = (\psi_1^{(0)}, \ldots, \psi_p^{(0)})$ represent the 0-th iteration of the vector of maximum likelihood incidence and survival estimates.

For a given age at prevalence $x_i$ and parameter estimate $\hat{\psi}_s$, we apply the definition of derivative (limit of the incremental ratio) to smaller and smaller right intervals:

$$\Delta_{i,s}^{(k)+} = \frac{R_{i,s}^{(k)+} - R_i}{h_s^{(k)}},$$

where $h_s^{(k)}$ is the k-th increment from the original s-th parameter estimate: $\psi_s^{(0)} + h_s^{(k)}$, and $R_{i,s}^{(k)+} = R_{x_i}(L;\underline{\psi}_s^{(k)+})$ is the completeness index computed for age group $x_i$ and estimated parameter vector $\underline{\psi}_s^{(k)+} = (\psi_1^{(0)}, \ldots, \psi_s^{(0)} + h_s^{(k)}, \ldots, \psi_p^{(0)})$. Notice that in $R_{i,s}^{(k)+}$ first subscript refers to age group, second subscript refers to the iterating parameter component, superscript refers to iteration number, + or - refer to right or left iteration.

For fixed $\epsilon$ and $\delta$, when

$$|\Delta_{i,s}^{(k)+} - \Delta_{i,s}^{(k-1)+}| < \epsilon \qquad (1^{st} \text{ condition})$$

we have reached convergence and $\Delta_{i,s}^{(k)+}$ could be the derivative;
we further verify the continuity of the derivative by checking if the right and

the left derivative coincide

$$|\Delta_{i,s}^{(k)+} - \Delta_{i,s}^{(k)-}| < \delta \qquad (2^{nd} \text{ condition}).$$

Here $\Delta_{i,s}^{(k)-} = \dfrac{R_{i,s}^{(k)-} - R_i}{h_s^{(k)}}$, where $R_{i,s}^{(k)-} = R_{x_i}(L; \underline{\psi}_s^{(k)-})$ is the completeness index computed for age group $x_i$ and estimated parameter vector $\underline{\psi}_s^{(k)-} = (\psi_1^{(0)}, \ldots, \psi_s^{(0)} - h_s^{(k)}, \ldots, \psi_p^{(0)})$.

Notice that, since the denominators of $\Delta_{i,s}^{(k)+}$ and $\Delta_{i,s}^{(k)-}$ are the same, the second condition becomes:

$$|R_{i,s}^{(k)+} - R_{i,s}^{(k)-}| < \delta.$$

If both condition holds, the approximation to the derivative related to the $s$-th component of the vector $\underline{\hat{\psi}}$ is

$$\frac{\partial R_{x_i}}{\partial \psi_s} \approx \Delta_{i,s}^{(k)+} = \frac{R_{i,s}^{(k)+} - R_i}{h_s^{(k)}}.$$

If not, we proceed with a further iteration, by setting $h_s^{(k+1)} = h_s^{(k)}/2$.
We repeat the procedure for every parameter $\hat{\psi}_s$ and every age group $x_i$, to compute the whole covariance matrix (2). Therefore the numerical approximation to the variance of $R_x$ will be

$$\tilde{\text{var}}_x = \left(\Delta_x^{(k)+}\right)^T \hat{\boldsymbol{V}} \left(\Delta_x^{(k)+}\right), \tag{3}$$

where $\Delta_x^{(k)+}$ is a vector of numerical approximations to the partial derivatives of $R_x$ with respect to the components of the vector $\psi$.

# 3 The algorithm

1. Input:
   a) the number of parameters $p$;
   b) the number of age classes $n$;
   c) the length of the registry $L$;
   d) the maximum likelihood estimates of incidence and survival parameters: $\underline{\psi}^{(0)} = (\hat{\psi}_1, \ldots, \hat{\psi}_s, \ldots, \hat{\psi}_p)$;
   e) the covariance matrix of the mle: $V = \text{cov}(\underline{\psi}^{(0)})$.

2. Fix the age at prevalence $x_i$ and the component $\hat{\psi}_s$ of the vector $\underline{\psi}^{(0)}$
3. Compute $R_{x_i} = R_{x_i}(L; \underline{\psi}^{(0)})$, the original completeness index
4. Fix $k = 1, \epsilon = \delta = 10^{-6}$
5. Fix $h_s^{(1)} = \mathrm{sd}(\psi_s^{(0)})/100$, i.e. one-hundredth of the standard deviation of the s-th component of the parameter estimate $\mathrm{sd}(\psi_s^{(0)})$
6. Compute $\underline{\psi}_s^{(1)+} = (\psi_1^{(0)}, \ldots, \psi_s^{(0)} + h_s^{(1)}, \ldots, \psi_p^{(0)})$ - where only the s-th component of the vector changes
7. Compute $R_{i,s}^{(1)+} = R_{x_i}(L; \underline{\psi}_s^{(1)+})$
8. Compute $\Delta_{i,s}^{(1)+} = \dfrac{R_{i,s}^{(1)+} - R_i}{h_s^{(1)}}$
9. Let $k \to k + 1$
10. Let $h_s^{(k)} = h_s^{(k-1)}/2$
11. Compute $\underline{\psi}_s^{(k)+} = (\psi_1^{(0)}, \ldots, \psi_s^{(0)} + h_s^{(k)}, \ldots, \psi_p^{(0)})$
12. Compute $R_{i,s}^{(k)+} = R_{x_i}(L; \underline{\psi}_s^{(k)+})$
13. Compute $\Delta_{i,s}^{(k)+} = \dfrac{R_{i,s}^{(k)+} - R_i}{h_s^{(k)}}$
14. If $|\Delta_{i,s}^{(k)+} - \Delta_{i,s}^{(k-1)+}| \geq \epsilon$ then goto 9
15. Compute $\underline{\psi}_s^{(k)-} = (\psi_1^{(0)}, \ldots, \psi_s^{(0)} - h_s^{(k)}, \ldots, \psi_p^{(0)})$
16. Let $R_{i,s}^{(k)-} = R_{x_i}(L; \underline{\psi}_s^{(k)-})$
17. If $|R_{i,s}^{(k)+} - R_{i,s}^{(k)-}| \geq \delta$ then goto 9
18. The s-th partial derivative of $R_{x_i}$ is: $\dfrac{\partial R_{x_i}}{\partial \psi_s}\Big|_{\psi_s = \hat{\psi}_s} = \Delta_{i,s}^{(k)+}$.
    Repeat 2–17 for each parameter $\psi_s \in (\psi_1, \ldots, \psi_p)$ and for each age group $x_i \in (x_1, \ldots, x_n)$.
19. Compute $\mathrm{cov}(R_i, R_j)$ as in formula (2):

$$\mathrm{cov}(R_i, R_j) = (\Delta_{i,s}^{(k)+})^T \hat{\boldsymbol{V}} (\Delta_{i,s}^{(k)+})$$

# 4 Results

The algorithm has been applied to the following cancer sites: all sites, anus, brain, colorectal for males and females, and to breast cancer for females (all races/ethnicities). Data come from the SEER-9 registries and refer to the period 1975–1999 (i.e. 25 years of observations).

To evaluate the performances of the old and new estimator two indicators are used:

1. the *coefficient of variation* (CV), which is the ratio between the standard deviation of the completeness index R and the completeness index itself
$$CV_x = \frac{\mathrm{sd}(R_x(L;\hat{\underline{\psi}}))}{R_x(L;\hat{\underline{\psi}})};$$
$CV_x$ is a dimensionless indicator, independent of the sample size. It is computed for analytical and numerical standard deviations which are then compared in absolute value:
$$\Delta CV_x = \left| \frac{\hat{\mathrm{sd}}_x - \tilde{\mathrm{sd}}_x}{R_x(L;\hat{\underline{\psi}})} \right|, \tag{4}$$
where $\hat{\mathrm{sd}}_x = \sqrt{\hat{\mathrm{var}}_x}$, as obtained in (1), and $\tilde{\mathrm{sd}}_x = \sqrt{\tilde{\mathrm{var}}_x}$, as obtained in (3);

2. the percent difference between analytical numerical standard deviation
$$\%\mathrm{diff}_x = \frac{\hat{\mathrm{sd}}_x - \tilde{\mathrm{sd}}_x}{\hat{\mathrm{sd}}_x} \times 100; \tag{5}$$
this is not a dimensionless indicator and may depend on the sample size, however it provides an idea of the relative difference between the two standard deviations.

Tables 1–9 describe the results for 5-year age classes.

The agreement between the two estimators is quite satisfactory, with the exception of brain cancer (tables 5 and 6), for which further investigation is required.

The numerical estimator is systematically smaller than the analytical estimator (hence %diff $\geq$ 0), however the difference of the coefficient of

variations is very small (never larger than $O(10^{-4})$). The maximum difference between the standard deviations (in percentage) is generally less than 2% and varies between 1.03% for anus cancer males and 18.87% for all sites cancer females (in older ages, where data is scarcer and variance is wider).

Brain cancer has a completely different behaviour: both analytical and numerical standard deviations are very high and very different between themselves. Although data on brain cancer is quite scarse, this alone does not explains the phenomenon. Further study is necessary to evaluate the goodness-of-fit of the incidence and survival models, which provide the basis for the standard deviation calculations.

# References

Capocaccia R, De Angelis R (1997). Estimating the completeness of prevalence based on cancer registry data. *Statistics in Medicine*, **16**, 425–440.

Clegg LX, Gail MH, Feuer EJ (2002). Estimating the variance of disease-prevalence estimates from population-based registries. *Biometrics*, **55**, 1137–1144.

Gail MH, Kessler L, Midthune D, Scoppa S (1999). Two approaches for estimating disease prevalence from population-based registries of incidence and total mortality. *Biometrics*, **55**, 1137–1144.

Gigli A, Mariotto A, Clegg LX, Tavilla A, Corazziari I, Hachey M, Scoppa S, Capocaccia R (2006). Estimating the variance of cancer prevalence from population-based registries. *Statistical Methods for Medical Research*, **15**, p. 235–253.

Verdecchia A, Capocaccia R, Egidi V, Golini A (1989). A method for the estimation of chronic disease morbidity and trends from mortality data. *Statistics in Medicine,* **8***, 201–216.*

Table 1: Completeness index, analytical and numerical standard deviation, $\Delta CV$, %diff. Data on male all sites cancer, SEER-9 registries in period 1975-1999.

| age interval | $R$ | $\hat{\text{sd}}$ | $\tilde{\text{sd}}$ | $\Delta CV$ | %diff |
|---|---|---|---|---|---|
| 0 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 1–4 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 5–9 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 10–14 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 15–19 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 20–24 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 25–29 | 0.95 | 0.74 | 0.74 | 0.0000000 | 0.00 |
| 30–34 | 0.93 | 1.11 | 1.11 | 0.0000011 | 0.01 |
| 35–39 | 0.93 | 0.96 | 0.96 | 0.0000032 | 0.03 |
| 40–44 | 0.94 | 0.72 | 0.72 | 0.0000043 | 0.06 |
| 45–49 | 0.94 | 0.50 | 0.50 | 0.0000053 | 0.10 |
| 50–54 | 0.95 | 0.33 | 0.33 | 0.0000063 | 0.18 |
| 55–59 | 0.96 | 0.22 | 0.22 | 0.0000083 | 0.37 |
| 60–64 | 0.97 | 0.15 | 0.15 | 0.0000104 | 0.66 |
| 65–69 | 0.97 | 0.12 | 0.12 | 0.0000113 | 0.91 |
| 70–74 | 0.97 | 0.11 | 0.11 | 0.0000165 | 1.43 |
| 75–79 | 0.96 | 0.12 | 0.12 | 0.0000238 | 1.90 |
| 80–84 | 0.95 | 0.15 | 0.15 | 0.0000388 | 2.47 |
| 85+ | 0.94 | 0.20 | 0.20 | 0.0000533 | 2.48 |

Table 2: Completeness index, analytical and numerical standard deviation, $\Delta CV$, %diff. Data on female all sites cancer, SEER-9 registries in period 1975-1999.

| age interval | $R$ | $\hat{\text{sd}}$ | $\tilde{\text{sd}}$ | $\Delta CV$ | %diff |
|---|---|---|---|---|---|
| 0 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 1–4 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 5–9 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 10–14 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 15–19 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 20–24 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 25–29 | 0.94 | 0.03 | 0.03 | 0.0000074 | 2.17 |
| 30–34 | 0.91 | 0.04 | 0.04 | 0.0000055 | 1.17 |
| 35–39 | 0.92 | 0.08 | 0.08 | 0.0000076 | 0.85 |
| 40–44 | 0.93 | 0.09 | 0.09 | 0.0000183 | 1.89 |
| 45–49 | 0.93 | 0.09 | 0.08 | 0.0000300 | 3.20 |
| 50–54 | 0.93 | 0.09 | 0.08 | 0.0000397 | 4.32 |
| 55–59 | 0.92 | 0.09 | 0.08 | 0.0000541 | 5.61 |
| 60–64 | 0.91 | 0.10 | 0.09 | 0.0000780 | 7.16 |
| 65–69 | 0.89 | 0.12 | 0.11 | 0.0001186 | 9.10 |
| 70–74 | 0.87 | 0.14 | 0.13 | 0.0001845 | 11.36 |
| 75–79 | 0.85 | 0.18 | 0.15 | 0.0002876 | 13.86 |
| 80–84 | 0.82 | 0.22 | 0.18 | 0.0004444 | 16.52 |
| 85+ | 0.79 | 0.28 | 0.23 | 0.0006667 | 18.87 |

Table 3: Completeness index, analytical and numerical standard deviation, $\Delta CV$, %diff. Data on male anus cancer, SEER-9 registries in period 1975-1999.

| age interval | $R$ | $\hat{\text{sd}}$ | $\tilde{\text{sd}}$ | $\Delta CV$ | %diff |
|---|---|---|---|---|---|
| 0 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 1–4 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 5–9 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 10–14 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 15–19 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 20–24 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 25–29 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 30–34 | 1.00 | 0.01 | 0.01 | 0.0000000 | 0.00 |
| 35–39 | 1.00 | 0.05 | 0.05 | 0.0000000 | 0.00 |
| 40–44 | 0.99 | 0.13 | 0.13 | 0.0000010 | 0.08 |
| 45–49 | 0.98 | 0.23 | 0.23 | 0.0000031 | 0.13 |
| 50–54 | 0.97 | 0.35 | 0.35 | 0.0000082 | 0.23 |
| 55–59 | 0.96 | 0.47 | 0.47 | 0.0000157 | 0.32 |
| 60–64 | 0.94 | 0.60 | 0.59 | 0.0000266 | 0.42 |
| 65–69 | 0.92 | 0.71 | 0.71 | 0.0000412 | 0.53 |
| 70–74 | 0.90 | 0.82 | 0.82 | 0.0000597 | 0.65 |
| 75–79 | 0.89 | 0.93 | 0.92 | 0.0000824 | 0.78 |
| 80–84 | 0.87 | 1.03 | 1.02 | 0.0001072 | 0.90 |
| 85+ | 0.85 | 1.13 | 1.12 | 0.0001366 | 1.03 |

Table 4: Completeness index, analytical and numerical standard deviation, $\Delta CV$, %diff. Data on female anus cancer, SEER-9 registries in period 1975-1999.

| age interval | $R$ | $\hat{\text{sd}}$ | $\tilde{\text{sd}}$ | $\Delta CV$ | %diff |
|---|---|---|---|---|---|
| 0 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 1–4 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 5–9 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 10–14 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 15–19 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 20–24 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 25–29 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 30–34 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 35–39 | 1.00 | 0.03 | 0.03 | 0.0000000 | 0.00 |
| 40–44 | 0.99 | 0.08 | 0.08 | 0.0000000 | 0.00 |
| 45–49 | 0.99 | 0.15 | 0.15 | 0.0000000 | 0.00 |
| 50–54 | 0.98 | 0.23 | 0.23 | 0.0000031 | 0.13 |
| 55–59 | 0.96 | 0.32 | 0.32 | 0.0000083 | 0.25 |
| 60–64 | 0.95 | 0.40 | 0.40 | 0.0000168 | 0.40 |
| 65–69 | 0.94 | 0.49 | 0.49 | 0.0000353 | 0.68 |
| 70–74 | 0.92 | 0.57 | 0.57 | 0.0000652 | 1.05 |
| 75–79 | 0.91 | 0.65 | 0.64 | 0.0001115 | 1.55 |
| 80–84 | 0.89 | 0.73 | 0.72 | 0.0001796 | 2.18 |
| 85+ | 0.88 | 0.82 | 0.79 | 0.0002804 | 3.01 |

Table 5: Completeness index, analytical and numerical standard deviation, $\Delta CV$, %diff. Data on male brain cancer, SEER-9 registries in period 1975-1999.

| age interval | $R$ | $\hat{\text{sd}}$ | $\tilde{\text{sd}}$ | $\Delta CV$ | %diff |
|---|---|---|---|---|---|
| 0 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 1–4 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 5–9 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 10–14 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 15–19 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 20–24 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 25–29 | 0.91 | 0.43 | 0.43 | 0.0000154 | 0.33 |
| 30–34 | 0.77 | 0.93 | 0.90 | 0.0002978 | 2.48 |
| 35–39 | 0.71 | 1.23 | 1.16 | 0.0009658 | 5.61 |
| 40–44 | 0.70 | 1.44 | 1.30 | 0.0019874 | 9.65 |
| 45–49 | 0.71 | 1.64 | 1.41 | 0.0033209 | 14.27 |
| 50–54 | 0.72 | 1.89 | 1.53 | 0.0050651 | 19.36 |
| 55–59 | 0.74 | 2.23 | 1.67 | 0.0075037 | 25.02 |
| 60–64 | 0.76 | 2.69 | 1.84 | 0.0111398 | 31.42 |
| 65–69 | 0.77 | 3.35 | 2.05 | 0.0169049 | 38.94 |
| 70–74 | 0.79 | 4.35 | 2.26 | 0.0265148 | 47.98 |
| 75–79 | 0.80 | 5.93 | 2.46 | 0.0432398 | 58.54 |
| 80–84 | 0.82 | 8.72 | 2.63 | 0.0739977 | 69.88 |
| 85+ | 0.85 | 14.38 | 3.02 | 0.1343128 | 79.01 |

Table 6: Completeness index, analytical and numerical standard deviation, $\Delta CV$, %diff. Data on female brain cancer, SEER-9 registries in period 1975-1999.

| age interval | $R$ | $\hat{\text{sd}}$ | $\tilde{\text{sd}}$ | $\Delta CV$ | %diff |
|---|---|---|---|---|---|
| 0 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 1–4 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 5–9 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 10–14 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 15–19 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 20–24 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 25–29 | 0.90 | 0.49 | 0.49 | 0.0000212 | 0.39 |
| 30–34 | 0.74 | 1.01 | 0.99 | 0.0003047 | 2.23 |
| 35–39 | 0.68 | 1.39 | 1.32 | 0.0009660 | 4.76 |
| 40–44 | 0.68 | 1.78 | 1.60 | 0.0025918 | 9.87 |
| 45–49 | 0.70 | 2.31 | 1.89 | 0.0060049 | 18.28 |
| 50–54 | 0.74 | 3.04 | 2.12 | 0.0122441 | 30.03 |
| 55–59 | 0.79 | 3.97 | 2.22 | 0.0221362 | 43.97 |
| 60–64 | 0.83 | 5.14 | 2.16 | 0.0360449 | 57.96 |
| 65–69 | 0.86 | 6.65 | 2.00 | 0.0542797 | 69.99 |
| 70–74 | 0.88 | 8.72 | 1.81 | 0.0783444 | 79.20 |
| 75–79 | 0.90 | 11.89 | 1.68 | 0.1132879 | 85.91 |
| 80–84 | 0.92 | 17.57 | 1.61 | 0.1740393 | 90.86 |
| 85+ | 0.93 | 30.40 | 1.72 | 0.3096149 | 94.35 |

Table 7: Completeness index, analytical and numerical standard deviation, $\Delta CV$, %diff. Data on female breast cancer, SEER-9 registries in period 1975-1999.

| age interval | $R$ | $\hat{\text{sd}}$ | $\tilde{\text{sd}}$ | $\Delta CV$ | %diff |
|---|---|---|---|---|---|
| 0 | 1.00 | 0.00 | 0.00 | 0.000000 | 0.00 |
| 14 | 1.00 | 0.00 | 0.00 | 0.000000 | 0.00 |
| 5–9 | 1.00 | 0.00 | 0.00 | 0.000000 | 0.00 |
| 1014 | 1.00 | 0.00 | 0.00 | 0.000000 | 0.00 |
| 1519 | 1.00 | 0.00 | 0.00 | 0.000000 | 0.00 |
| 2024 | 1.00 | 0.00 | 0.00 | 0.000000 | 0.00 |
| 2529 | 1.00 | 0.00 | 0.00 | 0.000000 | 0.00 |
| 3034 | 1.00 | 0.00 | 0.00 | 0.000000 | 0.00 |
| 35–39 | 1.00 | 0.00 | 0.00 | 0.000000 | 0.00 |
| 4044 | 1.00 | 0.00 | 0.00 | 0.000000 | 0.00 |
| 45–49 | 1.00 | 0.01 | 0.01 | 0.000000 | 0.00 |
| 5054 | 1.00 | 0.02 | 0.02 | 0.000000 | 0.00 |
| 55–59 | 0.99 | 0.04 | 0.04 | 0.000001 | 0.28 |
| 60–64 | 0.97 | 0.07 | 0.07 | 0.000008 | 1.20 |
| 65–69 | 0.94 | 0.11 | 0.10 | 0.000026 | 2.28 |
| 7074 | 0.91 | 0.14 | 0.14 | 0.000057 | 3.62 |
| 75–79 | 0.88 | 0.18 | 0.17 | 0.000107 | 5.37 |
| 8084 | 0.86 | 0.21 | 0.19 | 0.000175 | 7.29 |
| 85+ | 0.83 | 0.25 | 0.22 | 0.000252 | 8.51 |

Table 8: Completeness index, analytical and numerical standard deviation, $\Delta CV$, %diff. Data on male colorectal cancer, SEER-9 registries in period 1975-1999.

| | | | | | |
|---|---|---|---|---|---|
| 0 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 14 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 5–9 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 1014 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 1519 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 2024 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 2529 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 3034 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 35–39 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 4044 | 1.00 | 0.01 | 0.01 | 0.0000000 | 0.00 |
| 45–49 | 0.99 | 0.03 | 0.03 | 0.0000000 | 0.00 |
| 5054 | 0.98 | 0.04 | 0.04 | 0.0000000 | 0.00 |
| 55–59 | 0.97 | 0.06 | 0.06 | 0.0000000 | 0.00 |
| 60–64 | 0.96 | 0.08 | 0.08 | 0.0000010 | 0.13 |
| 65–69 | 0.95 | 0.09 | 0.09 | 0.0000021 | 0.21 |
| 7074 | 0.94 | 0.11 | 0.11 | 0.0000032 | 0.27 |
| 75–79 | 0.93 | 0.12 | 0.12 | 0.0000054 | 0.40 |
| 8084 | 0.92 | 0.14 | 0.14 | 0.0000087 | 0.59 |
| 85+ | 0.91 | 0.15 | 0.15 | 0.0000121 | 0.75 |

Table 9: Completeness index, analytical and numerical standard deviation, $\Delta CV$, %diff. Data on female colorectal cancer, SEER-9 registries in period 1975-1999.

| age interval | $R$ | $\hat{\text{sd}}$ | $\tilde{\text{sd}}$ | $\Delta CV$ | %diff |
|---|---|---|---|---|---|
| 0 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 1–4 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 5–9 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 10–14 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 15–19 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 20–24 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 25–29 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 30–34 | 1.00 | 0.00 | 0.00 | 0.0000000 | 0.00 |
| 35–39 | 1.00 | 0.01 | 0.01 | 0.0000000 | 0.00 |
| 40–44 | 0.99 | 0.02 | 0.02 | 0.0000000 | 0.00 |
| 45–49 | 0.99 | 0.04 | 0.04 | 0.0000000 | 0.00 |
| 50–54 | 0.97 | 0.06 | 0.06 | 0.0000010 | 0.17 |
| 55–59 | 0.96 | 0.08 | 0.08 | 0.0000000 | 0.00 |
| 60–64 | 0.95 | 0.10 | 0.10 | 0.0000011 | 0.10 |
| 65–69 | 0.93 | 0.12 | 0.12 | 0.0000021 | 0.17 |
| 70–74 | 0.92 | 0.14 | 0.14 | 0.0000033 | 0.22 |
| 75–79 | 0.90 | 0.15 | 0.15 | 0.0000055 | 0.33 |
| 80–84 | 0.89 | 0.17 | 0.17 | 0.0000079 | 0.42 |
| 85+ | 0.87 | 0.18 | 0.18 | 0.0000114 | 0.55 |