Asymptomatic Incidence and Duration of Prostate Cancer

Ruth Etzioni,1 Raymond Cha,1 Eric J. Feuer,2 and Ori Davidov1

Prostate cancer is known as a disease with an extremely high prevalence relative to its clinical incidence in the population. The combination of preclinical incidence and duration that could yield this phenomenon is of tremendous interest to researchers trying to understand the natural history of the disease and to develop efficient screening strategies. In this article, the authors present estimates of the age-specific asymptomatic incidence and average preclinical duration of prostate cancer. The methodological approach is to first estimate the age-specific incidence of new (stage AI) prostate cancers using preclinical prevalence data from autopsy studies performed between 1941 and 1964 and clinical incidence data for the years 1960–1986 from the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute. Then, the preclinical prevalence estimates are divided by the derived preclinical incidence estimates to yield estimates of the average duration of asymptomatic disease. The estimated mean duration among white men is between 11 and 12 years and appears to be approximately 1 year shorter for blacks than for whites. Comparison of the lifetime risks of preclinical and clinical disease suggests that approximately 75% of prostate cancers will never become diagnosed if clinical incidence remains at levels observed in 1984–1986, prior to the introduction of prostate-specific antigen (PSA) screening in the population. Am J Epidemiol 1998; 148:775–85.

disease progression; natural history; prevalence; prostatic neoplasms; SEER program

The duration of the asymptomatic or latent period in prostate cancer has important public health implications, particularly for the design and analysis of screening strategies. Indirect evidence, including high prevalence of the disease at autopsy relative to its clinical incidence, suggests that the time from preclinical onset to diagnosis of prostate cancer can be relatively lengthy. For example, the proportion of men aged 65–69 years with histologic prostate cancer has been estimated from autopsy studies to be at least 10–20 percent (1, 2). In contrast, the annual clinical incidence of prostate cancer in 1984–1988 was 0.37 percent for black males aged 60–64 years and 0.73 percent for those aged 65–69 years; the analogous figures for white males were 0.23 percent and 0.46 percent, respectively (3).

The goal of this paper is to estimate the length of the asymptomatic period in prostate cancer, that is, the time from onset of the disease until the appearance of symptoms leading to its diagnosis. We also estimate the duration of the preclinical period, which we define as the time from onset of the disease until its clinical diagnosis, whether due to symptoms or not. In calendar periods in which no interventions exist to diagnose asymptomatic prostate cancer, the preclinical duration and the asymptomatic duration are the same. However, modalities that can be used to detect asymptomatic disease have been developed more recently. In this article, we use the more general term “preclinical period” to refer to the time from disease onset to diagnosis; when necessary, we distinguish between time to diagnosis and time to symptoms.

Previous estimates of the preclinical duration of prostate cancer have been presented by Whittemore et al. (4). These authors used autopsy data to estimate by age and race the mean volume of low-grade, latent cancer among men in the general population. They then hypothesized that the incidence of high-grade cancer among men at a given age would be proportional to the mean volume of low-grade cancer when the men were a few years younger. Based on a linear regression analysis, the average time to detection was estimated to be approximately 7 years.

Work on longitudinal changes in prostate-specific antigen (PSA) in men with a subsequent diagnosis of prostate cancer also provides information about the preclinical duration. Indeed, Morrell et al. (5) have estimated that PSA curves in prostate cancer patients

Received for publication July 30, 1997, and accepted for publication March 20, 1998.

Abbreviations: PSA, prostate–specific antigen; SEER, Surveillance, Epidemiology, and End Results.

1 Fred Hutchinson Cancer Research Center, Seattle, WA.

2 National Cancer Institute, Bethesda, MD.

Reprint requests to Dr. Ruth Etzioni, Fred Hutchinson Cancer Research Center, Program in Biostatistics, 1100 Fairview Avenue N, MPE-665, P. O. Box 19024, Seattle, WA 98109–1024.
show a marked increase in slope beginning at roughly 6.4 to 10.7 years prior to clinical diagnosis, depending on the stage at diagnosis. Whittemore et al. (6), in a much larger longitudinal PSA study, estimated that the increase in slope occurs roughly 13 years prediagnosis for men with subsequent stages B through D prostate cancer.

There are clearly some differences between previously derived estimates of the preclinical duration in prostate cancer. In this article, we use a different methodological approach in an attempt to resolve these differences and to provide more reliable information about the natural history of prostate cancer.

**MATERIALS AND METHODS**

Our methodological approach extends methods for estimating age-specific incidence from age-specific prevalence in the case of an irreversible disease that is lifelong after diagnosis (7).

We begin by estimating the number of new cases of asymptomatic disease in any given age interval. Figure 1 illustrates the steps involved, which are described in more detail in the rest of this section. The newly incident cases in a specific age interval consist of those who survive to the end of the interval and those who do not. Formally, we denote the total number of new cases by $N$; $N$ is equal to $n + m$, where $n$ is the number of surviving new cases and $m$ is the number of new cases who do not survive. We derive an expression for $n$, which draws on mathematical development used in estimating the lifetime probability of a cancer diagnosis (8). We then develop an approximation for $m$, which allows us to estimate the total number of new cases in the interval and the associated incidence. Finally, we divide the prevalence of asymptomatic disease by the derived asymptomatic incidence to yield an estimate of the average sojourn time. More precisely, we denote the age-specific preclinical incidence rate in age interval $i$ by $I_i$ and the prevalence by $P_i$. An estimate of the average preclinical duration is then given by the ratio of population prevalence to incidence (9), namely,

$$M = \frac{\sum_{i=1}^{k} I_i P_i}{\sum_{i=1}^{k} I_i}$$

where age interval 1 is the lowest age interval at which the incidence and prevalence of disease are nonnegligible, $k$ is the highest age interval in the population, and $f_i$ denotes the fraction of the population in age group $i$. In our calculations, the interval length is 1 year and the lowest and highest age intervals are 30–31 and 89–90 years, respectively.

To estimate $n$, the number of surviving new cases in a given age interval, we extend the methods of Leske et al. (7), who estimated $n$ as

$$n = N_i P_i - N_o P_o + N_{i} P_{i} m_o,$$

where $N_o$ and $P_o$ are the number who are alive and the prevalence of the condition at the beginning and end of the interval, $N_i$ and $P_i$ are the number who are alive and the prevalence of the condition at the end, and $m_o$ is the probability of death due to other causes during the interval. Thus, the number of surviving new cases equals the number of prevalent cases at the end minus the number of prevalent cases at the start plus the number of cases prevalent at the start who were lost during the interval because of competing mortality. Cowen et al. (10) used a similar idea, but their calculation neglected competing mortality. Note that we use the term "prevalent" to refer to persons with asymptomatic disease.

Expression 2 is appropriate when the condition of interest is lifelong after diagnosis. In our case, the condition of interest is preclinical prostate cancer, which is terminated when the disease is diagnosed. Termination of the condition of interest is a mecha-
nism for removing prevalent cases from the interval in addition to competing mortality. In this case, the number of surviving new cases is given by

\[ n = N_1 P_1 - N_0 P_o + L, \]  

(3)

where \( L \) is the number of prevalent cases at the start who are lost during the interval either through competing mortality or through clinical diagnosis of cancer, i.e.,

\[ L = a_o + N_o P_o m_o. \]  

(4)

Here, \( a_o \) is the number of clinical diagnoses in the interval and \( m_o \) is the probability of mortality due to other causes during the interval. Note that \( m_o \) in this expression is not exactly the same as it is in expression 2, since in this case clinical diagnosis is a competing risk for other-cause mortality. Precise expressions for \( a_o \) and \( m_o \) in expression 4 are given below.

This formulation assumes that all clinically diagnosed cases are among the \( N_o P_o \) prevalent cases at the start of the interval. We show in the Appendix that this assumption does not lead to any loss of generality.

The prevalence estimates \( P_o \) and \( P_1 \) used in our application were obtained from Carter et al. (1), who studied 5,250 autopsies from the US literature (figure 2). These estimates apply to a symptom-free population (men with a diagnosis of prostate cancer during their lifetimes were excluded). Therefore, to estimate the number of prevalent cases at the start and end of each age interval, we require the size of the symptom-free population (i.e., those without a prior prostate cancer diagnosis) at these times. We denote the symptom-free population size at the start and end of a given age interval by \( N_{oo} \) and \( N_{lo} \), respectively. Expressions 3 and 4 still apply but with \( N_o \) and \( N_1 \) replaced by \( N_{oo} \) and \( N_{lo} \) and \( m_o \) now equal to the mortality probability in the symptom-free population. We assume that \( N_o = N_{oo} \) in the initial time interval.

Wun et al. (8) present expressions for \( a_o m_o \) and \( N_{lo} \), which take into account that death from other causes and clinical incidence are competing risks acting on the symptom-free population (11). These authors assume exponential distributions of event times in each age interval so that the probability of an event in an interval of length \( i \) is \( 1 - \exp(-ri) \), where \( r \) is the rate of the event. Given \( N_{oo} \) symptom-free individuals at the start of an interval, we have

\[ N_{lo} = N_{oo} \exp\left( - (q_o + r_o) \right), \]  

(5)

where \( q_o \) is the death rate from causes other than prostate cancer and \( r_o \) is the clinical incidence rate, respectively, among persons without symptoms at the start of the interval. The rate \( q_o \) is approximated by the rate in the total population (8); \( q_o = m_o - m_p \), where \( m_o \) is the all-cause mortality rate and \( m_p \) is the prostate-cancer mortality rate. Thus, deaths due to prostate cancer are assumed to occur among only those individuals previously diagnosed with the disease and not among persons without symptoms. This assumption appears to be adequate, because the incidence of previously undiagnosed cases with prostate cancer listed as a cause of death is extremely low (for example, the Surveillance, Epidemiology, and End Results (SEER) age-adjusted incidence of prostate cancer in 1984–1988 was 92.24 per 100,000, whereas the incidence of death-certificate-only cases was 0.49 per 100,000 (3)). The rate \( r_o \) of clinical diagnosis in the symptom-free population is computed by multiplying \( r \), the population clinical incidence rate, by \( (N_o/N_{oo}) \), where \( N_o \) is the population at the start of the interval. All clinically diagnosed cases in a given year are derived from the pool of asymptomatic, prevalent cases at the start; that this assumption leads to no loss of generality is shown in the Appendix.

The number of clinically diagnosed cases \( a_o \) is given by

\[ a_o = N_{oo} (1 - \exp\left( - q_o + r_o \right)) \frac{q_o}{q_o + r_o}, \]  

(6)

and the probability of noncancer death by

\[ m_o = 1 - \exp\left( - (q_o + r_o) \right) \frac{q_o}{q_o + r_o}. \]  

(7)

The total number of new cases \( N \) is approximately equal to \( n + m \), where \( m \) is the number of new cases who do not survive to the end of the interval. We estimated \( m \) by using the approximation proposed by
Leske et al. (7), namely, 

\[ m = nu_o/(1 - u_o)(1/2), \]

where \( u_o \) is the probability of dying from causes other than prostate cancer during the year, i.e., 

\[ u_o = 1 - \exp(-q_o). \]

A heuristic explanation for the approximation is the following: If \( n \) is the number of surviving new cases, then \( n/(1 - u_o) \) represents the total new cases. Of these, \( nu_o/(1 - u_o) \) are the "nonsurviving new cases," which includes those who did not survive to their time of asymptomatic onset as well as those who survived their time of onset but died before the end of the interval. On average, half of the nonsurviving new cases have a time of death that precedes their time of onset; the approximation is exact if time to preclinical onset and time to other-cause mortality are uniformly distributed within the interval. The number of incident new cases is then approximately

\[ N = n + m = n(1 + (1 - u_o))/2(1 - u_o). \]

(8)

\( N \) may be divided by an appropriate population denominator to yield the age-specific rate of asymptomatic onset in the interval. In our calculations, we used the population size at the start of the interval, although the population size at the midpoint of the interval could also be used. The average preclinical duration is then estimated by using expression 1.

In principle, our entire method can be applied in reverse if the goal is to generate the age-specific prevalence of preclinical disease given its incidence. Specifically, given \( I_o \), the asymptomatic incidence rate in an interval, the number of new cases \( N \) is approximately \( N_o(1 \exp(-I_o)) \). The number of new cases who survive to the end of the interval is \( n = 2N(1 - u_o)/(1 + (1 - u_o)) \) (compare with expression 8). Finally, the number of prevalent cases at the end of the interval given the number prevalent at the start is

\[ N_oP = N_oP_o - a_o - N_oP_o m_o + n \]

(compare with expression 2).

In the next section, we illustrate application of both the original and the reverse calculation. The original calculation enables us to estimate the incidence of preclinical disease by using data on preclinical prevalence from a time when relatively few interventions were available for early diagnosis of prostate cancer. However, it does not seem reasonable to use the same prevalence data for later time periods. Indeed, recent changes in medical practice patterns have had a substantial impact on the frequency of a prostate cancer diagnosis and, therefore, presumably also on the prevalence of preclinical disease. In the application, we assume that preclinical incidence is relatively constant over time, since this is not likely to be affected by medical interventions that lead to early diagnosis. Using our estimate of preclinical incidence from the earlier time period, we apply the reverse calculation to estimate preclinical prevalence and mean sojourn time in the later time periods.

**RESULTS**

Figure 3 presents estimates of the age-specific incidence of asymptomatic prostate cancer in white men. Table 1 shows the input data used in our calculations, and table 2 shows some intermediate results. We begin with 5 million individuals at age 30 years, when \( N_o = N_oo \). The clinical incidence and mortality rates given in 5-year intervals in table 1 are translated into 1-year intervals by linearly interpolating between the midpoints of each interval, except for ages 30–32 years and ages 87–89 years, where the rates are assumed to be constant because of a lack of precise information on men younger than age 30 years and older than age 89 years.

As an example of the sequence of calculations, consider the age interval 52–53 years. There are 4,650,509 symptom-free individuals at the start of this interval (\( N_{oo} \)). The clinical incidence rate among these individuals, to five decimal places, is 0.00014 (\( r_o = rN_o/N_{oo} \), where \( r = 0.000137 \) from table 1). The number of symptom-free individuals at the end of the interval is 4,618,287, since \( N_o = N_oo \exp(-\{q_o + r_o\}) \); \( q_o \) is taken directly from table 1. The number of clinically diagnosed cases, \( a_o \), is given by expression 6 and is equal to 635; the number of deaths due to other causes among cases prevalent at the start is \( N_{oo}P_o m_o = 1,708 \), where \( m_o \) is the probability of non-cancer death in the interval, given by expression 7.

Thus, the number of prevalent cases lost during the interval because of clinical diagnosis or other cause of death is $L = 2,343 = 635 + 1,708$. To obtain $n$, the number of surviving new cases, we solve $N_1 = N_o P_1 + n - L$ to get $n = 18,304$; the total number of incident cases is $N = 18,372$ given by expression 8, and the incidence in the interval (not shown in the table) is $I = N/N_o = 0.00395$.

The estimates in figure 3 are based on autopsy prevalence data from Carter et al. (1), clinical incidence from the Connecticut SEER tumor registry for 1960–1964 (12), all-cause mortality from 1986 US life tables (13), and prostate cancer mortality from 1984 to 1988 (3). We assumed a negligible incidence of asymptomatic disease prior to age 30 years. Our goal was to estimate the average asymptomatic duration in a population similar in terms of life expectancy to the US population in the 1980s. However, we used historical clinical incidence data in an effort to eliminate the impact of interventions leading to incidental diagnosis of cancer, such as transurethral resection of the prostate. Based on these input data, the average sojourn time in a population with an age distribution similar to that of the US population in 1980 (14) is 11.57 years.

Since the Connecticut SEER data from 1960 to 1964 do not contain sufficient information on black cases, we obtained sojourn time estimates for blacks based on clinical incidence in several other calendar time periods, namely 1975–1979, 1984–1988, and 1990–1992. This also allowed us to assess the impact of differing medical practice patterns over time. We used the reverse calculation to estimate how the changing incidence of clinical disease would affect the age-specific prevalence of preclinical disease. In doing so, we assumed that the prevalence of preclinical disease at age 30 years would be the same as in Carter et al. (1), since medical practice patterns have changed little.

### TABLE 1. Input parameters used to estimate the asymptomatic incidence of prostate cancer in blacks and whites, by age in years

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Blacks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30–34</td>
<td>1,121.9</td>
<td>0</td>
<td>0.2</td>
<td>0</td>
<td>448.0</td>
</tr>
<tr>
<td>35–39</td>
<td>1,627.3</td>
<td>0</td>
<td>1.0</td>
<td>0.2</td>
<td>589.8</td>
</tr>
<tr>
<td>40–44</td>
<td>2,360.4</td>
<td>1.8</td>
<td>2.7</td>
<td>4.2</td>
<td>797.0</td>
</tr>
<tr>
<td>45–49</td>
<td>3,423.9</td>
<td>17.3</td>
<td>12.4</td>
<td>15.6</td>
<td>1,034.4</td>
</tr>
<tr>
<td>50–54</td>
<td>4,966.4</td>
<td>49.4</td>
<td>63.9</td>
<td>103.5</td>
<td>1,052.6</td>
</tr>
<tr>
<td>55–59</td>
<td>7,203.8</td>
<td>155.2</td>
<td>162.1</td>
<td>323.5</td>
<td>2,075.4</td>
</tr>
<tr>
<td>60–64</td>
<td>10,449.2</td>
<td>313.8</td>
<td>370.2</td>
<td>679.6</td>
<td>3,075.6</td>
</tr>
<tr>
<td>65–69</td>
<td>15,168.8</td>
<td>620.3</td>
<td>727.6</td>
<td>1,220.7</td>
<td>4,074.1</td>
</tr>
<tr>
<td>70–74</td>
<td>21,965.1</td>
<td>874.3</td>
<td>1,019.6</td>
<td>1,733.1</td>
<td>5,797.0</td>
</tr>
<tr>
<td>75–79</td>
<td>31,889.8</td>
<td>1,166.5</td>
<td>1,331.0</td>
<td>1,940.6</td>
<td>8,038.1</td>
</tr>
<tr>
<td>80–84</td>
<td>46,256.7</td>
<td>1,310.6</td>
<td>1,556.7</td>
<td>2,047.5</td>
<td>11,654.7</td>
</tr>
<tr>
<td>85–89</td>
<td>67,096.1</td>
<td>1,439.3</td>
<td>1,486.4</td>
<td>1,840.2</td>
<td>15,488.1</td>
</tr>
<tr>
<td>Whites</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30–34</td>
<td>1,121.9</td>
<td>0.2</td>
<td>0.2</td>
<td>0</td>
<td>180.6</td>
</tr>
<tr>
<td>35–39</td>
<td>1,627.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0</td>
<td>212.3</td>
</tr>
<tr>
<td>40–44</td>
<td>2,360.4</td>
<td>0.7</td>
<td>1.3</td>
<td>2.4</td>
<td>452.0</td>
</tr>
<tr>
<td>45–49</td>
<td>3,423.9</td>
<td>6.0</td>
<td>4.6</td>
<td>6.9</td>
<td>12.2</td>
</tr>
<tr>
<td>50–54</td>
<td>4,966.4</td>
<td>13.7</td>
<td>23.4</td>
<td>30.2</td>
<td>2.6</td>
</tr>
<tr>
<td>55–59</td>
<td>7,203.8</td>
<td>40.0</td>
<td>73.0</td>
<td>95.4</td>
<td>11.6</td>
</tr>
<tr>
<td>60–64</td>
<td>10,449.2</td>
<td>113.9</td>
<td>176.5</td>
<td>235.5</td>
<td>31.2</td>
</tr>
<tr>
<td>65–69</td>
<td>15,168.8</td>
<td>233.5</td>
<td>336.0</td>
<td>462.5</td>
<td>31.2</td>
</tr>
<tr>
<td>70–74</td>
<td>21,965.1</td>
<td>409.8</td>
<td>556.6</td>
<td>725.2</td>
<td>139.2</td>
</tr>
<tr>
<td>75–79</td>
<td>31,889.8</td>
<td>554.9</td>
<td>760.5</td>
<td>978.5</td>
<td>426.0</td>
</tr>
<tr>
<td>80–84</td>
<td>46,256.7</td>
<td>773.2</td>
<td>1,006.8</td>
<td>1,155.4</td>
<td>380.5</td>
</tr>
<tr>
<td>85–89</td>
<td>67,096.1</td>
<td>934.3</td>
<td>1,110.6</td>
<td>1,149.0</td>
<td>616.8</td>
</tr>
</tbody>
</table>

* Computed at the midpoint of each age interval; based on autopsy studies published from 1941 to 1966 (1).
† Annual figures.
‡ Source: Miller et al. (3).
§ Source: US life tables (13).
¶ Source: US Census data (14).
§ Source: Connecticut Surveillance, Epidemiology, and End Results registry (12); not available for blacks.
** Source: Carter et al. (1), since medical practice patterns have changed little.
for this group in regard to prostate cancer care. Figure 4 shows how the different clinical diagnosis rates affect the estimates of age-specific preclinical prevalence in different calendar time periods.

Figure 4 and table 3 show that the asymptomatic prevalence and the average preclinical duration vary with age-specific clinical incidence, which has changed considerably over time. In general, higher rates of clinical diagnosis lead to lower prevalence estimates. For whites, the average sojourn time based on the clinical incidence from 1960 to 1964 is higher than the estimate based on later calendar periods. The earlier estimate is probably closest to the true asymptomatic duration. The duration estimate based on the clinical incidence from 1975 to 1979 is about 6 months longer than that based on the clinical incidence from 1984 to 1988. Both of these latter estimates are more reflective of time to detection than of a true asymptomatic duration. The incidence calculation uses data on the asymptomatic incidence and duration of prostate cancer. The incidence calculation uses data on the asymptomatic incidence and duration of prostate cancer. The average sojourn time for black males is almost 1 year shorter, supporting the belief that prostate cancer, on average, may be more aggressive and have higher growth rates in blacks than in whites.

DISCUSSION

This paper presents a novel approach to estimating the asymptomatic incidence and duration of prostate cancer. The incidence calculation uses data on the age-specific prevalence of preclinical disease from autopsy studies and on the age-specific incidence of clinical disease from SEER. Our results suggest that the average asymptomatic duration for whites is 11–12 years, which is between the published estimates (5, 6) of the time from initial acceleration of the PSA growth curve to clinical diagnosis of prostate cancer. Thus, these results, taken together, support the notion that the initial acceleration of PSA growth probably occurs at roughly the time of disease onset.

The methods presented here extend the work of
Leske et al (7), who estimated incidence from prevalence when the disease of interest is lifelong after diagnosis. In our case, the “disease of interest” is the asymptomatic phase of prostate cancer; since the asymptomatic period is terminated by clinical diagnosis, it is not necessarily a lifelong condition. Our methods directly address this problem by considering clinical diagnosis and other-cause mortality as competing risks for removal of asymptomatic cases from the prevalence pool. An alternative approach would be to use methods developed by Podgor and Leske (16) for estimating disease incidence when mortality rates differ for prevalent and nonprevalent cases. In this case, other-cause mortality and clinical diagnosis would be considered collectively as “mortality” in the prevalent group. This approach requires solution of a nonlinear equation and so may be computationally more complex than our approach, which may be im-

**FIGURE 4.** Age-specific Surveillance, Epidemiology, and End Results (SEER) program incidence and corresponding asymptomatic prevalence by race and calendar period. Top left: incidence, whites; bottom left: prevalence, whites; top right: incidence, blacks; and bottom right: prevalence, blacks. Asymptomatic incidence is as shown in figure 3; asymptomatic prevalence at age 30 years is from Carter et al. (1); all-cause mortality is from 1986 US life tables (13); and cancer mortality is from SEER (1984–1988) (8).

**TABLE 3.** Estimates* of the average sojourn time, by period of clinical incidence, among populations of blacks and whites with age distributions similar to those in the US population in 1980.

<table>
<thead>
<tr>
<th>Population</th>
<th>Mean preclinical duration (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whites</td>
<td></td>
</tr>
<tr>
<td>1960–1964</td>
<td>11.57</td>
</tr>
<tr>
<td>1975–1979</td>
<td>11.45</td>
</tr>
<tr>
<td>1984–1988</td>
<td>10.98</td>
</tr>
<tr>
<td>1990–1992</td>
<td>9.09</td>
</tr>
<tr>
<td>Blacks</td>
<td></td>
</tr>
<tr>
<td>1975–1979</td>
<td>10.40</td>
</tr>
<tr>
<td>1990–1992</td>
<td>7.63</td>
</tr>
</tbody>
</table>

* Estimates are based on prevalence from autopsy studies published between 1941 and 1966 (1), Surveillance Epidemiology, and End Results clinical incidence data from several calendar periods (3, 12), all-cause mortality from 1986 US life tables (13), and prostate cancer mortality from 1984 to 1988 (8).
plemented by using a spreadsheet program such as EXCEL (Microsoft Corporation, Redmond, Washington). Figure 5 shows that both approaches yield similar results in the present problem.

The mathematical relation, average duration equals prevalence divided by incidence, has a long history in the epidemiologic literature (9). An implicit assumption is that the condition of interest is progressive in the sense that it will terminate unless prevented, for example, by competing mortality (17). Therefore, this approach is not valid for nonprogressive diseases or for diseases that can regress. Given the possibility that prostate cancer cases may exist in whom the tumor might remain indolent no matter how long they lived (infinitely indolent), this is a limitation of our approach. Based on this scenario, the quantity of interest is presumably the mean asymptomatic duration among progressive cases. Our mean duration estimates may overestimate this quantity, but the extent of overestimation is impossible to gauge; little is known about the fraction of prevalent, asymptomatic cases that are infinitely indolent. Certainly, given an estimate of the age-specific prevalence of infinitely indolent disease, the methods presented in our paper may be applied using appropriately adjusted prevalence estimates to yield an estimate of the mean asymptomatic duration among progressive cases.

A second assumption in using the prevalence/incidence approach is that incidence and mean duration are constant over time, so that the current number of prevalent cases, which derive from prior incidence, may be compared with the current number of incident cases to give a valid result. In other words, the prevalent cases in the numerator must have arisen as a result of an incidence process that is the same as that in the denominator. The need for this second assumption invalidates age-specific sojourn time estimates that derive from the ratio of age-specific prevalence to incidence, unless age-specific incidence and duration are constant. The assumption is implicit in the calculations leading to table 3, although the duration of time to detection of cancer may not be constant in the later time periods. Therefore, strictly speaking, these estimates are hypothetical, representing a steady-state situation in which asymptomatic incidence, clinical incidence, and mean time to detection are not changing over time.

The identity relating prevalence, incidence, and duration specifies that the prevalence of the condition of interest should be divided by the incidence of new cases of that condition. This yields an estimate of the average duration of the condition. Thus, to estimate the asymptomatic period for any cancer, the prevalence of asymptomatic disease should be divided by the incidence of asymptomatic disease. If the average asymptomatic duration is fairly short, for example, in lung cancer, most preclinical cases will progress to become clinically diagnosed within their lifetimes. In this case, observed or expected clinical incidence will be a reasonable surrogate for preclinical incidence. However, in cancers such as prostate cancer that may progress slowly in a nontrivial fraction of cases, many asymptomatic cases will not become diagnosed during their lifetimes; therefore, it will not be adequate to approximate the asymptomatic incidence by the observed clinical incidence. Indeed, the use of clinical incidence will lead to a denominator that is too low relative to the numerator, and the mean sojourn time will be overestimated. In recognition of this issue, much of the effort reported on in this paper has focused on obtaining an unbiased estimate of asymptomatic incidence, which we assume is more or less unchanging over time.

The use of autopsy data to estimate prevalence leads to an estimate of average sojourn time that can be interpreted as the time from disease onset to symptoms, under the assumption that an autopsy will identify disease if it is present. This is in contrast to the use of prevalence rates as estimated from screening data, which leads to an estimate that can be interpreted as the time from screen-detectable disease onset to symptoms. Therefore, it would be expected that studies based on screening data would yield slightly shorter duration estimates than those based on autopsy studies.

FIGURE 5. Comparison of age-specific asymptomatic incidence with that estimated using the method of Podgor and Leske (16). Estimates are based on prevalence from Carter et al. (1); clinical incidence from the Connecticut Surveillance, Epidemiology, and End Results (SEER) program (1960-1964) (12); mortality from 1986 US life tables (13); and cancer mortality from SEER (1984-1988) (8). For comparison purposes, the denominator of our age-specific asymptomatic incidence is the population size in the middle of the age interval, approximated by the average of the population sizes at the start and end of the interval.
The accuracy of autopsies and the autopsy prevalence of prostate cancer are topics of some debate. For example, while Carter et al. (1) found a prevalence of approximately 2–4 percent among men aged 40–49 years, a recent article (18) reported small foci of histologic cancer in 34 percent of patients in this age group. In addition, Carter et al. did not find a dissimilar prevalence in black men and white men, which is in contrast to the investigation of Guileyardo et al. (2) but is consistent with the findings of Sakr et al. (18). Our methods are applicable to any autopsy dataset that presents age-specific prevalence within sufficiently fine age intervals, and the paper by Carter et al. was, to our knowledge, the best source from this point of view. However, we truncated the asymptomatic incidence at age 89 years, since we were not comfortable extrapolating the Carter et al. prevalence curve beyond this bound. Given the years in which the autopsies in the Carter et al. study were performed, we assume that the prevalences represent the prevalence of prostate cancer in the (living) population in the absence of diagnostic interventions such as transurethral resection of the prostate or screening. In general, application of autopsy prevalence to the living population assumes that death due to causes other than prostate cancer does not happen with a greater or lesser frequency in prevalent cases than in cancer-free individuals. Incidentally, this assumption is also needed to be able to interpret the ratio of prevalence to incidence as an estimate of the mean of the duration that would have been observed if not for competing mortality.

Given the uncertainty of the prevalence figures from Carter et al. (1), we conducted a sensitivity analysis in which we evaluated the impact of inflating the prevalence on our results by using the data for whites and the clinical incidence from Connecticut SEER (1960–1964). Figure 6 indicates that the asymptomatic incidence results are highly sensitive to the prevalence values. When prevalence was 20 percent higher than our original input, asymptomatic incidence increased by 15–20 percent; when prevalence was 50 percent higher, asymptomatic incidence increased by 40–50 percent. However, as might be expected, the estimates of average duration are rather less affected. We obtained 11.64 years instead of 11.57 years for an increase in prevalence of 20 percent and 11.72 years for an increase of 50 percent. We also evaluated the sensitivity of our results to the assumption that the other-cause mortality rate in the symptom-free population could be approximated by the rate in the population. Specifically, we set other-cause mortality in the symptom-free population to a value 20 percent higher and then 20 percent lower than that in the population as a whole. Average duration estimates corresponding to clinical incidence levels among whites from the Connecticut registry in 1960–1964 were 11.42 and 11.72, respectively, which show relatively little sensitivity to this assumption. Finally, we considered the sensitivity of the results to the population age distribution weights \( f_i \). In our example, the weights were chosen to reflect the approximate age distribution in the United States in 1980 leading, in a sense, to an estimate of sojourn time standardized to the 1980 US population. It turns out that the sojourn time estimate is relatively insensitive to the exact choice of these weights, because the ratio of prevalence to incidence within each age stratum is fairly constant across ages in our example. In fact, it can be shown analytically that if \( P_i = cI_i \) for all \( i \), then the ratio in expression 1 will be independent of the weights \( f_i \).

The asymptomatic incidence estimates are useful in understanding the relative frequency of diagnosed and undiagnosed prostate cancers. For instance, the methods of Feuer et al. (19) can be applied to our preclinical incidence estimates to compute the lifetime probability of developing preclinical prostate cancer. The lifetime probability for whites (up to age 90 years) is approximately 36 percent. The same methods may also be used to compute the lifetime probability of a cancer diagnosis; for example, based on the incidence of prostate cancer in 1984–1986, the lifetime probability (up to age 90 years) of a prostate cancer diagnosis was estimated to be approximately 9.3 percent. This suggests that roughly 75 percent of asymptomatic prostate cancers will never be diagnosed if diagnostic capabilities remain at levels similar to those in 1984–1986, just prior to the introduction of PSA screening in the population. While this may seem staggeringly high, we note that based on the observed incidence in 1960–1964 (3), the lifetime probability of a prostate...
cancer diagnosis was only 6 percent. This suggests that 83 percent of prostate cancers would never be diagnosed if few or no diagnostic interventions existed. It is interesting to note that the lifetime probability of a cancer diagnosis based on the observed incidence in 1990–1992 is approximately 18 percent; thus, with diagnosis occurring at a rate similar to that observed in recent years, the fraction of asymptomatic prostate cancers that will never be diagnosed is reduced to 50 percent. However, the rates of diagnosis observed in these later years are highly unstable because of the introduction of prostate cancer screening in the population during this time.

The asymptomatic incidence and sojourn time estimates are biologically plausible and are consistent with the literature on PSA growth in prostate cancer cases. They confirm what has already been suspected for some time, namely, that prostate cancer is a relatively slow-growing neoplasm, and they suggest that among whites, 50–75 percent of new cases are unlikely to surface clinically. The estimates should be useful to researchers studying the natural history of the disease and designing effective and cost-effective screening programs.

ACKNOWLEDGMENTS

Research was supported by National Institutes of Health grant R29 CA70227 (R. Etzioni), by contract NCI N01-CN-05230 from the National Cancer Institute (R. Etzioni and R. Cha), and by National Research Service Award 5 F32 CA 71133–02 (O. Davidov).

The authors thank Benjamin F. Hankey from the Surveillance Program, Division of Cancer Prevention and Control, National Cancer Institute, for suggesting the calculation of lifetime risk of asymptomatic prostate cancer that will never be detected.

REFERENCES


APPENDIX

Consider first the situation in which it is assumed that new cases may not become clinically diagnosed. Let \( u_o \) be the probability of competing mortality in the interval. By assumption, \( u_o \) is the same for new cases as for the population in general. To simplify mathematical development, we redefine our notation as follows: Let \( n(1 - u_o) \) be the number of new cases who survive to the end of the interval (previously \( n \)). Assume that new cases cannot develop and become clinical in the same interval. Then, as in Leske et al. (7), the number of new cases who survive to their time of preclinical onset but die before the end of the interval is approximately

\[
m = \frac{u_o}{2}.
\]

Therefore, the total number of new cases who survive to their time of preclinical onset is approximately

\[
N = n(1 - u_o) + m = n(1 - u_o/2).
\]

(A1)

Now suppose that new cases may become clinically diagnosed. Let \( c_o \) be the fraction of new cases who become clinically diagnosed by the end of the interval,
and let \( n(1 - u_o)(1 - c_o) \) be the number of new cases who survive to the end of the interval without dying or being clinically diagnosed. The number of new cases who survive to their time of preclinical onset but then die of other causes without being clinically diagnosed is approximately \( n(1 - c_o)u_o/2 \), and the number of new cases who survive to their time of preclinical onset but then become clinically diagnosed before possibly dying of other causes is approximately \( n(1 - u_o/2)c_o \). Therefore, the total number of new cases who survive to their time of preclinical onset is

\[
N = n(1 - u_o)(1 - c_o) + n(1 - c_o)u_o/2
\]

\[+ n(1 - u_o/2)c_o = n(1 - u_o/2), \quad (A2)\]

which is the same as expression 1. Therefore, regardless of the probability that a new case may develop and become clinically diagnosed in the same interval, the number of new cases in the numerator of the asymptomatic incidence remains the same. In the notation of expression 8, \( n \) and \( m \) may change, but their sum remains the same. Intuitively, this result may be thought of as follows: The cases making up the asymptomatic incidence in a given interval consist of all those newly developing cases who survive to their time of preclinical onset. The only event that will prevent a new case from reaching his or her time of preclinical onset is competing mortality. However, this is independent of the probability \( c_o \) that a new case will become clinically diagnosed in the same interval as his or her preclinical onset. Consequently, the number of new cases who survive to their time of preclinical onset is effectively independent of \( c_o \).