# Effect of Age on Incidence of Breast Cancer in Females [1, 2]

Suresh H. Moolgavkar, [3] Richard G. Stevens, [3] *and* John A. H. Lee [4]

**ABSTRACT—Incidence and mortality data on breast cancer in females from various Occidental (Western) and Japanese populations were analyzed. After adjustment for birth cohort and year of event, the age curves from all the Western populations were very similar in shape. The age curve for Osaka incidence was very similar in shape to the Western incidence curves. The Osaka curve continued to increase after menopause; the postmenopausal decrease in rates in individual cross sections was the result of a strong cohort effect. The premenopausal mortality curve for Japan was very similar in shape to the Western curves; however, the postmenopausal Japanese mortality curve had a smaller slope at each point.—J Natl Cancer Inst 62:493-501, 1979.**

The temporal trend in incidence of breast cancer in females in Iceland was shown by Bjarnason et al. (*1*) and Breslow and Day (*2*) to be due to a strong cohort effect. This paper extends and refines the methodology presented by Breslow and Day (*2*) and applies it to the analysis of female breast cancer incidence and mortality data with the following objectives: 1) to describe temporal trends in the populations studied; 2) to adjust for these temporal trends to extract an age curve that represents the "true" effect of age on breast cancer incidence and mortality and to compare these statistically adjusted age curves between populations.

## DATA SOURCES

Incidence data for female breast cancer were examined for Connecticut, years 1935-72 (*3-11*); for Denmark, years 1943-67 (*12-14*); for Iceland, years 1911-69 (*2*); and for Osaka, years 1963-75. (These data were provided by Dr. I. Fujimoto, Osaka Cancer Registry, in "Cancer in Osaka Annual Report, Osaka Prefectural Health Department, 1967-1975"). Mortality data were examined for the United States white female population, years 1931-70 (*15-18*); for England and Wales, years 1911-70 (*19*); for Canada, years 1931-70 (*20*); and for Japan, years 1947-75 (*21-23*).

## MODELS

The data can be organized in the form of a two-way contingency table with missing observations and unequal populations at risk in each cell. The top half of table 1 presents the age-specific incidence of breast cancer in Denmark from 1943 to 1967. The bottom half of the table shows the populations at risk. The experience of birth cohorts is shown in horizontal rows of the table, whereas cross sections are shown diagonally. There are five complete cross sections. Thus there are no more than five observations on any given cohort.

The generation-effects model (*24, 25*) postulates that mortality curves are completely determined by age and cohort in a simple manner. The age-specific mortality is assumed to be $M(c,t)=f(c)g(t)$, where $M(c,t)$ is the mortality at age $t$ in cohort $c$, $f(c)$ is a function that depends on cohort alone, and $g(t)$ is a function that depends on age alone. This multiplicative model has an important role in medical statistics under the guise of proportional hazards. It was rigorously tested against the incidence of female breast cancer in Iceland by Breslow and Day (*2*) and was found to fit the data very well. However, of all the populations we examined, the model fit only the Iceland and Osaka incidence data sets. Therefore, we formulated and tested a hierarchy of other models.

Each model is based on the following assumption: The numbers of events in individual cells are independent Poisson variates with parameters $N_{ij}E_{ij}$, where $N_{ij}$ is the population at risk in the cell representing age group $i$ and birth cohort $j$, and $E_{ij}$ is specified by the model being tested. In this contingency table setting, the generation-effects model postulates independence of the age effect from the cohort effect. The following list details the hierarchy of models tested against the data. Model I is the simple generation-effects model (*25*). Each model thereafter introduces a specific type of interaction between the age and cohort effects.

*Model I.*—In this model, $E_{ij}=a_ib_j$, where $a_i$ is the effect of age $i$ and $b_j$ is the effect of cohort $j$. The expected number of events in the cell $(i, j)$ is $a_ib_jN_{ij}$.

*Model II.*—In this model $E_{ij}=a_ib_jc_k$; $a_i$ and $b_j$ have the same interpretation as in model I. The parameter $c_k$ represents a year of incidence effect; it is constant along given cross sections. Birth cohort and age completely determine the year of incidence, and therefore, introduction of a third index $k$ to represent cross section is unnecessary. However, the notation would get cumbersome, so index $k$ has been used because little danger of confusion exists. This model is tantamount to introducing an interaction between age and birth cohort that is constant along cross sections.

*Model III.*—In this model $E_{ij}=a_i\exp(x_iy_j)$, where the $a_i$ and $x_i$ are age effects and the $y_j$ are cohort effects. This model is a special case of model IV with all the $b_j$ equal to 1.

*Model IV.*—In this model $E_{ij}=a_ib_j\exp(x_iy_j)$, where the $a_i$ and $x_i$ are age effects and the $b_j$ and $y_j$ are cohort effects. Model IV is a natural extension of model I, the simple multiplicative model, and postulates that on a

[1] Received June 15, 1978; accepted September 18, 1978.

[2] Supported in part by Public Health Service (PHS) contract N01-CB74100-34 from the Division of Cancer Biology and Diagnosis, National Cancer Institute (NCI); and supported in part by PHS grant CA23321-01 from NCI. Computer use was supported in part by PHS grants CA06927 and CA22780 from NCI.

[3] The Fox Chase Cancer Center, 7701 Burholme Ave., Philadelphia, Pa. 19111.

[4] The Fred Hutchinson Cancer Center, 1102 Columbia St., Seattle, Wash. 98104.

TABLE 1.—*Age-specific incidence and person-years at risk for female breast cancer in Denmark, 1943–67*

| Midyear of birth cohort | Age, yr[a] | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 20–24 | 25–29 | 30–34 | 35–39 | 40–44 | 45–49 | 50–54 | 55–59 | 60–64 | 65–69 | 70–74 | 75–79 | 80–84 |
| **Age-specific incidence** | | | | | | | | | | | | | |
| 1863 | | | | | | | | | | | | | 289 |
| 1868 | | | | | | | | | | | | 315 | 368 |
| 1873 | | | | | | | | | | | 476 | 406 | 461 |
| 1878 | | | | | | | | | | 667 | 602 | 494 | 624 |
| 1883 | | | | | | | | | 699 | 659 | 676 | 660 | 828 |
| 1888 | | | | | | | | 670 | 738 | 766 | 773 | 741 | |
| 1893 | | | | | | | 559 | 659 | 777 | 823 | 901 | | |
| 1898 | | | | | | 634 | 663 | 719 | 862 | 996 | | | |
| 1903 | | | | | 457 | 711 | 733 | 790 | 982 | | | | |
| 1908 | | | | 268 | 523 | 797 | 787 | 931 | | | | | |
| 1913 | | | 119 | 272 | 519 | 852 | 908 | | | | | | |
| 1918 | | 26 | 99 | 294 | 531 | 932 | | | | | | | |
| 1923 | 2 | 32 | 83 | 291 | 659 | | | | | | | | |
| 1928 | 10 | 30 | 91 | 289 | | | | | | | | | |
| 1933 | 7 | 26 | 94 | | | | | | | | | | |
| 1938 | 4 | 33 | | | | | | | | | | | |
| 1943 | 4 | | | | | | | | | | | | |
| **Person-years at risk** | | | | | | | | | | | | | |
| 1863 | | | | | | | | | | | | | 125,515 |
| 1868 | | | | | | | | | | | | 163,000 | 145,460 |
| 1873 | | | | | | | | | | | 252,500 | 185,700 | 168,820 |
| 1878 | | | | | | | | | | 351,300 | 292,800 | 217,780 | 201,955 |
| 1883 | | | | | | | | | 435,800 | 392,700 | 331,700 | 252,675 | 243,440 |
| 1888 | | | | | | | | 507,300 | 475,200 | 432,325 | 371,400 | 287,475 | |
| 1893 | | | | | | | 575,800 | 554,100 | 521,915 | 479,540 | 417,115 | | |
| 1898 | | | | | | 662,400 | 642,900 | 629,285 | 588,740 | 452,875 | | | |
| 1903 | | | | | 722,500 | 708,000 | 688,850 | 666,050 | 635,595 | | | | |
| 1908 | | | | 784,500 | 771,300 | 757,655 | 740,710 | 718,220 | | | | | |
| 1913 | | | 808,600 | 796,200 | 782,335 | 770,230 | 754,750 | | | | | | |
| 1918 | | 797,100 | 788,900 | 773,175 | 764,620 | 755,485 | | | | | | | |
| 1923 | 811,500 | 799,100 | 790,325 | 782,180 | 777,065 | | | | | | | | |
| 1928 | 737,300 | 735,380 | 727,055 | 724,885 | | | | | | | | | |
| 1933 | 714,355 | 703,365 | 702,225 | | | | | | | | | | |
| 1938 | 750,610 | 749,245 | | | | | | | | | | | |
| 1943 | 921,865 | | | | | | | | | | | | |

[a] Five-year age groups.

logarithmic scale, the interaction between age and cohort is multiplicative.

*Model V.*—In this model $E_{ij} = a_i b_j \exp(x_i z_k)$, where the $a_i$ and $x_i$ are age effects, the $b_j$ are cohort effects, and the $z_k$ are year-of-incidence effects. Model II is an extension of model II. It postulates that the interaction

TABLE 2.—*Incidence and mortality rates per 100,000 female population, age adjusted to the world population (1960)*[a]

| Location | Year | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Incidence** | | | | | | | | | | | |
| Denmark | | | | | 1943 | 1948 | 1953 | 1958 | 1963 | | |
| | | | | | 70.3 | 70.7 | 71.5 | 73.1 | 79.9 | | |
| Connecticut | | | | 1935 | 1940 | 1945 | 1950 | 1955 | 1960 | 1965 | 1970 |
| | | | | 81.7 | 79.7 | 85.9 | 93.8 | 92.6 | 96.2 | 108.7 | 117.3 |
| Osaka, Japan | | | | | | | | | 1963 | 1966 | 1971 |
| | | | | | | | | | 17.3 | 18.9 | 21.2 |
| Iceland | 1911–19 | 1920–29 | | 1930–39 | | 1940–49 | | 1950–59 | | 1960–69 | |
| | 30.8 | 30.0 | | 47.9 | | 54.0 | | 58.0 | | 63.8 | |
| **Mortality** | | | | | | | | | | | |
| United States | | | 1931 | 1936 | 1941 | 1946 | 1951 | 1956 | 1961 | 1966 | 1971 |
| | | | 34.9 | 36.5 | 35.7 | 35.6 | 34.9 | 34.8 | 34.5 | 36.2 | |
| England and Wales | | 40.3 | 41.5 | 41.0 | 38.4 | 38.4 | 37.8 | 37.8 | 39.2 | 41.0 | |
| Canada | | | 34.9 | 37.6 | 37.7 | 37.4 | 37.0 | 38.3 | 38.1 | | |
| Japan | | | | | | 6.4 | 6.5 | 6.2 | 6.2 | 6.7 | 7.8 |

[a] From (*26*).

introduced by the year-of-incidence effect in model II is modulated by age.

*Model VI.*—In this model $E_{ij}=a_ib_jc_k\exp(x_iz_k)$, where the $a_i$ and $x_i$ are age effects, the $b_j$ are cohort effects, and the $c_k$ and the $z_k$ are year-of-incidence effects. Model VI is an extension of model V.

Maximum likelihood estimates of the parameters of the model were obtained by a Newton-Raphson iterative procedure. These parameters are uniquely determined after normalization (*2*). Our main interest is the shape of the extracted age curve. Therefore, the absolute magnitudes of the age parameters are of secondary interest. Of principal concern to us is their internal relationship to one another (within a given data set). Specifically, if each age parameter is multiplied by the same non-zero constant, then on a semilogarithmic plot the age curves so obtained will simply be (vertical) translates of each other. Hence we are interested in the age parameters only up to a non-zero multiplicative constant. When comparisons of extracted age curves are made among populations, the appropriate normalizations are described in the text-figure legends. Goodness of fit was judged by means of a statistic $X^2$ (tables 4-6), which has an asymptotic chi-squared distribution. Details are in the "Statistical Appendix."

## OBSERVATIONS

Table 2 presents the incidence and mortality rates for ages 20-84 adjusted to the world population of Segi (*26*). The most graphic change in this statistic occurred in the Connecticut incidence data; incidence has gone from 81.7/100,000 during 1935-39 to 117.3/100,000 during 1970-72. The incidence in Denmark has also shown a steady increase from 1943 to 1967. Only three time periods of observation are available on the incidence in Osaka, but again, it shows a steady increase. The mortality picture is quite different. The only population to suggest a real increase recently is Japan. Examination of the age-specific incidence rates in the populations studied shows a relatively large increase in the very young ages, the paramenopausal ages (40-44 to 55-59), and the elderly (table 3).

TABLE 3.—*Age-specific rates for the first and last cross sections reported for each population*

| Location and years | Age, yr | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 | 60 | 65 | 70 | 75 | 80 |
| **Incidence**[a] | | | | | | | | | | | | | |
| Denmark | | | | | | | | | | | | | |
| 1943–47 | 0.25 | 3.3 | 14.7 | 34.2 | 63.2 | 95.7 | 97.1 | 135.1 | 160.4 | 189.9 | 188.5 | 193.2 | 230.2 |
| 1963–67 | 0.43 | 4.4 | 13.4 | 39.9 | 84.8 | 123.4 | 120.3 | 129.6 | 154.5 | 183.5 | 216.0 | 257.8 | 340.1 |
| | (72%) | (35%) | (−9%) | (17%) | (34%) | (29%) | (24%) | (−2%) | (−4%) | (−3%) | (13%) | (33%) | (48%) |
| Connecticut | | | | | | | | | | | | | |
| 1935–39 | 0.78 | 5.9 | 14.5 | 35.8 | 72.5 | 93.8 | 116.1 | 172.0 | 190.0 | 214.4 | 211.2 | 247.3 | 331.7 |
| 1970–72 | 1.07 | 8.0 | 21.1 | 52.5 | 131.4 | 190.1 | 200.2 | 196.2 | 227.3 | 267.5 | 283.7 | 309.0 | 325.3 |
| | (37%) | (36%) | (46%) | (47%) | (81%) | (103%) | (72%) | (14%) | (20%) | (25%) | (34%) | (25%) | (−2%) |
| Osaka, Japan | | | | | | | | | | | | | |
| 1963–65 | 0.63 | 1.4 | 5.0 | 14.2 | 22.9 | 29.8 | 33.0 | 31.0 | 27.6 | 32.2 | 21.3 | 31.9 | 25.5 |
| 1971–75 | 0.41 | 1.9 | 7.3 | 15.6 | 24.6 | 42.2 | 37.6 | 39.2 | 35.3 | 34.7 | 33.0 | 31.0 | 32.7 |
| | (−35%) | (36%) | (46%) | (10%) | (7%) | (42%) | (14%) | (26%) | (28%) | (8%) | (55%) | (−3%) | (28%) |
| Iceland[b] | | | | | | | | | | | | | |
| 1911–19 | 2.5 | | 12.3 | | 57.0 | | 56.4 | | 48.8 | | 33.1 | | |
| 1960–69 | 6.3 | | 28.9 | | 76.2 | | 125.6 | | 127.1 | | 126.1 | | |
| | (152%) | | (135%) | | (34%) | | (123%) | | (160%) | | (281%) | | |
| **Mortality**[a] | | | | | | | | | | | | | |
| United States | | | | | | | | | | | | | |
| 1931–35 | 0.28 | 1.3 | 5.0 | 12.5 | 24.2 | 39.5 | 55.2 | 70.5 | 82.2 | 90.6 | 107.6 | 129.0 | 151.2 |
| 1966–70 | 0.19 | 1.5 | 5.6 | 13.3 | 26.4 | 44.0 | 61.4 | 75.4 | 83.0 | 90.4 | 102.5 | 117.5 | 136.3 |
| | (−32%) | (15%) | (12%) | (6%) | (9%) | (11%) | (11%) | (7%) | (1%) | (0%) | (5%) | (9%) | (−10%) |
| Canada | | | | | | | | | | | | | |
| 1931–35 | 0.13 | 1.3 | 5.2 | 12.9 | 23.3 | 41.3 | 55.5 | 71.9 | 78.6 | 89.2 | 107.0 | 123.9 | 162.0 |
| 1966–70 | 0.10 | 1.3 | 4.7 | 14.5 | 28.1 | 46.9 | 67.1 | 79.4 | 86.8 | 94.8 | 102.1 | 124.1 | 157.3 |
| | (−23%) | (0%) | (−10%) | (12%) | (21%) | (14%) | (21%) | (10%) | (10%) | (6%) | (−5%) | (0%) | (−3%) |
| England and Wales | | | | | | | | | | | | | |
| 1911–15 | 0.7 | 0.64 | 4.0 | 12.3 | 25.8 | 43.2 | 55.7 | 67.9 | 79.5 | 87.7 | 120.6 | 136.4 | 169.5 |
| 1966–70 | 0.24 | 1.57 | 5.9 | 14.2 | 30.7 | 50.2 | 68.6 | 85.7 | 94.3 | 103.8 | 113.8 | 133.9 | 165.1 |
| | (243%) | (145%) | (48%) | (15%) | (19%) | (16%) | (23%) | (36%) | (19%) | (18%) | (−6%) | (−2%) | (−3%) |
| Japan | | | | | | | | | | | | | |
| 1946–50 | 0.10 | 0.45 | 1.7 | 3.6 | 6.8 | 9.1 | 10.8 | 12.4 | 12.8 | 14.3 | 14.6 | 16.0 | 14.1 |
| 1971–75 | 0.12 | 0.54 | 2.4 | 4.5 | 8.1 | 12.0 | 15.4 | 15.4 | 14.8 | 14.2 | 14.0 | 15.9 | 20.5 |
| | (20%) | (20%) | (41%) | (25%) | (19%) | (32%) | (43%) | (24%) | (16%) | (−1%) | (−4%) | (0%) | (45%) |

[a] Differences in these rates are expressed *in parentheses* as percentages of the rates in the earliest cross section.
[b] Reported in 10-yr age groups.

Table 4 shows that the multiplicative model fit the incidence in Iceland and Osaka well and provided a marginal fit to the mortality in Canada. The other data sets were poorly fit. Table 5 shows that the three-factor multiplicative model (model II) fit the Canadian (mortality) data well. Other data sets that were marginally fit are shown in table 5. The remaining data sets (except Iceland and Osaka, on which only model I was tried) were poorly fit by model II. The Denmark incidence data were well fit by model III, whereas they were poorly fit by models I and II. This indicates that the underlying effect of age in Denmark is modulated by an interaction of age and cohort. Table 6 summarizes the extent to which various data sets were fit by the different models.

Text-figure 1 shows the age parameters obtained by fitting the models to the data sets. The largest data sets

TABLE 4.—*Chi-square values for fit of the generation-effects model as applied by Breslow and Day (2) to incidence of female breast cancer in Iceland*

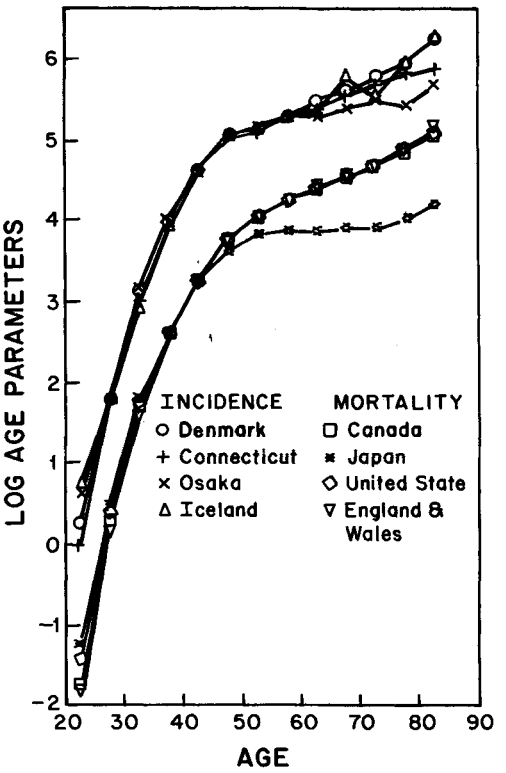
| Population | $X^2$ | Degrees of freedom | *P* |
|---|---|---|---|
| Iceland[a] | 48.97 | 54 | 0.69 |
| Osaka[a] | 9.24 | 12 | 0.60 |
| Denmark[a] | 59.888 | 36 | 0.0091 |
| Connecticut[a] | 116.32 | 72 | 0.0006 |
| Canada[b] | 88.077 | 72 | 0.095 |
| Japan[b] | 73.552 | 36 | 0.0001 |
| United States[b] | 496.00 | 72 | — |
| England and Wales[b] | 1,053.00 | 120 | — |

[a] Incidence data.
[b] Mortality data.

TABLE 5.—*Fit of model II to incidence and mortality from female breast cancer in various populations*

| Population | $X^2$ | Degrees of freedom | *P* |
|---|---|---|---|
| Connecticut[a] | 91.7 | 65 | 0.015 |
| Denmark[a] | 40.2 | 32 | 0.15 |
| Canada[b] | 59.3 | 65 | 0.60 |
| Japan[b] | 43.0 | 32 | 0.09 |

[a] Incidence data.
[b] Mortality data.

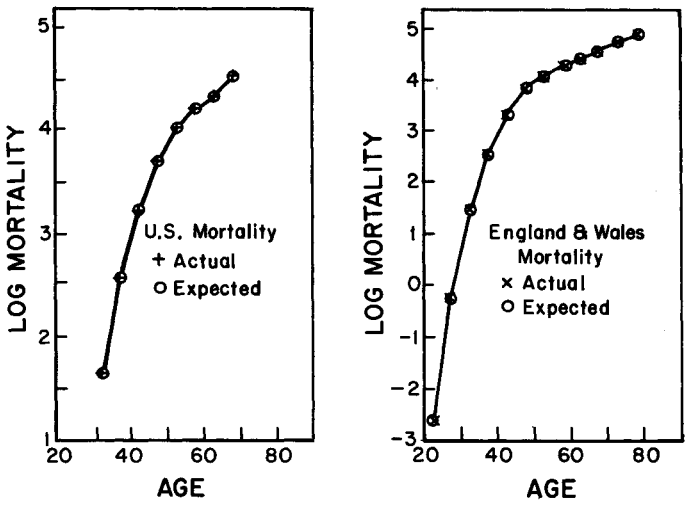



TEXT-FIGURE 1.—Extracted age curves for best-fitting model to each population. Incidence curves normalized to the 40- to 44-year-old rate in Connecticut averaged over all cross sections. Mortality curves normalized to the 40- to 44-year-old rate in the United States averaged over all cross sections.

TABLE 6.—*Extracted age parameters (×100,000) from best-fitting model to each population*

| Location[a] | Age, yr | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 20–24 | 25–29 | 30–34 | 35–39 | 40–44 | 45–49 | 50–54 | 55–59 | 60–64 | 65–69 | 70–74 | 75–79 | 80–84 | Model |
| **Incidence** | | | | | | | | | | | | | | |
| Iceland (49, 54, 0.6) | 0.99 | 2.71 | 8.59 | 22.6 | 47.7 | 74.6 | 79.5 | 92.8 | 99.0 | 153 | 121 | 183 | 249 | I |
| Denmark (22, 24, 0.6) | 0.83 | 3.77 | 14.8 | 34.9 | 65.2 | 104 | 109 | 130 | 156 | 174 | 210 | 256 | 336 | III |
| Connecticut (50, 46, 0.32) | 0.87 | 5.7 | 18.1 | 43.8 | 90.6 | 132 | 144 | 178 | 198 | 232 | 266 | 306 | 324 | VI |
| Osaka, Japan | 0.38 | 1.24 | 4.74 | 11.8 | 20.5 | 32.2 | 36.1 | 40.0 | 40.2 | 45.4 | 48.4 | 46.6 | 60.2 | I |
| **Mortality** | | | | | | | | | | | | | | |
| Canada (59, 65, 0.6) | 0.18 | 1.47 | 5.38 | 13.8 | 26.8 | 45.0 | 58.7 | 73.5 | 86.0 | 98.5 | 110 | 135 | 162 | II |
| Japan (27, 26, 0.5) | 0.07 | 0.44 | 1.70 | 3.74 | 7.10 | 10.2 | 12.6 | 13.2 | 13.1 | 13.7 | 13.7 | 15.2 | 18.2 | VI |
| United States (50, 46, 0.32) | 0.24 | 1.55 | 5.62 | 13.3 | 25.8 | 41.5 | 56.6 | 69.4 | 79.7 | 90.5 | 106 | 128 | 154 | VI |
| England and Wales (115, 86, 0.02) | 0.17 | 1.25 | 5.22 | 13.7 | 28.3 | 47.0 | 62.0 | 75.1 | 88.2 | 101 | 120 | 147 | 186 | VI |

[a] *Numbers in parentheses* are the $X^2$-value, degrees of freedom, and *P*-value, respectively.




TEXT-FIGURE 2.—Actual mortality rates for female breast cancer plotted against those predicted by model VI for the United States birth cohort of 1901 and the England and Wales cohort of 1891.

and the two least well fit by the models were the mortality data sets from the United States and from England and Wales. The most complete birth cohorts have information on 12 age groups in the England and Wales data set and on 8 age groups in the United States data set. One such cohort was chosen from each data set, and text-figure 2 is a plot of actual and expected rates for these cohorts. Text-figure 1 shows that the extracted age-incidence curves are all very similar in shape. The same figure shows that this is true for the extracted mortality curves from all the Occidental (Western) populations but that the post-menopausal Japanese mortality curve deviates below the Western paradigm.

## DISCUSSION

High-quality cancer data have been collected for several years by registries in many parts of the world. Examination of these data shows that large and striking differences exist in the incidence of, and mortality from, cancer in different parts of the world. In addition, there have been large temporal trends in cancer incidence and mortality. Careful examination of these trends provides insights into the etiology and dynamics of specific cancers. This examination is also of vital importance in understanding the impact of a changing environment on these cancers. Whereas studies on the geographic variations in cancer are abundant, the study of temporal trends has received comparatively little attention.

It is well known that the incidence rates of many chronic diseases are strongly influenced by factors that operate early in life (cohort, generation, or year of birth effects). The importance of birth cohort was first recognized by Derrick (*24*) in 1927 and first cast into mathematical terms by Kermack et al. (*25*) in 1934. The Kermack formulation is particularly simple: It postulates that the age-specific mortality (from all causes) is $M(c,t)=f(c)g(t)$, where $f(c)$ is a function of birth cohort alone and $g(t)$ is a function of age alone. This is the generation-effects model. The classic example of such an effect was considered to be tuberculosis (*27, 28*), although a recent paper (*29*) showed that tuberculosis did not fit the strict generations-effect model. More recently, the decline in mortality from tumors of the bone has been shown to be due to cohort effects in both the United States and Britain (*30*), whereas cohort effects have also been shown to be responsible for the rise in incidence of breast cancer in the female population of Iceland (*1, 2*). However, the simple generation-effects model does not hold for many cancers. Examples of such tumors are lung cancer and Hodgkin's disease (*31*). In fact, the simple generation-effects model is highly unlikely to fit the incidence of any tumors that are affected by the adult environment. In such instances, a term for the effects of year of incidence (or mortality) must be included in the analysis.

There is another, though related, reason why the careful analysis of cancer incidence and mortality data is important. That the incidence and mortality curves of cancer are strongly age-dependent has been known for a long time. This observation has led to the search for pathogenetic models for the process of carcinogenesis that could explain this fact. Whereas the multistage models of carcinogenesis were originally proposed (*32*) to explain the age-incidence curves of the so-called log/log cancers (*33*)—in which the logarithm of the age-specific incidence rate increases linearly with the logarithm of age—extensions of these models have more recently been proposed for female breast cancer (*34, 35*). The incidence curves generated by these models are then tested against the age-specific incidence rates in a population. However, the age curve for cancer is modulated not only by year-of-birth effects but also by year-of-incidence effects. At best, it is unsatisfactory to fit these models to individual cross-sectional curves (*36, 37*), and at worst it can be totally misleading (*38*). What is needed is a careful statistical analysis of the data in order to extract, to the extent possible, the "pure age-effects curve" and to isolate it from the temporal trends in the disease. One of the objectives of this paper is to extract such adjusted age curves and compare these across all the populations studied. It is these extracted age parameters that should be used to test pathogenetic models for the effect of age on cancer incidence. In this approach, all the available data are used.

Previous attempts at disentangling the effects of birth cohort, year of incidence, and age on cancer incidence have used either graphic methods or quantitative methodology such as cohort slopes (*31, 39*) and analysis of variance (*40, 41*).

The point of departure for this paper is the work of Breslow and Day on breast cancer incidence in Iceland. They treated the analysis of incidence data as a problem in the analysis of contingency tables with missing observations and observed that the generation-effects model explained the data well. Inasmuch as cancer is a rare event, the hypothesis that the events in individual cells of the contingency table (table 1) are

independently Poisson distributed is reasonable. The choice of age and birth cohort as axes of classification and our assumptions about the parameters of the models reflect our belief that age and birth cohort are the most important determinants of cancer incidence and that other factors may be considered to be interactions between these two factors. With the more complicated models considered by us, exact interpretations of the parameters estimated are difficult to give. However, our main interest is a study of the age parameters ($a_i$ in each of the models) to compare these "extracted age curves" between populations. Hence it is reasonable to use a stringent test of goodness of fit and to use the age parameters from the best-fitting model for these comparisons. With the large populations in each of the cells (millions in the mortality data from the United States and England), the chi-squared criterion used by us is very stringent. It should come as no surprise that different models are necessary to describe the data sets. These models are not mathematical formulations of underlying pathogenetic processes for breast cancer, but are statistical models designed only to describe the data to which they are applied. The temporal experience of each population is probably different, and the statistical models merely reflect this fact. When more than one model fit the data well, we chose the model with the larger number of degrees of freedom (or performed a likelihood ratio test if appropriate). However, we found that in such instances the age parameters were internally consistent; i.e., the age parameters from two models that fit the same data set equally well were very similar.

The simple multiplicative model that postulates the independence of age and birth cohort effect (model I) fit only the incidence data in Iceland and Osaka. Possibly, this finding indicates no fundamental difference between Iceland and Osaka and the other data sets but simply reflects the fact that we have only three periods of observation in Osaka and the population at risk in each cell in Iceland is small. Alternatively, this difference may be real as discussed below. An examination of the cohort parameters shows that a striking increase occurred in the incidence of breast cancer in these two populations (table 7).

The (unnormalized) age parameters extracted from the Osaka incidence data are much smaller than the age parameters from the Occidental incidence data. However, the general shape of the extracted age curves is very similar in all the countries examined. In particular, the extracted age curve does not decrease after menopause, as has been reported in the literature (*42*) for individual cross sections (text-figs. 1, 3). Our analysis shows that this fall in (cross-sectional) rates after menopause is due to a strong cohort effect. Whereas cross-sectional curves have different shapes in Osaka and the Occidental countries (text-figs. 3, 4), the smoothed curves extracted from the data are very similar in shape (text-fig. 1). Bjarnason et al. (*1*) demonstrated in 1974 that the shape of the cross-sectional incidence curves for female breast cancer in Iceland evolved from one very similar to the shape of the curve in a low-risk population such as in Japan to one that resembles the shape in a high-risk population such as in Connecticut, and that this change in shape was due to a strong cohort effect. Our analysis suggests that a similar phenomenon may be occurring in Osaka. Table 7 shows an approximately fourfold increase in breast cancer incidence in both Iceland and Osaka from the cohorts of 1880 to 1940. The absence of such strong cohort effects in the other Western populations studied is consistent with the idea that cohort influences are most prominent in a population in which breast cancer incidence is changing from a low-risk profile to a high-risk profile. In a population already at high risk, secular trends in breast cancer incidence are caused by more subtle interactions of age and cohort.

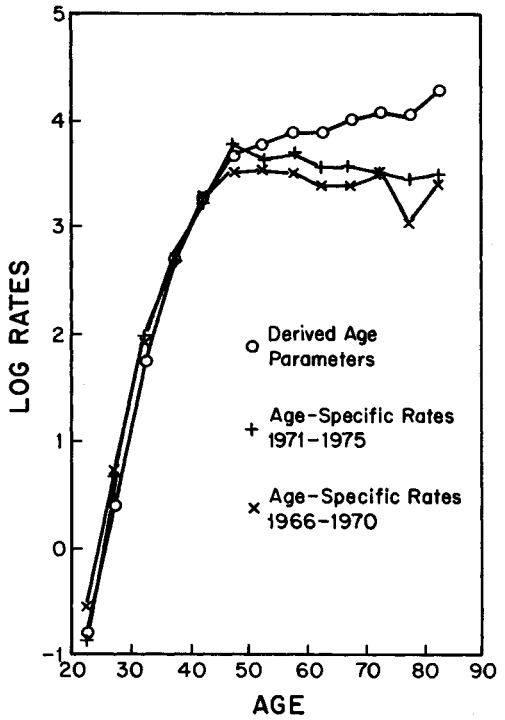
DeLisi (*35*) proposed a multistage model for female breast cancer and obtained good fits to both Western and Japanese age incidence curves. However, some of his conclusions should be reexamined because his data were not adjusted for temporal trend. In particular, he discussed the postmenopausal fall in rates in Japan in the light of his model and suggested that ". . .a lower overall incidence and a higher interstage transition rate in Japan would follow if the at risk fraction is smaller than in Scandinavia, but has a greater predisposition than the corresponding at risk fraction in Scandinavia toward developing breast cancer."

Our analysis in this paper shows that this conclusion is based on an artifact of the data. Moreover, inasmuch as individual cross-sectional curves are different from the adjusted age curves (text-figs. 3, 4), the parameters estimated by DeLisi for his model are probably strongly influenced by his choice of cross sections, and the use of "adjusted" curves may lead to different estimates of the parameters. Also, DeLisi used the density, rather than the hazard function, derived from his model to fit the data. The dangers implicit in this method have been discussed in (*43, 44*).

The extracted age curves for mortality are similar in shape among all the Western countries. The premenopausal extracted mortality curve from the Japanese data is similar in shape to the premenopausal (ex-

TABLE 7.—*Extracted cohort parameters for simple multiplicative model (model I) applied to incidence of female breast cancer in Iceland and Osaka, Japan*

| Location | Midyear of cohort followed by extracted cohort parameters applied to incidence of female breast cancer | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Iceland | | 1845, | 1855, | 1865, | 1875, | 1885, | 1895, | 1905, | 1915, | 1925, | 1935, | 1945, | | | |
| | | 0.24 | 0.33 | 0.53 | 0.85 | 0.95 | 1.02 | 1.20 | 1.50 | 1.48 | 2.29 | 4.17 | | | |
| Osaka, Japan | 1881, | 1886, | 1891, | 1896, | 1901, | 1906, | 1911, | 1916, | 1921, | 1926, | 1931, | 1936, | 1941, | 1946, | 1951, |
| | 0.42 | 0.57 | 0.48 | 0.69 | 0.67 | 0.75 | 0.86 | 0.95 | 1.05 | 1.23 | 1.22 | 1.36 | 1.58 | 1.54 | 1.08 |

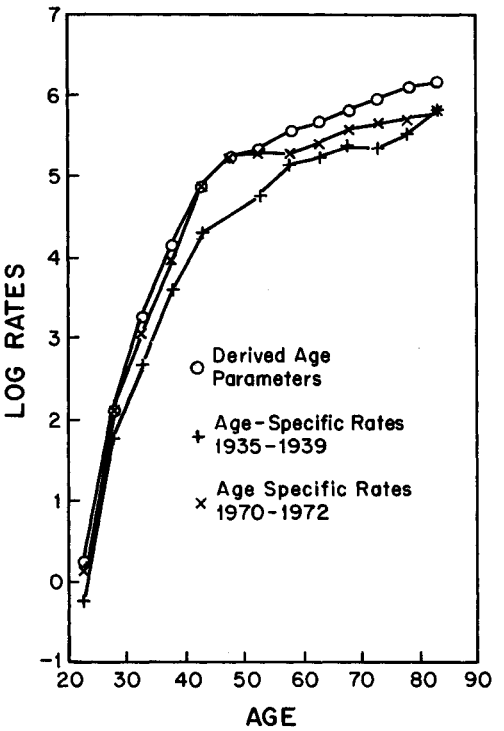



TEXT-FIGURE 3.—Age-specific rates and adjusted age curve for incidence of female breast cancer in Osaka. Adjusted age curve was normalized to the 40- to 44-year-old rate in 1971-75.

tracted) curves from the Occidental countries. After the menopause the extracted curve for Japan deviates below the curve for the Western populations. However, there is no indication of an actual fall after menopause, as has been reported for mortality rates (*42*). On the basis of graphic techniques, Armstrong (*39*) drew qualitative conclusions on the influence of cohort and year of event on the incidence and mortality from breast cancer in the United States and in England and Wales. However, his methods are incapable of yielding an adjusted age curve for cancer incidence. Our analysis confirmed Armstrong's conclusion that both cohort and cross-sectional effects are important in determining temporal trends in breast cancer incidence and mortality in the United States and in England and Wales; however, subtle interactions between age and year of event are also involved.

The "hook," first described by Clemmesen in 1948 (*45*), is an important feature of breast cancer incidence data in Denmark. Clemmesen noticed that the incidence curve of breast cancer in females in Denmark rose steadily until about the age of menopause; there was then a leveling off (the hook), followed by an increase in rates into old age. MacMahon (*46*) suggested that this hook was an artifact in the cross-sectional data due to a cohort effect. Our analysis reveals the presence of the hook in all incidence data sets examined by us. The menopausal hook is actually seen to be a "point of inflection" in the extracted age curve, i.e., a point where the curve changes its shape from being concave downward to being concave upward. Such an inflection point is seen in all the extracted mortality curves as well, though it occurs about 10 years later than in the incidence curves. However, the hook is much less prominent in these adjusted age curves than in the original cross-sectional data.

Our analysis discussed in this paper has allowed us to extract complete age curves for female breast cancer incidence and mortality in various geographic locations. These curves measure the effect of age in breast cancer incidence and mortality after adjustment for birth cohort and year of incidence (mortality). Although the shapes of cross-sectional curves between populations or between time periods in the same population are dissimilar, the shapes of all the extracted incidence age curves are remarkably similar, notwithstanding the fact that breast tumors in Japanese women seem to differ histologically from those in Western women (*47-49*). This perhaps reflects the fact that, whereas the basic biologic mechanisms for tumorigenesis are similar all over the world, Japanese women are somehow less susceptible or react differently, which leads to lower rates and a different histologic type. In contrast to the extracted incidence curves, the slope of the extracted mortality curve after menopause in Japanese women is smaller than the slopes for those curves in Western women. It is interesting to note in this regard that survival rates of breast cancer patients in Japan were better than those of breast cancer patients in Scotland and the United States. This difference in rates persisted after adjustment for age, extent of disease, histologic type, and treatment modality (*47, 49*).




TEXT-FIGURE 4.—Age-specific rates and adjusted age curve for incidence of female breast cancer in Connecticut. Adjusted age curve was normalized to the 40- to 44-year-old rate in 1970-72.

The methods presented in this paper have allowed us to isolate age effects for female breast cancer. A more stringent test of the models used would be provided by tumors that do not fit the strict generation-effects model of Kermack but do show strong temporal trends (e.g., Hodgkin's disease and lung cancer). The operation of temporal trends in these tumors is easy to perceive. However, the fitting of statistical models as done here should prove more of a challenge. A similar analysis for these tumors would be of interest.

In 1969, Cook et al. (*36*) undertook a comprehensive test of the Armitage-Doll approximation to the multistage model for carcinogenesis. They found that more than half the tumors they considered showed deviations below the prediction made by the model. They suggested that this could be due to a strong cohort effect. More recently, one of us (*43, 44*) has pointed out that this deviation could occur if the transition rates in the multistage model were not small enough. Cook et al. used individual cross sections for their analysis. The methods of analyses presented in this paper should distinguish between the two possibilities mentioned above.

## STATISTICAL APPENDIX

### The Statistic $X^2$

$X^2=\Sigma(E-O)^2/E$, where $E$ is the expected number of events, $O$ is the observed number of events, and the sum is taken over all the (nonempty) cells of the contingency table. With the models considered by us, one can easily see that maximum likelihood estimates of the parameters are identical under the Poisson and multinomial sampling schemes. As usual, let $R^n$ denote $n$-dimensional Euclidean space and let ${}_nP\subset R^n$ be defined as $P=\{(P_1,P_2,\ldots,P_n)\epsilon R^n \mid P_i\geqslant o$ and $\sum_{i=1}^{n} P_i=1\}$. Let $M(N;P_1,P_2\ldots,P_n)$ be a multinomial distribution, let $S\subset R^k$ be an open set, and let $f:S\rightarrow P$ be a differentiable function [i.e., each $P_i$ depends differentiably on the parameters $(x_1,x_2\ldots,x_k)$ in $S$]. If $k<n$, and if the matrix $(\partial P_i/\partial x_j)_{1\leq i\leq n,\ 1\leq j\leq k}$ has maximal rank (i.e., it has rank $k$), then it is well known (*50*) that the statistic $X^2$ is asymptotically chi-squared distributed with $(n-k-1)$ degrees of freedom (providing the true value of the parameters is in $S$). If, however, the matrix has rank $m<k$, the rank theorem of advanced calculus guarantees the existence (at least locally) of $m$ independent parameters, and it follows that $X^2$ is still asymptotically chi-squared distributed with $(n-m-1)$ degrees of freedom. It is then easily seen that our statistic $X^2$ is asymptotically chi-squared distributed with the appropriate number of degrees of freedom as asserted in the text of the paper.

### Maximum Likelihood Estimation

Maximum likelihood estimates of the parameters were obtained by Newton-Raphson iteration. These estimates are not unique; they depend on the initialization. In each instance, the parameters were initialized identically. For comparison, the age parameters were normalized by multiplication by a suitable non-zero constant. The appropriate normalizations are described in the text-figure legends.

## REFERENCES


(*1*) BJARNASON O, DAY N, SNAEDAL G, et al: The effect of year of birth on the breast cancer age-incidence curve in Iceland. Int J Cancer 13:689-696, 1974

(*2*) BRESLOW NE, DAY NE: Indirect standardization and multiplicative models for rates, with reference to the age adjustment of cancer incidence and relative frequency data. J Chronic Dis 28:289-303, 1975

(*3*) EISENBERG H: Cancer in Connecticut, Incidence and Rates 1935-1962. Hartford: Department of Health, 1966

(*4*) EISENBERG H, CAMPBELL PC, FLANNERY JT: Cancer in Connecticut, Incidence Characteristics 1935-1962. Hartford: Department of Health, 1967

(*5*) EISENBERG H: Cancer in Connecticut 1963. Hartford: Department of Health, 1966

(*6*) ———: Cancer in Connecticut 1964. Hartford: Department of Health, 1967

(*7*) CHRISTINE BW, FLANNERY JT, SULLIVAN PD, et al: Cancer in Connecticut 1965. Hartford: Department of Health, 1969

(*8*) CHRISTINE BW, FLANNERY JT, SULLIVAN PD: Cancer in Connecticut 1966-1968. Hartford: Department of Health, 1971

(*9*) ———: Cancer in Connecticut 1969. Conn Health Bull 86:103, 1972

(*10*) ———: Cancer in Connecticut 1970. Conn Health Bull 86:345, 1972

(*11*) CHRISTINE B: Cancer in Connecticut 1972. Conn Health Bull 88: 127, 1974

(*12*) CLEMMESEN J: Statistical studies in malignant neoplasms. II. Basic tables, Denmark 1943-1957. Acta Pathol Microbiol Scand 174:2-3, 1965

(*13*) ———: Statistical studies in malignant neoplasms. III. Basic tables 1958-1962. Acta Pathol Microbiol Scand 209:3 and 78, 1969

(*14*) ———: Statistical studies in malignant neoplasms. IV. Denmark, 1943-1967. Acta Pathol Microbiol Scand 247:119 and 194, 1974

(*15*) U.S. Department Of Commerce, Bureau of the Census: Mortality Statistics 1931-1936. Washington, D.C.: U.S. Govt Print Off, 1935-1938

(*16*) ———: Vital Statistics of the United States 1937-1943. Washington, D.C.: U.S. Govt Print Off

(*17*) Federal Security Agency, Public Health Service: Vital Statistics of the United States 1944-1949. Washington, D.C.: U.S. Govt Print Off

(*18*) U.S. Department of Health, Education, and Welfare: Vital Statistics of the United States 1950-1970. Washington, D.C.: U.S. Govt Print Off

(*19*) The Registrar General's Statistical Review of England and Wales: Office of Population Censuses and Surveys. Cancer Mortality in England and Wales 1911-70. Studies on Medical and Population Subjects, No. 29. London: Her Majesty's Stationery Office, 1975

(*20*) Dominion Bureau of Statistics: Causes of Death: Canada. Ottawa: Government of Canada, 1975

(*21*) Health and Welfare Statistics Division, Ministry of Health and Welfare: Vital Statistics Japan Series, vol 2, 1963-1974. Tokyo: Ministry of Health and Welfare, 1964, 1975

(*22*) SEGI M, KURIHARA M, MATSUYAMA T: Cancer Mortality in Japan 1899-1962. Tohoku University School of Medicine, Department of Public Health, 1965

(*23*) Health and Welfare Statistics Division, Ministry of Health and Welfare: Vital Statistics Japan Series, vol 2, 1975. Tokyo: Ministry of Health and Welfare, 1976

(*24*) DERRICK VP: Errors in the population statistics, and changes in mortality. J Inst Actuaries 58:117-146, 1927

(25) KERMACK W, MCKENDRICK A, MCKINLAY P: Death rates in Great Britain and Sweden: Expression of specific mortality rates as products of two factors, and some consequences thereof. J Hyg (Camb) 34:433-457, 1934
(26) SEGI M: Cancer Mortality for Selected Sites in 24 Countries (1950-57). Sendai: Department of Public Health, Tohoku University School of Medicine, 1960
(27) ANDVORD KF: What can we learn by studying tuberculosis by generations? Nor Mag Laegevidensk 91:642-660, 1930
(28) FROST WH: The age selection of mortality from tuberculosis in successive decades. Am J Hyg 30:91-96, 1939
(29) STEVENS RG, LEE JA: Tuberculosis: Generation effects and chemotherapy. Am J Epidemiol 107:120-126, 1978
(30) GARDNER MJ, BARKER DJ: Mortality from malignant tumors of bone and Paget's disease in the United States and in England and Wales. Int J Epidemiol 7:121-130, 1978
(31) SZEKELY DR, LEE JA: Trends of fatal Hodgkin's disease as a consequence of immune experience. Cancer Res 37:4568-4571, 1977
(32) ARMITAGE P, DOLL R: The age distribution of cancer and a multistage theory of carcinogenesis. Br J Cancer 8:1-12, 1954
(33) PETO R: Epidemiology, Multistage Models and Short-Term Mutagenicity Tests in Origins of Human Cancer, Cold Spring Harbor Conferences on Cell Proliferation (Hiatt HH, Watson JD, Winsten JA, eds), vol 4. New York: Cold Spring Laboratory, 1977, pp 1403-1428
(34) LEE JA, CHIN PG, KUKULL WA, et al: Relationship of age to incidence of breast cancer in young women. J Natl Cancer Inst 57:753-756, 1976
(35) DELISI C: The age incidence of female breast cancer—simple models and analysis of epidemiological patterns. Math Biosci 37:245-266, 1977
(36) COOK PJ, FELLINGHAM SA, DOLL R: A mathematical model for the age distribution of cancer in man. Int J Cancer 4:93-112, 1969
(37) CASE RA: Cohort analysis of mortality rates as an historical or a narrative technique. Br J Prev Soc Med 10:159-171, 1956
(38) LEE JA: Mathematical models of age and ultraviolet effects on the incidence of skin cancer among whites in the United States. Am J Epidemiol 107:259, 1978
(39) ARMSTRONG B: Recent trends in breast cancer incidence and mortality in relation to changes in possible risk factors. Int J Cancer 17:204-211, 1976
(40) SACHER GA: Analysis of life tables with secular terms. *In* The Biology of Aging (Strehler B, ed), No. 6. Washington, D.C.: Am Inst Biol Sci, 1960, pp 253-257
(41) BARRETT JC: Age, time and cohort factors in mortality from cancer of the cervix. J Hyg (Camb) 71:253-259, 1973
(42) THOMAS DB, LILIENFELD AM: Geographic, reproductive and sociobiological factors. *In* Risk Factors in Breast Cancer (Stoll BA, ed). Chicago: Year Book Medical Publishers, 1976, pp 25-53
(43) MOOLGAVKAR SH: The multistage theory of carcinogenesis. Int J Cancer 20:730, 1977
(44) ———: The multistage theory of carcinogenesis and the age distribution of cancer in man. JNCI 61:49-52, 1978
(45) CLEMMESEN J: Results from statistical research (symposium). Br J Radiol 21:583-590, 1948
(46) MACMAHON B: Breast cancer at menopausal ages: An explanation of observed incidence changes. Cancer 10:1037-1043, 1957
(47) MORRISON AS, LOWE CR, MACMAHON B, et al: Survival of breast cancer patients related to incidence risk factors. Int J Cancer 9:470-476, 1972
(48) MORRISON AS, BLACK MM, LOWE CR, et al: Some international differences in histology and survival in breast cancer. Int J Cancer 11:261-267, 1973
(49) MORRISON AS, LOWE CR, MACMAHON B, et al: Some international differences in treatment and survival in breast cancer. Int J Cancer 18:269-273, 1976
(50) BISHOP YM, FEINBERG SE, HOLLAND PW: Discrete Multivariate Analysis: Theory and Practice. Cambridge: The MIT Press, 1975, pp 513-530