Partitioned prostate cancer prevalence estimates: an informative measure of the disease burden

R M Merrill

Abstract

Study objectives—Public health burden of disease is often measured using prevalence statistics. Prevalence of invasive prostate cancer in the United States is presented according to age at diagnosis, time from diagnosis, geographical area, and two races (white and black).

Design—Invasive prostate cancer data from the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute is used for obtaining prevalence estimates.

Main results—Despite falling prostate cancer incidence rates, the prevalence of this disease continues to rise for both white and black men. Black men diagnosed at ages 60 years and older experience lower levels of prevalence of prostate cancer than white men because of poorer survival and a smaller proportion of black men living to older ages where the disease becomes common. Black men require fewer years of follow up than white men to capture over 99% of prevalent cases (that is, 14 years versus 16 years, respectively).

Prevalence estimates in the United States are traditionally based on Connecticut data. On 1 January 1997, United States prostate cancer prevalence estimates based on Connecticut are overestimated for white men and underestimated for black men.

Conclusions—Partitioned prevalence estimates may provide a more meaningful and informative measure of the disease burden than conventional prevalence estimates. Prostate cancer prevalence estimates based on SEER rather than Connecticut data are better representative of the United States.

In this paper, prevalence of invasive prostate cancer in white and black men is estimated for the first day of specific calendar years according to age at diagnosis, time from diagnosis, and geographical location. The study is based on data from the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute.1 2 Also reported are the estimated number of invasive prostate cancer prevalent cases in the United States and the number of years of follow up data required to obtain complete cancer prevalence.

Methods

Prevalence estimates in this analysis reflect the number of living persons with a previously diagnosed prostate cancer on the first day of each year. Point prevalence proportions are obtained by dividing these numbers by corresponding population values. Population estimates are obtained by averaging successive midyear population values reported by the United States Bureau of the Census.1 4 Direct computation of complete prostate cancer point prevalence proportions require that these proportions be based on tumour registry data that have been collected over a sufficiently long period of time to capture all prevalent cases of the disease. The Connecticut Tumor Registry (CTR), which began collecting and annually reporting cancer data in the mid-1930s,5 6 is currently the only registry in the United States with sufficient years of follow up to compute complete prevalence. A method to derive point prevalence proportions from these data, adjusted for cases lost to follow up, has been previously developed.7

Although the CTR allows direct computation of prevalence, a recent study showed that complete prevalence estimates based on CTR do not necessarily represent those of the United States, particularly for black people.1 In 1973, the CTR was one of six other cancer registries (Hawaii, Iowa, New Mexico, Utah, San Francisco-Oakland, Detroit) to be incorporated into the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute. In 1974 and 1975 tumour registries in metropolitan Atlanta and Seattle-Puget Sound, respectively, were added to SEER. These registries represent about 10% of the United States population and provide a better representation of the national cancer experience, although the combined registries do not contain enough years of follow up data to directly compute complete prevalence for several cancer sites. Nevertheless, there is a sufficient number of years available in the SEER data to effectively partition prostate cancer prevalence across recent years.
Computation of prostate cancer prevalence involves the following: a matrix $a_i$ ($i =$ year of diagnosis) of prostate cancer cases; an upper triangular matrix $b_{ij}$ ($i =$ year of diagnosis and $j =$ year of death) of those dying; and an upper triangular matrix $c_{ij}$ ($i =$ year of diagnosis and $j =$ year of last follow up) of those lost to follow up. An upper triangular matrix $d_{ij}$ is then constructed from these data that indicates for each year of diagnosis the number still alive and not lost to follow up for subsequent years; that is,

- $d_{i,j} = a_i$ when $i = j$
- $d_{i,j} = d_{i+1,j} - b_{i+1} - c_{i,j}$ when $i < j$
- $d_{i,j} = 0$ when $i > j$

For example, the number of prostate cancer cases diagnosed in 1985 is adjusted according to the number dying or lost to follow up in subsequent years to get the number of diagnosed cases remaining in the registry at the beginning of 1986, 1987, etc. 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 $b_i$ is divided by $d_i$ to give the matrix $e_i$ of annual death hazards; then $1-e_i$ is estimated to give the matrix $f_i$; actuarial survival (matrix $g_i$) is obtained by multiplying cells in $f_i$ cumulatively over columns within each row; and the number alive adjusted for loss to follow up (matrix $h_i$) is obtained by multiplying $a_i$ by each column of $g_i$. We report point prevalence proportion estimates based on SEER data by United States population estimates.

Crude rather than age adjusted point prevalence proportions of invasive prostate cancer are reported, because they better capture the public health burden of the disease. Proportions are expressed per 100 000. All tests of significance and confidence intervals are based on the 0.01 level.

**Results**

Figure 1 shows estimated white and black crude invasive prostate cancer point prevalence proportions for the first day of calendar year $y$ by 0–$x$ cumulative years from diagnosis. Prevalence estimates vary considerably across calendar years and according to cumulative years from diagnosis. The majority of living prevalent cases were diagnosed within five years. The overall trends in complete point prevalence proportions are captured by the trends based on 0–15 years of diagnosed cases. Despite decreasing 0–1 point prevalence proportion estimates beginning in 1993 for white men and 1995 for black men, complete point prevalence proportion estimates continue to increase, although not as rapidly as during the early 1990s. On the basis of these data, complete point prevalence proportion estimates should level off and begin to decline in the near future for white men. A future levelling off and possible decline is also anticipated for black men, although not as soon as for white men. The percentage difference between 0–$x$ curves for black men is slightly less than that observed for white men. Comparing the 0–$x$ curves between white and black men also show the magnitude of the curves to generally be lower for black men.

The 1 January 1997 point prevalence proportions based on 0–$x$ ($x \leq 3$) cumulative years from diagnosis are significantly lower for white men than black men. Prevalence proportions based on 0–$x$ ($x \geq 5$) cumulative years from diagnosis are significantly higher for white men than black men. Prevalence proportions (per 100 000) based on 0–22 cumulative years from diagnosis are 1084 (99% CI: 1075, 1092) for white men and 886 (99% CI: 865, 907) for black men.

Figure 2 shows estimates for white and black men of invasive prostate cancer prevalence at the beginning of calendar years 1988 through 1997 by 0–15 cumulative years of diagnosed cases and age at diagnosis. Trends in proportions are influenced by age at diagnosis. Increasing trends persist over the study period.
for ages 0–59 years at diagnosis, ages 60–69 years at diagnosis, and ages 70–79 years at diagnosis, although the increasing trends become less pronounced with increasing age groups. For ages 80 years and older at diagnosis, the trends actually began to fall, beginning on 1 January 1995. These patterns occur for both white and black men. Lower levels of point prevalence proportions for black men than white men are seen in each age group except in ages 0–59 years at diagnosis, where they are higher for black men. The lower levels in black men are most pronounced in ages 70 and older at diagnosis. The 1 January 1997 point prevalence proportions (per 100 000) in age groups 0–59, 60–69, 70–79, and 80+ are 126 (123, 129), 407 (402, 413), 417 (412, 423), and 97 (94, 100) for white men and 154 (151, 157), 367 (361, 372), 284 (279, 289), and 58 (55, 60) for black men, respectively.

Figure 3 shows estimates of crude invasive prostate cancer point prevalence proportions at the beginning of calendar years from 1988 through 1997 by 0–15 cumulative years of diagnosed cases, two races (white and black), and area.

In 1996, the population in Connecticut was about 12.8% (14.3% for white men and 10.4% for black men) of the SEER population. Although the white population is represented well in each of the SEER areas, this is not true for black men. Approximately 77.2% of the black population in SEER is represented by the metropolitan areas of Atlanta (27.2%), Detroit (33.5%), and San Francisco (16.5%). Because these three areas reflect the majority of the
black population in SEER, included in the figure is the crude point prevalence proportion for black men based solely on these areas. The magnitude of the estimates based on data from these three areas is significantly higher than when based on SEER data. For example, the 1 January 1997 point prevalence proportion (per 100,000) is 993 (984, 1002).

Table 1 shows the estimated number of invasive prostate cancer prevalence in the United States on 1 January 1997, according to 0–x cumulative years from diagnosis (Dx) and area (Connecticut, SEER) upon which the estimates were based.

<table>
<thead>
<tr>
<th>0–x Years From Dx</th>
<th>White men</th>
<th>Black men</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conn</td>
<td>SEER</td>
</tr>
<tr>
<td>1</td>
<td>161 710</td>
<td>141 706</td>
</tr>
<tr>
<td>2</td>
<td>322 695</td>
<td>276 521</td>
</tr>
<tr>
<td>3</td>
<td>478 898</td>
<td>408 978</td>
</tr>
<tr>
<td>4</td>
<td>628 365</td>
<td>552 098</td>
</tr>
<tr>
<td>5</td>
<td>784 213</td>
<td>703 996</td>
</tr>
<tr>
<td>6</td>
<td>893 047</td>
<td>823 135</td>
</tr>
<tr>
<td>7</td>
<td>964 077</td>
<td>902 210</td>
</tr>
<tr>
<td>8</td>
<td>1 012 025</td>
<td>958 796</td>
</tr>
<tr>
<td>9</td>
<td>1 046 016</td>
<td>1 004 617</td>
</tr>
<tr>
<td>10</td>
<td>1 074 733</td>
<td>1 041 495</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>0–x Years From Dx</th>
<th>White men</th>
<th>Black men</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conn</td>
<td>SEER</td>
</tr>
<tr>
<td>1</td>
<td>22 221</td>
<td>50 229</td>
</tr>
<tr>
<td>2</td>
<td>41 893</td>
<td>99 464</td>
</tr>
<tr>
<td>3</td>
<td>59 115</td>
<td>149 392</td>
</tr>
<tr>
<td>4</td>
<td>75 863</td>
<td>203 825</td>
</tr>
<tr>
<td>5</td>
<td>90 820</td>
<td>260 528</td>
</tr>
<tr>
<td>6</td>
<td>101 276</td>
<td>305 551</td>
</tr>
<tr>
<td>7</td>
<td>108 309</td>
<td>336 613</td>
</tr>
<tr>
<td>8</td>
<td>113 611</td>
<td>388 734</td>
</tr>
<tr>
<td>9</td>
<td>118 213</td>
<td>438 548</td>
</tr>
<tr>
<td>10</td>
<td>122 643</td>
<td>471 508</td>
</tr>
</tbody>
</table>

Data source: Surveillance, Epidemiology, and End Results.
The percentage of 0–22 year prevalent cases of white men in each age category is: 12.6% for ages 0–59 years at diagnosis, 39.4% for ages 60–69 years at diagnosis, 39.1% for ages 70–79 at diagnosis, and 9.0% for ages 80 years and older at diagnosis. Corresponding percentages for black men are: 18.2%, 42.8%, 32.5%, and 6.5%. Hence, the prevalent pool of black versus white cases represent a younger age distribution at diagnosis ($\chi^2 p < 0.0001$).

Figure 4 shows the prevalence proportion of invasive prostate cancer on the first day of calendar year $y$ by 0–15 cumulative years of diagnosed cases according to the effects of survival and incidence. The height of the grey area of the figure shows prevalence proportions assuming prostate cancer incidence rates and death hazards remained constant at the 1986 levels. The year 1986 was selected because it immediately precedes the time of rapid changes in prostate cancer incidence rates in the United States. The height of the black area of the figure shows the effect of improved survival for prostate cancer patients. Finally, the height of the white section of the figure shows the effect of higher prostate cancer incidence rates. By 1 January 1997, percentages of the prevalence proportion attributed to improved survival and higher incidence rates were 6.8% and 48.2%, respectively.

**Discussion**

In past years, despite the considerable attention given to the dramatic swings in prostate cancer incidence rates over time, little attention has been given to the prevalence of this disease. This paper identified the influences of age at diagnosis, years from diagnosis, two races (white and black), and geographical area (SEER versus Connecticut) on prostate cancer prevalence estimates. Incidence and survival rates for prostate cancer patients and the ages where this disease becomes common directly influence the level of prevalent cases and how the prevalent pool of diagnosed cases is partitioned by years from diagnosis. Such information may be useful for identifying the health care burden for this disease, given that the demand for treatment and care is strongly related to the patient’s age at diagnosis and is greatest shortly after diagnosis, declining with years already survived. It further shows that even though black men experience considerably higher incidence rates of the disease, the prevalence of the disease is lower than in white men because of their poorer survival and because a smaller proportion of black men are living to ages where the disease becomes common.

Increasing prostate cancer incidence and improved survival among patients in the United States explain the growing pool of prevalent cases. Prostate cancer incidence rose gradually between the mid-1970s to the 1980s, followed by an unprecedented increase in rates beginning in 1990. After a peak in rates in 1992 for white men and in 1993 for black men, the rates declined through 1996. Currently prostate cancer is the most frequently diagnosed cancer in men in the United States.

Prostate cancer survival has also improved in the United States. Consistent improvements in
survival have been observed in the past few decades for both white and black men, for all age groups. 

As a result of sharp changes in prostate cancer incidence and survival, the prevalence of prostate cancer has increased dramatically in recent years. Prostate cancer prevalence is currently higher than that of any other cancer in the United States. Estimates of the relative influence of prostate cancer incidence and survival (fig 4), based on the assumption that prostate cancer incidence and survival remained constant at 1986 levels. Despite falling incidence rates of prostate cancer 1993 to 1996, trends in complete point prevalence proportions continued to increase, albeit at a slower rate than in the early 1990s. Trends in 0–1 point prevalence proportion estimates closely resemble trends in the incidence rates, although slightly lower in magnitude because the prevalence estimates remove cases dying from the disease. The dramatic increase in complete prevalence estimates suggested by the results will continue, although a peak is expected within five or six years of the study period.

A comparison of the 0–x point prevalence proportions between white and black cases illustrates the combined influences of incidence, survival, and the age structure on the proportions. It is well established that prostate cancer incidence rates are higher for black than white men. Over the study period, age adjusted (to the 1970 United States population) prostate cancer incidence rates in black men ranged from 1.3 to 1.7 times that of white men. Nevertheless, the crude point prevalence proportions reported in this paper were lower for black men. There are two reasons for this. Firstly, black cases experience poorer survival. This fact is well established in the literature, but also evident in this study by the fewer number of follow up years required for black men than white men to obtain complete prevalence. Secondly, life expectancy for black men is considerably lower than that of white men, such that black men are less likely to live to older ages where the incidence of prostate cancer becomes common.

Conventional point prevalence proportions in the United States have traditionally been based on the method of Feldman et al applied to long term Connecticut Tumor Registry Data. SEER data cover a much larger proportion of the United States and provide a better representative sample than only Connecticut. Although using SEER data results in underestimation of complete prevalence, it can be used to provide more representative 0–x point prevalence proportion estimates. Nevertheless, the 22 years of follow up data available from all nine registries in the SEER programme were seen to be sufficient to explain almost all prevalent cases of prostate cancer. A comparison of SEER and Connecticut based prevalence estimates were made, suggesting that Connecticut may be inadequate to represent the prevalence of prostate cancer in the United States, for both white and black men.

In contrast with incidence rates, which are useful for identifying risk factors for disease, prevalence is useful from a public health policy perspective for identifying the burden of the disease and influencing health care planning in terms of allocation of resources and services. The value of this paper is that it illustrates a method for partitioning prevalence. This seems particularly important given that health care procedure and service costs are directly influenced by the patient’s age at and time from diagnosis. In the 1990s, the majority of cases were treated within four months of diagnosis: above 35% with radical prostatectomy, above 30% with radiation therapy, and above 8% with hormonal therapy. Morbidity associated with aggressive treatment of this disease is also common, increasingly so with older age. It seems that partitioned prevalence estimates, combined with information on medical costs, may provide a more meaningful and useful measure of the disease burden than traditional prevalence estimates.

Funding: none.

Conflicts of interests: none.