Cancer Epidemiology 34 (2010) 667–681

Contents lists available at ScienceDirect

Cancer Epidemiology

The International Journal of Cancer Epidemiology, Detection, and Prevention

journal homepage: www.cancerepidemiology.net

Cancer survival as a function of age at diagnosis: A study of the Surveillance, Epidemiology and End Results database

Mena N. Bassily [a,b,c,*], Richard Wilson [c], Francesco Pompei [c], Dimitriy Burmistrov [d]

[a] Harvard School of Public Health, 401 Park Drive, Boston, MA 02215, USA
[b] Menoufiya University, Shebin El-kom, Menoufiya, Egypt
[c] Harvard University, Department of Physics, Jefferson Laboratory, 17 Oxford Street, Cambridge, MA 02138, USA
[d] ITA software, 141 Portland Street, Cambridge, MA 02139, USA

ARTICLE INFO

Article history:
Accepted 14 April 2010
Available online 12 May 2010

Keywords:
Cancer survival
Cause-specific survival
Cancer in old age
SEER
Kaplan–Meier
Survival phases

ABSTRACT

Background: Recent research suggested that cancer survival has improved in recent cohorts. Improvement in cancer survival is considered a valid indicator of the quality of care introduced to the patients. The aim of this study is to investigate the changes in the survival profile over age for patients with the most incident cancers. *Methods*: Survival data of 3.94 million patients diagnosed with 23 primary-site cancers within the periods of 1979–1983, 1989–1993, and 1999–2003 were adopted from the Surveillance, Epidemiology and End Results database. Gender and cause-specific survival probabilities were estimated at one, three, and five years after diagnosis using the Kaplan–Meier survival estimate. Survival was presented for each of the studied cancers, cohorts, and sexes in the form of line graphs as a function of age at diagnosis. Error bars demonstrated the probability of error at 95% confidence level. *Results*: The graphs demonstrated that cancer survival was improved over the successive cohorts for most cancers, with several exceptions such as brain and lung cancers. The relation between survival and the age at diagnosis was generally described in the form of a gradual decline phase and a rapid fall-off phase at 70–80 years of age, with few exceptions as in leukemia and Hodgkin lymphoma. Patients who survived for three years were more likely to live for five years after diagnosis, but this prediction could not be extrapolated to the one-year survivors. *Conclusion*: Further studies on tumor-specific characteristics and treatment modalities of these patients are suggested for clarification of the possible causes of variations in patient's survival profile over age.

1. Introduction

Studying survival of different cancers has important practical value for patients, providers, and researchers [1,2]. Cancer patients may wish to know how their prognosis is changing over time, and what is their life expectancy based on the disease status. The proper understanding of prognosis may help both of the physicians and the patients decide on treatment options, balancing the personal values for quality versus quantity of life [3]. Knowledge of cancer survival provides a more objective basis to deem a patient "cured" of their disease. Providers can make use of survival information to more objectively determine an appropriate frequency of follow-up visits and aggressiveness of surveillance testing based on patient's current risk profile. When designing clinical trials, clinical researchers may also find it useful in helping to determine sufficient follow-up times for trial endpoints [4].

Period survival more accurately describes patients' prognosis since the overall survival projections are often discouraging and not necessarily pertinent for patients who have survived the initial treatment period, as prognosis after initial management is not static. Patients who have survived an interval of time after diagnosis have a different probability of surviving for the following five years from that which was estimated at the time of diagnosis [5].

Two forms of net (non-crude) survival analyses are available in the SEER*Stat; relative and cause-specific survivals. Both of these methods present the likelihood that cancer patients will not die from causes associated with their cancer. Relative survival is derived by comparing the survival of all causes of death in a group of cancer patients to survival of all causes of death in a cancer-free age- and sex-matched population. This means that in relative survival, unlike cause-specific survival, both of the numerator and the denominator are derived from different populations. Besides,

Abbreviations: DNA, deoxyribonucleic acid; KM, Kaplan–Meier; NCI, National Cancer Institute; SEER, Surveillance, Epidemiology, and End Results; SEER*Stat, Surveillance, Epidemiology, and End Results statistical software.

* Corresponding author at: Harvard University, Jefferson Laboratory, Room 257-A, 17 Oxford Street, Cambridge, MA 02138, USA. Tel.: +1 617 332 4823; fax: +1 617 495 3387.
E-mail addresses: mbassily@hsph.harvard.edu, mnf_prof@hotmail.com (M.N. Bassily).

1877-7821/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.canep.2010.04.013

relative survival is entirely based on the assumption that the other causes of death in the Surveillance, Epidemiology and End Results database (SEER) cohorts and in the general U.S. population are highly comparable. This assumption may be misleading if one factor is present at a cohort that may increase the risk of dying from causes other than cancer, compared to the general population. For example, smoking may be highly represented in a lung cancer cohort who often die due to causes related to smoking other than lung cancer e.g. heart disease. Relative survival cannot separate the risk of death from lung cancer from the risk of dying of non-cancer causes due to smoking. Therefore, relative survival fraction for those with lung cancer is an overestimate of the effect of lung cancer alone [6].

This article provides an overview of cause-specific survival of many cancers as a function of age at diagnosis to assess whether there is a specific pattern or trend which describes the survival of all cancer types, specially at old age groups (85+ years). To provide a detailed modeling of survival at old ages, the 85+ age group provided in the SEER database were re-classified into smaller five-year age groups. This study may also be used by clinicians as a benchmark to tell the patients of different cancers about the probability of surviving for one, three, and five years. We were also concerned with studying the progress in cancer survival over the last three decades to detect if the advances in cancer screening, diagnosis, and treatment were interpreted in the form of improvement of cancer patients' longevity.

2. Materials and methods

The Surveillance, Epidemiology and End Results database of the National Cancer Institute (NCI) is the largest population-based cancer registry in the United States, which geographically encompasses approximately 26% of the U.S. population. The latest SEER17 cancer registry collects data on cancer incidence and survival from seventeen population-based cancer registries which are Connecticut, Iowa, New Mexico, Utah, Hawaii, the metropolitan areas of Detroit, California (San Francisco-Oakland, San Jose-Monterey, and Los Angeles), Atlanta, Seattle-Puget, rural Georgia, Arizona, New Orleans, Louisiana, New Jersey, Puerto Rico, Kentucky, in addition to American Indian Alaska natives [7]. These registries were chosen for their completeness and their adequate representation of all races and minority populations [8]. The SEER program standard for data completeness is 98% [9].

The SEER program registries routinely collect data on demographic characteristics of the patients, tumor characteristics and staging, as well as a follow-up for the survival status. An epidemiological study comparing SEER areas with non-SEER areas in the United States concluded that the age and sex distributions of these areas were comparable [8]. Data about survival are actively collected by SEER cancer registries and reported to the NCI, data are then ascertained from hospital records, private laboratories, radiotherapy units, other health care service units, and from death certificates when cancer is listed as a cause of death [10].

2.1. SEER data confidentiality

All available data in the SEER database are retrospective in nature. Personal identifiers are absent from the database. All variables that might lead to reidentification such as the date of birth have been removed or transposed. Any remaining risk of reidentification has been minimized by the governing agency in not allowing the data to be available as public-use information. Investigators have to sign a legally binding data-use agreement with the Centers for Medicare and Medicaid Services and SEER [11].

2.2. Statistical methods

SEER*stat software (version 6.5.2, National Cancer Institute, Bethesda, MD) was used to download and analyze patients' cause-specific survival. Cause-specific survival is defined as the probability of surviving certain type of cancer for certain period of time, excluding death from causes other than the cancer of interest (as reported in the death certificate and/or autopsy). In this study, only primary cancers were considered. Consequently, the cause-specific survival presented in this study is an estimate of the likelihood that primary cancer patients will not die from primary cancer only, not from a secondary cancer or other associated causes. To calculate cause-specific survival we used the Kaplan–Meier product-limit method available in SEER*Stat program. The Kaplan–Meier (KM) estimator is the nonparametric maximum likelihood estimate of survival function [12]. In the SEER database, survival is available on each cancer patient from the time of initial diagnosis to the date of last contact or the date of death if the patient had died. The KM estimator calculates the survival probability at a defined period of time based on calculation of the survival estimate at the end of each month of this period [7]. This method allows for early exclusion of those who deceased during the specified time interval and for prompt censoring of the cases lost from follow-up, with regular adjustment of the at-risk group (denominator) on a monthly basis, in order to introduce an accurate net survival estimate for the defined period of time.

One, three, and five-year survival data for approximately 3.94 million patients diagnosed with different primary-site cancers (21 sites for females and 20 sites for males) were analyzed using the KM method. The KM analysis allows estimation of survival over time, even when patients drop out or are studied for different lengths of time. For each interval, survival probability is calculated as the number of patients surviving for certain time divided by the number of patients at risk. Patients who have died, dropped out, or did not reach the time of follow-up are not counted as "at-risk" and are considered censored. Eventually, the probability of surviving to any point is estimated from the cumulative probability of surviving each of the preceding time intervals. When the population is large enough (such as in the SEER database), the estimated Kaplan–Meier survival approaches, to a large extent, the true survival of that population [12,26,27].

Cancer cases were grouped according to age at diagnosis into twenty-three 5-year age groups ranging from 0–4 to 110–114 years old. Age and cause-specific survival rates for each type of cancer were calculated separately for each sex and over three different cross-sectional cohorts (1979–1983, 1989–1993, and 1999–2003). We used three cohorts derived from three successive decades to study the possible time to time variability in cancer survival. To investigate the changes in cancer survival over decades, the age and sex-specific five-year survival fraction for the three studied cohorts was demonstrated in line graphs (Fig. 3). To show the survival probability at different periods after diagnosis, the one-, three-, and five-year survival fractions for each type of cancer and sex were plotted with the age at diagnosis in the form of line graphs (Figs. 1 and 2).

For graphical clarity reasons, only the five-year survival plots show error bars. In case of appearance of a very large error bar (>75% of the entire scale of the *y*-axis), the corresponding data point was considered statistically unreliable and the error bars were not shown at such points. The points which have statistical uncertainty were connected to the other points of the line graph with a grey line. Two types of error bars are presented in the plots. The first is the 95% confidence bands provided by SEER*Stat program. These confidence intervals were presented whenever SEER*Stat was able to calculate them in the survival-session outputs. In cases when the last patient is dead of the cancer of