Assessing Changes in the Impact of Cancer on Population Survival Without Considering Cause of Death

Barry W. Brown, Christopher Brauner, Lawrence B. Levy*

Background: We have previously argued against the calculation of cancer-specific death rates as philosophically undefined and biased. Deaths attributed to cancer during a particular year occur in patients diagnosed during an unknown distribution of past times, so cancer-specific death rates cannot be used to assess changes in the impact of cancer on survival of the population at specific periods of diagnosis. Purpose: Our goal was to develop and analyze three measures of the impact of cancer on population survival that do not use the attributed cause of death: 1) the age-adjusted proportion of the population diagnosed with cancer in a particular year and projected to be dead of any cause by a particular age; 2) the same measure corrected for population mortality; and 3) the expected years of life lost to a 20-year-old individual because of the possibility of a diagnosis of cancer. Methods: Data on all adults diagnosed with any cancer during the period from January 1973 through December 1990 were obtained from the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute. The measures were calculated separately for various combined sex and race groups. Three 2-year diagnosis periods spaced 5 years apart were considered: 1975-1976, 1980-1981, and 1985-1986. A statistical model was used to extrapolate survival beyond observation; the same model was used for patients diagnosed in the three time periods to minimize the effect of possible model misspecification on changes. Results: Cancer incidence increased for three of the four sex-race groups; age-adjusted changes from 1975 through 1976 to 1985 through 1986 were +11.5% for white men, +6.9% for white women, +15.1% for black men, and −9.2% for black women. Human immunodeficiency virus (HIV)-related cancers were responsible for an increased cancer incidence at early ages in white men in the latest time period studied. There was a decrease in the incidence of gynecologic cancers at early ages; the decrease was greater among black women (−55.1%) than among white women (−39.3%). Age- and incidence-adjusted 5-year survival increased by 17.9% for white men, 2.3% for white women, and 7.4% for black men and decreased by 14.1% for black women. When the data from 1985 through 1986 were compared with those from 1975 through 1976, the expected number of years lost to a 20-year-old individual because of cancer changed as follows for the various sex-race groups: +1.4% for white men (−4.0% if HIV-related cancers were not included in the calculation), +2.1% for white women, +12.2% for black men, and +8.8% for black women. Conclusions: For white men and women, there has been an increase in both the incidence of and survival following the diagnosis of cancer; the two effects nearly cancel in our measures. The experience of black men and women has worsened because of increasing incidence or decreased survival.

The age-adjusted, cancer-specific death rate is commonly used to assess progress against this disease (1-5). We are skeptical of this measure because we believe that there are usually several factors contributing to an outcome of death. The relative importance of these factors to mortality is unobservable (6). We have found that the rate of noncancer deaths among cancer patients exceeds the rate of noncancer deaths among the general population, and that the excess noncancer mortality in cancer patients is largely concentrated in the first few years following the diagnosis of cancer. This finding has led us to suspect that these deaths, although not attributed to cancer, may have been associated with cancer treatments and that they should be included in any assessment of the effect of cancer on population mortality.

Deaths attributed to cancer during a particular year occur in patients diagnosed during an unknown distribution of past times, so cancer-specific death rates cannot be used to assess the changes in the impact of cancer on survival of the population at specific periods of diagnosis. We propose, therefore, that this assessment be based only on cancer incidence and consequent survival and not utilize the attributed cause of death.

In this study, we compared the population survival experience from cancer (all types combined) in three 2-year time periods of cancer diagnosis: 1975-1976, 1980-1981, and 1985-1986. This assessment was performed separately by sex and race. Three measures of survival experience are calculated: 1) the age-adjusted proportion of the population that was diagnosed with cancer in a particular year and was projected to be dead of any cause by a particular age, 2) the age-adjusted proportion of the population that was diagnosed with cancer and projected to be dead in excess of U.S. population experience by a particular age, and 3) the expected number of years of life lost to a 20-year-old individual because of the probability of being diagnosed with cancer and of being exposed to its increased mortality rate.

*Affiliation of authors: Department of Biomathematics, The University of Texas M. D. Anderson Cancer Center, Houston. Correspondence to: Barry W. Brown, Ph.D., Department of Biomathematics, Box 237, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX 77030.

See "Notes" following "References."
Subjects and Methods

Data

The incidence of cancer and survival following cancer diagnosis was obtained from the Surveillance, Epidemiology, and End Results (SEER) Program tapes on cases diagnosed from 1973 through 1990, with follow-up complete through December 1990 (7). The same tapes contained age-specific population estimates by race and sex; these were used in calculating incidence. The SEER Program consists of nine population-based registries covering nine geographic regions of the United States (8).

Excluded from this study were patients with in situ tumors, patients in whom the only data on cancer was from a death certificate, and patients diagnosed with cancer when younger than 20 years old. The population size estimates combined ages greater than 85 years, so incidence by age can only be calculated up to this value. Consequently, patients diagnosed with cancer at ages more than 85 years were also excluded. The only race groups considered were whites and blacks, since there were not enough cases in other race groups to allow a similar analysis.

The population rates of death were obtained from the National Center for Health Statistics of the United States (9). Before 1979, mortality rates were available for only two race categories: white and nonwhite. In 1979, a third category was added for blacks. To estimate black population mortality rates for the 5-year time period of 1973-1978, the ratio of black to nonwhite mortality rates was calculated for each sex and age group for each year within the range of 1979-1990. This ratio was linearly extrapolated to the years 1973-1978, and black population mortality was estimated as this extrapolated ratio multiplied by the nonwhite mortality rate.

The U.S. population is aging with time; for this reason alone, the incidence of cancer is increasing. Since increased longevity should not be considered a failure in cancer containment, age adjustment is used to extrapolate the experience of an observed population to that of a standard population with a fixed age distribution. We take the standard age distribution to be that of the population of the United States in 1980. Age-adjusted cancer incidence estimates the overall incidence of cancer in a population with the chosen standard age distribution assuming an age-specific incidence rate from a time period possibly different from that used for the standard population. To obtain a similar estimate for the proportion of cancer patients in the standard population that would survive 5 years following diagnosis, adjustments must be made for both age and incidence.

Since most cancers occur in the elderly, their 5-year survival experience must be weighted more than that of the young.

Computations

The first criterion, the age-adjusted proportion of the population that is diagnosed with cancer in a particular year and is projected to be dead by age \( A \),

\[ P_{\text{adj}}(A) = \sum_{d=20}^{85} P_d(d)(1 - S_d(A|d)). \]

In this formula, \( d \) is the age at diagnosis. \( P_d(d) \), the age-adjustment factor, is the proportion of the 1980 U.S. population between the ages of 20 and 85 years inclusive who are age \( d \). \( S_d(A|d) \) is the probability that an individual will survive to age \( A \) or longer given a diagnosis of cancer at age \( d \); thus, \( 1 - S_d(A|d) \) is the probability that an individual will be dead by age \( A \), given diagnosis at age \( d \).

The second criterion, the age-adjusted proportion of individuals diagnosed in a particular year and projected to be dead in excess of population experience by age \( A \),

\[ P_{\text{adj}}(A), \]

is calculated using the same formula, but with \( S_d \) replacing the survival function \( S_d \). \( S_d \) is the decrement in survival for cancer patients compared with the U.S. population.

The calculations of \( S_c \) and \( S_{cc} \) use the observed survival experience of those diagnosed at age \( d \). For each year, \( i \), following diagnosis, let the number of individuals who die in that year be \( m_i \) and let the number who are lost to follow-up be \( c_i \). The usual reason for being lost to follow-up was that the patient was alive at the end of 1990. The Kaplan–Meier formula (10) was used to calculate \( S_c \) from the \( m_i \) and \( c_i \). With the use of population mortality tables, a calculation was made of the number of deaths, \( e_i \), that would be expected in year \( i \) from a general population matched to the cancer population by age, sex, and race. The observed number of deaths, \( m_i \), was decreased by \( e_i \) to correct for population experience, and the number of observations considered lost to follow-up during year \( i \) was increased by the same amount to maintain the correct sample size. The Kaplan–Meier method was then used to calculate \( S_{cc} \) from these modified data.

These criteria measure the survival experience of cancer patients diagnosed during 1 year. There is no reason, however, that the incidence and cancer patient survival must be estimated using only the data from 1 year. Indeed, to decrease variability in this investigation, the incidence and survival experience from 2 consecutive years were combined to obtain one composite year of diagnosis. The composite year of diagnosis demonstrates the effect of the appropriate average of the experience of these 2 years.

The first two criteria examine the impact of cancer on survival in a single year. The third criterion examines this impact on a 20-year-old individual who was exposed for a lifetime to the cancer incidence, survival, and population mortality of that single year. This criterion is the expected number of years lost to this individual due to the possibility of being diagnosed with cancer before age 86 years. This number, \( L_c \), is:

\[ L_c = \sum_{d=20}^{85} P_d(d)(d) \sum_{j=1}^{20} (j s_j(d + j f_j) - s_c(d + j f_j)). \]

In this formula, \( s_j(d + j f_j) \) is the probability that an individual in the general population who was alive at age \( d \) will survive exactly \( j \) years, i.e., death occurs at age \( d + j \). Similarly, \( s_c(d + j f_j) \), is the probability that a cancer patient will survive precisely \( j \) years following diagnosis at age \( d \).

The expected-years criterion is not a one-number summary of either of the other two measures. These two measures consider only the impact of diagnosis in a single year, while the expected-years measure extrapolates this impact over a lifetime. Also, the expected-years criterion differentially weights a death at a young age much more than death at an old age. The other two measures count only the number of deaths before a particular age.

Although the formulas are presented in terms of single-year age groupings for comprehensibility, age groupings of 5 years were used in the calculations, corresponding to those of the SEER population files. All survival calculations were performed by month rather than by year. In calculating \( L_c \), hazard of death in the population was taken to be constant after age 85 years. Separate mortality tables were used for each year prior to 1987; 1987 tables are used for all years after 1987.

Extrapolating Survival

The difficulty in applying these formulas lay in the estimation of survival experience beyond observation. The proposed criteria use survival to age 85 years, while diagnosis can occur as early as age 20 years. The SEER data, on the other hand, cover a time span of only 15 years. There are two ways of dealing with this problem. One way is to consider deaths only within some delineated period following diagnosis; 15 years is a natural boundary. The other way is to use a statistical model to extrapolate survival beyond that observed. For men, it matters little which method is used, because few men survive 15 years beyond a diagnosis of cancer. Women, however, do survive beyond this point in appreciable numbers. We chose extrapolation, because the proposed measures have an inherent interest over the full range of survival. The same statistical model is used for the several years of diagnosis considered, so errors in prediction will largely cancel in examining changes over time.

Originally, it was hoped that the very flexible log-F model described by Kalbfleisch and Prentice (11) would serve to describe cancer survival experience. The model was fit to 5-year age-at-diagnosis intervals for white male cancer patients diagnosed in 1975-1976, and the estimated values were compared with the observed survival for the 15 available years of follow-up time. Agreement was good for a few years immediately following diagnosis, but the model did not adequately describe later survival experience. A fit of the same model to survival experience at 5-10 years following diagnosis provided a reasonable description, not only for this time period but of the 5 additional years of observation. Fig. 1 shows predicted and observed survival for individuals in several age-at-diagnosis groups. On the basis of the adequacy of this fit, the log-F model was fit to the 10- to 15-year experience, and the results were used to predict survival of individuals for more than 15 years. In a few cases, particularly the youngest age groups in which there were few observations, the predicted \( S_c(A|d) \) was better than population survival. In these cases, the prediction was curtailed to match the population. This method was applied sepa-
rately to the 13 5-year age-at-diagnosis groups for patients diagnosed in 1975-1976, and the same model was used for all diagnosis periods considered.

Actual survival experience of case subjects was used when available, even if it reflected the experience of an earlier period of diagnosis. Thus, the 5- to 10-year survival experience of patients diagnosed in 1980-1981 was used for patients diagnosed in 1985-1986, and the experience of the 10- to 15-year survival of patients diagnosed in 1975-1976 was used for patients diagnosed in 1980-1981 and 1985-1986. In applying earlier survival experience to later diagnosis periods, the mortality rate from the earlier period was applied to the observed cumulative survival to obtain an extended cumulative survival function.

Only observed changes in survival of subjects were used to calculate changes in the measures over periods of diagnosis. For example, in calculating the change from 1980-1981 to 1985-1986, only the survival experience of subjects for the first 5 years following diagnosis differs; the same survival experience was used for both groups of subjects beyond 5 years. Consequently, changes in their long-term survival experiences are not reflected in the calculations. However, there are no data from which to estimate such changes.

Lack of Statistical Considerations

This work contains neither confidence intervals for estimates nor P values for differences. In part, this omission is made because methods for obtaining these quantities for the complex estimators described in the previous section have not been completed. However, we feel that the use of these estimators of variation in studies with a large number of case subjects, such as the current one, is generally optimistic. Confidence intervals and P values may be too small, because unconsidered sources of systematic error may be at least as important as random variability in such studies. Such sources of systematic error might include errors in age-specific population size estimates, underascertainment of case subjects, systematic losses of subjects to follow-up, and slight misspecification of the form of long-term survival of subjects.

Results

Before showing the results of our three measures, we examined the changes that occurred in cancer incidence and survival during the three time periods studied.

Fig. 2 shows the incidence of cancer in white men, black men, white women, and black women. The left panels (A, D, G, and J) show the number of new case subjects per year per 100 000 people. The middle panels (B, E, H, and K) show how the incidence of cancer differs in the 1980-1981 and 1985-1986 time periods from that in 1975-1976. The right panels (C, F, I, and L) display the same differences when men with human immunodeficiency virus (HIV)-related cancers and women with gynecologic cancers are not included in the calculations. The age-adjusted change in incidence between 1975-1976 and 1985-1986 was +11.5% for white men, +6.9% for white women, +15.1% for black men, and −9.2% for black women. The incidence of cancer increased with time for people older than 55 years in all race-sex groups; this increase was greatest for black men. The increase in incidence of cancer in black men was noted at most cancer sites, with the largest increases occurring in cancers of the lung, prostate, and colon and rectum (data not shown).

The peak in incidence of cancer among white men ages 30-45 years during the 1985-1986 time period was primarily caused by two HIV-related cancers: nodal non-Hodgkin’s lymphoma and nonepithelial skin cancer. These cancers constituted 2.4% of the age-adjusted total diagnosed in 1975-1976, but 4.6% of those diagnosed in 1985-1986. The peak in incidence disappeared when these cancers were eliminated from the incidence calculations. There was almost no effect of HIV-related cancers on cancer incidence in black men of the same age range; these constituted 1.6% of the age-adjusted total diagnosed in 1975-1976 and 2.6% of those diagnosed in 1985-1986.

Decreases in incidence of gynecologic cancer, particularly cancer of the cervix, were responsible for the overall lessening of cancer incidence in young women (Fig. 2, H and K). This incidence decreased by 39.3% in white women and 55.1% in black women from 1975-1976 to 1985-1986. The incidence of all other types of cancer combined increased in most age groups (Fig. 2, I and L); the primary sites of increase were lung and breast cancer (data not shown). Excluding gynecologic cancer from the calculation, the age-adjusted incidence rate increased by 22.9% in white women in the 10 years examined; the corresponding figure for black women was an increase of 23.5%.
Fig. 3 shows 5-year survival rates. The panels are analogous to those of Fig. 2. The age- and incidence-adjusted change between 1975-1976 and 1985-1986 was +17.9% for white men, +2.3% for white women, +7.4% for black men, and −14.1% for black women.

The decrease in the proportion of women under the age of 50 years surviving 5 years is largely due to the decrease in incidence of gynecologic cancer as well as the increase in incidence of lung cancer. It may appear aberrant that the decrease in gynecologic cancers could, by itself, cause a decrease in survival. However, gynecologic cancers are associated with a higher survival rate than the other cancers, so the decrease in incidence of gynecologic cancer increased the proportion of other cancers and worsened the overall survival. When gynecologic cancers were not considered in the calculations, a distinct improvement in the overall 5-year survival rate with time was seen for white women (+8.5%) but not for black women (+1.8%).

The decrease in survival of white men at younger ages appeared to be caused almost entirely by HIV-related cancers (Fig. 3, B and C), because when these case subjects were removed from the analysis, survival increased with time. The proportion of 5-year survivors aged 30-35 years decreased from 67.3% to 58.9% over the 10 years studied when individuals with HIV-related cancers were eliminated from the calculation; the proportion increased from 66.9% to 71.9% when these case subjects were excluded. The effect of HIV-related cancer in black men was
much smaller than in white men; among black men, the change in the proportion of 5-year survivors aged 30-35 years during 1985-1986 with and without these cancers was only 0.7%.

We return to our proposed measures of the impact of cancer on survival. The upper set of lines in the left column of Fig. 4 show, by race and sex, the proportion of the population diagnosed in a particular year with cancer and projected to be dead of any cause by age, $P_{DD}(A)$. The lower set of lines shows the proportion of the population diagnosed with cancer in a particular year and projected to be dead in excess of population experience, $P_{DE}(A)$. This figure is not intended to reveal detailed changes with time, but rather to show the effect of age, $A$, on these measures.

The middle and right columns of Fig. 4 show the changes that occurred in these measures from the diagnosis period of 1975-1976 to those of the 1980-1981 and 1985-1986 periods. When HIV-related cancers were not included in the calculations for white men, $P_{DD}(A)$ was nearly constant until approximately age $A = 72$. After age 72 years, this proportion increased with time of diagnosis. When only deaths in excess of population experience were considered, the measure decreased very slightly for all ages over successive diagnosis periods. Including individuals with HIV-related cancers increased both the raw and population-corrected measures in the latest diagnosis period to a level greater than that for the two previous periods for all ages greater than 30 years.
The incidence of cancer among black men increased between 1975-1976 and 1980-1981 and the changes in survival did not offset these increases. Hence, $P_{DD}(A)$ increased between these two periods of diagnosis, as did the same measure corrected for population mortality. Both measures remained approximately constant between 1980-1981 and 1985-1986.

For white women, both measures showed a slight improvement between 1975-1976 and 1980-1981, but by 1985-1986 the measures worsened at ages over 70 years. These changes reflected the joint effect of an overall increase in survival, a decrease in the incidence of gynecologic cancers, and an increase in incidence of breast and lung cancer (data not shown).

The proportion of black women diagnosed with cancer in a particular year and projected to be dead by a particular age decreased from 1975-1976 to 1980-1981 and then remained approximately constant. The change was primarily attributable to a large decrease in the incidence of gynecologic cancers. The improvement occurred at advanced ages. Since everyone dies eventually, the proportion of the population diagnosed and dead at advanced ages is close to the age-adjusted incidence of cancer. Because the death rate associated with gynecologic cancers did not greatly exceed that of the population, this decrease in incidence had less of an effect on the excess deaths. The changes with time in this measure were much the same for black and white women.

Fig. 5 shows the expected years of life lost to a 20-year old.
individual because of the possibility of being diagnosed with cancer and the subsequent increase in mortality rate. If cancer were cured, so that those diagnosed with it lived at population rates, the population life expectancy at age 20 years would increase by 2.5-3.5 years, depending on race and sex.

When HIV-related cancers were not included in calculations for white men, there was a 4% decrease in the expected years of life lost between the first and last diagnosis periods; if individuals with HIV-related cancers were included in the calculation, there was a 1.4% increase in expected years lost. For white women, the expected years lost increased by 2.1% over the period considered. For black men, there was a 12.2% increase in the expected years lost, and for black women the corresponding figure was an increase of 8.8%.

In summary, there has been an overall increase in the incidence of cancer between 1975 and 1985 for black and white men and women, with the exception of a decrease in gynecologic cancers in young women of both races. For whites, the increase in incidence has been largely offset by improvement in survival; for blacks, this is not so. The net effect for whites ranged from a minimal decrease for men to a minimal increase for women. The impact of cancer on survival for blacks, particularly men, increased over the time period studied.

**Discussion**

Various measures of change on the impact of cancer on mortality are either evident or have been proposed. These include, in addition to the measures introduced here, changes in incidence rate, survival, relative survival, cancer-specific mortality (2), and a variant on the expected years of life lost to cancer. This variant (5) sums the expected years of life remaining to a member of the general population at the time of death attributable to cancer.

Incidence, survival, and relative survival individually are incomplete measures of the impact of cancer on mortality. Survival and relative survival do not consider incidence. Thus, these factors may interact in peculiar fashions, as exemplified by the decreasing incidence of gynecologic cancer contributing to an apparent decrease in survival.

We have already stated that we were suspicious of any assessment that uses the attributed cause of death. Several factors other than attributed cause of death can bias measures of the impact of cancer on survival; these include lead-time bias, false-positive error, and false-negative error.

**Lead-Time Bias**

Early cancer detection will by itself increase the time between diagnosis and death, even without any change in the time of death. Cancer-specific mortality is immune to this bias because the time of diagnosis is not used. However, cancer-specific mortality cannot be used to infer changes in incidence or survival in specific-diagnosis time periods.

Our measures would be similarly immune to lead-time bias were follow-up long enough to observe all deaths of a population diagnosed in a particular time period. However, long-term survival rates from earlier times are used to extrapolate survival in those with short follow-up. Early detection will by itself change these long-term survival rates, leading to some bias because of incomplete observation of changes. To estimate the possible extent of this bias in an extreme case, we decreased the age at diagnosis of a random half of the 1985-1986 population by 2.5 years (leaving the age at death unchanged). The probability of diagnosis in a particular year and death by a particular year was recalculated; the changes from the original were small. The maximum change for white men (including HIV-related cancers) was an increase of 43 per 100 000 at age 67 years; for
during different times. The measures all use the observed incidence of the impact at specified times and comparisons of impact survival are proposed. These measures allow assessment of these groups. Differential ascertainment of cancer over time in minority groups could result in overestimates of the incidence of cancer. Consequently, undercounting of individuals in a particular year and projected to be dead of any cause by a certain age is subject to lead-time bias, since only the age at death is considered. The age-adjusted proportion of the population diagnosed with cancer in a particular year and projected to be dead of any cause by a particular age is subject to false-positive bias; this bias is corrected by the adjustment for population mortality.

False-Negative Error

The effect of false-negative error would be to improve any measure of the impact of cancer on survival by lessening both incidence and the death rate because of cancer. It is ironic that, because of this effect, improved access to medical care could lessen the apparent improvement in cancer survival experience. Improved screening methodology or access would make the measures described look worse for a while, from the period of immediate increase in incidence until the effect of increased survival because of early treatment is seen.

Accurate population counts are essential to an assessment of incidence of cancer. Consequently, undercounting of individuals in minority groups could result in overestimates of the incidence of these groups. Differential ascertainment of cancer over time could likewise bias observed changes in incidence and perhaps survival. Although the magnitude of these effects cannot be assessed using the current data, we believe that it is small.

Summary

In summary, three measures of the impact of cancer on population survival are proposed. These measures allow assessment of the impact at specified times and comparisons of impact during different times. The measures all use the observed incidence of cancer and recently available observed survival experience of cancer patients. A statistical model is used only for long-term survival for which observations are not available, and the same model is used for all diagnosis periods.

None of the proposed measures uses the attributed cause of death with its unknown bias. All measures are largely immune to lead-time bias, since only the age at death is considered. The age-adjusted proportion of the population diagnosed with cancer in a particular year and projected to be dead of any cause by a particular age is subject to false-positive bias; this bias is corrected by the adjustment for population mortality.

References

7. Surveillance, Epidemiology, and End Results (SEER) Program public use tape (1973-90), National Cancer Institute, Division of Cancer Prevention and Control, Surveillance Program, Cancer Statistics Branch.

Notes

1Editor’s note: SEER is a set of geographically defined, population-based central tumor registries in the United States, operated by local nonprofit organizations under contract to the National Cancer Institute (NCI). Each registry annually submits its cases to the NCI on a computer tape. These computer tapes are then edited by the NCI and made available for analysis.