Chapter 1
Introduction

Lynn A. Gloeckler Ries, Marie-Josephe D. Horner, and John L. Young, Jr.

INTRODUCTION

The Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute (NCI) has devoted this monograph to examining cancer survival by patient and tumor characteristics for cancers diagnosed during the period 1988-2001. The analyses focus on cancer survival in adults aged 20 years and older, with the exceptions of acute lymphoblastic leukemia (all ages), placenta (ages 15+), and Hodgkin lymphoma (ages 15+). This chapter describes the sources of the data and the methods used. It also provides a summary of the results. Each subsequent chapter focuses on a distinct anatomical site and associated histologies.

DATA SOURCES

Surveillance, Epidemiology, and End Results (SEER) Program

The Surveillance, Epidemiology, and End Results (SEER) Program was established in 1973 as part of the National Cancer Institute (NCI). A sequel to two earlier NCI initiatives (the End Results Program and the Third National Cancer Survey), the SEER Program has evolved in response to the mandate of the National Cancer Act of 1971, which requires the collection, analysis, and dissemination of data relevant to the prevention, diagnosis, and treatment of cancer. The SEER Program (http://seer.cancer.gov) collects cancer incidence, treatment, and survival data which are used to monitor the burden of cancer on the population of the United States. The NCI contracts with medically-oriented nonprofit institutions, such as universities and state health departments, to obtain data on all in situ and invasive cancers diagnosed in residents of the SEER geographic areas, except for basal cell and squamous cell carcinomas of the skin and in situ cervical cancer.

The analyses in this monograph are based on data from 12 geographic areas representing approximately 14% of the United States population: the States of Connecticut, Iowa, New Mexico, Utah, and Hawaii; the metropolitan areas of Detroit, Atlanta, San Francisco, San Jose, Los Angeles, and Seattle; and ten counties in rural Georgia. Cases were diagnosed during the period 1988-2001 and followed through 2002. All registries contributed data for diagnosis years 1988-2001, except Los Angeles, which contributed data for 1992-2001.

Table 1.1: All Cancers: Number of Cases and Exclusions, 12 SEER Areas, 1988-2001

Number Selected/Remaining	Number Excluded	Reason for Exclusion/Selection
2,246,603	0	Select 1988-2001 diagnosis (Los Angeles for 1992-2001 only)
1,925,529	321,074	Select first primary only
1,901,067	24,462	Exclude death certificate only or at autopsy
1,874,432	26,635	Exclude unknown race
1,870,229	4,203	Active follow-up and exclude alive with no survival time
1,846,162	24,067	Exclude children (000-019)
1,736,210	109,952	Exclude in situ cancers for all except breast & bladder cancer
1,660,376	75,834	Exclude no or unknown microscopic confirmation
1,629,964	30,412	Exclude sarcomas

A total of 1,629,955 primary cancers were used in analyses. Survival rates are calculated on demographic and tumor information. Cases of second or later primaries, cases identified by death certificate or autopsy only, cases of unknown race, and those alive with no follow-up were excluded from the analysis (Table 1.1).

The SEER data are available for analyses by researchers. See www.seer.cancer.gov for further information.

SEER*Stat Software

The SEER*Stat statistical software, a convenient, intuitive mechanism for the analysis of SEER and other cancer-related databases, was used for analyses. It is a powerful PC tool to view individual cancer records and to produce statistics for studying the impact of cancer on a population. It is available at the following website: http://seer.cancer.gov/seerstat/

Tumor Information

The SEER program collects the month and year of diagnosis, primary tumor site, behavior, histology, extent of disease at diagnosis, and, starting in 1990, breast cancer receptor status. The International Classification of Diseases for Oncology, 2nd edition (ICD-O-2) (1) was the standard reference for classifying primary site, histology, behavior and grade. The ICD-O-2 tumor site and morphology codes allow for precise coding of tumor location (including sub-location within an organ) and histology. For 2001 cases, the third edition of ICD-O (ICD-O-3) was used and all prior histology data were converted to ICD-O-3 (2).

The histologic grade of malignant tumors is also collected: *grade I* is well differentiated; *grade II* is moderately differentiated; *grade III* is poorly differentiated, and *grade IV* is undifferentiated or anaplastic (1, 2). For leukemias and lymphomas, the grade code can reflect T-cell, B-cell, and N-K cell phenotype.

Extent of Disease

SEER has collected extent of disease (EOD) information on all cancers since the inception of the program. Extent of disease information since 1988, consists of five data items: tumor size where applicable, extension (within the primary site or contiguous or metastatic), highest involved lymph node chain, number of regional lymph nodes found positive (with certain exceptions), and number of regional nodes examined (with certain exceptions). The extension and lymph node fields are specific to the site of the primary tumor. The detail and amount of information collected for EOD have varied over time.

Stage

Stage of disease is determined from EOD information. In this monograph several different staging systems were used depending on the extent of disease information available. The American Joint Committee on Cancer's (AJCC) Staging Manual for the third edition (3), the fifth edition (4), and sixth edition (5) TNM: *tumor size/extent* (T), *node involvement* (N), and *distant metastases* (M) and then combines TNM into stages. Sometimes additional information is needed such as grade.

Since 1988, the tumor extension information in EOD is collected utilizing only one variable (except for prostate since 1995) and is based on the best information available on the furthest extension of the tumor. For some AJCC schemas, there is both a clinical T and a pathologic T. Therefore, in the conversion from EOD to AJCC, the T information is based on a combination of clinical and pathologic information. If there are distant metastases, the SEER EOD conversion will be TX M1, i.e. the T information is not recorded. Similarly, if distant nodes are involved, the information on regional nodes is not recorded in SEER. For many primary sites AJCC tumor extension classifications can range from T0 to T4 with subcategories, node involvement classifications can range from N0 to N3 with subcategories, and metastasis classifications can range from M0 to M1. The AJCC T, N, and M are then combined into stage ranging from Stage 0 through Stage IV. There are some primary sites for which there is no TNM and/or no AJCC stage. For all cancer sites except bladder and breast, in situ lesions were excluded from the analyses. For most cancer sites, this means that Stage 0 is excluded, but for breast and colon/rectum, Stage 0 includes more than in situ alone. For colon/rectum, Stage 0 also includes cases confined to the lamina propria with no nodes and for breast, Paget disease with no underlying tumor.

To perform the analyses in this monograph covering data from 1988-2001, it was necessary to achieve consistency of the stage variable over time. Changes to EOD were made in 1988 to be compatible with the AJCC third edition. In 1998, some of the EOD schemas were changed to be compatible with the fifth edition of AJCC so that SEER EOD information could be easily converted into the TNM staging classifications based on the fifth edition of the AJCC Manual for Staging of Cancer. Therefore, depending

on the cancer site and the changes between the third and fifth editions of AJCC, some chapters present data according to the AJCC third, AJCC fifth, or a different stage definitions (see below). Except for lymphomas, the AJCC staging criteria were applied to all histologies for each primary site. In some chapters, a *SEER modified AJCC* stage was used. The main difference between the SEER modified and AJCC versions, is that NX was combined with N0 in the conversion of TNM to AJCC stage.

SEER has also used a more simplistic stage with five levels: *In situ* tumors are those that have not yet broken through the adjacent basement membrane. For most cancer sites treated in this monograph, in situ tumors are excluded from the analysis; the urinary bladder and the female breast are exceptions. The term *localized* describes tumors, regardless of size, that are confined to the organ of origin. *Regional* tumors are those that have metastasized to the regional lymph nodes or have extended directly from the organ of origin. *Distant* describes a tumor whose metastases have traveled to other parts of the body. (Leukemia and myeloma are considered distant at diagnosis.) When information is not sufficient to assign a stage, a cancer is said to be *Unstaged or Unknown*. Most of the chapters which use stages of localized, regional, and distant are based on the SEER Summary Stage (1977) (6). Based on the same principles as Summary Stage 1977, SEER has used more historical definitions that are more consistent over time for historical trends back to 1973. In a few places the SEER historic stage is used. The SEER Summary Staging Manual 2000 lists the definitions for SEER Summary Stage 2000 and in the footnotes for each site describes how the SEER Summary Stage 1977 and the SEER historic stage differ from it (7).

SURVIVAL METHODS

The *observed survival rate,* obtained using the actuarial (life table) method, is the proportion of cancer patients surviving for a specified time interval after diagnosis. The *expected survival rate* for a hypothetical cohort of persons of the same sex, age, and race as the patient cohort is the proportion, based on the 1990 life table, of the given cohort that will survive to the end of the given time interval. For some sites, median survival times are presented. The median survival time is based on the observed survival rate and is defined as the point at which 50% have died and 50% are alive.

Most of the survival analyses in this monograph is based on the *relative survival rate* (8), except in Chapter 31 on race and ethnicity, where the *cause-specific survival rate* (9) is used.

Relative survival is a *net survival* measure representing cancer survival in the absence of other causes of death. Relative survival is defined as the ratio expressed as a percent, of the proportion of *observed* survivors in a co-

Table 1.2: Ten Most Common Cancer Sites: 1-, 2-, 3-, 5-, 8- & 10-Year Relative Survival Rates by Site, Ages 20+, 12 SEER Areas, 1988-2001

Site	Cases	Percent	Relative Survival Rate (%)					
			1-Year	2-Year	3-Year	5-Year	8-Year	10-Year
			Percent	Percent	Percent	Percent	Percent	Percent
All sites (except male and female breast in situ)	1,584,884	100.0	79.5	72.3	68.7	64.4	60.6	58.6
Prostate	275,280	17.4	100.0	99.5	98.9	97.6	94.5	91.7
Breast (female, in situ)	44,875	2.8	100.0	100.0	100.0	100.0	100.0	100.0
Breast (female, invasive)	257,888	16.3	97.8	94.8	91.9	87.1	81.9	79.2
Lung	201,067	12.7	42.6	25.9	20.0	15.5	12.4	11.0
Colon/Rectum	182,589	11.5	83.3	75.1	69.9	63.6	59.2	57.7
Melanoma	55,039	3.5	97.1	94.4	92.4	90.0	88.2	87.9
Urinary Bladder	67,528	4.3	91.5	87.1	84.8	81.9	78.9	77.4
Non-Hodgkin Lymphoma	65,932	4.2	74.2	66.3	62.1	56.3	49.9	47.0
Uterine Corpus	48,642	3.1	93.5	89.5	87.0	84.7	83.1	82.6
Leukemia (all ages)	42,678	2.7	67.0	58.0	53.4	47.2	40.7	38.1
Kidney and Renal Pelvis	32,583	2.1	80.8	73.8	70.4	65.5	60.9	57.9

hort of cancer patients (the observed survival rate defined above) to the proportion of expected survivors (the expected survival rate defined above). Thus, a relative survival of 100% means that a cancer patient cohort is just as likely to survive the given interval as a cohort in the general population of the same sex, age, and race. It does not mean that everyone will survive their cancer. For example, in a group of screening found cancers, many of the people seek medical care on a more routine basis than the general population and may have better non-cancer survival than the general population. In this case the expected life table is too low which makes the relative rate too high. On the other hand, lung cancer patients who smoke may be at excess risk of dying of other smoking related causes than the general population and the calculated expected rate would be too high which means that the relative survival rate may be lower than it would be if life tables based on smoking could be used.

While many times 5-year relative survival rates are presented, a five year rate may be less informative than a survival rate over a shorter time frame for a site or group with poor survival or over a longer time frame for a site or characteristic with good survival. Up to 10-year survival rates are shown for many sites.

The conditional survival rate, while difficult to explain, may be the most clinically informative of the survival rates. Instead of evaluating survival from diagnosis, for example a 5-year relative survival rate from diagnosis, the conditional survival rate can start anytime after diagnosis, i.e., it is conditioned on the cohort surviving to that point of time and then a survival rate is calculated for the patients who have survived to that point. For this monograph, 5-year relative survival rates are presented for some sites conditioned on specific times after diagnosis. For some sites where survival is very poor, the eight year survival rate may obscure that for the small group of patients who have already survived 3 years, their probability of surviving the next 5 years may be quite high.

For certain racial and ethnic groups, the life tables that are typically used for calculating expected survival do not accurately represent the experience of that specific racial/ethnic population. Since the calculation of relative survival rates needs accurate life tables, the relative survival rates are not shown for race/ethnic groups other than white or black in the individual site chapters. In order to present information for race/ethnic groups other than white patients or black patients, a cause-specific (c-s) survival rate was used. Since survival calculated under different methods can not be compared to one another, the survival rates for more specific racial/ethnic groups were put in a special chapter on race-and-ethnicity, Chapter 31. The c-s rate is dependent on knowledge not only of the date of death but also accurate information on the cause of death. The c-s rate is similar to the observed survival rate except that only patients who died of their cancer are considered as deaths and patients who died of other causes are 'censored' at the time of death. This method avoids problems of finding appropriate expected survival rates which are needed for the relative survival rate, but is dependent on which cause of deaths are considered due to the cancer. The cause-specific rate, however, is dependent on accurate cause of death (COD) information. When the population used in calculating the expected survival is similar to the population of cancer patients except for the latter's cancer experience, the relative survival rate and the cause-specific survival rate will

Table 1.3: Ten Most Common Cancer Sites: Five-Year Relative Survival Rates by Sex and Race, Ages 20+, 12 SEER Areas, 1988-2001

Site	Total	Male	Female	White Male	White Female	Black Male	Black Female
All sites (except male and female Breast in situ)	64.4	63.6	65.3	65.3	66.5	55.8	52.9
Prostate	97.6	97.6	n/a	98.4	n/a	93.5	n/a
Breast (female, in situ)	100.0	n/a	100.0	n/a	100.0	n/a	100.0
Breast (female, invasive)	87.1	n/a	87.1	n/a	88.3	n/a	74.5
Lung	15.5	13.6	18.0	13.9	18.4	10.9	15.0
Colon/Rectum	63.6	63.7	63.5	64.6	64.4	55.3	54.9
Melanoma	90.0	88.2	92.1	88.4	92.4	70.1	76.3
Urinary Bladder	81.9	84.0	75.9	84.8	77.3	69.3	55.4
Non-Hodgkin Lymphoma	56.3	52.5	60.9	53.4	61.5	43.4	54.8
Uterine Corpus	84.7	n/a	84.7	n/a	86.4	n/a	61.8
Leukemia (ages 0-19 and 20+)	47.2	48.0	46.2	49.6	47.6	37.2	37.9
Kidney and Renal Pelvis	65.5	65.2	66.0	65.9	66.2	61.4	64.8

Table 1.4: Number of Cases by Leading Cancer Site and Stage at Diagnosis, 5-Year Relative Survival Rates, Ages 20+, 12 SEER Areas, 1988-2001

	Relative Survival						
Site	Localized	5-year percent (localized)	Regional	5-year percent (regional)	Distant	5-year percent (distant)	Unstaged
Prostate	@	@	236,377	100.0	17,953	35.8	20,950
Breast (female, invasive)	160,105	97.4	78,299	79.2	14,359	24.4	5,125
Lung	32,709	50.5	75,551	15.8	78,510	1.9	14,297
Colon/Rectum	70,343	90.6	69,942	66.2	34,756	9.4	7,548
Melanoma	44,969	97.2	5,869	61.1	1,931	14.6	2,270
Urinary Bladder	50,331	93.9	12,686	48.5	2,166	5.8	2,345
Non-Hodgkin Lymphoma	19,971	69.4	9,098	61.1	30,468	44.3	6,395
Uterine Corpus	35,646	95.7	7,237	66.2	3,993	26.0	1,766
Kidney and Renal Pelvis	17,591	90.4	7,316	60.2	6,598	8.2	1,078

@ Local combined with Regional for Prostate

be nearly equal. That is, the relative survival rate closely indicates the probability that a patient will not die due to cancer-related causes within the given time interval. When the population used for the expected survival is dissimilar to the population of cancer patients, the relative survival may differ from the cause-specific survival rate by tumor and patient characteristics. Comparisons of survival rates should be based on the same survival method for calculating rates.

RESULTS

Relative survival up to 10 years after diagnosis of invasive cancer is shown in Table 1.2 for patients diagnosed in the 12 SEER catchment areas during 1988-2001. Survival rates vary substantially according to the cancer site. Among the most frequently diagnosed cancers, the sites with the highest 10-year relative survival rates are prostate, female breast in situ, uterine corpus, and melanoma, which have 10-year relative survival rates of 83% (uterine corpus) to 100% (female breast in situ). Lung cancer has the least favorable survival across the 10-year period following diagnosis (11%).

Survival by sex and race is presented in Table 1.3 for select cancer sites. For all cancers combined, excluding male and female breast in situ, there is only a small difference by sex in terms of 5-year relative survival rates. However, a survival advantage by sex varies by cancer site as well as within race groups. For example, five-year survival for non-Hodgkin lymphoma among white women is 62% compared to 53% in white males. Among black women the non-Hodgkin lymphoma 5-year survival rate (55%) is twelve percentage points higher than in black men (43%). Among white males, the 5-year relative survival rate for urinary bladder is 85% compared to 77% in white females. Blacks seem to fare worse with this disease, where the 5-year survival rate is 69% among black males and 55% among black females.

Survival by summary stage is presented in Table 1.4 for select cancers. The differences in 5-year survival by stage are notable. The earlier the stage at diagnosis, the more favorable is the 5-year survival. For screenable cancer sites, survival ranges from 91% at localized stage to 9% at distant stage for colorectal cancer, and 97% at localized stage to 24% at distant stage for female invasive breast cancer. Other cancer sites are as extreme in terms of survival by stage of diagnosis (urinary bladder, melanoma).

DISCUSSION

Many times in population-based statistics the emphasis is on incidence and mortality statistics. While these are important in measuring cancer, they are not as relevant to the medical community concerned about prognosis. The focus of this monograph is to present descriptive analyses of cancer survival by patient and tumor characteristics.

Since the emphasis is on the influence of patient and tumor characteristics on survival and not on how survival rates have changed over time, a discussion of biases in survival trends is not presented here. See the introduction of the SEER Cancer Statistics Review for a discussion of survival biases (10). In comparing any two groups, one should consider whether any differences in survival may be due to the two groups being different by some other characteristic than the comparison. For example, in a cohort of patients over 85 years of age, due to co-morbid conditions some may not have had as extensive staging work-up as a younger age group.

The analyses presented in this monograph did not test for statistical significance of observed differences between population groups, therefore neither confidence intervals nor p- values are provided. Any comparisons of survival rates between age, sex, race groups, or tumor characteristics are based on point estimates, and thus, issues related to small case numbers need to be considered when making or interpreting comparisons. The numbers of cases are given in most cases so that one has a general idea about the variability of the point estimates. Survival rates were not calculated for fewer than 25 cases.

An attempt was made to include all cancer sites. A chapter on rare cancers contains information on cancers not included in the site-specific chapters.

REFERENCES

1. Percy C, Van Holten V, Muir C, editors. International classification of diseases for oncology, 2nd ed. Geneva: World Health Organization; 1990.
2. Fritz A, Percy C, Jack A, Shanmugaratnam K, Sobin L, Parkin DM, Whelan S. International Classification of Diseases for Oncology--Third Edition. Geneva: World Health Organization, 2000.
3. Beahrs, OH, Henson DE, Hutter RVP, Myers MH (eds). AJCC Cancer Staging Manual, Third edition. American Joint Committee on Cancer. Philadelphia: Lippincott, 1988.
4. Fleming ID, Cooper JS, Henson DE, Hutter RVP, Kennedy BJ, Murphy, GP, O'Sullivan, B, Sobin LH and Yarbro, JW (eds) AJCC Cancer Staging Manual, Fifth Edition, Lippincott-Raven, Philadelphia, 1998.
5. Greene FL, Page DL, Fleming ID, Fritz AG, Balch CM, Haller DG, Morrow M (eds). AJCC Cancer Staging Manual, Sixth edition. American Joint Committee on Cancer. New York: Springer 2002.
6. Shambaugh, EM, Weiss, MA. Axtell, LM (eds), The 1977 Summary Staging Guide for the Cancer Surveillance, Epidemiology and End Results Reporting Program. National Cancer Institute, SEER Program, Bethesda, MD, April 1977.
7. Young JL Jr, Roffers SD, Ries LAG, Fritz AG, Hurlbut AA (eds). SEER Summary Staging Manual – 2000; Codes and Coding Instructions, National Cancer Institute, NIH Pub. No. 01-4969, Bethesda, MD, 2001.
8. Ederer F, Axtell LM, Cutler SJ. The relative survival rate: a statistical methodology. J Natl Cancer Inst Monogr 1961;6:101 121.
9. Clegg LX, Li FP, Hankey BF, Chu K, Edwards BK. Cancer survival among US whites and minorities: a SEER (Surveillance, Epidemiology, and End Results) Program population-based study. Arch Intern Med 2002;162:1985-1993.
10. Ries LAG, Harkins D, Krapcho M, Mariotto A, Miller BA, Feuer EJ, Clegg L, Eisner MP, Horner MJ, Howlader N, Hayat M, Hankey BF, Edwards BK (eds). SEER Cancer Statistics Review, 1975-2003, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2003/, based on November 2005 SEER data submission, posted to the SEER web site, 2006.