Prevalence-corrected Prostate Cancer Incidence Rates and Trends

Ray M. Merrill1,2 and Mark K. Morris3

A correction is made of prostate cancer incidence rates based on data from the Surveillance, Epidemiology, and End Results Program of the United States National Cancer Institute. Unlike conventional incidence rates reported by the Program, corrected rates remove from the population the estimated number already diagnosed with the disease. The corrected rates reflect the average prostate cancer risk for men in the at-risk population. Because of the high incidence of and relatively good survival for prostate cancer, the prevalence of this disease is high. Corrected prostate cancer incidence rates were higher in magnitude, particularly in older age groups and among Black men. For example, in 1997 for Whites, the corrected rates were 3.8 percent higher in cases aged 60–69 years, 9.3 percent higher in cases aged 70–79, and 13.1 percent higher in cases aged 80 or more. Corresponding percentages for Blacks were 5.9, 18.9, and 16.9 percent, respectively. Percent changes over calendar time were very similar between corrected and uncorrected prostate cancer incidence rates according to age and race (White and Black). Failure to account for high levels of prostate cancer prevalence in conventional incidence rates of the disease results in underestimation of the rates but little temporal difference in the trends. Am J Epidemiol 2002;155:148–152.

Conventional cancer incidence rates reported in the United States represent the number of newly diagnosed cases in a given year divided by the midyear population for that year (1). Incidence rates have been reported on an annual basis in the United States since 1973, when the National Cancer Institute initiated the Surveillance, Epidemiology, and End Results (SEER) Program (2). Population-based, site-specific cancer incidence rates reflect the average risk of developing the disease (3). However, this assumes that the rate calculation includes new cases of the cancer in the numerator and the population at risk for developing the cancer in the denominator. In other words, the risk of developing a given cancer should be new cases of the disease divided by those who have never had the disease but are at risk of developing it (4).

Unlike other cancers, such as those of the breast, in which multiple primary tumors may be diagnosed at the same site, multiple diagnosed cancers of the prostate are extremely rare (4). On the other hand, the prevalence of prostate cancer exceeds that of any other cancer in men in the United States (5). The prevalence of this disease is high because of a combination of high incidence and good survival (1). There has also been a steady increase observed in prostate cancer prevalence in recent years (6). The extensive pool of prostate cancer cases in the population and the increasing number of prevalent cases suggest that failure to remove these men from the denominator in the rate calculation may give inaccurate prostate cancer incidence rates and trends. Correcting the denominator in the rate calculation for prevalent cases of the disease will increase the magnitude of the rates and may also change the temporal trend in rates. The purpose of this study is to obtain corrected prostate cancer incidence rates and to assess the change in magnitude and the trend of the rates according to age and race.

MATERIALS AND METHODS

Method of estimation

The numerator in the age-specific prostate cancer incidence rate reflects the first diagnosis of the cancer. Previous cancer diagnoses involving other sites may have occurred. The denominator in the rate represents the general population minus the number of previously diagnosed prostate cancer cases that are still alive. Age-specific corrected prostate cancer incidence rates are calculated as follows:

\[
R_{ik} = \frac{C_{ik}}{L(1 - P_{ik})}, \quad l = \text{year} \quad k = \text{age group},
\]

where \(C\) is the number of first primary cases, \(L\) is the midyear population for a given calendar year, and \(P\) is the
midyear point prevalence proportion of prostate cancer for a given calendar year. The subscript \( l \) denotes the years currently available in the SEER Program from 1973 through 1997. The subscript \( k \) denotes 5-year age groups at last follow-up, from ages 50–54 to 80–84 years, and then age 85 or more. Age-adjusted rates also are reported, standardized to the 2000 United States standard million population.

Data sources

Prostate cancer data for this study are taken from nine cancer registries in the National Cancer Institute’s SEER Program (1). The registries began collecting and reporting cancer data in 1973 (San Francisco-Oakland, California; metropolitan Detroit, Michigan; Connecticut; Hawaii; Iowa; New Mexico; and Utah); 1974 (Seattle and Puget Sound, Washington); and 1975 (metropolitan Atlanta, Georgia). These SEER areas represent approximately 10 percent of the United States population.

Analysis is based on malignant prostate cancer cases and is restricted to White men and Black men. These two racial groups provide sufficient numbers for analysis and display sufficient differences in their incidence, survival, and age distribution (i.e., for the Blacks population, a smaller proportion living to older ages) to allow us to make interesting comparisons (6, 7).

Population and point prevalence estimates

Population estimates from the United States Bureau of the Census are combined with cancer cases collected by SEER to calculate incidence rates. Cancer point prevalence proportion estimates calculated for the first day of each year are averaged across successive 2-year periods to provide midyear estimates. These estimates are derived using SEER data and include living persons previously diagnosed with the disease. SEER data have been shown to be representative of the United States (8). An extension of a previous method (6) to consider age at last follow-up was used to generate prevalence (see appendix 1).

RESULTS

Prostate cancer point prevalence estimates in White men and Black men by year and age at last follow-up are shown in figure 1. The prevalence estimates vary according to age and race. The prevalence estimates increase sharply with age and are consistently higher for Blacks, especially in the age groups 70–74 and 75–79 years. Trends in the prevalence estimates graphed on a log-scale (data not shown) indicated similar percent changes in slopes through the late 1980s across the age groups, which were slightly more pronounced for Whites. The increase in slope, beginning in 1989 (when prostate-specific antigen (PSA) screening began to be widespread), was highest for those aged 50–54 years, then for those aged 55–59, and so on, until for those aged 75–79 years the percent change leveled off and, in the oldest age groups, declined. For the time period 1989–1992, when PSA screening had its greatest effect on prostate cancer incidence rates, the percentage increase was 42 percent for White men and 35 percent for Black men.

Age-adjusted incidence rates for prostate cancer for White men and Black men aged 50 years and older in SEER


FIGURE 1. Prostate cancer point prevalence proportion (>100) for White men and Black men in SEER according to year and age, 1975–1997.
between 1975–1997 are shown in figure 2. The conventional and corrected prostate cancer incidence rates show a steady increase through the late 1980s, a sharp increase through 1992, and a decline and leveling off thereafter. The corrected rates are consistently higher in magnitude, whereas the trends remain similar.

Prostate cancer age-adjusted incidence rates for White men and Black men in SEER between 1975 and 1997 are shown for the age groups 50–59, 60–69, 70–79, and 80 years or more in figures 3 and 4. Very little difference in magnitude and trend is observed between the conventional and corrected incidence rates for those aged 50–59 years.

---


However, the differences increase consistently with age. Very little difference (in terms of percentage change) was observed within each age group between corrected and uncorrected trends over calendar time.

**DISCUSSION**

Conventional incidence rates include prevalent cases of the disease in the denominator. Since the prevalence of prostate cancer in the United States exceeds that of any other cancer in men (5), failure to remove from the denominator prevalent cases that are no longer at risk of a diagnosis will underestimate the true rates. Hence, this paper explored the extent to which conventional rates underestimate the true prostate cancer incidence rates and influence trends in the rates.

Dramatic increases in prostate cancer incidence and improvements in survival explain the increasing prevalence of prostate cancer cases in recent decades (6). Improved survival may be explained, at least in part, by PSA screening advancing the point of diagnosis without changing the future course of the disease (lead-time bias). Prostate cancer prevalence increased with age, although the trend peaked and fell in the oldest age groups. This is primarily explained by trends in the age-specific incidence rates (9). As a result, correcting prostate cancer incidence rates for prevalent cancer increased the rates, more so among older age groups.

Age-adjusted incidence rates are used for monitoring disease over time. Changing trends may reflect changes in risk factors and/or medical interventions. An association between the trends in prostate cancer incidence rates and the widespread adoption of PSA screening is well established (10–12). The important influence of transurethral resection of the prostate on prostate cancer incidence trends has also been addressed in the literature (13). Correction of the incidence rates influenced the trends, more so in the older age groups. This is largely explained by the increasing influence of PSA screening on the incidence and prevalence of prostate cancer with increasing age.

A primary purpose of the annually reported cancer incidence rates by the SEER Program is to monitor cancer trends. Although the temporal trends changed as a result of correcting the population in the rate calculations, these changes are small.

This study suggests that corrected prostate cancer incidence rates differ noticeably in magnitude from conventionally derived rates, more so in older age groups, among Black men, and in recent years because of PSA screening. The American Cancer Society calculates annual estimates of new cancer cases for the United States by using conventional SEER-based cancer incidence rates. Hence, these national estimates of new cases will be underestimated for prostate cancer.

**REFERENCES**

APPENDIX 1

Computation of prostate cancer prevalence involves the following: a matrix $a_{ik}$ of prostate cancer cases diagnosed in year $i$ and alive, dead, or lost to follow-up in age group $k$; an upper triangular matrix $h_{ijk}$ of cases dying in year $j$; and an upper triangular matrix $c_{ijk}$ of cases lost to follow-up in year $j$. An upper triangular matrix $d_{ijk}$ is then constructed from these data that indicates for each year of diagnosis and age group the number still alive and not lost to follow-up for subsequent years; that is,

$$d_{ijk} = a_{ik} \text{ when } i = j$$

$$d_{ijk} = d_{i-1,k} - h_{i-1,k} - c_{i-1,k} \text{ when } i < j$$

$$d_{ijk} = 0 \text{ when } i > j.$$

For example, the number of prostate cancer cases diagnosed in 1985 is adjusted according to the number who die or are lost to follow-up in subsequent years to obtain the number of diagnosed cases remaining in the registry at age group $k$ in the midyear 1986, 1987, . . . , 1997. Not everyone lost to follow-up has died. Hence, additional steps are needed to provide an estimate of the number alive adjusted for loss to follow-up: the matrix $h_{ijk}$ is divided by $d_{ijk}$ to give the matrix $e_{ijk}$ of annual death hazards; then $1 - e_{ijk}$ is estimated to give the matrix $f_{ijk}$; actuarial survival (matrix $g_{ijk}$) is obtained by multiplying cells in $f_{ijk}$ cumulatively over columns within each row; and the number alive adjusted for loss to follow-up (matrix $h_{ijk}$) is obtained by multiplying $a_{ik}$ by each column of $g_{ijk}$.

From the matrix $h_{ijk}$, another matrix was obtained that gives for each year 1973–1997 the number of prevalent cases based on the available years of follow-up data in SEER. For example, prevalence in 1997 is based on 25 years of follow-up. Prevalence in 1996 is based on 24 years of follow-up. Twenty-five years of follow-up explained approximately 100 percent of all prevalent cases of prostate cancer. To obtain an estimate of total prostate cancer prevalence for years with fewer than 25 years of follow-up, a back-calculation method was used. To illustrate, consider the following, which was applied to each age group at last follow-up $k$:

$$P_{lm,k} = \left( \frac{P_{l+1,m,k}}{P_{l+1,m-1,k}} \right) \times P_{lm-1,k}$$


$$m = 1, 2, \ldots, 24, 25$$

$$k = 50 - 4, 55 - 59, \ldots, 80 - 84, 85+$$

For example, the estimated prevalence in 1996 had there been 25 years of follow-up is obtained by the following calculation:

$$P_{1996.25,k} = \left( \frac{P_{1997.25,k}}{P_{1997.24,k}} \right) \times P_{1996.24,k}.$$