

Comparison of Age Distribution Patterns for Different Histopathologic Types of Breast Carcinoma

William F. Anderson,¹ Ruth M. Pfeiffer,¹ Graça M. Dores,³ and Mark E. Sherman²

¹Biostatistics Branch and ²Hormonal and Reproductive Epidemiology Branch, Division of Cancer Epidemiology and Genetics, and ³Office of Preventive Oncology, Division of Cancer Prevention, National Cancer Institute, NIH, Department of Health and Human Services, Bethesda, Maryland

Abstract

Background: Historically, female breast carcinoma has been viewed as an etiologically homogeneous disease associated with rapidly increasing incidence rates until age 50 years, followed by a slower rate of increase among older women. More recent studies, however, have shown distinct age incidence patterns for female breast cancer when stratified by estrogen receptor (ER) expression and/or histopathologic subtypes, suggesting etiologic heterogeneity.

Materials and Methods: To determine if different age incidence patterns reflect etiologic heterogeneity (more than one breast cancer type within the general breast carcinoma), we applied "smoothed" age histograms at diagnosis (density plots) and a two-component statistical mixture model to all breast carcinoma cases ($n = 270,124$) in the Surveillance, Epidemiology, and End Results Program of the National Cancer Institute. These overall patterns were then reeval-

uated according to histopathologic type, race, and ER expression.

Results: A bimodal age distribution at diagnosis provided a better fit to the data than a single density for all breast carcinoma populations, except for medullary carcinoma. Medullary carcinomas showed a single age distribution at diagnosis irrespective of race and/or ER expression.

Conclusions: Distinct age-specific incidence patterns reflected bimodal breast cancer populations for breast carcinoma overall as well as for histopathologic subtypes, race, and ER expression. The one exception was medullary carcinoma. Of note, medullary carcinomas are rare tumors, which are associated with germ-line mutations in the *BRCA1* gene. These descriptive and model-based results support emerging molecular data, suggesting two main types of breast carcinoma in the overall breast cancer population. (Cancer Epidemiol Biomarkers Prev 2006;15(10):1899–905)

Introduction

Breast carcinomas among women are extremely diverse in clinical and histopathologic features, suggesting that these tumors also vary in etiology. Descriptive studies support this view. For example, data from the Danish Breast Cancer Cooperative Group and the Surveillance, Epidemiology, and End Results (SEER) Program posit that the classically recognized inflection point in age-specific breast cancer incidence rates overall around menopause ("Clemmesen's Hook"; ref. 1) may reflect the superimposition of two different rate curves (2-5), corresponding to estrogen receptor (ER)-negative and ER-positive tumors. Rates for ER-negative tumors increase rapidly until age 50 years, then flatten or decrease, whereas ER-positive tumors increase rapidly until age 50 years, then continue to increase at a slower pace.

In addition to the *two* age incidence patterns for ER expression, SEER data show *three* incidence rate curves according to histopathologic subtype (6-8). First, there is a rapid increase in incidence until age 50 years, then a slower increase for ductal, tubular, and lobular carcinomas, as for ER-positive tumors. Second, there is a rapid increase until age 50 years followed by leveling for medullary and inflammatory carcinomas, similar to curves for ER-negative tumors. Finally, there is a steady increase with aging for papillary and mucinous carcinomas, as is seen for epithelial tumors such as colorectal carcinoma (9, 10).

To explore further the age incidence patterns for different histopathologic types, we applied descriptive techniques (age-density plots) and a model-based analysis using two-component statistical mixture and single-density models.

Materials and Methods

We used the SEER Cancer Incidence Public-Use Database (November 2004 submission) to analyze invasive breast carcinoma among women, diagnosed during the years 1992 to 2002 (11). A total of 270,124 breast cancer cases were obtained from 13 SEER registries, including the Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, San Francisco-Oakland, Seattle-Puget Sound, Utah, Los Angeles, San Jose-Monterey, rural Georgia, and Alaskan Native Tumor Registries. Data were stratified by histopathologic subtype, using codes from the International Classification of Diseases for Oncology, 3rd edition (ICD-O-3), and the WHO (12-14): duct carcinoma of no special type (duct NST; ICD-O-3 code 8500), tubular carcinoma (ICD-O-3 code 8211), lobular carcinoma (ICD-O-3 code 8520), medullary carcinoma (ICD-O-3 code 8510), inflammatory carcinoma (ICD-O-3 code 8530), papillary carcinoma (ICD-O-3 codes 8050, 8260, and 8503), and mucinous carcinoma (ICD-O-3 code 8480); other or unknown included all other ICD-O-3 codes.

Racial groups included White, Black, and Asian or Pacific Islanders. SEER did not record method of ER assay although immunohistochemical staining was probably used during the time period for this study (1992-2002). There was no centralized laboratory for hormone receptor expression for SEER; thus, each registry recorded hormone receptor expression as positive, negative, missing, borderline, or unknown. We combined missing, borderline, and unknown data into one group, designated as other or unknown.

Received 3/13/06; revised 7/10/06; accepted 7/18/06.

Grant support: Intramural Research Program of the NIH/National Cancer Institute.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: William F. Anderson, Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Department of Health and Human Services, EPS, Room 8036, 6120 Executive Boulevard, Bethesda, MD 20892-7244. Phone: 301-594-9125; Fax: 301-402-0081. E-mail: wanderso@mail.nih.gov

Copyright © 2006 American Association for Cancer Research.

doi:10.1158/1055-9965.EPI-06-0191

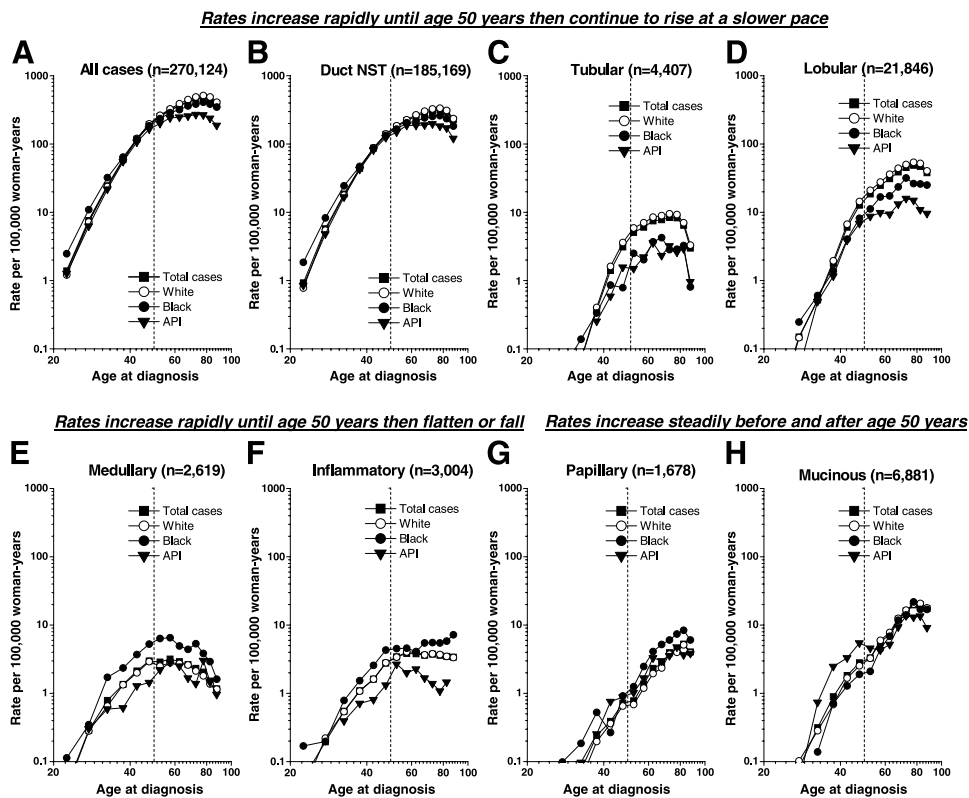


Figure 1. Age-specific incidence rates in SEER 13 Registry Database among female breast carcinoma cases collected during the years 1992 to 2002. Each plot includes a reference line at age 50 years and one histopathologic subtype with four superimposed charts: (i) total breast carcinoma cases for each histopathologic subtype; (ii) breast carcinoma cases defined by White race; (iii) breast carcinoma cases defined by Black race; and (iv) breast carcinoma cases defined by Asian or Pacific Islander (API) race.

Age-Specific Incidence Rates and Age Distribution. Age-standardized incidence rates (2000 U.S. standard population) were calculated using SEER*Stat 6.1.4⁴ and expressed per 100,000 woman-years. Age-specific incidence rates were charted for each histopathologic subtype for all women and by race (Fig. 1). Rates in Fig. 1 were plotted on a log-log scale, as originally described by Armitage and Doll (9, 10). We also approximated age incidence patterns with histograms using nonparametric probability density function curves in Figs. 2 and 3, as previously described (4, 15). In brief, the probability density function reflected a smoothed age distribution of cases at the time of primary breast carcinoma diagnosis. These density plots or curves are presented for each histopathologic subtype overall and by race (Fig. 2) as well as for each histopathologic subtype and ER expression (Fig. 3).

Statistical Model. We used a two-component mixture model to determine the probability density function for an early-onset versus a late-onset breast cancer population (or density; ref. 16). The probability density is defined as the derivative of a cumulative probability distribution function. To remove skewness, we used a power (or Box-Cox) transformation, $y = (x^\lambda - 1)/\lambda$ for $\lambda \neq 0$ and $y = \log(x)$ for $\lambda = 0$, where x denotes the age at diagnosis.

The probability density function g of the transformed age at diagnosis (in years) is then given by the mixture model $g(y;\theta) = f(y;\alpha_0)(1 - p) + f(y;\alpha_1)p$. In the mixture model, we interpreted $f(y;\alpha_0)$ to be the probability density reflecting the early age at diagnosis population, whereas $f(y;\alpha_1)$ was the probability density corresponding to the late age of onset population. The mixing probability p represents the proportion of women within the breast carcinoma population with early age distribution. All the parameters in the model are denoted by θ , and include the parameters of the component densities, the mixing proportion, and the parameter of the power transformation i.e., $\theta = (\alpha_0, \alpha_1, p, \lambda)$.

An extension of the two-component mixture model let p depend on covariates such as race or ER expression through a logistic regression function. We employed two different parameterizations for the component densities $f(y;\alpha)$: normal probability densities $f = \phi(y;\mu, \sigma)$ with mean μ and SD σ , and semi-nonparametric densities, which multiply the normal density with a polynomial component, allowing for skewness and heavier tails than the normal density (17). We chose a polynomial of degree one, yielding $f(y;\alpha) = \phi(y;\mu, \sigma)(\alpha_0 + \alpha_1 y)^2$, properly standardized.

To assess if the age distribution patterns were truly heterogeneous, we compared the fit of a mixture model with two components to that of a single density using the Akaike information criterion (AIC), which penalizes the log-likelihood by the number of parameters in the model (18, 19). Different parameterizations of the mixture model were compared using likelihood ratio test statistics, as these models are nested. Parameters in all models were estimated via a maximum likelihood procedure implemented in SAS 9.0 (SAS Institute, Cary, NC).

Results

Descriptive Statistics (Table 1A and B). The SEER 13 Registry Database collected information for 270,124 invasive female breast carcinoma cases, diagnosed during the years 1992 to 2002. Similar to our previous study (8), infiltrating duct NST and lobular breast carcinomas were the most common histopathologic subtypes, accounting for 68.5% and 8.1%, respectively. Median age at diagnosis ranged from 51 years for medullary breast cancer to 71 years for mucinous breast carcinoma. The relative risk (RR, expressed as a rate ratio) of breast cancer for Black compared with White women was greatest for medullary [RR, 2.0; 95% confidence interval (95% CI), 1.8-2.2], papillary (RR, 1.8; 95% CI, 1.6-2.1), and inflammatory (RR, 1.4; 95% CI, 1.3-1.6) histologic subtypes. RR for Asian Pacific Islanders compared with White women was greatest for papillary breast carcinomas (RR, 1.2; 95% CI, 0.9-1.6). ER expression differed by histopathologic type. The RR of

⁴ <http://seer.cancer.gov/>.

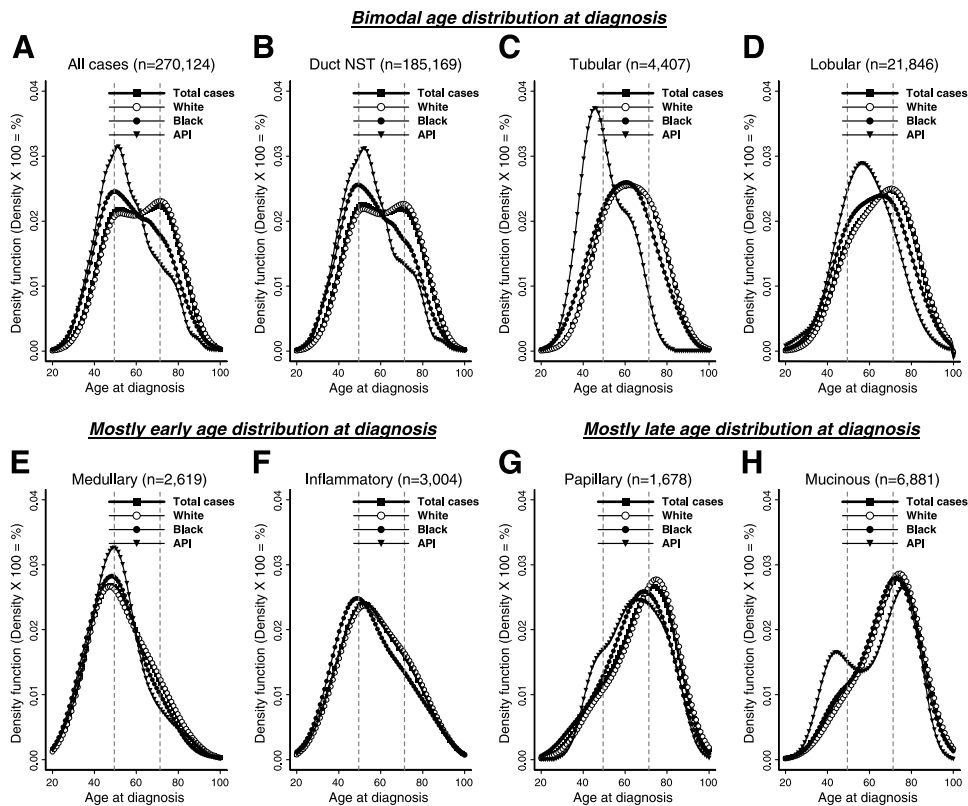


Figure 2. Age at diagnosis-density plots (or smoothed histograms) of female breast cancer diagnosed in SEER 13 Registry Database (1992-2002), stratified by race. Each plot includes a reference line at age 50 years, a reference line at age 70 years, and one histopathologic subtype with four superimposed charts: (i) total breast carcinoma cases for each histopathologic subtype; (ii) breast carcinoma cases defined by White race; (iii) breast carcinoma cases defined by Black race; and (iv) breast carcinoma cases defined by Asian or Pacific Islander race.

ER-negative compared with ER-positive breast cancer ranged from 0.1 for mucinous type to 3.8 for medullary breast carcinoma.

Age-Specific Incidence Rates Curves. Age-specific incidence rates for all breast cases combined ($n = 270,124$) and for each histopathologic type were stratified by White, Black, and

Asian or Pacific Islander races. Similar to total rates for all breast cases combined (Fig. 1A), total rates for duct NST, tubular, and lobular breast carcinomas increased rapidly until age 50 years and then continued to increase at a slower pace (Fig. 1B-D). Rates among White women were generally similar to total rates, which is not surprising given that the White race accounted for the majority of breast cancer cases in SEER.

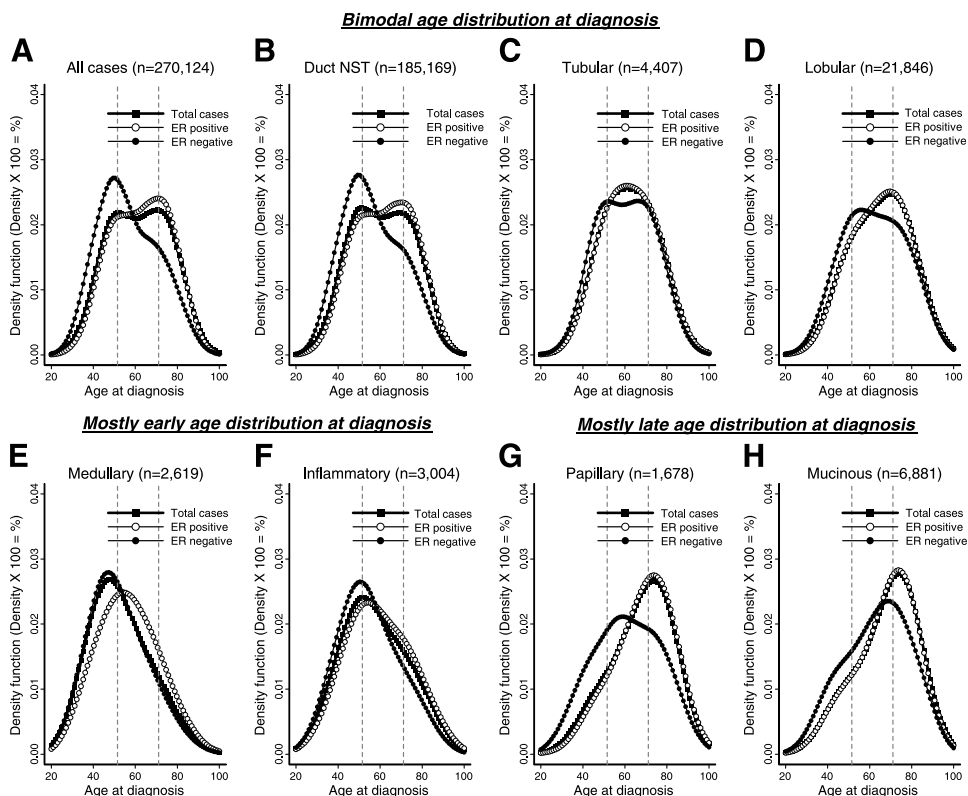


Figure 3. Age at diagnosis-density plots (or smoothed histograms) of female breast cancer diagnosed in SEER 13 Registry Database (1992-2002), stratified by ER-positive and ER-negative expression. Each plot includes a reference line at age 50 years, a reference line at age 70 years, and one histopathologic subtype with three superimposed charts: (i) total breast carcinoma cases for each histopathologic subtype; (ii) breast carcinoma cases defined ER-positive expression; and (iii) breast carcinoma cases defined by ER-negative expression.

Downloaded from <http://aacrjournals.org/cebp/article-pdf/15/10/1899/2265160/1899.pdf> by guest on 06 November 2024

In contrast to rates among White women, rates among Black women rose more slowly after age 50 years, whereas rates among Asian or Pacific Islander women plateaued or decreased after age 50 years for ductal, tubular, and lobular carcinomas.

Rates for medullary and inflammatory breast carcinomas increased until age 50 years, and then flattened or decreased, irrespective of race (Fig. 1E-F). Rates for papillary and

mucinous breast carcinomas increased steadily with age, irrespective of race (Fig. 1G-H).

Age Distribution. The age distribution at diagnosis for all breast cases combined ($n = 270,124$) showed a bimodal pattern with incidence peaks (or modes) near ages 50 and 70 years (Figs. 2A and 3A). In addition, total cases for every histopathologic type except for medullary carcinoma were bimodally

Table 1. Descriptive statistics for standard tumor characteristics in SEER 13 Registry Database among female breast cancer cases collected during 1992 to 2002

Table 1A:

| | All breast cancer cases | | | Duct NST | | | Tubular | | | Lobular | | |
|---------------------------------|-------------------------|-------------------|--------------------------|--------------|-------|-----------------|-------------|------|------------------|-------------|------|------------------|
| Sample size | 270,124 | | | 185,169 | | | 4,407 | | | 21,846 | | |
| % Total cases | 100.0 | | | 68.5 | | | 1.6 | | | 8.1 | | |
| Median age (y) | 62.0 | | | 61.0 | | | 62.0 | | | 66.0 | | |
| Median tumor size (centimeters) | 1.6 | | | 1.6 | | | 0.8 | | | 2.0 | | |
| Rate (SE) | 132.4 (0.26) | | | 91.0 (0.21) | | | 2.19 (0.03) | | | 10.7 (0.07) | | |
| Variable | ¹ N (%) | ² Rate | ³ RR (95% CI) | N (%) | Rate | RR (95% CI) | N (%) | Rate | RR (95% CI) | N (%) | Rate | RR (95% CI) |
| Demographics | | | | | | | | | | | | |
| Age (y) | | | | | | | | | | | | |
| <50 | 63,118 (23) | 42.5 | 1.0 | 46,073 (25) | 31.0 | 1.0 | 754 (17) | 0.5 | 1.0 | 3,249 (15) | 2.2 | 1.0 |
| 50-59 | 58,524 (22) | 280.0 | 6.6 (6.5-6.7) | 41,154 (22) | 196.9 | 6.4 (6.3-6.4) | 1,147 (26) | 5.5 | 10.6 (9.7-11.6) | 4,445 (20) | 21.3 | 9.6 (9.1-10.0) |
| 60-69 | 57,866 (21) | 390.8 | 9.2 (9.1-9.3) | 39,614 (21) | 267.6 | 8.6 (8.5-8.8) | 1,125 (26) | 7.6 | 14.7 (13.4-16.1) | 5,144 (24) | 34.7 | 15.6 (14.9-16.3) |
| 70-79 | 57,269 (21) | 470.4 | 11.1 (10.9-11.2) | 37,911 (20) | 311.4 | 10.1 (9.9-10.2) | 1,010 (23) | 8.3 | 16.0 (14.6-17.6) | 5,694 (26) | 46.8 | 21.1 (20.2-22.0) |
| 80+ | 33,347 (12) | 430.7 | 10.1 (10.0-10.3) | 20,417 (11) | 264.3 | 8.5 (8.4-8.7) | 371 (8) | 4.9 | 9.4 (8.3-10.6) | 3,314 (15) | 42.9 | 19.3 (18.4-20.2) |
| Race | | | | | | | | | | | | |
| White | 225,187 (83) | 138.3 | 1.0 | 153,090 (83) | 94.4 | 1.0 | 4,022 (91) | 2.5 | 1.0 | 19,636 (90) | 12.0 | 1.0 |
| Black | 22,987 (9) | 120.2 | 0.9 (0.9-0.9) | 15,844 (9) | 82.1 | 0.9 (0.9-0.9) | 179 (4) | 0.9 | 0.4 (0.3-0.4) | 1,219 (6) | 6.7 | 0.6 (0.5-0.6) |
| API | 19,253 (7) | 92.8 | 0.7 (0.7-0.7) | 14,376 (8) | 69.0 | 0.7 (0.7-0.7) | 172 (4) | 0.8 | 0.3 (0.3-0.4) | 823 (4) | 4.0 | 0.3 (0.3-0.4) |
| Other | 1,243 (<1) | | | 916 (<1) | | | 11 (<1) | | | 59 (<1) | | |
| Unknown | 1,454 (1) | | | 943 (1) | | | 23 (1) | | | 109 (<1) | | |
| ER receptors | | | | | | | | | | | | |
| ER | | | | | | | | | | | | |
| ER positive | 164,444 (61) | 80.7 | 1.0 | 114,666 (62) | 56.3 | 1.0 | 3,216 (73) | 1.6 | 1.0 | 16,347 (75) | 8.0 | 1.0 |
| ER negative | 48,905 (18) | 24.1 | 0.3 (0.3-0.3) | 37,680 (20) | 18.6 | 0.3 (0.3-0.3) | 178 (4) | 0.1 | 0.1 (0.0-0.1) | 1,470 (7) | 0.7 | 0.1 (0.1-0.1) |
| Unknown | 56,775 (21) | | | 32,823 (18) | | | 1,013 (23) | | | 4,029 (18) | | |

Table 1B:

| | Medullary | | | Inflammatory | | | Papillary | | | Mucinous | | |
|------------------------|-------------|------|---------------|--------------|------|---------------|-------------|------|------------------|-------------|------|------------------|
| Sample size | 2,619 | | | 3,004 | | | 1,678 | | | 6,881 | | |
| % Total cases | 1.1 | | | 1.2 | | | 0.7 | | | 2.8 | | |
| Median age (y) | 51.0 | | | 56.0 | | | 70.0 | | | 71.0 | | |
| Median tumor size (cm) | 2.0 | | | 5.3 | | | 1.5 | | | 1.5 | | |
| Rate (SE) | 1.29 (0.03) | | | 1.50 (0.03) | | | 0.81 (0.02) | | | 3.31 (0.04) | | |
| Variable | N (%) | Rate | RR (95% CI) | N (%) | Rate | RR (95% CI) | N (%) | Rate | RR (95% CI) | N (%) | Rate | RR (95% CI) |
| Demographics | | | | | | | | | | | | |
| Age (y) | | | | | | | | | | | | |
| <50 | 1,204 (46) | 0.8 | 1.0 | 988 (33) | 0.7 | 1.0 | 228 (14) | 0.2 | 1.0 | 926 (13) | 0.6 | 1.0 |
| 50-59 | 630 (24) | 3.0 | 3.8 (3.4-4.2) | 754 (25) | 3.6 | 5.5 (5.0-6.0) | 217 (13) | 1.0 | 6.8 (5.6-8.2) | 902 (13) | 4.3 | 6.9 (6.3-7.6) |
| 60-69 | 413 (16) | 2.8 | 3.5 (3.1-3.9) | 547 (18) | 3.7 | 5.6 (5.1-6.2) | 369 (22) | 2.5 | 16.3 (13.8-19.2) | 1,431 (21) | 9.6 | 15.4 (14.1-16.7) |
| 70-79 | 268 (10) | 2.2 | 2.8 (2.4-3.1) | 447 (15) | 3.7 | 5.6 (5.0-6.2) | 499 (30) | 4.1 | 26.8 (22.9-31.3) | 2,167 (31) | 17.8 | 28.5 (26.3-30.8) |
| 80+ | 104 (4) | 1.3 | 1.7 (1.4-2.1) | 268 (9) | 3.5 | 5.2 (4.6-6.0) | 365 (22) | 4.7 | 30.9 (26.2-36.4) | 1,455 (21) | 18.8 | 30.1 (27.6-32.7) |
| Race | | | | | | | | | | | | |
| White | 1,896 (72) | 1.2 | 1.0 | 2,403 (80) | 1.5 | 1.0 | 1,246 (74) | 0.7 | 1.0 | 5,672 (82) | 3.4 | 1.0 |
| Black | 498 (19) | 2.4 | 2.0 (1.8-2.2) | 416 (14) | 2.1 | 1.4 (1.3-1.6) | 237 (14) | 1.3 | 1.8 (1.6-2.1) | 516 (7) | 2.9 | 0.9 (0.8-1.0) |
| API | 195 (7) | 0.9 | 0.8 (0.5-1.3) | 161 (5) | 0.9 | 0.6 (0.4-1.0) | 178 (11) | 0.9 | 1.2 (0.9-1.6) | 634 (9) | 3.1 | 1.1 (0.9-1.2) |
| Other | 24 (1) | | | 19 (1) | | | 10 (1) | | | 24 (<1) | | |
| Unknown | 6 (<1) | | | 5 (<1) | | | 7 (<1) | | | 35 (1) | | |
| ER receptors | | | | | | | | | | | | |
| ER | | | | | | | | | | | | |
| ER positive | 447 (17) | 0.2 | 1.0 | 1,081 (36) | 0.5 | 1.0 | 982 (59) | 0.5 | 1.0 | 5,128 (75) | 2.5 | 1.0 |
| ER negative | 1,697 (65) | 0.8 | 3.8 (3.4-4.2) | 1,011 (34) | 0.5 | 0.9 (0.9-1.0) | 166 (10) | 0.1 | 0.2 (0.1-0.2) | 265 (4) | 0.1 | 0.1 (0.0-0.1) |
| Unknown | 475 (18) | | | 912 (30) | | | 530 (32) | | | 1,488 (22) | | |

NOTE: ¹N, sample size; ²Rate, age-adjusted (2000 U.S. standard) incidence rate per 100,000 woman-years; ³RR, expressed as a rate ratio where a high-risk characteristic is compared to a low-risk characteristic with an assigned RR of 1.0.

Table 2. Estimates for early-onset and late-onset modes and group probabilities given a specified histopathologic subtype of breast carcinoma, race, or ER expression

| | All breast cancer cases | | Duct NST | | Tubular | | Lobular | |
|---|-------------------------|-------|--------------|-------|------------|-------|-------------|-------|
| Sample size | 270,124 | | 185,169 | | 4,407 | | 21,846 | |
| Median age in years | 62 | | 61 | | 62 | | 66 | |
| Rate (SE) | 132.4 (0.26) | | 91.0 (0.21) | | 2.2 (0.03) | | 10.7 (0.07) | |
| Variable | Result | SE | Result | SE | Result | SE | Result | SE |
| Mode in years | | | | | | | | |
| Early-onset group | 50.27 | 0.16 | 50.93 | 0.16 | 52.76 | 6.79 | 52.01 | 1.09 |
| Late-onset group | 72.96 | 0.26 | 73.45 | 0.17 | 71.75 | 10.13 | 73.23 | 1.64 |
| ¹ p for the early-onset group | 0.49 | <0.01 | 0.56 | <0.01 | 0.49 | <0.01 | 0.38 | <0.01 |
| ² AIC mixture | -1100201.0 | | -720027.2 | | -17053.0 | | -86327.3 | |
| AIC single density | -1107039.0 | | -752190.4 | | -17295.5 | | -87582.1 | |
| ³ p for the early-onset group given race | | | | | | | | |
| White | 0.54 | <0.01 | 0.59 | <0.01 | 0.49 | 0.01 | 0.36 | 0.10 |
| Black | 0.71 | <0.01 | 0.76 | <0.01 | 0.61 | 0.07 | 0.49 | 0.01 |
| API | 0.76 | <0.01 | 0.80 | <0.01 | 0.59 | 0.07 | 0.59 | 0.01 |
| ³ p for the early-onset group given ER | | | | | | | | |
| ER positive | 0.37 | <0.01 | 0.41 | <0.01 | 0.50 | <0.01 | 0.36 | 0.01 |
| ER negative | 0.61 | <0.01 | 0.64 | <0.01 | 0.56 | 0.06 | 0.48 | 0.02 |
| | Medullary | | Inflammatory | | Papillary | | Mucinous | |
| Sample size | 2,619 | | 3,004 | | 1,678 | | 6,881 | |
| Median age in years | 51 | | 56 | | 70 | | 71 | |
| Rate (SE) | 1.3 (0.03) | | 1.5 (0.03) | | 0.8 (0.02) | | 3.3 (0.04) | |
| Variable | Result | SE | Result | SE | Result | SE | Result | SE |
| Mode in years | | | | | | | | |
| Early-onset group | ~ | ~ | 49.74 | 5.50 | 52.25 | 10.71 | 50.31 | 2.19 |
| Late-onset group | ~ | ~ | 73.61 | 9.24 | 74.99 | 17.46 | 74.72 | 3.53 |
| ¹ p for the early-onset group | ~ | ~ | 0.67 | 0.04 | 0.32 | 0.06 | 0.28 | 0.02 |
| ² AIC mixture | -10515.1 | | -12196.4 | | -6756.5 | | -27300.7 | |
| AIC single density | -10504.2 | | -12343.3 | | -6797.8 | | -27792.0 | |
| ³ p for the early-onset group given race | | | | | | | | |
| White | ~ | ~ | 0.65 | 0.06 | 0.44 | 0.10 | 0.22 | 0.01 |
| Black | ~ | ~ | 0.73 | 0.03 | 0.62 | 0.06 | 0.26 | 0.02 |
| API | ~ | ~ | 0.85 | 0.05 | 0.75 | 0.09 | 0.54 | 0.02 |
| ³ p for the early-onset group given ER | | | | | | | | |
| ER positive | ~ | ~ | 0.55 | 0.07 | 0.27 | 0.06 | 0.28 | 0.02 |
| ER negative | ~ | ~ | 0.73 | 0.02 | 0.58 | 0.05 | 0.44 | 0.04 |

Key: Rate, age-adjusted (2000 US standard) incidence rate per 100,000 woman-years; SE, standard error; ¹p, probability of being in the early-onset group; ²AIC, Akaike information criteria where the larger number confirms the better model fit; ³p, probability of being in the early-onset group with covariates included in logistic regression model (see Materials and Methods), ~, not calculated since the data fit a single density model better than a two-component mixture model.

distributed (Figs. 2B-H and 3B-H), although more so for duct NST and tubular carcinomas than for other subtypes. For example, total cases for duct NST and tubular carcinomas were equally distributed between early-onset and late-onset modes (Figs. 2B-C and 3B-C). Total cases for lobular, papillary, and mucinous had less early-onset and more late-onset disease (Figs. 2D, G-H and 3D, G-H). Medullary and inflammatory had predominant early-onset breast cancer populations (Figs. 2E-F and 3E-F).

Similarly, all breast cases and every histopathologic type except for medullary carcinoma were best described by a two-component mixture model with modes at ages ~50 and 70 years (Table 2). For example, among duct NST and tubular carcinomas, the probability of a single case belonging to the early-onset distribution was 0.56 (SE <0.01) and 0.49 (SE <0.01), respectively. In contrast, the probability of a case belonging to an early-onset distribution was lower for lobular (0.38; SE <0.01), papillary (0.32; SE = 0.06), and mucinous (0.28; SE = 0.02) carcinomas. Inflammatory carcinomas had a higher likelihood of belonging to the early-onset group (0.67; SE = 0.04). Larger (less negative) AIC values for all histopathologic subtypes (except for medullary) confirmed a better fit for the mixture model than for a single density. In contrast, medullary carcinoma had only a single early-onset mode at age 51 years.

When total cases for each histopathologic subtype were stratified by race and ER expression, the density plots showed varying mixtures of bimodal breast carcinoma types with early-onset and late-onset modes or peak frequencies (Figs. 2B-H and 3B-H). These findings were also confirmed when we fit the mixture models stratified by race and ER. However, whereas the fraction of cases in the early-onset and late-onset groups changed by race and ER among each histopathologic subtype, the modal ages remained unchanged, near ages 50 and 70 years.

Given that all of the parameters in the mixture model, with the exception of the mixing proportion, were unaffected by the covariates, we incorporated race and ER expression into the model via logistic mixing proportions. For all histopathologic subtypes, except tubular carcinomas, Asian or Pacific Islanders were more likely than either Whites or Blacks to have early-onset tumors with peak incidence at age 50 years. For example, among duct NST (Table 2), the probability of a single Asian or Pacific Islander case belonging to the early-onset group was 0.80 (SE < 0.01), compared with 0.59 (SE < 0.01) for a single White case and 0.76 (SE < 0.01) for a single Black case. Similarly, ER-negative tumors were more likely than ER-positive tumors to show early-onset modes at age 50 years.

Paraphrasing, we have observed similar bimodal incidence patterns according to tumor size, axillary lymph nodes, tumor grade, and progesterone receptor expression (20). That is, large tumors were more likely than small tumors to show early-onset modes at age 50 years. Tumors with positive axillary lymph nodes high grade, and negative progesterone receptor expression were also more likely to show early-onset modes at age 50 years than were breast cancers with negative nodes, low grade, and positive progesterone receptor expression. In sum, high-risk tumors (large size, positive lymph nodes, high grade, negative ER and progesterone receptor expression) have predominantly early-onset age distributions at diagnosis, whereas low-risk tumors (small size, negative lymph nodes, low grade, positive ER and progesterone receptor expression) have mostly late-onset age distributions at diagnosis.

Discussion

Emerging molecular analyses of female breast carcinoma suggest the existence of two main pathways for mammary carcinogenesis based on epithelial cellular origin (luminal or basal) and/or ER expression (positive or negative; refs. 5, 21). There is (a) derivation of tumors from stem cells committed to luminal differentiation and ER expression and (b) development of neoplasms from stem cells programmed to display basal differentiation and lacking ER expression (22-24). If confirmed, these observations may form the basis for revised conceptual frameworks and, therefore, should be validated in the general population.

Indeed, our analysis of population-based SEER data provided evidence for two main breast cancer populations according to age at onset, mixed within breast cancer overall. One breast cancer type was mostly early-onset with peak incidence near age 50 years. The second breast cancer type was largely late-onset with its mode occurring at age ~70 years. Although covariates had varying proportions of early-onset and late-onset age distributions, modal ages remained near 50 and 70 years, respectively.

The age structure of the breast cancer population (Figs. 2 and 3) seemed to define the shape of the age-specific incidence rate curve (Fig. 1). For example, medullary and inflammatory carcinomas with predominant early-onset age distributions (Figs. 2E-F and 3E-F) produced age-specific incidence rate curves that flattened or decreased after age 50 years (Fig. 1E-F). Papillary and mucinous carcinomas with late-onset age distributions (Figs. 2G-H and 3G-H) yielded age-specific rate curves that rose steadily before and after age 50 years (Fig. 1G-H). Finally, duct NST, tubular, and lobular carcinomas with bimodal breast cancer populations (Figs. 2B-D and 3B-D) produced intermediate age-specific rate curves (i.e., age-specific rates increased rapidly until age 50 years, and then continued to increase at a slower pace; Fig. 1B-D).

In their seminal study, Sorlie et al. (25) used a 534 "intrinsic" gene set to subdivide ER-positive and ER-negative tumors into five subtypes: luminal A, luminal B, basal-like, HER2-overexpressing, and "normal-like" subgroups. Luminal A and luminal B are ER-positive subtypes, whereas basal, HER2, and the normal-like subgroups are ER-negative subtypes. A straightforward interpretation of their results regards each molecular signature as a distinct breast cancer entity (24, 25). To test this concept, we constructed and analyzed age-density plots using breast cancer cases ($n = 122$) from Sorlie et al.'s data set (Fig. 4). After matching molecular portraits with ages at diagnosis from their supplementary data (25), luminal A and B cases had a predominant late-age distribution with a mode at age 74 years (SE = 1.7), as did ER-positive tumors in our SEER analysis. On the other hand, basal-like and HER2-overexpressing phenotypes showed a dominant early-onset mode at age 52

years (SE = 2.9), as did ER-negative tumors in SEER. Therefore, and despite their small sample size, these results show only two modes with small variances, occurring at ages similar to those in the SEER population. Thus, where SEER's two age-specific incidence rate patterns for ER and three rate patterns for histopathologic subtypes reflected bimodal mixtures of early-onset and late-onset breast cancer populations, Sorlie et al.'s molecular phenotypes also displayed bimodal breast cancer populations.

Our study was limited by the usual concerns related to analyses of registry data: nonstandardization of histopathologic diagnosis and/or ER testing and incomplete data collection. However, it is reassuring that the age distribution for different histopathologic subtypes and known ER expression is similar to that in other reported series (26). In previous studies, we have also shown that cases with unknown data are generally similar to all cases combined (20, 27-29). Moreover, in a sensitivity analysis, we examined age incidence patterns in different SEER registries, over different time periods, and among different birth cohorts. SEER bimodal age incidence patterns were robust to geographic location as well as to calendar period and/or birth cohort effects (data not shown). Tarone and Chu (3) also showed that SEER's distinct ER-positive and ER-negative incidence patterns were unaffected by adjustments for calendar period and birth cohort effects or artifacts, implying a true age-related biological phenomenon. Moreover, bimodal female breast cancer populations were first described by von Pirquet in 1930 (30) and have been observed worldwide in Africa (31), Taiwan (32), Italy (33), New Zealand (34), Europe, and America (35), suggesting a universal breast cancer phenomenon.

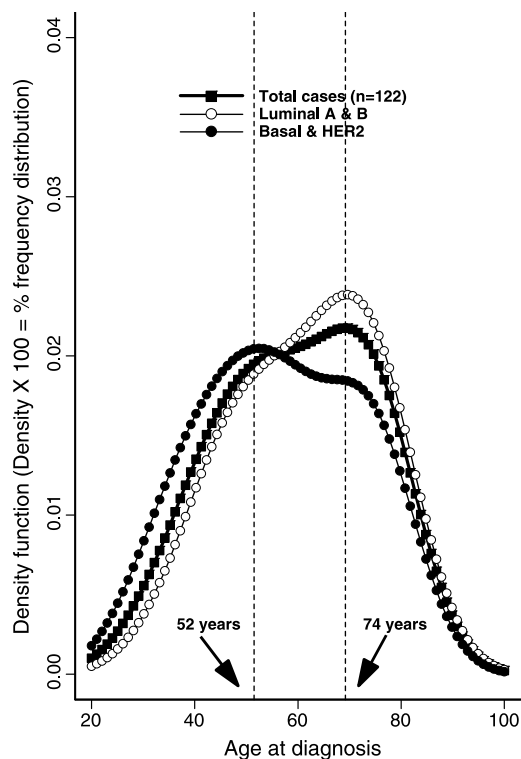


Figure 4. Age at diagnosis-density plots (or smoothed histograms) were applied to data provided in Supplementary Table S2 (<http://www.pnas.org>) for all tissue samples in the extended Norway/Stanford extended data set, described by Sorlie et al. (25): (i) all tissue samples ($n = 122$) in the Norway/Stanford data set; (ii) luminal A and luminal B cases in the Norway/Stanford data set; and (iii) basal-like and HER2-overexpressing cases in the Norway/Stanford data set.

In conclusion, age incidence patterns showed bimodal early-onset and late-onset breast cancer types, mixed within SEER's overall breast cancer population. Bimodal age distribution—the occurrence of two incidence peaks or modes—is always of interest. It suggests etiologic heterogeneity due to at least two biological subtypes or causal pathways (36). Although bimodal age incidence patterns are acknowledged for malignancies such as Hodgkin's lymphoma (37, 38), bimodality is not very well established for breast cancer. Moreover, whereas the two modes for Hodgkin's disease are widely separated in early and late adulthood, bimodal breast cancer has a narrower range of ages 50 and 70 years.

Of note, the modal ages for breast cancer of 50 and 70 years were robust across molecular portraits, race, ER, as well as other tumor characteristics (20); however, the relative distributions for early-onset and late-onset breast cancer types varied by covariate. In general, high-risk tumors such as ER-negative expression were surrogates for bimodal breast cancer populations, shifted towards the early-onset mode near age 50 years. Low-risk tumors such as ER-positive expression were surrogates for bimodal breast cancer, weighted towards the late-onset mode at age 70 years.

Additional etiologic clues may be provided by medullary carcinoma, the one histopathologic subtype with a single mode at age 51 years. Medullary carcinomas are rare tumors that have been associated with early-onset hereditary breast carcinomas due to the *BRCA1* gene (39, 40). Given that age at onset is the strongest predictor for a genetic versus an acquired breast cancer phenotype (41, 42), bimodality in the general breast cancer population might be a reflection of early-onset hereditary (or familial) versus late-onset nonhereditary (acquired or sporadic) breast cancer types. Because age 50 years is both the early-onset mode and a valid menopausal surrogate (43, 44), we further speculate that premenopausal exposures will have greater effect on early-onset than on late-onset breast cancers. Future population-based studies that include assessment of family history, breast cancer risk factors, environmental exposures, standardized histopathology review, and molecular characterization are needed to examine the conceptual framework of a two-class or bimodal breast cancer model.

References

- Clemmesen J. Carcinoma of the breast. *Br J Radiol* 1948;21:583–90.
- Yasui Y, Potter JD. The shape of age-incidence curves of female breast cancer by hormone-receptor status. *Cancer Causes Control* 1999;10:431–7.
- Tarone RE, Chu KC. The greater impact of menopause on ER- than ER+ breast cancer incidence: a possible explanation (United States). *Cancer Causes Control* 2002;13:7–14.
- Anderson WF, Chatterjee N, Ershler WB, Brawley OW. Estrogen receptor breast cancer phenotypes in the Surveillance, Epidemiology, and End Results database. *Breast Cancer Res Treat* 2002;76:27–36.
- Anderson WF, Matsuno RK. Breast cancer heterogeneity: a mixture of at least two main types. *J Natl Cancer Inst* 2006;98:948–51.
- Stalsberg H, Thomas DB. Age distribution of histologic types of breast carcinoma. *Int J Cancer* 1993;54:1–7.
- Anderson WF, Chu KC, Chang S. Inflammatory breast carcinoma and non-inflammatory locally advanced breast carcinoma: distinct clinicopathologic entities? *J Clin Oncol* 2003;21:2254–9.
- Anderson WF, Chu KC, Chang S, Sherman ME. Comparison of age-specific incidence rate patterns for different histopathologic types of breast carcinoma. *Cancer Epidemiol Biomarkers Prev* 2004;13:1128–35.
- Armitage P, Doll R. A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. *Br J Cancer* 1957;11:161–9.
- Armitage P, Doll R. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br J Cancer* 1954;8:1–12.
- SEER. Surveillance, Epidemiology, and End Results (SEER) Program. Public-Use Database (1973–2002), National Cancer Institute DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2005, based on the November 2004 submission. Available from: <http://www.seer.cancer.gov>.
- Berg JW, Hutter RV. Breast cancer. *Cancer* 1995;75:257–69.
- SEER. ICD-O-3 coding materials. 2004. Available from: <http://seer.cancer.gov/icd-o-3/>.
- Tavassoli FA, Devilee P, editors. Pathology and genetics of tumours of the breast and female genital organs. World Health Organization classification of tumours. Lyon: IARC Press; 2003.
- Anderson WF, Chu KC, Chatterjee N, Brawley OW, Brinton LA. Tumor variants by hormone receptor expression in white patients with node-negative breast cancer from the Surveillance, Epidemiology, and End Results database. *J Clin Oncol* 2001;19:18–27.
- Pfeiffer RM, Gail MH, Brown LM. Probability of helicobacter pylori infection based on IgG levels and other covariates using a mixture model. *J Epidemiol Biostat* 2000;5:267–75.
- Gallant AR, Nychka DW. Semi-nonparametric maximum likelihood estimation. *Econometrica* 1987;55:363–90.
- Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, editors. 2nd International Symposium on Information Theory. Budapest: Akademiai Kiado; 1973. p. 267–81.
- Schwarz G. Estimating dimension of a model. *Ann Statist* 1978;6:461–4.
- Anderson WF, Jatoi I, Devesa SS. Distinct breast cancer incidence and prognostic patterns in the NCI's SEER program: suggesting a possible link between etiology and outcome. *Breast Cancer Res Treat* 2005;90:127–37.
- Birnbaum D, Bertucci F, Ginestier C, Tagett R, Jacquemier J, Charafe-Jauffret E. Basal and luminal breast cancers: basic or luminous? (review). *Int J Oncol* 2004;25:249–58.
- Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature* 2000;406:747–52.
- Boecker W, Buerger H. Evidence of progenitor cells of glandular and myoepithelial cell lineages in the human adult female breast epithelium: a new progenitor (adult stem) cell concept. *Cell Prolif* 2003;36 Suppl 1: 73–84.
- Sorlie T, Wang Y, Xiao C, et al. Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: Gene expression analyses across three different platforms. *BMC Genomics* 2006;7:127.
- Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 2003;100:8418–23.
- Rosen PP, Oberman HA. Tumors of the mammary gland. Atlas of tumor pathology. Washington (DC): Armed Forces Institute of Pathology; 1993.
- Anderson WF, Althuis MD, Brinton LA, Devesa SS. Is male breast cancer similar or different than female breast cancer? *Breast Cancer Res Treat* 2004; 83:77–86.
- Anderson WF, Chu KC, Devesa SS. Distinct incidence patterns among in-situ and invasive breast carcinomas, with possible etiologic implications. *Breast Cancer Res Treat* 2004;88:149–59.
- Anderson WF, Devesa SS. In-situ male breast carcinoma in the Surveillance, Epidemiology, and End Results Database of the National Cancer Institute. *Cancer* 2005;104:1733–41.
- von Pirquet C. Allergie des lebensalters, die bosartigen geschwulste [Allergy of age, the malignant tumors]. Leipzig (Germany): Thieme Verlag; 1930.
- Muguti GI. Experience with breast cancer in Zimbabwe. *J R Coll Surg Edinb* 1993;38:75–8.
- Chie WC, Chen CF, Lee WC, Chen CJ, Lin RS. Age-period-cohort analysis of breast cancer mortality. *Anticancer Res* 1995;15:511–5.
- Sant M, Gatta G, Micheli A, et al. Survival and age at diagnosis of breast cancer in a population-based cancer registry. *Eur J Cancer* 1991;27: 981–4.
- Newman PD, Mason BH, Holdaway IM, Kay RG, Arthur JF, Hitchcock GC. Incidence and clinical features of breast cancer in the Auckland region. *N Z Med J* 1992;105:117–20.
- De Waard F, Baanders-van Halewijn EA, Huizinga J. The bimodal age distribution of patients with mammary carcinoma; evidence for the existence of 2 types of human breast cancer. *Cancer* 1963;17:141–51.
- MacMahon B, Pugh TF. Characteristics of persons, age. In: MacMahon B, Pugh TF, editors. Epidemiology principles and methods. Boston: Little, Brown and Company; 1970. p. 103–9.
- Cole P, MacMahon B, Aisenberg A. Mortality from Hodgkin's disease in the United States. Evidence for the multiple-aetiology hypothesis. *Lancet* 1968;2: 1371–6.
- MacMahon B. Epidemiology of Hodgkin's disease. *Cancer Res* 1966;26: 1189–201.
- Lakhani SR, Gusterson BA, Jacquemier J, et al. The pathology of familial breast cancer: histological features of cancers in families not attributable to mutations in *BRCA1* or *BRCA2*. *Clin Cancer Res* 2000;6:782–9.
- Claus EB, Risch N, Thompson WD, Carter D. Relationship between breast histopathology and family history of breast cancer. *Cancer* 1993; 71:147–53.
- Claus EB, Risch NJ, Thompson WD. Age at onset as an indicator of familial risk of breast cancer. *Am J Epidemiol* 1990;131:961–72.
- Lynch HT, Watson P, Conway TA, Lynch JF. Natural history and age at onset of hereditary breast cancer. *Cancer* 1992;69:1404–7.
- Morabia A, Flandre P. Misclassification bias related to definition of menopausal status in case-control studies of breast cancer. *Int J Epidemiol* 1992;21:222–8.
- Morabia A, Costanza MC. International variability in ages at menarche, first livebirth, and menopause. World Health Organization collaborative study of neoplasia and steroid contraceptives. *Am J Epidemiol* 1998;148: 1195–205.