Istituto di Ricerche sulla Popolazione e le Politiche Sociali - CNR

IRPPS Working Papers

ISSN 2240-7332

Estimation of cancer incidence and survival parameters from population-based registries. Data on white, black and hispanic race/ethnicity

Anna Gigli e
Arianna Simonetti

What is IRPPS?

IRPPS is an Interdisciplinary Research Institute that conducts studies on demographic and migration issues, welfare systems and social policies, on policies regarding science, technology and higher education, on the relations between science and society, as well as on the creation of, access to and dissemination of knowledge and information technology.

www.irpps.cnr.it

IRPPS WPs n. 2/2002

Abstract

Cancer prevalence is the proportion of people in a population diagnosed with cancer in the past and still alive. One way to estimate prevalence is via population-based registries, where data on diagnosis and life status of all incident cases occurring in the covered population are collected. In this paper some methods regarding the fitting of the incidence and survival models to data collected by the SEER9 and SEER11 registries in the United States on various cancer sites, various race/ethnicitiy categories and two sexes, are applied. Such estimates are then used to estimate complete prevalence, as described by Gigli (2001).

Section 1 describes two survival models which will be fitted to the data, section 2 describes two incidence models, and finally section 3 makes brief comments on the results. The tables list the survival and incidence parameter estimates for each cancer site, sex, race/ethnicity and model implemented. Two further tables describe the survival parameter values for the white race/ethnicity estimated from the SEER9 and SEER11 registries in the period 1992–1998. [1]

keywords: limited duration prevalence, complete prevalence, cancer registries

1 Introduction and data description

In Gigli (2001) we illustrated some methods regarding the fitting of the incidence and survival models to data on various cancer sites and two sexes. In Gigli, Tavilla and Capocaccia (2002) we applied these methods to SEER9 registries data not differentiated by race/ethnicity. In this report we apply the mentioned methods to SEER9 and SEER11 cancer registries, for the race/ethnicity categories of whites, black and hispanics.

SEER9 registries contain information on all, white, black and Asian Pacific Islanders (API) patients and have been used for the estimates for whites and blacks. Data is available for the following periods:

• incidence data is available for the period 1975–1999,

[1]This manuscript is part of a research contract with the National Cancer Institute, Bethesda, USA, order no. 263-MQ-117231-1 of September 25, 2001.

• survival data is available for the period 1975–1998.

SEER11 registries contain information on all, white, black and hispanic patients and have been used for the estimation related to hispanics. Data is available for the following periods:
• incidence data is available for the period 1988–1999 for whites and blacks and for the period 1990–1999 for hispanics (since population data for hispanics, needed to compute the frequencies, is only available from 1990),
• survival data is available for the two periods 1992–1994 and 1995–1998 for every race/ethnicity.
Due to data problems on hispanics, the Detroit and Hawaii registries have been excluded. Therefore the estimations for each race/ethnicity have been based on 9 registries.

2 Survival models

Let

$$S(x-t,t;\psi) = \{(1-Q)+Q\exp[-[\lambda(x-t)]^{\beta}]\}^{\exp[\gamma_1(t-t_0)+\gamma_2(t+c-s_0)+\gamma_3\delta_1+\gamma_4\delta_2]}. \quad (1)$$

be the most general model for the cumulative relative survival function for a patient belonging to the birth cohort c, diagnosed in year $t+c$ at age t, who survives until age x (De Angelis *et al.*, 1999). This class of models assumes that only a portion Q of the individuals with cancer will die with a relative survival following a Weibull distribution with parameters λ, β, where λ is the scale parameter, which measures the mortality rate of the fatal cases (the higher λ the faster the mortality), β is the shape parameter (β =1 implies a constant mortality rate, greater than 1 implies mortality more concentrated at the beginning of the diagnosed period, vice versa when β is less than 1); the remaining $1-Q$ patients are supposed to be cured; they have the same mortality rate as the general population, and consequently their relative survival is 1. In the model considered here the death hazard due to cancer is assumed to be linearly dependent, on a logarithmic scale, of age at diagnosis and calendar time of diagnosis; the corrections due to the risk of being diagnosed one year older than the reference age t_0 and one year later than the reference year s_0 are parameterized by γ_1 and γ_2, respectively; the

reference age t_0 is computed as the average age at diagnosis for all patients and varies for cancer site and sex; the reference year s_0 is fixed at year 83.5. The reference race/ethnicity is white and the corrections due to the risk of belonging to hispanic or black race/ethnicity are parameterized as γ_3 and γ_4, respectively, while δ_1 and δ_2 are dummy variables for the race/ethnicity.

2.1 Choice between two survival models

In the applications we make the following simplifing assumptions:

1. When data for each cancer site is large enough (such as for whites and blacks in SEER9) we adopt an independent *stratified* model for each race/ethnicity, therefore we omit γ_3 and γ_4 and obtain independent vectors of parameter estimates:

$$S(x-t,t;\psi) = \{(1-Q)+Q\exp[-[\lambda(x-t)]^{\beta}]\}^{\exp[\gamma_1(t-t_0)+\gamma_2(t+c-s_0)]}. \quad (2)$$

2. When data is scarse, such as for hispanics in SEER11, we use also information on whites (but not on blacks), and implement a *race risk* (or race parameter) model, which is a special case of model (1) with $\gamma_4 = 0$:

$$S(x-t,t;\psi) = \{(1-Q)+Q\exp[-[\lambda(x-t)]^{\beta}]\}^{\exp[\gamma_1(t-t_0)+\gamma_2(t+c-s_0)+\gamma_3\delta_1]}. \quad (3)$$

Such general strategy has been validated both graphically and by looking at the parameter estimates and their standard deviations, and some modifications have been implemented where the results were not satisfactory.

These special cases will be treated in more depth in the results section.

3 Incidence models

For the incidence hazard we generally use the exponential model, which is based on the strong assumption of independence between age and cohort, and has been shown to have a biological rationale for a general class of cancers (Armitage and Doll, 1954); in the situations where the exponential model does not fit properly the data, such as hormonal factors-related tumors (Capocaccia *et al.*, 1990), a polynomial model is proposed.

In one situation (testis cancer in black males) data was too scarse and therefore a polynomial model combining white and black data has been proposed. This case will be illustrated in the results section.

3.1 Exponential incidence

Let

$$\mathrm{I}(x;\psi) = \exp(a_c)x^b \tag{4}$$

be the cancer incidence for a person of current age x who belongs to the $c-th$ birth cohort. In presence of rare events (such as cancer) $I \to 0$ and (4) can be approximated by

$$\mathrm{I}(x;\psi) = \frac{1}{1+\exp\{-[a_c + b\log(x)]\}},$$

which in turn leads to

$$\mathrm{logit}(\mathrm{I}(x;\psi)) = a_c + b\log(x).$$

The parameter b is to be estimated, while a_c is a multiplicative constant depending on the birth cohort, which is not of interest for our purposes, because it cancels out in the computation of the completeness index.

3.2 Polynomial incidence

Various studies have reported that, in those cancers for which the exponential model fails, the best fit of the incidence hazard is obtained by a 6-degree polynomial. Here we use the notation proposed by Merrill *et al.*(2000): let

$$I(x;\psi) = \frac{1}{1+\exp\{-[a_c + b_1(x/k - t_0/k) + b_2(x/k - t_0/k)^2 + \ldots + b_6(x/k - t_0/k)^6]\}},$$

where a_c is the logit of the incidence at the c-th birth cohort, when age is equal to the reference age t_0 (chosen to be equal to 55) and the constant k (= 30) is an arbitrary scale factor which avoids numerical instability; $a_c, b_1, \ldots, b_6$ are the parameters to be estimated. The corresponding logit is

$$\mathrm{logit}(I(x;\psi)) = a_c + b_1\left(\frac{x-t_0}{k}\right) + b_2\left(\frac{x-t_0}{k}\right)^2 + \ldots + b_6\left(\frac{x-t_0}{k}\right)^6.$$

4 Results

Tables 1 to 8 illustrate the parameter estimates for incidence and survival for each of the race/ethnicities and sex. In particular tables 1 and 2 report the incidence parameter estimates when the polynomial incidence model is fitted to the data, tables 3 and 4 the incidence parameter estimates when the model chosen is exponential. They contain cancer site, race/ethnicity, parameter estimates, cohort width and total number of cases; usually 10-year cohort width has been used, as it provides smoother estimates, unless the fit was not satisfactory, and in that case we used 5-year cohort width. Tables 5–8 illustrate the results related to survival estimation, and contain cancer site, race/ethnicity, model fitted, parameter estimates and total number of cases.

The choice between the two models for incidence (exponential or polynomial) is based on graphical diagnosis and on comparison of standard deviations, since more sophisticated statistical tests are not available and the parameters do not hold a special meaning (especially in the polynomial model). For every cancer site the same model is fitted for each race/ethnicity and sex. Generally speaking the results regarding incidence estimates seem rather satisfactory.

Some problems arise for the survival estimation, especially for the SEER11 registries: due to data scarcity, a race risk model is fitted to white and hispanic data in order to obtain estimates for hispanics. However, the reliability of SEER11 database has been carefully checked, by comparing the survival estimates for the whites (stratified model) from SEER11 and SEER9 (restricted to the same period) and unfortunately the results are often quite different, as illustrated in tables 9 and 10. As a consequence, the reliability of the survival estimates for hispanics cannot be guaranteed. Nevertheless those estimates are reported.

The following list briefly describes the behaviour of the convergence.

1. **Acute Lymphocytic Leukemia**:
   the exponential incidence model fits well data on whites and blacks, both sexes; regarding hispanics it fits well the females, but the mle's reach only a relative maximum for the males.
   The survival model fits well data on whites and black (both sexes), but gives quite different results for hispanics (especially in the estimation of the proportion of cured cases).

2. **All sites**:
   the polynomial incidence model fits well the data.
   The survival model converge and fits well the white and black data, but gives quite different results for hispanics.

3. **Brain and Other Nervous System**:
   the 6th degree polynomial model fits well every subset, except the black males, for which a 4th degree polynomial has been fitted. Generally speaking, the model appears to capture quite well the young ages.
   The survival model converges and fits well the white and black data, but gives quite different results and large standard deviations for hispanics.

4. **Breast**: the polynomial incidence model fits well the data of every race/ethnicity.
   The survival model fits well the white and black females, but gives quite different estimates for the hispanic women (especially in the proportion of cured cases).

5. **Cervix**: the polynomial incidence model converges and fit well the data.
   The survival model fits well data on white and black females, but gives quite different estimates for the hispanic women.

6. **Colon and Rectum**: the exponential incidence model converges and fits well the data.
   The survival model fits well the data for every subgroup.

7. **Corpus and Uterus**: the polynomial incidence model converges, although the standard deviations of the 4th and 6th degree parameter estimates are quite large for blacks and hispanics.
   The survival model fits well the white and black data, but gives a very different estimate of the proportion of cured among black women; this might be caused by the lack of data.

8. **Esophagus**: the exponential incidence model converges and fits well the data.
   The survival model fits well the data, except for hispanic females, where a very different estimate of the scale parameter is given; this corresponds to a mortality rate of fatal cases less concentrated at the be-

ginning of the illness for the hispanic females, which requires further investigation.

9. **Hodgkins Disease**: the polynomial incidence model converges, but generally speaking the standard deviations of the parameters are quite large (especially for the higher degrees of the polynomial).
   The survival estimates are quite different according to the race/ethnicity, and the model for hispanic females does not converge; a further investigation into this model is needed.

10. **Kidney and Renal Pelvis**: the polynomial incidence model converges and fits well the data.
    The survival estimates are quite different according to the race/ethnicity, and therefore a further investigation into this model is needed.

11. **Larynx**: the polynomial incidence model converges, but the parameter standard deviations of black and hispanic data are quite high.
    The survival model for white and black males fits well, for white and black females does not converge, for hispanics gives quite different results from those for white and black males.

12. **Leukemias**: the polynomial incidence model converges and fits well the data.
    The survival model fits well the data, although the estimation of the proportion of cured cases seems quite different for each race/ethnicity; this fact needs further investigation.

13. **Lung and Bronchus**: the polynomial incidence model converges, but gives quite high parameter standard deviations, for every subgroup; the incidence model gives worse results.
    The survival model fits well the data for every subgroup.

14. **Melanomas of the Skin**: the exponential incidence model converges and fits well the data, especially the whites, which are the overwhelming majority for this cancer.
    The survival model fits well the subgroup of whites and hispanics; some care is needed to assess the estimates related to the blacks, due to data scarcity.

15. **Non-Hodgkins Lymphomas**: the polynomial incidence model converges and fits well the data; some larger-than-usual standard deviations are found for the black males parameters, especially for the higher degrees of the polynomial.
    The survival model converges, but some of the results are not homogeneous among subgroups; the estimates for hispanics are different from the other races/ethnicities (especially the scale estimate) and therefore a further investigation is required.

16. **Oral Cavity and Pharynx**: the polynomial incidence model converges and fits well the data.
    The survival model converges, but there is a difference between the estimates for males and females which need further investigation; also the estimates for hispanics are quite different from whites and blacks.

17. **Ovary**: the polynomial incidence model converges and fits well the data.
    The survival model converges and fits well the data.

18. **Pancreas**: the exponential incidence model converges and fits well the data.
    The survival model converges and fits well the data.

19. **Prostate**: the exponential incidence model converges to a relative maximum for blacks and hispanics, and gives no problems for the whites.
    The survival model converges and fits well the data for whites and blacks, but gives completely different results for hispanics.

20. **Stomach**: the exponential incidence model converges and fits well the data.
    The survival model converges and fits well the data.

21. **Testis**: for the incidence model some problems arise when analyzing black data, so a combined white+black model is used: a race parameter, independent of age at diagnosis, is introduced in the polynomial model and the corresponding estimate is added to the intercept (for the blacks), to provide two parallel curves for whites and blacks; therefore the estimates for whites and blacks are the same except for the intercept. Some large errors are found in the fitting of the hispanics, too.

The survival model converges and fits well the data for whites and blacks, but gives completely different results for hispanics.

22. **Thyroid**: the exponential incidence model converges and fits well the data.
    The survival model converges only for black males. Probably the lack of fitting is due to the difficulties in capturing a different behaviour in age: the younger ages, up to 55 years survive almost completely, while the older ages die more; there is a need to fit a better model.

23. **Urinary Bladder**: the polynomial incidence model converges; some larger-than-usual errors are found for the black and hispanic females, which are probably due to data scarcity.
    The survival model converges and fits well the data for whites and blacks, but gives completely different results for hispanics.

References

P. Armitage and R. Doll (1954). The age distribution of cancer and a multi-stage theory of carcinogenesis. *British Journal of Cancer* **8**, 1–15.

R. Capocaccia, A. Verdecchia, A. Micheli, M. Saint, G. Gatta, F. Berrino (1990). Breast cancer incidence and prevalence estimated from survival and mortality. *Cancer Causes Control*, **1**, 23–29.

R. De Angelis, R. Capocaccia, T. Hakulinen, B. Soderman, A. Verdecchia (1999). Mixture models for cancer survival analysis: application to population-based data with covariates. *Statistics in Medicine*, **18**, 441–454.

A. Gigli (2001). The variance of the completeness index. *Working paper IRP*, **W.P. 3/2001**.

A. Gigli, A. Tavilla, R. Capocaccia (2002). Estimation of cancer incidence and survival parameters from population-based registries. *Working paper IRPPS*, **W.P. 1/2002**.

Table 1: Incidence parameter estimates for males - polynomial incidence function. Data on various cancer sites, white and black race/ethnicities (r/e) collected by SEER9 registries in period 1975–1999; data on various cancer sites, hispanic race/ethnicity (r/e) collected by SEER11 registries in period 1990–1999.

cancer site	r/e	a_c	b_1	b_2	b_3	b_4	b_5	b_6	cohort	cases
All Sites	w	-5.08	3.62	-0.30	-2.33	-0.77	1.14	0.58	10	914565
All Sites	b	-4.61	3.42	-1.64	-1.04	0.38	0.37	0.11	10	98884
All Sites	h	-5.47	3.74	0.23	-3.49	-1.37	1.59	0.81	10	47524
Brain	w	-8.94	1.39	-0.55	-0.36	-0.02	-0.11	-0.04	10	16777
Brain	b	-9.47	1.47	-0.78	-0.98	-0.12	0.00	0.00	10	1060
Brain	h	-9.34	0.85	-0.50	1.12	0.20	-1.18	-0.46	10	1115
Hodgkins Disease	w	-10.24	0.58	0.48	-2.54	-0.78	1.81	0.69	10	8113
Hodgkins Disease	b	-10.35	-0.37	0.25	0.49	-0.13	-0.43	-0.30	10	754
Hodgkins Disease	h	-10.41	1.12	0.71	-0.73	-0.48	0.06	-0.03	10	641
Kidney and Renal Pelv.	w	-8.37	2.40	-2.56	1.59	1.33	-1.31	-0.56	5	22871
Kidney and Renal Pelv.	b	-8.24	2.23	-2.37	1.40	1.32	-1.20	-0.56	5	2221
Kidney and Renal Pelv.	h	-8.31	2.34	-4.46	3.21	2.55	-2.45	-1.13	5	1557
Larynx	w	-8.60	2.28	-3.38	0.41	0.53	0.00	0.04	5	15659
Larynx	b	-8.01	1.81	-2.89	1.29	-1.01	-0.04	0.59	5	2256
Larynx	h	-9.21	3.36	-4.29	0.70	3.07	-0.95	-1.29	5	623
Leukemias	w	-8.65	2.47	-0.39	-0.58	0.31	-0.11	-0.12	10	29583
Leukemias	b	-8.71	1.88	-0.56	-0.39	0.34	0.02	-0.05	10	2249
Leukemias	h	-9.12	2.03	0.63	-0.24	-0.54	-0.74	-0.25	10	2130
Lung and Bronchus	w	-6.61	2.98	-2.31	0.22	0.11	-0.18	0.04	5	155601
Lung and Bronchus	h	-7.53	3.58	-1.83	-2.15	1.10	0.66	-0.69	5	4548
Lung and Bronchus	b	-5.93	2.23	-2.59	1.02	-0.24	-0.11	0.25	5	20770
Non-Hodgkins Lymph.	w	-8.15	2.25	-0.47	0.40	0.11	-0.41	-0.16	10	35813
Non-Hodgkins Lymph.	b	-8.42	2.04	-1.20	-0.01	0.14	0.08	0.04	10	2811
Non-Hodgkins Lymph.	h	-8.28	1.42	-0.04	-0.46	-0.62	0.24	0.22	10	2746
Oral Cavity and Phar.	w	-7.98	1.96	-2.75	-0.39	1.07	0.22	-0.10	5	32227
Oral Cavity and Phar.	b	-7.26	0.69	-4.00	2.58	2.57	-1.60	-1.02	5	4123
Oral Cavity and Phar.	h	-8.66	2.93	-1.15	-4.23	-0.95	2.96	1.26	5	1317
Testis	w	-10.36	-1.21	-0.58	-3.91	-1.23	4.75	2.26	10	13363
Testis	b	-12.11	-1.21	-0.58	-3.91	-1.23	4.75	2.26	10	13363
Testis	h	-10.78	-1.30	-1.53	-6.83	-0.70	7.65	3.34	10	1228
Urinary Bladder	w	-7.78	3.34	-1.12	-0.81	-0.06	0.34	0.15	10	61854
Urinary Bladder	b	-8.46	3.11	-1.51	0.09	0.21	-0.16	0.01	10	2442
Urinary Bladder	h	-8.61	3.51	-1.81	-1.08	0.25	0.45	0.13	10	1660

Table 2: Incidence parameter estimates for females - polynomial function. Data on various cancer sites, white and black race/ethnicities (r/e) collected by SEER9 registries in period 1975–99; data on various cancer sites, hispanic race/ethnicity (r/e) collected by SEER11 registries in period 1990–99.

cancer site	r/e	a_c	b_1	b_2	b_3	b_4	b_5	b_6	cohort	cases
All Sites	w	-5.03	1.94	-0.81	-0.07	-0.18	0.10	0.17	10	879716
All Sites	b	-5.03	1.84	-0.95	0.34	0.02	-0.09	0.08	10	81312
All Sites	h	-5.33	1.49	-0.84	0.50	-0.01	-0.29	-0.01	10	47987
Brain	w	-9.42	1.37	-0.19	-0.29	-0.14	-0.21	-0.06	10	13222
Brain	b	-9.98	1.51	-0.54	-0.06	0.37	-0.59	-0.33	10	883
Brain	h	-9.73	1.04	-0.46	0.57	0.47	-0.45	-0.22	10	955
Breast	w	-6.00	1.38	-0.70	1.19	-1.38	0.02	0.30	10	265204
Breast	b	-6.15	1.25	-0.82	1.04	-0.09	0.05	-0.51	10	23359
Breast	h	-6.34	0.84	-1.57	2.01	0.21	-0.94	-0.64	10	13652
Cervix and Uteri	w	-8.86	-0.37	0.14	-1.50	-1.69	1.87	0.79	5	21662
Cervix and Uteri	b	-8.10	-0.83	-0.64	1.00	0.88	-0.21	-1.07	5	4437
Cervix and Uteri	h	-7.77	-0.69	-0.34	0.48	0.08	0.08	-0.96	5	3822
Corpus and Uterus	w	-7.43	1.90	-3.59	0.31	1.70	-0.06	-0.52	5	62417
Corpus and Uterus	b	-8.09	3.39	-1.40	-4.15	-0.24	3.14	0.34	5	3719
Corpus and Uterus	h	-8.09	2.30	-2.14	-1.77	-0.28	1.64	0.53	5	2531
Hodgkins Disease	w	-11.04	0.92	1.13	-2.50	-0.80	1.46	0.38	10	6572
Hodgkins Disease	b	-11.12	1.37	-0.47	-3.35	-0.68	1.83	0.56	10	554
Hodgkins Disease	h	-11.15	3.12	1.11	-4.51	-1.59	2.23	0.92	10	468
Kidney and Renal Pelv.	w	-9.18	2.55	-1.71	1.26	0.96	-1.28	-0.53	5	14148
Kidney and Renal Pelv.	b	-9.07	2.19	-0.98	2.18	0.72	-1.94	-0.76	5	1557
Kidney and Renal Pelv.	h	-9.01	2.63	-3.98	1.98	3.05	-1.68	-1.07	5	1118
Larynx	w	-10.05	2.35	-3.86	-0.35	3.00	0.11	-1.38	5	3575
Larynx	b	-9.40	1.52	-4.50	2.04	2.38	-0.78	-0.56	5	577
Larynx	h	-11.27	2.15	0.86	2.94	-2.39	-4.88	-1.58	5	116
Leukemias	w	-9.22	2.11	-0.31	-0.23	0.40	-0.16	-0.12	10	22292
Leukemias	b	-9.47	1.65	-0.30	-0.45	0.34	0.21	0.03	10	1889
Leukemias	h	-9.42	1.67	-0.58	-0.15	0.84	-0.04	-0.17	10	1600
Lung and Bronchus	w	-7.12	3.13	-2.01	0.46	0.33	-0.37	-0.14	5	92835
Lung and Bronchus	b	-6.93	2.45	-1.84	1.68	-0.69	-0.60	0.18	5	9345
Lung and Bronchus	h	-7.97	2.84	-1.30	0.06	-1.87	0.72	0.83	5	3106
Non-Hodgkins Lymph.	w	-8.52	2.61	-0.49	-0.17	-0.00	-0.18	-0.08	10	31063
Non-Hodgkins Lymph.	b	-8.90	2.64	-0.52	-1.00	-0.53	0.72	0.44	10	1976
Non-Hodgkins Lymph.	h	-8.63	1.83	-1.06	-0.40	0.49	0.47	0.12	10	1901
Oral Cavity and Phar.	w	-8.98	1.89	-1.93	-1.01	0.83	0.70	0.08	5	15047
Oral Cavity and Phar.	b	-8.48	0.47	-3.22	1.47	1.90	-0.58	-0.47	5	1608
Oral Cavity and Phar.	h	-9.85	2.22	-0.64	-0.90	0.02	0.18	-0.04	5	592
Ovary	w	-8.03	1.40	-1.72	0.60	1.30	-0.50	-0.62	10	37466
Ovary	b	-8.56	2.02	-1.41	-0.59	1.06	0.14	-0.34	10	2633
Ovary	h	-8.21	0.93	-1.21	-0.51	1.00	0.65	-0.09	10	2403
Urinary Bladder	w	-9.05	3.16	-1.06	-1.00	-0.06	0.51	0.23	10	21219
Urinary Bladder	b	-9.54	3.55	-2.03	-0.28	1.44	-0.15	-0.40	10	1326
Urinary Bladder	h	-10.23	3.43	-0.04	-1.68	-0.43	0.95	0.13	10	591

Table 3: Incidence parameter estimates for males - exponential function. Data on various cancer sites, white and black race/ethnicities (r/e) collected by SEER9 registries in period 1975–1999; data on various cancer sites, hispanic race/ethnicity (r/e) collected by SEER11 registries in period 1990–1999.

cancer site	r/e	b	cohort	cases
Acute Lymph. Leuk.	w	-0.52	10	3499
Acute Lymph. Leuk.	b	-0.32	10	269
Acute Lymph. Leuk.	h	-0.49	10	840
Colon and Rectum	w	4.59	10	111456
Colon and Rectum	b	4.81	10	9425
Colon and Rectum	h	4.83	10	4644
Esophagus	w	5.08	10	10728
Esophagus	b	2.76	10	2843
Esophagus	h	6.47	10	551
Melanomas	w	3.21	5	31983
Melanomas	b	2.96	5	169
Melanomas	h	2.85	5	517
Pancreas	w	4.34	10	21365
Pancreas	b	4.26	10	2743
Pancreas	h	6.08	10	1013
Prostate	w	8.87	10	234749
Prostate	b	8.90	10	30517
Prostate	h	6.95	10	13212
Stomach	w	3.81	10	20936
Stomach	b	3.94	10	3217
Stomach	h	4.74	10	1919
Thyroid	w	2.02	10	6110
Thyroid	b	2.14	10	304
Thyroid	h	1.88	10	451

Table 4: Incidence parameter estimates for females - exponential function. Data on various cancer sites, white and black race/ethnicities (r/e) collected by SEER9 registries in period 1975–1999; data on various cancer sites, hispanic race/ethnicity (r/e) collected by SEER11 registries in period 1990–1999.

cancer site	r/e	b	cohort	cases
Acute Lymph. Leuk.	w	-0.63	10	2535
Acute Lymph. Leuk.	b	-0.54	10	234
Acute Lymph. Leuk.	h	-0.46	10	612
Colon and Rectum	w	4.17	10	112163
Colon and Rectum	b	4.61	10	10809
Colon and Rectum	h	4.09	10	4135
Esophagus	w	4.44	10	4122
Esophagus	b	2.75	10	1043
Esophagus	h	5.89	10	136
Melanomas	w	2.28	5	28484
Melanomas	b	2.62	5	201
Melanomas	h	2.26	5	689
Pancreas	w	4.93	10	21911
Pancreas	b	4.89	10	2945
Pancreas	h	4.84	10	1208
Stomach	w	3.84	10	12720
Stomach	b	4.28	10	1992
Stomach	h	3.92	10	1399
Thyroid	w	1.74	10	17255
Thyroid	b	2.08	10	1176
Thyroid	h	1.75	10	1944

Table 5: Survival parameter estimates for males. Data on various cancer sites, white and black race/ethnicities (r/e) collected by SEER9 registries in period 1975–1998; data on various cancer sites, hispanic race/ethnicity (r/e) collected by SEER11 registries in period 1992–1998.

cancer site	r/e	model	Q	λ	β	γ_1	γ_2	γ_3	cases
Acute Lymph. Leuk.	w	strat.	0.68	0.38	1.26	0.053	-0.062		3337
Acute Lymph. Leuk.	b	strat.	0.75	0.46	1.27	0.042	-0.062		257
Acute Lymph. Leuk.	h	race risk	0.83	0.59	1.46	0.058	-0.111	0.23	620
All Sites	w	strat.	0.71	0.31	0.51	0.010	-0.030		853053
All Sites	b	strat.	0.81	0.43	0.55	0.001	-0.028		91785
All Sites	h	race risk	0.47	0.58	0.62	0.003	-0.008	0.23	33353
Brain	w	strat.	0.97	0.55	0.56	0.040	-0.019		15656
Brain	b	strat.	0.88	0.80	0.64	0.030	-0.024		973
Brain	h	race risk	0.80	0.75	0.92	0.044	-0.007	0.16	742
Colon and Rectum	w	strat.	0.53	0.40	0.80	0.004	-0.024		105167
Colon and Rectum	b	strat.	0.66	0.42	0.78	0.010	-0.019		8851
Colon and Rectum	h	race risk	0.48	0.37	0.86	0.005	-0.006	0.09	3255
Esophagus	w	strat.	0.96	1.37	0.65	0.012	-0.026		9862
Esophagus	b	strat.	0.97	1.57	0.77	0.003	-0.023		2722
Esophagus	h	race risk	0.91	1.04	0.74	0.013	-0.002	0.07	377
Hodgkins Disease	w	strat.	0.95	0.03	0.70	0.042	-0.032		7754
Hodgkins Disease	b	strat.	0.69	0.07	0.64	0.024	-0.005		698
Hodgkins Disease	h	race risk	0.33	0.25	0.91	0.045	-0.049	0.72	436
Kidney and Renal Pelv.	w	strat.	0.87	0.09	0.42	0.020	-0.022		21118
Kidney and Renal Pelv.	b	strat.	0.79	0.23	0.51	0.016	-0.016		2019
Kidney and Renal Pelv.	h	race risk	0.52	0.39	0.62	0.019	-0.008	0.11	1080
Larynx	w	strat.	0.89	0.06	0.73	0.023	-0.002		14918
Larynx	b	strat.	0.77	0.17	0.81	0.004	0.003		2134
Larynx	h	race risk	0.47	0.26	1.02	0.006	-0.003	0.12	422
Leukemias	w	strat.	1.00	0.18	0.51	0.011	-0.028		27682
Leukemias	b	strat.	0.93	0.35	0.80	0.010	-0.010		2121
Leukemias	h	race risk	0.85	0.19	0.53	0.019	0.005	0.37	1508
Lung and Bronchus	w	strat.	0.95	1.33	0.52	0.012	-0.006		145783
Lung and Bronchus	b	strat.	0.96	1.37	0.56	0.008	-0.003		19439
Lung and Bronchus	h	race risk	0.90	1.24	0.73	0.010	-0.002	0.12	3128
Melanomas	w	strat.	0.26	0.23	1.03	0.004	-0.039		29481
Melanomas	b	strat.	0.65	0.16	0.62	0.023	-0.020		158
Melanomas	h	race risk	0.20	0.32	1.27	0.006	-0.023	0.73	345
Non-Hodgkins Lymph.	w	strat.	0.87	0.16	0.52	0.010	-0.005		33292
Non-Hodgkins Lymph.	b	strat.	0.90	0.22	0.54	0.005	0.013		2594
Non-Hodgkins Lymph.	h	race risk	0.77	0.84	0.62	0.001	-0.055	0.41	1978
Oral Cavity and Phar.	w	strat.	0.74	0.19	0.71	0.013	-0.000		30743
Oral Cavity and Phar.	b	strat.	0.94	0.40	0.68	0.011	-0.004		3917
Oral Cavity and Phar.	h	race risk	0.59	0.45	1.00	0.010	-0.022	0.34	945

Table 6: Survival parameter estimates for males *(cont.d from table 5)*. Data on various cancer sites, white and black race/ethnicities (r/e) collected by SEER9 registries in period 1975–1998; data on various cancer sites, hispanic race/ethnicity (r/e) collected by SEER11 registries in period 1992–1998.

cancer site	r/e	model	Q	λ	β	γ_1	γ_2	γ_3	cases
Pancreas	w	strat.	0.99	2.98	0.59	0.011	-0.012		19588
Pancreas	b	strat.	0.98	4.36	0.56	0.016	-0.004		2510
Pancreas	h	race risk	0.97	2.41	0.64	0.011	0.001	-0.03	711
Prostate	w	strat.	0.50	0.09	1.17	0.023	-0.095		216782
Prostate	b	strat.	0.69	0.11	1.10	0.021	-0.073		27921
Prostate	h	race risk	0.07	0.21	1.56	-0.082	-0.168	0.99	9535
Stomach	w	strat.	0.90	1.16	0.61	0.006	-0.009		19719
Stomach	b	strat.	0.92	1.12	0.52	0.008	-0.004		3049
Stomach	h	race risk	0.86	1.13	0.70	0.004	-0.005	0.03	1332
Testis	w	strat.	0.11	0.17	0.63	-0.006	-0.084		12362
Testis	b	strat.	0.11	0.17	0.63	-0.006	-0.084		263
Testis	h	race risk	0.09	0.62	1.40	-0.025	-0.079	0.59	846
Thyroid	w	strat.	non	con-	-ver-	-gen-	-ce-		5535
Thyroid	b	strat.	0.38	0.06	0.48	0.034	-0.032		275
Thyroid	h	race risk	non	con-	-ver-	-gen-	-ce-		1135
Urinary Bladder	w	strat.	1.00	0.02	0.63	0.043	-0.032		58392
Urinary Bladder	b	strat.	1.00	0.06	0.55	0.038	-0.023		2289
Urinary Bladder	h	race risk	0.29	0.16	0.76	0.034	-0.005	0.23	1135

Table 7: Survival parameter estimates for females. Data on various cancer sites, white and black race/ethnicities (r/e) collected by SEER9 registries in period 1975–1998; data on various cancer sites, hispanic race/ethnicity (r/e) collected by SEER11 registries in period 1992–1998.

cancer site	r/e	model	Q	λ	β	γ_1	γ_2	γ_3	cases
Acute Lymph. Leuk.	w	strat.	0.62	0.31	1.08	0.055	-0.053		2394
Acute Lymph. Leuk.	b	strat.	0.54	0.46	0.79	0.016	-0.037		224
Acute Lymph. Leuk.	h	race risk	0.86	0.49	0.85	0.059	-0.130	0.22	435
All Sites	w	strat.	0.66	0.18	0.52	0.024	-0.015		815109
All Sites	b	strat.	0.76	0.33	0.58	0.020	-0.008		74511
All Sites	h	race risk	0.51	0.35	0.61	0.029	-0.008	0.21	33021
Brain	w	strat.	0.98	0.53	0.47	0.039	-0.018		12216
Brain	b	strat.	0.89	0.84	0.54	0.025	-0.011		818
Brain	h	race risk	0.75	0.75	0.85	0.044	0.004	0.03	641
Breast	w	strat.	0.40	0.13	1.20	0.001	-0.038		246226
Breast	b	strat.	0.53	0.19	1.13	0.001	-0.022		21464
Breast	h	race risk	0.23	0.22	1.45	-0.008	-0.025	0.35	9453
Cervix and Uteri	w	strat.	0.48	0.15	0.71	0.037	0.001		20452
Cervix and Uteri	b	strat.	0.56	0.21	0.81	0.032	0.011		4148
Cervix and Uteri	h	race risk	0.36	0.42	1.23	0.031	-0.010	-0.10	2572
Colon and Rectum	w	strat.	0.53	0.41	0.71	0.008	-0.020		105080
Colon and Rectum	b	strat.	0.67	0.39	0.68	0.012	-0.012		9949
Colon and Rectum	h	race risk	0.49	0.38	0.74	0.013	-0.003	0.15	2849
Corpus and Uterus	w	strat.	0.17	0.31	0.84	0.049	0.001		59110
Corpus and Uterus	b	strat.	0.60	0.29	0.69	0.042	-0.009		3485
Corpus and Uterus	h	race risk	0.16	0.43	0.98	0.038	-0.002	0.49	1755
Esophagus	w	strat.	0.96	1.29	0.66	0.009	-0.019		3807
Esophagus	b	strat.	0.96	1.57	0.77	0.009	-0.017		976
Esophagus	h	race risk	0.94	0.68	0.68	0.025	0.016	-0.25	91
Hodgkins Disease	w	strat.	1.00	0.02	0.73	0.050	-0.045		6197
Hodgkins Disease	b	strat.	0.42	0.13	0.63	0.023	-0.028		512
Hodgkins Disease	h	race risk	non	con-	-ver-	-gen-	-ce		318
Kidney and Renal Pelv.	w	strat.	1.00	0.05	0.37	0.026	-0.023		12930
Kidney and Renal Pelv.	b	strat.	0.78	0.20	0.53	0.028	-0.020		1379
Kidney and Renal Pelv.	h	race risk	0.87	0.03	0.40	0.031	0.019	0.03	737
Larynx	w	strat.	non	con-	-ver-	-gen-	-ce		3319
Larynx	b	strat.	non	con-	-ver-	-gen-	-ce		526
Larynx	h	race risk	0.42	0.24	0.93	0.024	0.035	0.38	72
Leukemias	w	strat.	1.00	0.18	0.48	0.012	-0.018		20669
Leukemias	b	strat.	0.94	0.34	0.65	0.014	-0.012		1766
Leukemias	h	race risk	1.00	0.10	0.43	0.020	0.007	0.27	1111
Lung and Bronchus	w	strat.	0.94	1.03	0.52	0.014	-0.006		84479
Lung and Bronchus	b	strat.	0.94	1.11	0.60	0.009	-0.004		8474
Lung and Bronchus	h	race risk	0.87	1.02	0.74	0.014	0.001	0.10	2103

Table 8: Survival parameter estimates for females *(cont.d from table 7)*. Data on various cancer sites, white and black race/ethnicities (r/e) collected by SEER9 registries in period 1975–1998; data on various cancer sites, hispanic race/ethnicity (r/e) collected by SEER11 registries in period 1992–1998.

cancer site	r/e	model	Q	λ	β	γ_1	γ_2	γ_3	cases
Melanomas	w	strat.	0.19	0.15	1.05	0.021	-0.037		26046
Melanomas	b	strat.	1.00	0.03	0.57	0.013	0.001		183
Melanomas	h	race risk	0.16	0.21	1.09	0.027	-0.034	0.76	467
Non-Hodgkins Lymph.	w	strat.	1.00	0.09	0.53	0.025	-0.020		28711
Non-Hodgkins Lymph.	b	strat.	0.98	0.13	0.52	0.020	0.003		1779
Non-Hodgkins Lymph.	h	race risk	1.00	0.06	0.48	0.028	-0.012	0.27	1331
Oral Cavity and Phar.	w	strat.	1.00	0.07	0.61	0.030	-0.014		14175
Oral Cavity and Phar.	b	strat.	1.00	0.18	0.61	0.022	-0.005		1498
Oral Cavity and Phar.	h	race risk	0.43	0.32	0.92	0.028	0.014	-0.06	420
Ovary	w	strat.	0.72	0.39	0.81	0.040	-0.030		34930
Ovary	b	strat.	0.83	0.41	0.73	0.050	-0.025		2422
Ovary	h	race risk	0.72	0.30	0.84	0.045	-0.021	0.01	1683
Pancreas	w	strat.	0.99	2.88	0.60	0.014	-0.011		19867
Pancreas	b	strat.	0.99	2.64	0.75	0.019	-0.002		2666
Pancreas	h	race risk	0.97	2.00	0.73	0.017	0.004	-0.02	794
Stomach	w	strat.	0.89	1.20	0.49	0.009	-0.013		11791
Stomach	b	strat.	0.90	1.11	0.62	0.014	-0.018		1830
Stomach	h	race risk	0.87	1.23	0.62	0.009	-0.011	0.06	960
Thyroid	w	strat.	non	con-	-ver-	-gen-	-ce		15840
Thyroid	b	strat.	non	con-	-ver-	-gen-	-ce		1042
Thyroid	h	race risk	non	con-	-ver-	-gen-	-ce		413
Urinary Bladder	w	strat.	1.00	0.02	0.49	0.039	-0.014		19826
Urinary Bladder	b	strat.	1.00	0.15	0.50	0.024	-0.025		1205
Urinary Bladder	h	race risk	0.48	0.16	0.56	0.037	-0.018	0.46	413

Table 9: Differences in survival parameter estimates between SEER9 and SEER11 registries. Data on various cancer sites, males, white race/ethnicity collected by SEER9 and SEER11 reg. in period 1992–1998.

cancer site	registries	Q	λ	β	γ_1	γ_2
Acute Lymph. Leuk.	SEER11	0.82	0.57	1.37	0.058	-0.104
Acute Lymph. Leuk.	SEER9	0.98	0.63	1.16	0.045	-0.085
All sites	SEER11	0.46	0.58	0.62	0.003	-0.006
All sites	SEER9	0.86	0.40	0.48	0.014	-0.019
Brain	SEER11	0.80	0.76	0.96	0.045	-0.008
Brain	SEER9	0.99	1.10	0.50	0.035	-0.024
Esophagus	SEER11	0.91	1.02	0.74	0.013	-0.001
Esophagus	SEER9	0.99	2.16	0.62	0.010	-0.015
Hodgkins Disease	SEER11	0.33	0.28	0.93	0.046	-0.056
Hodgkins Disease	SEER9	1.00	0.04	0.68	0.041	-0.016
Kidney and Renal Pelv.	SEER11	0.52	0.38	0.61	0.020	-0.007
Kidney and Renal Pelv.	SEER9	0.95	0.12	0.40	0.019	-0.011
Non-Hodgkins Lymph.	SEER11	0.74	0.78	0.63	0.001	-0.046
Non-Hodgkins Lymph.	SEER9	0.94	0.21	0.55	0.016	-0.021
Oral Cavity and Phar.	SEER11	0.60	0.45	0.99	0.010	-0.023
Oral Cavity and Phar.	SEER9	0.86	0.13	0.63	0.020	-0.003
Pancreas	SEER11	0.97	2.42	0.66	0.011	-0.002
Pancreas	SEER9	0.99	3.46	0.59	0.009	-0.006
Prostate	SEER11	0.06	0.23	1.63	-0.088	-0.166
Prostate	SEER9	non	con-	-ver-	-gen-	-ce-
Testis	SEER11	0.07	0.59	1.42	-0.018	-0.047
Testis	SEER9	0.71	0.29	0.55	-0.010	-0.171

Table 10: Differences in survival parameter estimates between SEER9 and SEER11 registries. Data on various cancer sites, females, white race/ethnicity collected by SEER9 and SEER11 reg. in period 1992–1998.

cancer site	registries	Q	λ	β	γ_1	γ_2
Acute Lymph. Leuk.	SEER11	0.76	0.45	0.97	0.058	-0.104
Acute Lymph. Leuk.	SEER9	0.97	0.55	1.06	0.059	-0.092
All sites	SEER11	0.51	0.34	0.61	0.030	-0.007
All sites	SEER9	0.66	0.21	0.53	0.021	0.003
Brain	SEER11	0.75	0.76	0.86	0.045	0.003
Brain	SEER9	0.98	0.83	0.49	0.037	-0.010
Breast	SEER11	0.23	0.21	1.43	-0.007	-0.024
Breast	SEER9	0.51	0.13	1.13	0.005	-0.004
Cervix and Uteri	SEER11	0.34	0.42	1.15	0.033	-0.003
Cervix and Uteri	SEER9	0.46	0.09	0.64	0.039	0.024
Esophagus	SEER11	0.94	0.66	0.68	0.025	0.018
Esophagus	SEER9	0.94	1.26	0.70	-0.001	0.011
Hodgkins Disease	SEER11	1.00	0.00	0.59	0.068	-0.052
Hodgkins Disease	SEER9	1.00	0.02	0.72	0.047	0.009
Kidney and Renal Pelv.	SEER11	0.94	0.02	0.39	0.031	0.019
Kidney and Renal Pelv.	SEER9	0.95	0.12	0.36	0.022	-0.011
Non-Hodgkins Lymph.	SEER11	1.00	0.07	0.47	0.029	-0.014
Non-Hodgkins Lymph.	SEER9	1.00	0.20	0.54	0.023	-0.030
Oral Cavity and Phar.	SEER11	0.44	0.34	0.93	0.029	0.011
Oral Cavity and Phar.	SEER9	1.00	0.12	0.62	0.030	-0.021
Pancreas	SEER11	0.97	2.03	0.72	0.016	0.002
Pancreas	SEER9	0.99	3.31	0.61	0.012	-0.006