

Complete and Limited Duration Cancer Prevalence Estimates

Angela Mariotto¹, Anna Gigli², Riccardo Capocaccia³, Andrea Tavilla³, Limin X. Clegg¹, Michael Depry⁴, Steve Scoppa⁴, Lynn A. G. Ries¹, Julia H. Rowland¹, Gina Tesauro¹, Eric J. Feuer¹

¹National Cancer Institute, National Institutes of Health, Bethesda, USA

²Institute for Population Research and Social Policies, National Research Council, Rome, Italy

³Istituto Superiore di Sanita', Rome, Italy

⁴Information Management Services, Inc., Silver Spring, USA

Introduction

Prevalence is an indicator of primary interest in public health because it measures the burden of cancer in a population and on the health care system. Prevalence is defined as the number or percent of people alive on a certain date in a population who previously had a diagnosis of the disease. It includes new (incidence) and pre-existing cases and is a function of both past incidence and survival. Information on prevalence can be used for health planning, resource allocation and an estimate of cancer survivorship. In past reports of the Cancer Statistics Review, US cancer prevalence was estimated by multiplying the Connecticut cancer prevalence proportions to the US population. This year, US cancer prevalence is estimated by applying SEER-9 and SEER-11 prevalence proportions to the US population. SEER proportion rates are more representative of the US and permit estimation of prevalence by racial/ethnic groups. Other changes with respect to previous reports are in the methods for tumor inclusion and complete prevalence calculation.

The counting method

We used the *counting method* (Byrne et al., 1992) to estimate prevalence from incidence and follow-up data from the SEER cancer registries. Variance for these estimates is proposed and evaluated by Gail et al. (1999) and Clegg et al. (2001). The counting method estimates prevalence by dividing the estimated number of diagnosed persons in the prevalence cohort by the study population size at the prevalence point of time, taking into account loss to follow-up. For those in the prevalence cohort who are lost to follow-up, the following procedure is used to estimate the probability that each individual is alive as of the prevalence point. First, survival functions, stratified by age at diagnosis and year of diagnosis, are estimated from the prevalence cohort. Then, for each individual lost to follow-up in the prevalence cohort, his or her probability of being alive at the prevalence time is estimated from the appropriate (age and year at diagnosis) survival function, conditional on the time lost to follow-up.

Tumor inclusion criteria

Different methods can be used to determine which tumors are to be included in the prevalence statistics. For the results presented here only the *first malignant tumor ever* is counted. Thus, if a woman had a melanoma prior to a breast cancer diagnosis, her melanoma would contribute to the prevalence of melanoma and to the prevalence of all sites, but the breast cancer would not contribute to the prevalence of breast cancer. Counting only one cancer per individual avoids some ambiguity in prevalence counts, and allows the counts for individual sites to sum to the all sites total. Other selection criteria are possible and different criteria have been used in the past. For more information and to generate statistics using other tumor selection criteria refer to <http://srab.cancer.gov/prevalence>.

Complete Prevalence and Prevalence by Years Since Diagnosis

Complete prevalence (i.e., the proportion of persons alive who ever had a history of the disease) can be estimated using the counting method from registries of long duration. In the US, only the Connecticut Tumor Registry has information on cancer cases from 1940 and may be used to approximate complete prevalence. *Limited duration prevalence*, representing the proportion of people alive at the prevalence date with a diagnosis of cancer in the previous *L* years, can be calculated from registries of shorter duration. For example, SEER incidence and follow-up data from 1975 through 1998 can provide estimates of prevalent cases diagnosed up to 24 years prior to Jan 1, 1999, the most recent date for which we can estimate prevalence.

In order to obtain complete prevalence from registries of shorter duration, a method based on the statistical modeling of a *completeness index* (Capocaccia et al. 1997) can be used. A factor, $R(L)$ is estimated, which represents the proportion of complete prevalence that can be observed from a registry length of length L . $R(L)$ is estimated from registry data as a function of the observation time L , incidence and survival. Incidence and survival are modeled based on SEER data so that incidence and survival can be predicted for calendar years before cancer registration started (1975). The complete prevalence is estimated by

$$P_c = P(L) / R(L),$$

where $P(L)$ is the prevalence of cases diagnosed in the observation time L , and P_c is the complete prevalence. For a more detailed account of the methodology, see Capocaccia et al. 1997. Merrill et al. 2000 shows the validation of the completeness index based on Connecticut Tumor Registry data, for selected cancer sites. The same validation procedure was performed for the cancer sites in Table 1 by comparing the modeled index R to an 'empirical' R calculated from the historical Connecticut data. Details of the validation will be the subject of a future publication. For Hodgkin lymphoma, acute lymphocytic leukemia, melanoma, ovary and testis cancers, the validation was not satisfactory and we have used an 'empirical' R to adjust prevalence. The ratio between the prevalence based on 1973-1996 Connecticut data to prevalence based on 1940-1996 by age at prevalence was used to derive an empirical completeness. For the prevalence of acute lymphocytic leukemia in women we have used an overall (not by age) empirical completeness due to zero prevalent cases in the Connecticut data for some older age groups. One advantage of the modeled completeness index is that it is more stable than the empirical completeness, especially for rare cancer sites. However, the main advantage is that it permits estimation of completeness index by race/ethnicity from SEER data.

Complete prevalence for childhood cancers (i.e. cancers diagnosed at limited age ranges: 0-19 or 0-14) cannot be obtained by the completeness index or even estimated from the Connecticut registry as we have no way to estimate how many elderly survivors there are of childhood cancer. This is a topic for future research.

US cancer prevalence estimates

Table 1 represents the US cancer prevalence as of Jan 1, 1999, estimated by multiplying the SEER 9 registries prevalence proportions by the US population. Prevalence estimates for cases diagnosed 0-<5, 5-<10, 10-<15, and 15-<20 years respectively, prior to Jan 1, 1999, are computed by applying the counting method to the SEER-9 registries. Complete prevalence and prevalence for cases diagnosed >20 years prior to Jan 1, 1999, are calculated by applying the completeness index methodology to the 24-years prevalence counts from SEER-9 registries. Tables 2 and 3 give cancer prevalence for people, diagnosed in the previous 10 years (1989-1998). For whites, blacks and Asian/Pacific Islanders, prevalence was based on SEER-11 registries and for Hispanics it was based on SEER-11 excluding Hawaii and Detroit due to problems in these registries with reporting of ethnicity. Limited duration, 10-year prevalence is reported because it is the maximum amount of time for which all of the racial/ethnic groups could be consistently reported.

References

Byrne J, Kessler LG, Devesa SS. The prevalence of cancer among adults in the United States: 1987. *Cancer* 1992;68:2154-9.

Capocaccia R, De Angelis R. Estimating the completeness of prevalence based on cancer registry data. *Statistics in Medicine* 1997;16:425-40.

Merrill RM, Feuer EJ, Capocaccia R, Mariotto A. Cancer Prevalence Estimates Based on Tumor Registry Data in the SEER Program. *Int J Epidemiol* 2000;29:197-207.

Gail MH, Kessler L, Midthune D, Scoppa S. Two Approaches for Estimating Disease Prevalence from Population-Based Registries of Incidence and Total Mortality. *Biometrics* 1999;55:1137-44.

Clegg LX, Mitchell GH, Feuer EJ. Estimating the Variance of Disease Prevalence Estimates from Population-Based Registries. Technical report 2001. For a copy contact prevalence@ims.nci.nih.gov.

Table XXIX-1

Estimated United States Cancer Prevalence Counts†, January 1, 1999,
By Site, Sex and Years Since Diagnosis

Site/Sex	Years Since Diagnosis					Complete♦
	0 to <5	5 to <10	10 to <15	15 to <20	>=20♦	
All Sites	3,484,843	2,190,687	1,245,704	736,356	1,270,469	8,928,059
Males	1,737,334	1,069,743	481,417	255,959	385,062	3,929,515
Females	1,747,509	1,120,944	764,287	480,397	885,407	4,998,544
Brain and CNS	32,241	19,060	13,107	7,612	16,594	88,614
Males	18,335	10,656	6,701	3,955	8,351	47,998
Females	13,907	8,404	6,406	3,657	8,242	40,616
Breast(females)	724,510	495,499	326,501	173,627	331,143	2,051,280
Cervix	47,462	38,073	31,272	25,950	83,475	226,232
Colon and Rectum	379,531	233,948	159,212	92,037	144,042	1,008,770
Males	189,741	115,949	77,320	40,786	59,394	483,190
Females	189,791	117,999	81,891	51,250	84,649	525,580
Corpus and Uterus, NOS	132,210	99,873	79,423	68,266	153,500	533,272
Esophagus	13,693	3,331	1,344	490	1,012	19,870
Males	10,409	2,394	888	368	645	14,704
Females	3,284	937	457	122	366	5,166
Hodgkin Lymphoma	30,916	26,385	21,910	16,265	22,120	117,596
Males	16,320	13,702	11,280	8,930	11,153	61,385★
Females	14,596	12,683	10,630	7,334	10,968	56,211★
Kidney and Renal Pelvis	77,095	45,659	26,597	15,426	23,327	188,104
Males	46,743	26,977	15,633	9,106	13,347	111,806
Females	30,352	18,681	10,965	6,320	9,980	76,298
Larynx	33,066	22,713	15,834	10,635	11,682	93,930
Males	26,273	18,212	13,007	8,420	9,923	75,835
Females	6,793	4,501	2,827	2,215	1,759	18,095
Leukemia	72,223	39,658	23,009	13,066	18,403	166,359
Males	41,716	22,310	12,912	7,282	8,949	93,169
Females	30,507	17,348	10,098	5,784	9,453	73,190
Acute Lymphocytic Leukemia	12,786	9,216	7,140	5,351	2,157	36,650
Males	7,344	5,034	3,938	2,832	654	19,802★
Females	5,442	4,182	3,202	2,520	1,502	16,848†
Lung and Bronchus	193,595	61,466	31,699	17,126	23,009	326,895
Males	100,698	31,176	16,294	9,025	13,415	170,608
Females	92,897	30,289	15,405	8,101	9,595	156,287
Melanoma of the Skin	168,416	115,092	84,788	56,762	15,242	440,300
Males	89,280	58,083	39,332	24,491	614	211,800★
Females	79,137	57,009	45,456	32,271	14,627	228,500★
Non-Hodgkin Lymphoma	135,098	69,795	40,447	21,388	22,662	289,390
Males	70,959	35,184	20,443	10,751	11,714	149,051
Females	64,139	34,611	20,004	10,636	10,949	140,339
Oral Cavity and Pharynx	81,545	48,132	32,158	20,926	36,414	219,175
Males	54,074	30,310	19,647	13,003	23,973	141,007
Females	27,471	17,823	12,511	7,924	12,439	78,168
Ovary	67,114	39,703	23,389	16,606	29,851	176,663★
Pancreas	17,176	2,752	1,310	819	685	22,742
Males	8,418	1,126	511	332	355	10,742
Females	8,758	1,626	800	487	329	12,000
Prostate	759,939	520,702	126,812	41,349	28,357	1,477,159
Stomach	31,347	12,423	7,962	4,530	7,353	63,615
Males	18,636	6,502	4,163	2,594	3,994	35,889
Females	12,711	5,921	3,799	1,935	3,360	27,726
Testis	33,409	30,570	24,712	19,160	18,164	126,015★
Thyroid	74,189	54,427	42,600	32,452	74,517	278,185
Males	16,952	12,896	9,387	7,205	16,722	63,162
Females	57,237	41,531	33,213	25,247	57,795	215,023
Urinary Bladder	170,650	111,690	73,019	43,394	50,882	449,635
Males	127,434	83,438	54,200	31,129	34,667	330,868
Females	43,216	28,253	18,819	12,264	16,215	118,767
All Sites(Age DX: 0-19)	51,797	42,329	35,495	29,734	&	&
Males	27,710	22,313	17,921	15,055	&	&
Females	24,088	20,016	17,574	14,679	&	&

† U.S. 1999 cancer prevalence counts are based on 1999 cancer prevalence proportions from the 9 SEER registries and 1/1/1999 U.S. population estimates based on the average of 1998 and 1999 population estimates from the U.S. Bureau of the Census. Prevalence was calculated using only the First Malignant Tumor Ever for a person.

♦ The last two columns represent prevalence estimates using the completeness index method (Capocaccia et. al. 1997, Merrill et. al. 2000). Totals are obtained by summing males and females and not by modeling.

★ Completeness index was approximated using empirical data from historical Connecticut tumor registry: ★ by age at prevalence †for all ages combined due to instability of age specific estimates.

& Current methodology does not allow for the estimation of complete prevalence for childhood cancer.

Table XXIX-2

Estimated United States Cancer Prevalence Counts†, January 1, 1999,
By Site, Race/Ethnicity, Sex and Years Since Diagnosis

Years Since Diagnosis	Males		Females	
	0 to <5	5 to <10	0 to <5	5 to <10
<u>Site/Race/Ethnicity</u>				
All Sites				
Whites	1,490,055	962,725	1,521,638	1,013,982
Blacks	169,281	83,573	146,668	81,587
Hispanics*	73,073	37,054	80,512	48,622
Asian/Pacific Islander	31,161	14,853	40,363	22,002
All Sites(Age DX: 0-19)				
Whites	22,730	18,785	20,120	16,178
Blacks	3,129	2,375	2,723	2,392
Hispanics*	3,669	2,700	3,435	2,172
Asian/Pacific Islander	999	634	862	690
Colon and Rectum				
Whites	161,753	101,529	158,862	101,668
Blacks	15,748	8,874	19,612	10,395
Hispanics*	8,058	3,961	6,726	3,751
Asian/Pacific Islander	5,190	2,572	4,409	2,383
Lung and Bronchus				
Whites	81,284	26,933	79,931	27,534
Blacks	10,941	3,076	8,850	2,538
Hispanics*	2,607	728	2,049	799
Asian/Pacific Islander	2,274	579	1,424	361
Prostate				
Whites	637,159	462,941		
Blacks	96,246	49,568		
Hispanics*	30,528	14,899		
Asian/Pacific Islander	10,487	5,438		
Breast				
Whites			637,210	451,187
Blacks			61,689	36,882
Hispanics*			29,886	18,337
Asian/Pacific Islander			16,723	9,490

† U.S. 1999 cancer prevalence counts are based on 1999 cancer prevalence proportions from the 11 SEER registries and 1/1/1999 U.S. population estimates based on the average of 1998 and 1999 population estimates from the U.S. Bureau of the Census. SEER prevalence proportions were calculated using only the First Malignant Tumor Ever for a person.

* Hispanic counts are based on 1999 prevalence proportions from the SEER 11 areas excluding Hawaii and Detroit.

Table XXIX-3

Estimated Percent[†] of January 1, 1999, Male SEER 11 Population Diagnosed with Cancer in the Previous 10 Years (1989-1998)
By Age at Prevalence, Site and Race/Ethnicity

Age at Prevalence	Males										
	Crude										Age-Adjusted [◆]
	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80+	All Ages	All Ages
<u>Site/Race/Ethnicity</u>											
All Sites	0.089%	0.103%	0.206%	0.433%	0.806%	2.441%	7.397%	13.804%	14.311%	1.905%	2.343%
Whites	0.094%	0.110%	0.229%	0.475%	0.835%	2.476%	7.536%	13.934%	14.326%	2.035%	2.380%
Blacks	0.067%	0.081%	0.111%	0.248%	0.743%	2.989%	9.057%	15.306%	13.481%	1.544%	2.532%
Hispanics*	0.082%	0.090%	0.153%	0.270%	0.429%	1.420%	4.623%	8.893%	8.532%	0.657%	1.462%
Asian/Pacific Islander	0.074%	0.081%	0.116%	0.233%	0.500%	1.291%	3.922%	9.247%	11.012%	1.085%	1.496%
Colon and Rectum	0.000%	0.000%	0.004%	0.020%	0.074%	0.278%	0.839%	1.518%	1.900%	0.206%	0.260%
Whites	0.000%	0.000%	0.004%	0.019%	0.069%	0.268%	0.839%	1.534%	1.925%	0.217%	0.260%
Blacks	0.000%	0.001%	0.003%	0.019%	0.097%	0.347%	0.848%	1.342%	1.558%	0.151%	0.250%
Hispanics*	0.000%	0.000%	0.004%	0.016%	0.048%	0.198%	0.609%	0.977%	1.026%	0.071%	0.169%
Asian/Pacific Islander	0.000%	0.000%	0.005%	0.026%	0.078%	0.281%	0.787%	1.483%	1.863%	0.184%	0.256%
Lung and Bronchus	0.000%	0.000%	0.001%	0.005%	0.027%	0.141%	0.431%	0.650%	0.434%	0.086%	0.105%
Whites	0.000%	0.000%	0.001%	0.004%	0.025%	0.135%	0.429%	0.644%	0.442%	0.089%	0.103%
Blacks	0.000%	0.000%	0.001%	0.005%	0.052%	0.237%	0.603%	0.721%	0.369%	0.086%	0.134%
Hispanics*	0.000%	0.000%	0.001%	0.002%	0.010%	0.056%	0.178%	0.305%	0.239%	0.020%	0.047%
Asian/Pacific Islander	0.000%	0.000%	0.001%	0.006%	0.023%	0.107%	0.310%	0.669%	0.405%	0.068%	0.092%
Prostate	0.000%	0.000%	0.000%	0.001%	0.042%	0.711%	3.684%	8.075%	8.245%	0.856%	1.103%
Whites	0.000%	0.000%	0.000%	0.001%	0.038%	0.693%	3.663%	7.998%	8.031%	0.894%	1.086%
Blacks	0.000%	0.000%	0.000%	0.002%	0.096%	1.332%	5.860%	11.056%	9.574%	0.885%	1.560%
Hispanics*	0.000%	0.000%	0.000%	0.000%	0.020%	0.404%	2.363%	5.553%	5.146%	0.264%	0.723%
Asian/Pacific Islander	0.000%	0.000%	0.000%	0.001%	0.017%	0.197%	1.450%	4.633%	6.152%	0.404%	0.612%

[†] SEER population estimates are the average of 1998 and 1999 population estimates from the U.S. Bureau of the Census. SEER prevalence percentages were calculated using only the First Malignant Tumor Ever for a person.
^{*} Hispanic counts are based on 1999 prevalence proportions from the SEER 11 areas excluding Hawaii and Detroit.
[◆] Percentages are age-adjusted to the 2000 U.S. standard population by 5-year age groups.

Table XXIX-3 - continued

Estimated Percent[†] of January 1, 1999, Female SEER 11 Population Diagnosed with Cancer in the Previous 10 Years (1989-1998)
By Age at Prevalence, Site and Race/Ethnicity

Age at Prevalence	Females										
	Crude										Age-Adjusted [◆]
	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80+	All Ages	All Ages
<u>Site/Race/Ethnicity</u>											
All Sites	0.077%	0.097%	0.254%	0.679%	1.682%	3.591%	5.557%	7.387%	7.280%	1.902%	1.903%
Whites	0.080%	0.103%	0.273%	0.725%	1.740%	3.749%	5.904%	7.769%	7.495%	2.079%	1.995%
Blacks	0.061%	0.072%	0.173%	0.486%	1.353%	2.969%	4.428%	5.961%	5.843%	1.230%	1.527%
Hispanics*	0.074%	0.085%	0.179%	0.494%	1.244%	2.383%	3.608%	4.528%	4.181%	0.805%	1.254%
Asian/Pacific Islander	0.067%	0.078%	0.174%	0.482%	1.456%	2.739%	3.818%	4.826%	5.187%	1.277%	1.382%
Colon and Rectum	0.000%	0.000%	0.003%	0.018%	0.066%	0.217%	0.562%	1.045%	1.551%	0.197%	0.191%
Whites	0.000%	0.000%	0.003%	0.017%	0.061%	0.204%	0.557%	1.053%	1.582%	0.210%	0.190%
Blacks	0.000%	0.000%	0.002%	0.020%	0.080%	0.301%	0.658%	1.165%	1.364%	0.158%	0.212%
Hispanics*	0.000%	0.000%	0.002%	0.013%	0.045%	0.154%	0.351%	0.628%	0.828%	0.064%	0.117%
Asian/Pacific Islander	0.000%	0.000%	0.004%	0.019%	0.079%	0.223%	0.522%	0.857%	1.312%	0.151%	0.173%
Lung and Bronchus	0.000%	0.000%	0.001%	0.006%	0.027%	0.122%	0.327%	0.444%	0.278%	0.078%	0.078%
Whites	0.000%	0.000%	0.002%	0.005%	0.026%	0.129%	0.351%	0.469%	0.283%	0.086%	0.082%
Blacks	0.000%	0.000%	0.001%	0.007%	0.037%	0.136%	0.330%	0.414%	0.240%	0.060%	0.079%
Hispanics*	0.000%	0.000%	0.000%	0.001%	0.011%	0.049%	0.124%	0.176%	0.128%	0.017%	0.031%
Asian/Pacific Islander	0.000%	0.000%	0.001%	0.005%	0.016%	0.061%	0.165%	0.237%	0.245%	0.039%	0.044%
Breast	0.000%	0.000%	0.015%	0.164%	0.778%	1.848%	2.506%	3.108%	2.886%	0.808%	0.815%
Whites	0.000%	0.000%	0.014%	0.163%	0.796%	1.943%	2.689%	3.311%	3.013%	0.888%	0.857%
Blacks	0.000%	0.000%	0.022%	0.185%	0.682%	1.529%	1.938%	2.364%	2.205%	0.532%	0.663%
Hispanics*	0.000%	0.000%	0.014%	0.109%	0.521%	1.115%	1.515%	1.758%	1.484%	0.297%	0.490%
Asian/Pacific Islander	0.000%	0.000%	0.012%	0.137%	0.723%	1.395%	1.655%	1.868%	1.692%	0.534%	0.576%

[†] SEER population estimates are the average of 1998 and 1999 population estimates from the U.S. Bureau of the Census. SEER prevalence percentages were calculated using only the First Malignant Tumor Ever for a person.
^{*} Hispanic counts are based on 1999 prevalence proportions from the SEER 11 areas excluding Hawaii and Detroit.
[◆] Percentages are age-adjusted to the 2000 U.S. standard population by 5-year age groups.