

# Using Mortality Data to Estimate Radiation Effects on Breast Cancer Incidence

by David G. Hoel\*<sup>†</sup> and Gregg E. Dinse\*

In this paper we combine Japanese data on radiation exposure and cancer mortality with U.S. data on cancer incidence and lethality to estimate the effects of ionizing radiation on cancer incidence. The analysis is based on the mathematical relationship between the mortality rate and the incidence and lethality rates, as well as on statistical models that relate Japanese incidence rates to U.S. incidence rates and radiation risk factors. Our approach assumes that the risk of death from causes other than the cancer does not depend on whether or not the cancer is present, and among individuals with the cancer, the risk of death attributable to the cancer is the same in Japan and the U.S. and is not affected by radiation exposure. In particular, we focus on the incidence of breast cancer in Japanese women and how this incidence is affected by radiation risk factors. The analysis uses Japanese exposure and mortality data from the Radiation Effects Research Foundation study of atomic bomb survivors and U.S. incidence and lethality data from the Surveillance, Epidemiology, and End Results Registry. Even without Japanese incidence data, we obtain reasonable estimates of the incidence of breast cancer in unexposed Japanese women and identify the radiation risk factors that affect this incidence. Our analysis demonstrates that the age at exposure is an important risk factor, but that the incidence of breast cancer is not affected by the city of residence (Nagasaki versus Hiroshima) or the time since exposure.

## Introduction

The relationship between ionizing radiation and cancer mortality is fairly well understood. Quantitative risk models have been developed to assess the impact of environmental and occupational levels of ionizing radiation on cancer mortality and to establish radiation exposure standards (1,2). The epidemiological data on which this work has been based are primarily the mortality data from follow-up studies of atomic bomb survivors (3-7) and several large medical cohorts, such as patients with ankylosing spondylitis (8). Both the cancer mortality data and the exposure data have been of sufficient quality to enable the modeling of cancer mortality rates.

Recently the emphasis of quantitative risk assessment has shifted from cancer mortality to cancer incidence as researchers have become aware of the need to assess the relationship between ionizing radiation and the onset of cancer (9). The primary impediment to the development of quantitative incidence models is the lack of tumor registries with complete coverage and high-

quality data on radiation exposure and cancer incidence. Furthermore, though some analyses have used tumor registry data from Nagasaki alone (5-7), there have not been any analyses of cancer incidence data from both Hiroshima and Nagasaki.

Our analysis combines Japanese data on radiation exposure and cancer mortality with U.S. data on cancer incidence and cancer lethality to estimate the effects of ionizing radiation on cancer incidence. This approach is based on standard likelihood techniques, the mathematical relationship between the mortality rate and the incidence and lethality rates, and statistical models relating the Japanese incidence rates to both the U.S. incidence rates and various radiation risk factors. We focus on estimating the incidence of breast cancer in Japanese women and identifying the radiation risk factors that influence a woman's chances of developing breast cancer.

## Methods

### Definitions

The cancer incidence rate, the cancer mortality rate, and the cancer lethality rate are three important descriptors of the cancer onset/death process. Suppose that age is grouped into  $J$  intervals and let  $I_j$ ,  $M_j$ , and  $L_{jk}$  denote the incidence rate, the mortality rate, and the lethality rate associated with breast cancer ( $1 \leq k \leq j \leq J$ ). Among all women surviving to the beginning

\*Division of Biometry and Risk Assessment, National Institute of Environmental Health Sciences, P.O. Box 12233, Research Triangle Park, NC 27709.

<sup>†</sup>Radiation Effects Research Foundation, 5-2 Hijiyama Park, Minami-Ku, Hiroshima 732, Japan.

Address reprint requests to D. Hoel, National Institute of Environmental Health Sciences, P.O. Box 12233, Research Triangle Park, NC 27709.

of the  $j$ th interval,  $I_j$  represents the risk of developing breast cancer in the  $j$ th interval,  $M_j$  represents the risk of dying due to breast cancer in the  $j$ th interval, and  $L_{jk}$  is a conditional mortality rate that represents the risk of dying due to breast cancer in the  $j$ th interval, given the cancer was diagnosed in the  $k$ th interval. The incidence rate, the mortality rate, and the lethality rate are discrete hazard rates that correspond to the distributions of age at breast cancer onset, age at death due to breast cancer, and time from breast cancer onset to death due to breast cancer, respectively.

### Assumptions

The proposed methods rely on three assumptions. First, we assume that the cancer lethality rates in Japan and the U.S. are the same. The distribution of ages at which women develop breast cancer ( $I_j$ ) can differ between Japan and the U.S., as can the distribution of ages at which women die due to breast cancer ( $M_j$ ), but among women diagnosed with breast cancer the risk of death due to cancer ( $L_{jk}$ ) does not depend on their nationality. Second, we assume that risk factors related to radiation exposure do not affect the cancer lethality rate, although these risk factors can affect the cancer incidence rate and hence the cancer mortality rate. Third, we assume that the risk of death from causes other than breast cancer does not depend on whether or not breast cancer is present.

### Breast Cancer Lethality

Among all women diagnosed with breast cancer in the  $k$ th interval, let  $c_{jk}$  denote the number who died due to breast cancer in the  $j$ th interval and let  $b_{jk}$  denote the number who died due to other causes (or were lost to follow-up) in the  $j$ th interval. Apart from constants of proportionality, the log-likelihood of the breast cancer lethality data can be expressed in terms of the  $L_{jk}$  values as follows:

$$\lambda_L = \sum_{k=1}^J \sum_{j=k}^J \left\{ c_{jk} \log(L_{jk}) + \left[ \sum_{i=j}^I (b_{ik} + c_{ik}) - c_{jk} \right] \log(1 - L_{jk}) \right\}. \quad (1)$$

Maximum likelihood (ML) estimates of the  $L_{jk}$  values can be derived by maximizing  $\lambda_L$ .

Recall that the lethality rate,  $L_{jk}$ , is the probability of death due to breast cancer in the  $j$ th interval, given that the cancer was diagnosed in the  $k$ th interval and the woman was alive just before the  $j$ th interval ( $1 \leq k \leq j \leq J$ ). If the follow-up period for women diagnosed with breast cancer is relatively short, a yearly lethality rate would provide much more information than one based on intervals of several years. For example, if women are followed for up to 12 years after diagnosis, then typically they contribute to the estimation of only a few 5-year lethality rates.

Let  $G_{ik}$  denote the yearly lethality rate, which is defined as the probability of death due to breast cancer at age  $i$ , given that the cancer was diagnosed in the  $k$ th interval and the woman was alive at age  $i-1$ . It can be shown that  $L_{jk}$  and  $G_{ik}$  are related as follows:

$$1 - L_{jk} = \prod_i (1 - G_{ik}), \quad (2)$$

where the product is over all ages  $i$  in the  $j$ th interval. Apart from constants, the log-likelihood of the yearly lethality data (conditioned on the grouped diagnosis data) is  $\lambda_G = \sum_{k=1}^J \lambda_G(k)$ , where

$$\lambda_G(k) = \sum_{j=k}^J \sum_i \left\{ c_{ik} \log(G_{ik}) + \left[ \sum_{h>i} (b_{hk} + c_{hk}) - c_{ik} \right] \log(1 - G_{ik}) \right\}. \quad (3)$$

Among the women diagnosed with breast cancer in the  $k$ th interval,  $c_{ik}$  denotes the number who died due to breast cancer at age  $i$  and  $b_{ik}$  denotes the number who died due to other causes (or were lost to follow-up) at age  $i$ . The ML estimates of the  $G_{ik}$  values are obtained by maximizing  $\lambda_G$  and the  $L_{jk}$  values can be estimated from the ML estimates of the  $G_{ik}$  values by using Eq. (2).

We focus on three ways of estimating the yearly breast cancer lethality rates. First, in the absence of any parametric modeling assumptions, the nonparametric ML estimate of each  $G_{ik}$  is

$$\hat{G}_{ik} = \frac{\#\{\text{death due to breast cancer at age } i\}}{\#\{\text{alive at age } i-1 \text{ with breast cancer diagnosed in interval } k\}}, \quad (4)$$

where  $\#\{E\}$  denotes the number of women for whom event  $E$  is true. Practically speaking, the nonparametric analysis can be viewed as a fully saturated parametric analysis in the sense that a separate value of the lethality rate is estimated for each combination of the age at death and the intervals for age at diagnosis. Second, in order to reduce the number of quantities to be estimated, we can specify a parametric model for the  $G_{ik}$  values, such as the following logistic model (10):

$$\log\{G_{ik} / (1 - G_{ik})\} = \sum_{p=0}^2 (i - a_k + 1)^p \psi_p, \quad (5)$$

where  $a_k$  is the earliest age in the  $k$ th interval. This model assumes that the logit of the lethality rate is a quadratic function of the time since diagnosis. For all women diagnosed with breast cancer in the  $k$ th component of the log-likelihood,  $\lambda_G(k)$ , can be maximized as a function of the parameters  $\psi_p$  ( $p = 0, 1, 2$ ), and for each  $k$  the ML estimate of  $G_{ik}$  is calculated by substituting the ML estimates of the  $\psi_p$  values into Eq. (5). Third, for all women with breast cancer,  $\lambda_G$  can be maximized

as a function of the parameters  $\psi_{pq}$  ( $p = 0,1,2$ ;  $q = 0,1,2$ ) under the logistic model:

$$\log\{G_{ik} / (1 - G_{ik})\} = \sum_{p=0}^2 \sum_{q=0}^2 (i - a_k + 1)^p a_k^q \psi_{pq}, \quad (6)$$

which assumes that the lethality rate is a quadratic (logistic) function of the age at diagnosis and the time since diagnosis. In this case, a total of only nine parameters (i.e., the  $\psi_{pq}$  values) need to be estimated. Each  $G_{ik}$  is estimated by substituting the ML estimates of the  $\psi_{pq}$  values into Eq. (6).

### Breast Cancer Mortality and Incidence

We assumed a piecewise Exponential model for the breast cancer mortality rates, which allows the  $M_j$  values to vary with age (across intervals) even though they are constant within intervals. The log-likelihood of the breast cancer mortality data under the piecewise Exponential model is

$$\lambda_M = \sum_j \{c_j \log(M_j) - r_j M_j\}, \quad (7)$$

where  $c_j = \sum_{k=1}^j c_{jk}$  is the number of women who died due to breast cancer during the  $j$ th interval and  $r_j$  is the number of person-years at risk observed during the  $j$ th interval. Dinse and Hoel (manuscript in preparation) demonstrate that the cancer mortality rate can be expressed as the following function of the cancer incidence rates and the cancer lethality rates:

$$M_j = \frac{\sum_{k=1}^j \left\{ \prod_{i=1}^{k-1} (1 - I_i) \right\} I_k \left\{ \prod_{i=k}^{j-1} (1 - L_{ik}) \right\} L_{jk}}{\prod_{i=1}^j (1 - I_i) + \sum_{k=1}^j \left\{ \prod_{i=1}^{k-1} (1 - I_i) \right\} I_k \left\{ \prod_{i=k}^{j-1} (1 - L_{ik}) \right\}}. \quad (8)$$

Therefore, the log-likelihood  $\lambda_M$  can be expressed as a function of the  $I_i$  values (and the  $L_{ik}$  values), say  $\lambda_I$ , by substituting Eq. (8) for  $M_j$  in Eq. (7). Estimates of the  $I_i$  values can be obtained by first replacing the  $L_{ik}$  values with their estimates and then maximizing  $\lambda_I$  as a function of the  $I_i$  values.

### Linking U.S. and Japanese Incidence Rates

In order to relate the breast cancer incidence rates in Japan and the U.S., we specified the following model:

$$1 - I_j^{\text{JAPAN}} = (1 - I_j^{\text{USA}}) e^{\phi_0 + \phi_1 t_j}, \quad (9)$$

where  $t_j$  is the midpoint of the  $j$ th age interval. We make no parametric assumptions about the form of the incidence rates themselves, only the way in which  $I^{\text{JAPAN}}$  relates to  $I^{\text{USA}}$ . Based on a first-order approximation,

Eq. (9) can be viewed as a discrete analog of the proportional hazards model in which the ratio of U.S. and Japanese incidence rates is a log-linear function of age.

### Estimating Baseline Incidence in Japan

The incidence of breast cancer in unexposed Japanese women can be estimated as follows. Express  $\lambda_I$  as a function of  $I_j^{\text{JAPAN}}$ ,  $L_{jk}$ , and the Japanese mortality data. Replace the  $L_{jk}$  values with the ML estimates obtained from the U.S. lethality data and use Equation (9) to express  $I_j^{\text{JAPAN}}$  in terms of  $I_j^{\text{USA}}$ ,  $\phi_0$ , and  $\phi_1$ . Estimate  $I_j^{\text{USA}}$  from the U.S. incidence data as follows:

$$\hat{I}_j^{\text{USA}} = \frac{\#\{\text{diagnosed with breast cancer in the } j\text{th interval}\}}{\#\{\text{alive without breast cancer upon entering the } j\text{th interval}\}}. \quad (10)$$

If we treat the estimates of the U.S. lethality and baseline incidence rates as known, an estimate of the Japanese baseline incidence rate,  $I_j^{\text{JAPAN}}$ , can be calculated by maximizing  $\lambda_I$  as a function only of  $\phi_0$  and  $\phi_1$  and then substituting the estimates of  $I_j^{\text{USA}}$ ,  $\phi_0$ , and  $\phi_1$  into Eq. (9).

### Linking Incidence to Radiation Exposure

In order to relate information on radiation dose ( $D$ ) and a vector of covariates or exposure variables ( $Z$ ) to the breast cancer incidence in Japanese women, we specified the following model:

$$1 - I_j^{\text{JAPAN}}(D, Z) = (1 - I_j^{\text{JAPAN}})^{1 + D e^{\theta Z}}, \quad (11)$$

where  $\theta$  is a vector of parameters,  $I_j^{\text{JAPAN}}$  is the baseline breast cancer incidence rate, and  $I_j^{\text{JAPAN}}(D, Z)$  is the breast cancer incidence rate for women who were exposed to dose  $D$  with covariate vector  $Z$ . Based on a first-order approximation, Eq. (11) can be viewed as a discrete analog of the relative risk model in which the ratio of incidence rates for exposed and unexposed Japanese women is a simple linear function of dose and an exponentiated covariate term.

### Analyzing Radiation Effects on Incidence

The first step is to categorize the covariates. Next, assume that the cancer mortality function is piecewise Exponential with a different value,  $M_i$ , for the  $i$ th combination of the various covariate categories and the  $J$  age-at-death intervals. The log-likelihood of the data on cancer mortality and radiation exposure is

$$\lambda_M^* = \sum_i \{c_i \log(M_i) - r_i M_i\}, \quad (12)$$

where  $c_i$  is the number of women in the  $i$ th category who died due to breast cancer and  $r_i$  is the number of person-years at risk observed in the  $i$ th category. Rewrite this log-likelihood as a function of  $I_j^{\text{JAPAN}}$ ,  $I_j^{\text{JAPAN}}(D, Z)$ , and  $L_{jk}$ , say  $\lambda_I^*$ , by replacing each  $M_i$  in Eq. (12)

by a modification of Eq. (8) in which we substitute either  $I_j^{\text{JAPAN}}$  or  $I_j^{\text{JAPAN}}(D, Z)$  for  $I_j$ , depending on whether or not the exposure occurred after the  $j$ th age interval. A 10-year latency period was incorporated by including the covariate terms only for the age intervals that exceeded the age at exposure by at least ten years. Replace the  $L_{jk}$  values with the ML estimates obtained from the U.S. lethality data and use Eq. (11) to express  $I_j^{\text{JAPAN}}(D, Z)$  in terms of  $I_j^{\text{JAPAN}}$  and  $\theta$ . Treat the estimates of the lethality and baseline Japanese incidence rates as known so that  $\lambda_I^*$  reduces to a function in which only  $\theta$  is unknown. An estimate of the Japanese incidence rate,  $I_j^{\text{JAPAN}}(D, Z)$ , can be obtained by maximizing  $\lambda_I^*$  as a function of  $\theta$  and then substituting the estimates of  $I_j^{\text{JAPAN}}$  and  $\theta$  into Eq. (11).

The importance of various exposure variables can be assessed by means of the likelihood ratio test (11). The likelihood ratio statistic is equal to twice the difference between the maximized values of  $\lambda_I^*$  for models with and without the exposure variable(s) of interest. The significance level ( $p$ -value) associated with the exposure variable(s) is obtained by comparing the likelihood ratio statistic to the percentage points of the chi-squared distribution. The number of degrees of freedom equals the difference in the numbers of parameters estimated in the two models.

## Results

### Sources of Data

We obtained data on breast cancer incidence and lethality for unexposed U.S. women from the Surveillance, Epidemiology, and End Results (SEER) Registry (12). The SEER Registry gives information on women diagnosed with breast cancer between 1973 and 1984, as well as follow-up data on these women over the 12-year period from 1973 to 1985. For more than 125,000 women diagnosed with breast cancer, the SEER Registry provides the age at diagnosis, the age at the last follow-up visit, and an indicator of whether the woman was dead or lost to follow-up (e.g., still alive). If the woman died, the registry also gives the age at death and a cause-of-death indicator.

We used data on radiation exposure and breast cancer mortality for Japanese women from the Radiation Effects Research Foundation (RERF) Life Span Study (3,4). This study consists of a cohort of atomic bomb survivors followed since 1950, with each individual having a calculated radiation exposure that depends on distance and shielding. The women in the RERF study have contributed approximately 1.35 million person-years at risk; there have been more than 15,000 deaths, including 155 deaths due to breast cancer. Radiation dose is measured in milligrays (mGy) and represents the total free-in-air kerma. There are three exposure variables: city (1 = Hiroshima, 2 = Nagasaki), age at exposure, and time since exposure. The data on age at exposure, time since exposure, and attained age are recorded in 5-year intervals. For the  $i$ th combination of

city, age at exposure, time since exposure, and attained age, the data set gives the number of person-years at risk ( $r_i$ ), the number of deaths due to breast cancer ( $c_i$ ), and the mean radiation dose ( $D_i$ ).

### Breast Cancer Lethality

The women in the SEER study were followed for, at most, 12 years; thus we chose to model the yearly lethality rates (i.e., the  $G_{ik}$  values) rather than the lethality rates corresponding to the 5-year intervals (i.e., the  $L_{jk}$  values). As described earlier, the  $G_{ik}$  values were estimated in three ways. First, the nonparametric ML estimate in Eq. (4) was calculated for each of the 168 combinations of the 12 years of follow-up and the 14 intervals for age at diagnosis. Second, the three parameters in Eq. (5) were estimated separately for each of the 14 age-at-diagnosis intervals, giving in a total of 42 estimates. Third, the nine parameters in Eq. (6) were estimated on the basis of all the data.

These three sets of breast cancer lethality estimates are plotted in Figure 1a-c. For each age-at-diagnosis interval  $k$ , the ML estimates of the  $G_{ik}$  values are plotted against the number of years since diagnosis. For women diagnosed before age 60, the breast cancer death rate initially tends to increase as a function of time since diagnosis and then decreases toward zero, with a peak between 2 and 4 years after diagnosis. For women diagnosed at or after age 60, the breast cancer death rate seems to steadily decrease with time since diagnosis. The three figures suggest that Eq. (6) is adequate for summarizing the yearly lethality rates because the estimates based on the 9-parameter model provide roughly the same information as those based on the 42-parameter or 168-parameter models. The  $L_{jk}$  values can be estimated by substituting the ML estimates of the  $G_{ik}$  values into Eq. (2).

### Baseline Incidence

In order to estimate the baseline incidence of breast cancer in U.S. women, we applied Eq. (10) to the data from the SEER Registry. The resulting estimate of each  $I_j^{\text{USA}}$ , multiplied by 100,000 and rounded to the nearest integer, is listed in Table 1 and plotted on a log-scale in Figure 2. The incidence of female breast cancer in the U.S. is negligible for women under 25, increases rapidly for women between 25 and 60, and then begins to level off for women over 60.

In order to estimate the baseline incidence of breast cancer in Japanese women, we treated the mortality data in the lowest radiation exposure group in the RERF study (i.e., 0-4 mGy) as baseline mortality data. Next, we used these data to construct  $\lambda_I$ , treated the estimates of  $I_j^{\text{USA}}$  and  $L_{jk}$  as fixed, and then maximized  $\lambda_I$  as a function of  $\phi_0$  and  $\phi_1$ . The resulting estimates were:  $\hat{\phi}_0 = -0.1467$  and  $\hat{\phi}_1 = -0.0279$ . The corresponding estimate of each  $I_j^{\text{JAPAN}}$ , which is obtained by substituting  $\hat{\phi}_0$  and  $\hat{\phi}_1$  into Eq. (9), is listed in Table 1 and plotted in Figure 2. For comparison purposes, published

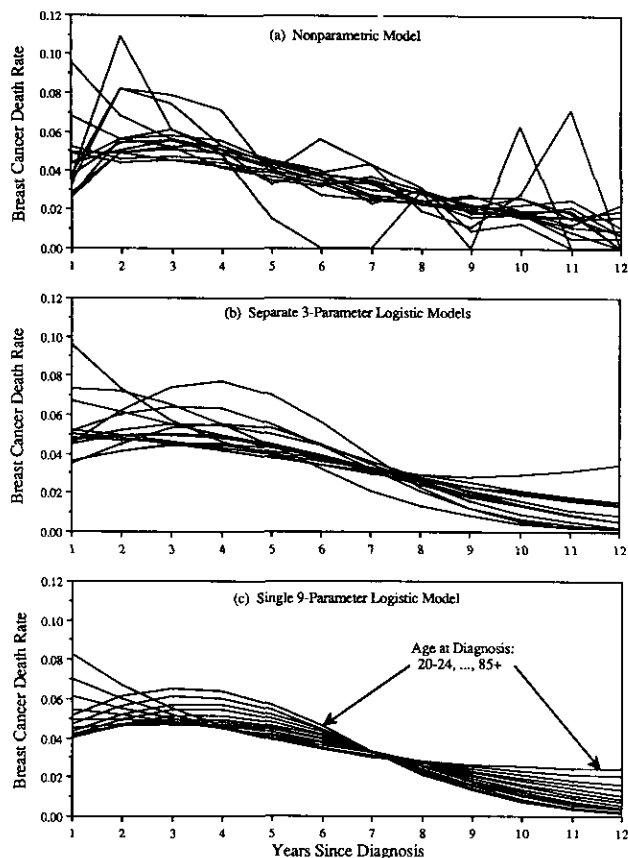


FIGURE 1. Breast cancer lethality. The rate of death due to breast cancer, conditional on the cancer being diagnosed in a particular age interval, is plotted against the number of years since diagnosis for each of 14 age-at-diagnosis groups. (a) Depicts the nonparametric estimates, (b) depicts the parametric estimates derived from fitting a separate quadratic logistic model to the data for each age-at-diagnosis group, and (c) depicts the parametric estimates derived from fitting a single logistic model that is quadratic in both the age at diagnosis and the time since diagnosis.

estimates (13) of the incidence rates of breast cancer for women in three Japanese prefectures (Osaka in 1973-7; Fukuoka in 1974-5; Miyagi in 1973-7) are also listed in Table 1 and plotted in Figure 2.

The most notable feature of Table 1 and Figure 2 is that the incidence of breast cancer in Japanese women is much lower than in U.S. women. The breast cancer incidence in Japan peaks at a level that is roughly an order of magnitude less than the highest incidence in the U.S. and then, in contrast to the U.S. incidence, appears to decrease slightly in the older age groups. The other remarkable result is that our estimates of the Japanese incidence rates, which are based on Japanese mortality data, U.S. incidence and lethality data, and the simple log-linear model in Eq. (9), are very similar to the observed incidences of breast cancer in Osaka, Fukuoka, and Miyagi.

**Radiation Effects**

We evaluated the effect of ionizing radiation on the incidence of breast cancer in Japanese women through

Table 1. Age-specific incidence of breast cancer among unexposed women.

Age interval	Breast cancer incidence per 100,000 unexposed women				
	United States <sup>a</sup>	Japan			Our estimate
		Osaka <sup>b</sup>	Fukuoka <sup>b</sup>	Miyagi <sup>b</sup>	
0-4	0	0	0	0	0
5-9	0	0	0	0	0
10-14	0	0	0	0	0
15-19	0	0	0	0	0
20-24	1	0	1	1	1
25-29	9	2	4	3	4
30-34	29	7	13	14	10
35-39	70	15	16	20	21
40-44	121	25	38	39	32
45-49	181	39	47	54	41
50-54	205	41	42	58	41
55-59	251	35	49	40	44
60-64	297	35	41	42	45
65-69	336	38	39	51	44
70-74	366	35	33	55	42
75-79	386	37	28	33	38
80-84	390	37	31	36	34

<sup>a</sup> U.S. rates were obtained by applying Eq. (10) to the data from the SEER Registry (12).

<sup>b</sup> Japanese rates were obtained from the IARC data (13).

the use of Eq. (11). The log-likelihood  $\lambda_j^*$  was expressed as a function of the U.S. lethality rates, the Japanese baseline incidence rates, and the parameter vector  $\theta$ . We then substituted the Japanese mortality data into  $\lambda_j^*$ , treated the estimates of  $L_{jk}$  and  $I_j^{JAPAN}$  as fixed, and maximized  $\lambda_j^*$  with respect to the elements of  $\theta$ . The effects of the exposure variables were assessed by comparing differences in the values of twice the maximized log-likelihoods, which are listed in Table 2 for various combinations of the exposure variables. In every case, the first element of the covariate vector  $Z$  is the constant 1, which allows for a simple dose effect.

Table 2 illustrates that age at exposure was the primary predictor of breast cancer incidence among exposed women; the incidence rate decreased as the age at exposure increased. In fact, the relevant information was captured by simply dichotomizing age at exposure into two groups: women under age 20 and women of age 20 or older. Further partitioning of the age at exposure into three groups (ages 0-19, 20-49, 50+) did not yield a significant improvement. As for the other covariates, city seemed to have no effect at all, and time since exposure had a minimal effect; the incidence rate tended to increase with the time since exposure, but the increase was not statistically significant.

**Conclusions**

We have shown that by borrowing information on cancer incidence and cancer lethality in the U.S. from the SEER data, age-specific cancer incidence rates in Japan can be estimated from Japanese cancer mortality data. In fact, our methods produced very good estimates of the Japanese incidence rates of breast cancer, even though the incidence of breast cancer in the U.S. is much

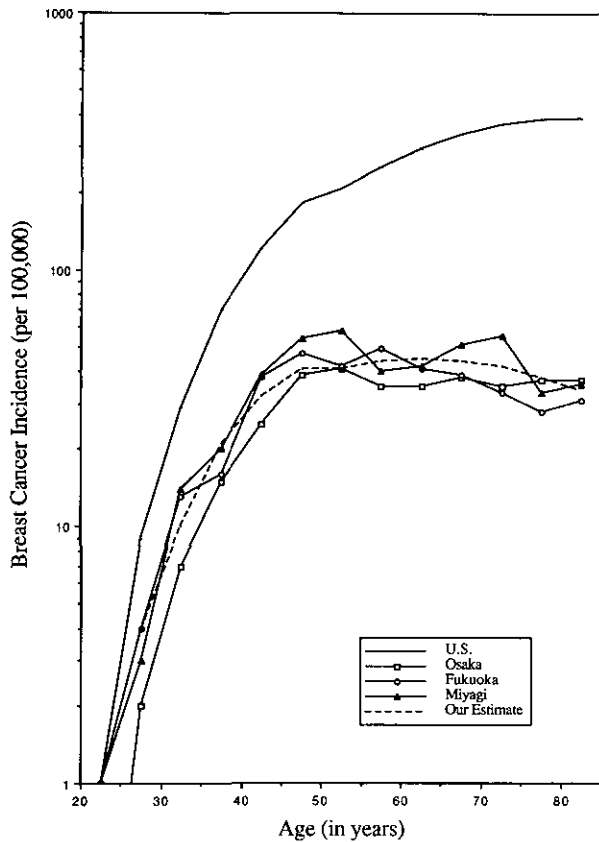


FIGURE 2. Baseline breast cancer incidence. The baseline incidence of breast cancer is plotted against age for women in the U.S. and Japan. The plain solid line depicts the general incidence rate in the U.S.; the lines connected by open squares, circles, and triangles depict the incidence rates in the Japanese prefectures of Osaka, Fukuoka, and Miyagi; and the dashed line depicts our estimate of the general incidence rate in Japan.

Table 2. Regression analysis of the effect of various exposure variables on the incidence of breast cancer in Japanese women.<sup>a</sup>

Exposure variables	-2 Log-likelihood	p-value <sup>b</sup>
Constant	3036.20	—
Constant, city	3036.18	0.888
Constant, time since exposure	3035.53	0.413
Constant, age at exposure	3030.30	0.015
Constant, I {age at exposure $\geq$ 20} <sup>c</sup>	3029.61	0.010
Constant, I {age at exposure $\geq$ 20}, I {age at exposure $\geq$ 50}	3028.89	0.026

<sup>a</sup> Assuming a 10-year latency period for the exposure effects.

<sup>b</sup> Significance level associated with the likelihood ratio test which compares a model involving one or more exposure variables with the model involving only the constant term.

<sup>c</sup> I {E} is an indicator variable which equals one if event E is true and 0 otherwise.

greater than in Japan. The accuracy of the incidence estimates provides support for our assumption of similar U.S. and Japanese breast cancer lethality rates. Currently we are refining the model for the cancer lethality rates, and we are in the process of applying our technique to other cancer sites.

The second aspect of this work is the application of statistical models that estimate cancer incidence induced by radiation. This approach permits incidence-based radiation risk estimation, even without adequate cancer incidence data. We are applying our modeling technique to various cancer sites and will compare our incidence estimates with other available estimates, even those based on direct estimation from actual incidence data.

The Japanese data on radiation and cancer mortality have been obtained from the Radiation Effects Research Foundation (RERF) in Hiroshima, Japan. RERF is a private foundation that is funded equally by the Japanese Ministry of Health and Welfare and the U.S. Department of Energy through the U.S. National Academy of Sciences. The conclusions reached in this report are those of the authors and do not necessarily reflect the scientific judgment of RERF or its funding agencies. The authors thank RERF for the use of their data and Frank Dilorio for his assistance with much of the computer programming.

## REFERENCES

- Pierce, D., and Preston, D. Hazard function modeling for dose-response analysis of cancer incidence in the A-bomb survivor data. In: *Atomic Bomb Survivor Data: Utilization and Analysis* (R. Prentice and D. Thompson, Eds.), Society for Industrial and Applied Mathematics, Philadelphia, PA 1984, pp. 51-66.
- Hoel, D., and Preston, D. Dose-response in radiation carcinogenesis: human studies. In: *Epidemiology and Quantitation of Environmental Risk in Humans from Radiation and Other Agents* (A. Castellani, Ed.), Plenum Press, London, 1985, pp. 105-122.
- Preston, D., Kato, H., Kopecky, K., and Fujita, S. Studies of the mortality of A-bomb survivors, Report 10. *Cancer mortality, 1950-82*. *Radia. Res.* 111: 151-78 (1987).
- Shimizu, Y., Kato, H., and Schull, W. Life Span Study, Report 11. Part II. *Cancer Mortality in the Years 1950-85 Based on the Recently Revised Doses (DS86)*. RERF TR5-88, Radiation Effects Research Foundation Technical Report Series, Hiroshima, Japan, 1988.
- Wakabayashi, T., Kato, H., Ikeda, T., and Schull, W. Studies of the mortality of A-bomb survivors, Report 7. Part III. *Incidence of cancer in 1950-78, based on the Tumor Registry, Nagasaki*. *Radiat. Res.* 93: 112-46 (1983).
- Monzen, T., and Wakabayashi, T. Tumor and tissue registries in Hiroshima and Nagasaki. GANN Monograph on Cancer Research No. 32. In: *Cancer in Atomic Bomb Survivors* (I. Shigematsu and A. Kagan, Eds.), Japan Scientific Societies Press, Tokyo, Japan, 1986, pp. 29-40.
- Ikeda, T., Hayashi, I., Matsuo, T., Maeda, H., and Shimokawa, I. The cancer registry in Nagasaki city, with atomic bomb survivor data, 1973-1977. GANN Monograph on Cancer Research No. 32. In: *Cancer in Atomic Bomb Survivors* (I. Shigematsu and A. Kagan, Eds.), Japan Scientific Societies Press, Tokyo, Japan, 1986, pp. 41-52.
- Darby, S., Doll, R., Gill, S., and Smith, P. Long term mortality after a single treatment course with X-rays in patients treated for ankylosing spondylitis. *Br. J. Cancer* 55: 179-90 (1987).
- Report of the National Institutes of Health Ad Hoc Working Group to Develop Radioepidemiological Tables, NIH Publication No. 85-2748. U.S. Department of Health and Human Services, National Institutes of Health, Bethesda, MD, 1985.
- Cox, D. *Analysis of Binary Data*. Chapman and Hall, London, 1970.
- Cox, D., and Hinkley, D. *Theoretical Statistics*. Chapman and Hall, London, 1974.
- 1987 Annual Cancer Statistics Review (Including Cancer Trends: 1950-1985). NIH Publication No. 88-2789. U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute, Bethesda, MD, 1988.

13. Waterhouse, J., Muir, C., Shanmugaratnam, K., and Powell, J.  
Cancer Incidence in Five Continents, Volume IV, IARC Scientific

Publications No. 42. International Agency for Research on Cancer,  
Lyon, 1982.