**DOI:**10.1093/jnci/dju055
First published online April 28, 2014



ARTICLE

# US Incidence of Breast Cancer Subtypes Defined by Joint Hormone Receptor and HER2 Status

Nadia Howlader, Sean F. Altekruse, Christopher I. Li, Vivien W. Chen, Christina A. Clarke, Lynn A. G. Ries, Kathleen A. Cronin

Manuscript received June 26, 2013; revised November 22, 2013; accepted February 10, 2014.

**Correspondence to:** Nadia Howlader, MS, Mathematical Statistician, Surveillance Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, 9609 Medical Center Dr, Bethesda MD 20892 (e-mail: howladern@mail.nih.gov)

**Background** In 2010, Surveillance, Epidemiology, and End Results (SEER) registries began collecting human epidermal growth factor 2 (HER2) receptor status for breast cancer cases.

**Methods** Breast cancer subtypes defined by joint hormone receptor (HR; estrogen receptor [ER] and progesterone receptor [PR]) and HER2 status were assessed across the 28% of the US population that is covered by SEER registries. Age-specific incidence rates by subtype were calculated for non-Hispanic (NH) white, NH black, NH Asian Pacific Islander (API), and Hispanic women. Joint HR/HER2 status distributions by age, race/ethnicity, county-level poverty, registry, stage, Bloom–Richardson grade, tumor size, and nodal status were evaluated using multivariable adjusted polytomous logistic regression. All statistical tests were two-sided.

**Results** Among case patients with known HR/HER2 status, 36810 (72.7%) were found to be $HR^+/HER2^-$, 6193 (12.2%) were triple-negative ($HR^-/HER2^-$), 5240 (10.3%) were $HR^+/HER2^+$, and 2328 (4.6%) were $HR^-/HER2^+$; 6912 (12%) had unknown HR/HER2 status. NH white women had the highest incidence rate of the $HR^+/HER2^-$ subtype, and NH black women had the highest rate of the triple-negative subtype. Compared with women with the $HR^+/HER2^-$ subtype, triple-negative patients were more likely to be NH black and Hispanic; $HR^+/HER2^+$ patients were more likely to be NH API; and $HR^-/HER2^+$ patients were more likely to be NH black, NH API, and Hispanic. Patients with triple-negative, $HR^+/HER2^+$, and $HR^-/HER2^+$ breast cancer were 10% to 30% less likely to be diagnosed at older ages compared with $HR^+/HER2^-$ patients and 6.4-fold to 20.0-fold more likely to present with high-grade disease.

**Conclusions** In the future, SEER data can be used to monitor clinical outcomes in women diagnosed with different molecular subtypes of breast cancer for a large portion (approximately 28%) of the US population.

JNCI J Natl Cancer Inst (2014) 106(5): dju055 doi:10.1093/jnci/dju055

Several distinct molecular subtypes of breast cancer have been defined based on gene expression patterns (1). Characterization of this heterogeneity has changed how patients with this complex malignancy are treated. The major subtypes of breast cancer are approximated by the joint expression of three tumor markers: estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor 2–neu (HER2), which are evaluated routinely because of their utility in guiding clinical care. Recent findings indicate that immunohistochemical protein expression profiles are surrogates for intrinsic gene-derived expression profiles defining molecular breast cancer subtypes (2). The most common subtypes are hormone receptor (ER or PR) positive (i.e., $ER^+$ or $PR^+$), comprising the luminal A and luminal B subtypes. Luminal B cancers and two other subtypes, triple-negative tumors ($ER^-/PR^-/HER2^-$ cancers, most of which are of the basal-like phenotype) and HER2– overexpressing tumors ($ER^-/HER2^+$), are known to be more clinically aggressive and have poorer prognoses compared with luminal A tumors (3–5). A growing body of evidence suggests that there are notable demographic differences across these subtypes. Triple-negative breast cancer has been shown to be more likely to occur among younger women and black women (6–11). The literature, however, is based largely on relatively small observational studies or confined to particular geographic regions (8,9,12–14), with the exception of cancer registry data covering the state of California (6,10,11). Information on HER2 status and its availability was collected on all breast cancer cases diagnosed in 2010 by the entire population-based Surveillance, Epidemiology, and End Results (SEER) program. This article presents the first report of nationally representative incidence rates for the major breast cancer subtypes based on joint ER/PR/HER2 status and an assessment of demographic and clinical differences across these subtypes using SEER data covering an estimated 28% of the US population (15)

## Methods

### Study Population

This study used data from 17 population-based cancer registries that participate in the SEER program (data from the Alaska Native

registry were excluded, n = 57), together comprising approximately 28% of the total population of the United States (16). Women diagnosed with invasive breast cancer in 2010 were included in the analysis. The year 2010 is the most recent year for which complete SEER data are available and is the first year for which data on HER2 status are available (data on ER and PR status have been collected since 1990). Case patients diagnosed by autopsy or death certificate (n = 229) or with sarcomas of the breast (based on histology codes 8800, 8801, 8805, 8815, 8830, 8850, 8858, 8890, 8935, 8980, 8982, 8983, 9120, 9180, 9181, 9260) were excluded (n = 84). The final analytic set consisted of 57 483 case patients.

All study data—including ER, PR, and HER2 status, demographic characteristics, and tumor stage and grade—were ascertained across SEER registries using standardized coding rules based on hospital medical records and pathology reports. Additionally, area-level poverty data (percentage of persons living below the poverty variable) were derived from the 2000 US Census, based on county at diagnosis, and were used as a surrogate for socioeconomic status (SES). Cutpoints based on empirical research and policy relevance (17,18) were used to create three levels for this variable (ie, poverty <10.0% for high SES, 10%–19.99% for medium SES, and >20% for low SES). The data on ER, PR, and HER2 status were recorded by the SEER program in the following categories: 1) test not done, 2) positive (+), 3) negative (–), 4) borderline, 5) test done but results missing, and 6) unknown. For each biomarker, the original six categories were combined into four categories: positive, negative, borderline, or unknown (Supplementary Table 1, available online). Detailed coding instructions for all three tumor markers can be found under the collaborative stage data collection system (19). The HER2 variable used in the analysis was based on a single summary derived variable created by the SEER program using five HER2-related site-specific factors from the Collaborative Stage data collection system. Details of the derived HER2 variable can be obtained from the SEER website (http://seer.cancer.gov/seerstat/databases/ssf/her2-derived.html).

ER and PR results were combined and analyzed jointly as hormone receptor (HR) status. $HR^+$ was defined as either $ER^+$, $PR^+$, or borderline (categories 2 and 4); $HR^-$ was defined as both $ER^-$ and $PR^-$ (category 3); and unknown HR was defined as test not done, test done but results missing, or unknown (categories 1, 5, and 6). Similarly, HER2 status was defined as $HER2^+$ (category 2), $HER2^-$ (category 3), and unknown HER2 (categories 1, 4, 5, and 6). Note that case patients with borderline ER or PR status were treated as having $ER^+$ or $PR^+$ status (borderline ER: n = 62, 0.1%; borderline PR: n = 191, 0.3%), whereas case patients with borderline HER2 status were treated as having unknown HER2 status (borderline HER2: n = 1566, 2.7%). ER/PR borderline case patients were grouped with positive case patients because recent guideline changes indicated that the borderline category most likely was classified as positive because lower cutoffs (such as 1%) were used for the ER/PR test, whereas cutoffs as high as 10% had previously been used for determining ER/PR positivity (20). Using tumor subtype definitions based on joint ER/PR/HER2 status (6,14,21), tumors were classified into four mutually exclusive categories: $HR^+/HER2^-$; $ER^-/PR^-/HER2^-$ (triple negative); $HR^+/HER2^+$; and $HR^-/HER2^+$. Details of how tumors with positive or negative expressions for ER/PR/HER2 were coded into the subtypes are presented in Supplementary Table 2 (available online). The SEER*Stat software (22) includes a variable to facilitate the analysis of trends in breast cancer molecular subtypes. The derived HER2 variable or the breast cancer subtype variable can be obtained from the custom database with extra Collaborative Stage site-specific factors upon request from the following URL: http://seer.cancer.gov/seerstat/databases/ssf/.

### Statistical Analysis

Age-specific incidence rates per 100 000 women by breast cancer subtypes were calculated based on 5-year age categories using the SEER*Stat software (22). New intercensal population estimates released by the US Census Bureau were used as the denominators in generating rates (23). Standard errors and 95% confidence intervals (CIs) for rates were calculated using the Tiwari method (24). The age-specific rates were presented for four mutually exclusive race/ethnicity groups: non-Hispanic white (NH white), non-Hispanic black (NH black), non-Hispanic Asian Pacific Islander (NH API), and Hispanic.

Unordered polytomous logistic regression was used to calculate odds ratios (ORs) and 95% confidence intervals to quantify associations between breast cancer subtypes and various demographic and clinical factors. These included age at diagnosis (<50, 50–64, 65–74, ≥75 years), race/ethnicity (NH white, NH black, NH API, Hispanic), the American Joint Committee on Cancer's *Cancer Staging Manual* (7th edition) (25) stage at diagnosis (I, II, III, IV), Bloom–Richardson tumor grade (low, medium, high), and SEER registry. Because of collinearity with stage, tumor size and lymph node status were not included with stage in the model. SAS version 9.3 statistical software was used to fit the unordered polytomous logistic regression (26). All odds ratios were adjusted for race/ethnicity, age, stage, tumor grade, and SEER region and based on patients having complete information for each of these covariables (ie, women missing data for one or more of these covariables were dropped from the regression analysis; n = 13 980). All statistical tests were two-sided.

## Results

Among 2010 case patients with known HR/HER2 status, 36 810 (72.7%) were found to be $HR^+/HER2^-$, 6193 (12.2%) were triple-negative ($HR^-/HER2^-$), 5240 (10.3%) were $HR^+/HER2^+$, and 2328 (4.6%) were $HR^-/HER2^+$; 6912 (12%) of the case patients had an unknown HR/HER2 status (Table 1). Subtype distributions varied by age, race/ethnicity, county-level poverty, stage, and grade. Compared with $HR^+/HER2^-$ case patients (the most common subtype), those diagnosed with the other three subtypes were somewhat more likely to be younger, belong to minority racial or ethnic groups, live in counties with higher poverty levels, and have later stage and higher Bloom-Richardson grade disease (Table 1). Subtype distribution also varied by SEER registry. Cases with missing HR/HER2 status tended to be black, Hispanic, older, and diagnosed with more advanced stage disease.

Figure 1 shows age-specific female breast cancer incidence rates per 100 000 by molecular subtype for four racial and ethnic groups. Incidence rates for $HR^+/HER2^-$ were higher than those for other subtypes across all racial/ethnic groups and all age groups

**Table 1.** Demographic and clinical characteristics of breast cancer subtypes in women with invasive breast cancer, SEER-18, excluding Alaska, 2010*

| | All case patients | Among case patients with known subtype (n = 50 571)† | | | | | | | | Among total case patients‡ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | HR+/HER2− | | Triple-negative | | HR+/HER2+ | | HR−/HER2+ | | Unknown subtype | |
| **Characteristic** | **n = 57 483** | **n = 36 810** | **72.7%** | **n = 6193** | **12.2%** | **n = 5240** | **10.3%** | **n = 2328** | **4.6%** | **n = 6912** | **12.0%** |
| Demographic characteristics | | | | | | | | | | | |
| Age at diagnosis, y | | | | | | | | | | | |
| <50 | 11 949 | 6902 | 64.8% | 1616 | 15.2% | 1528 | 14.4% | 599 | 5.6% | 1304 | 10.9% |
| 50–64 | 21 586 | 13 610 | 70.7% | 2540 | 13.2% | 2066 | 10.7% | 1032 | 5.4% | 2338 | 10.8% |
| 65–74 | 12 643 | 8641 | 77.8% | 1151 | 10.4% | 939 | 8.5% | 382 | 3.4% | 1530 | 12.1% |
| ≥75 | 11 305 | 7657 | 80.1% | 886 | 9.3% | 707 | 7.4% | 315 | 3.3% | 1740 | 15.4% |
| Race/ethnicityβ | | | | | | | | | | | |
| Non-Hispanic white | 40 744 | 27 165 | 75.5% | 3850 | 10.7% | 3532 | 9.8% | 1438 | 4.0% | 4759 | 11.7% |
| Non-Hispanic black | 6007 | 3169 | 60.2% | 1183 | 22.5% | 598 | 11.4% | 318 | 6.0% | 739 | 12.3% |
| Non-Hispanic Asian Pacific Islander | 4367 | 2748 | 71.1% | 376 | 9.7% | 475 | 12.3% | 265 | 6.9% | 503 | 11.5% |
| Hispanic | 5694 | 3361 | 68.2% | 727 | 14.7% | 564 | 11.4% | 280 | 5.7% | 762 | 13.4% |
| County-level poverty 2000ll | | | | | | | | | | | |
| High SES, poverty <10% | 22 454 | 14 800 | 74.0% | 2276 | 11.4% | 2073 | 10.4% | 859 | 4.3% | 2446 | 10.9% |
| Medium SES, poverty 10%–19.99% | 30 611 | 19 389 | 72.4% | 3359 | 12.6% | 2739 | 10.2% | 1284 | 4.8% | 3840 | 12.5% |
| Low SES, poverty >20% | 4398 | 2608 | 69.1% | 558 | 14.8% | 427 | 11.3% | 184 | 4.9% | 621 | 14.1% |
| SEER registry | | | | | | | | | | | |
| Atlanta, metropolitan | 2094 | 1340 | 73.3% | 233 | 12.8% | 179 | 9.8% | 76 | 4.2% | 266 | 12.7% |
| Connecticut | 3066 | 2101 | 76.1% | 280 | 10.1% | 282 | 10.2% | 98 | 3.6% | 305 | 10.0% |
| Detroit, metropolitan | 2899 | 1801 | 69.0% | 410 | 15.7% | 282 | 10.8% | 118 | 4.5% | 288 | 9.9% |
| Greater California | 12 852 | 8147 | 73.5% | 1306 | 11.8% | 1110 | 10.0% | 518 | 4.7% | 1771 | 13.8% |
| Hawaii | 1070 | 750 | 75.1% | 97 | 9.7% | 101 | 10.1% | 51 | 5.1% | 71 | 6.6% |
| Iowa | 2331 | 1584 | 74.1% | 254 | 11.9% | 193 | 9.0% | 106 | 5.0% | 194 | 8.3% |
| Kentucky | 3056 | 1963 | 72.2% | 383 | 14.1% | 248 | 9.1% | 125 | 4.6% | 337 | 11.0% |
| Los Angeles | 5768 | 3634 | 71.7% | 616 | 12.2% | 575 | 11.4% | 241 | 4.8% | 702 | 12.2% |
| Louisiana | 3094 | 1759 | 67.8% | 407 | 15.7% | 297 | 11.5% | 131 | 5.1% | 500 | 16.2% |
| New Jersey | 6627 | 4065 | 72.4% | 667 | 11.9% | 628 | 11.2% | 258 | 4.6% | 1009 | 15.2% |
| New Mexico | 1266 | 738 | 74.9% | 106 | 10.8% | 101 | 10.2% | 41 | 4.2% | 280 | 22.1% |
| Rural + greater Georgia | 3973 | 2435 | 69.0% | 503 | 14.3% | 404 | 11.5% | 186 | 5.3% | 445 | 11.2% |
| San Francisco–Oakland | 3124 | 2114 | 75.8% | 293 | 10.5% | 256 | 9.2% | 126 | 4.5% | 335 | 10.7% |
| San Jose–Monterey | 1556 | 1043 | 74.0% | 171 | 12.1% | 142 | 10.1% | 54 | 3.8% | 146 | 9.4% |
| Seattle, Puget Sound | 3439 | 2536 | 77.2% | 304 | 9.3% | 310 | 9.4% | 136 | 4.1% | 153 | 4.5% |
| Utah | 1268 | 800 | 69.1% | 163 | 14.1% | 132 | 11.4% | 63 | 5.4% | 110 | 8.7% |

*(Table continues)*

**Table 1 (Continued).**

| Characteristic | All case patients | Among case patients with known subtype (n = 50 571)† | | | | | | | | Among total case patients‡ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | HR+/HER2− | | Triple-negative | | HR+/HER2+ | | HR−/HER2+ | | Unknown subtype | |
| | n = 57 483 | n = 36 810 | 72.7% | n = 6193 | 12.2% | n = 5240 | 10.3% | n = 2328 | 4.6% | n = 6912 | 12.0% |
| Clinical characteristics | | | | | | | | | | | |
| AJCC 7th stage¶ | | | | | | | | | | | |
| I | 27 816 | 19 881 | 79.5% | 2214 | 9.0% | 2115 | 8.4% | 779 | 3.1% | 2827 | 10.2% |
| II | 17 494 | 10 873 | 69.3% | 2488 | 14.9% | 1783 | 11.0% | 776 | 4.8% | 1574 | 9.0% |
| III | 6505 | 3705 | 62.6% | 958 | 16.1% | 803 | 13.6% | 465 | 7.8% | 574 | 8.8% |
| IV | 3203 | 1532 | 61.2% | 379 | 15.1% | 370 | 14.8% | 223 | 8.9% | 699 | 21.8% |
| Unknown | 2390 | 818 | 66.2% | 152 | 13.5% | 167 | 13.7% | 80 | 6.6% | 1173 | 49.1% |
| Bloom–Richardson grade | | | | | | | | | | | |
| Low grade | 13 158 | 10 999 | 91.5% | 356 | 3.0% | 547 | 4.6% | 124 | 1.0% | 1132 | 8.6% |
| Medium grade | 20 562 | 15 561 | 82.4% | 967 | 5.1% | 1847 | 9.8% | 508 | 2.7% | 1679 | 8.2% |
| High grade | 14 157 | 5731 | 44.1% | 3948 | 30.4% | 2032 | 15.6% | 1288 | 9.9% | 1158 | 8.2% |
| Unknown | 9606 | 4519 | 67.8% | 922 | 13.8% | 814 | 12.2% | 408 | 6.1% | 2943 | 30.6% |
| Tumor size | | | | | | | | | | | |
| <2.0 cm | 30 763 | 21 852 | 79.0% | 2463 | 8.9% | 2424 | 8.8% | 932 | 3.4% | 3092 | 10.1% |
| 2.0–4.9 cm | 18 614 | 11 231 | 66.8% | 2677 | 15.9% | 2015 | 12.0% | 900 | 5.4% | 1791 | 9.6% |
| ≥5.0 cm | 5036 | 2730 | 61.2% | 817 | 18.3% | 557 | 12.5% | 355 | 8.0% | 577 | 11.5% |
| Unknown | 3070 | 997 | 61.6% | 236 | 14.6% | 244 | 15.1% | 141 | 8.7% | 1452 | 47.3% |
| Nodal status | | | | | | | | | | | |
| Positive | 16 085 | 10 185 | 69.05% | 1875 | 12.71% | 1800 | 12.20% | 890 | 6.03% | 1335 | 8.30% |
| Negative | 32 891 | 22 321 | 74.91% | 3592 | 12.05% | 2771 | 9.30% | 1115 | 3.74% | 3092 | 9.40% |
| Unknown | 8507 | 4304 | 71.47% | 726 | 12.06% | 669 | 11.11% | 323 | 5.36% | 2485 | 29.21% |

* AJCC = American Joint Committee on Cancer; HER2 = human epidermal growth factor 2; HR = hormone receptor; SEER = Surveillance, Epidemiology, and End Results; SES = socioeconomic status.

† Percentages are calculated among case patients with a known breast cancer subtype.

‡ Percentages are calculated among total case patients.

§ Totals do not add up because non-Hispanic American Indian/Alaska Native and non-Hispanic other race categories were not shown.

‖ Totals do not add up because several unknown counties were not shown.

¶ Totals do not add up because stage 0 was not shown.

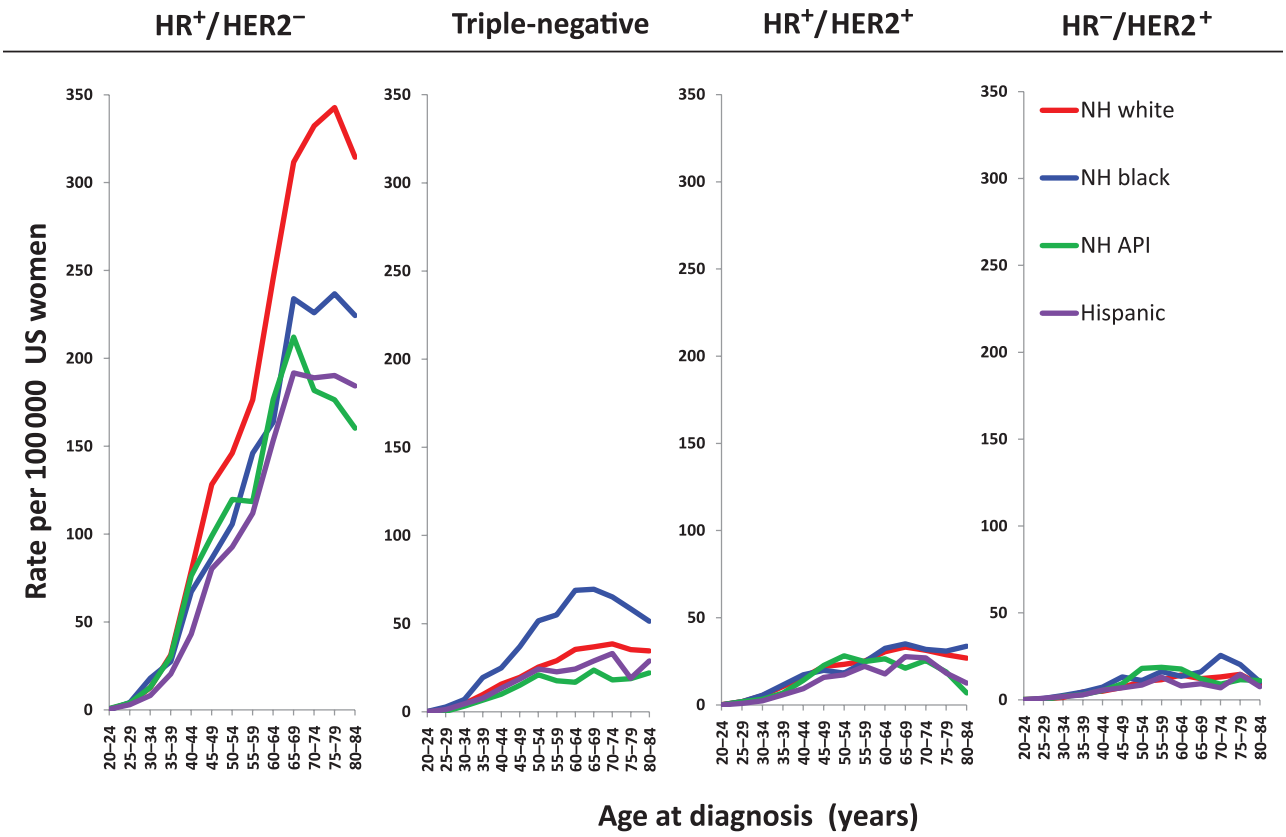


**Figure 1.** Age-specific incidence rates of breast cancer subtypes by race/ethnicity, Surveillance, Epidemiology, and End Resulsts 18, excluding Alaska, 2010. The 95% confidence intervals for incidence rates are presented in Supplementary Table 3 (available online). API = Asian Pacific Islander; HER = human epidermal growth factor; HR = hormone receptor; NH = non-Hispanic.

(Figure 1). NH white women had the highest rate for this subtype, followed by NH black women, and then NH API and Hispanic women. Racial and ethnic differences in HR+/HER2− rates peaked at 75 to 79 years of age, with higher rates among NH whites (342.7; 95% CI = 329.6 to 356.2), followed by NH blacks (236.8; 95% CI = 206.8 to 270), NH APIs (176.4; 95% CI = 150.8 to 205.1), and Hispanics (190.3; 95% CI = 165.4 to 217.9) (Supplementary Table 3, available online). NH black women had the highest incidence rates of triple-negative breast cancer across all age groups, with the difference in rates reaching its widest point at ages 60 to 64 and 65 to 69 years, when NH black women were much more likely to be diagnosed with this subtype than were the three other racial/ethnic groups. In particular, the peak triple-negative incidence rate among 65 to 69 year-old NH black women aged 65 to 69 years was 69.5 (95% CI = 57.5 to 83.3), with lower rates among women of the same age in other racial and ethnic groups (eg, NH whites: 36.8, 95% CI = 33.4 to 40.4; NH APIs: 23.6, 95% CI = 16.6 to 32.6; Hispanics: 28.8; 95% CI = 21.7 to 37.4). The HER2-overexpressing tumors (HR+/HER+ and HR−/HER2+) were less common subtypes with fewer observed variations by race/ethnicity compared with both the HR+/HER2− and triple-negative subtypes.

Results from the polytomous logistic regression model are summarized in Table 2. Based on the model results and using the HR+/HER2− tumors as the reference outcome and NH white as the reference covariable, NH blacks and Hispanics were more likely to be diagnosed with triple-negative (NH blacks: OR = 2.0, 95% CI = 1.8 to 2.2; Hispanics: OR = 1.3, 95% CI = 1.2 to 1.5) and HR−/HER2+ breast cancer (NH blacks: OR = 1.4, 95% CI = 1.2 to 1.6; Hispanics: OR = 1.4, 95% CI = 1.2 to 1.6); and NH APIs were less likely to be diagnosed with triple-negative tumors (OR = 0.8; 95% CI = 0.7 to 0.9) but more likely to be diagnosed with both HR+/HER2+ and HR−/HER2+ tumors (OR = 1.2, 95% CI = 1.1 to 1.4; OR = 1.8, 95% CI = 1.5 to 2.1, respectively) (Table 2). Compared with patients with HR+/HER2− breast cancer, those diagnosed with triple-negative, HR+/HER2+, and HR−/HER2+ were 10% to 30% less likely to be aged 65 to 74 or 75 years or older. This observation is consistent with the earlier age of onset seen in Figure 1. Triple-negative cancers had a similar stage distribution compared with HR+/HER2− cancers, but HR+/HER2+ and, in particular, HR−/HER2+ tumors were more likely to present at stage III or IV. Lastly, marked differences in tumor grade were observed across subtypes, with triple-negative, HR+/HER2+, and HR−/HER2+ tumors being 6.4-fold to 20.0-fold more likely to be high grade compared with HR+/HER2− tumors.

Given the large number of case patients with missing data on Bloom–Richardson grade, we conducted sensitivity analyses that included an additional 6118 case patients with an unknown grade. The only appreciable differences observed were with respect to stage and the comparison of triple-negative to HR+/HER2− case patients. Analyses adjusted for grade that included unknown grade as a separate category showed that, compared with HR+/HER2− case patients, triple-negative tumor patients had an elevated risk

**Table 2.** Adjusted odds ratios for patient and tumor characteristics by breast cancer subtypes, SEER-18, excluding Alaska, 2010*

| | HR+/HER2–† | Triple-negative | | HR+/HER2+ | | HR–/HER2+ | |
|---|---|---|---|---|---|---|---|
| | n = 31 500 | n = 5140 | | n = 4270 | | n = 1849 | |
| Characteristics | % Case patients | % Case patients | Odds ratio‡ (95% CI) | % Case patients | Odds ratio‡ (95% CI) | % Case patients | Odds ratio‡ (95% CI) |
| Race/ethnicity | | | | | | | |
| NH white (referent) | 75 | 62 | 1.0 | 68 | 1.0 | 62 | 1.0 |
| NH black | 8 | 19 | 2.0 (1.8 to 2.2) | 11 | 1.2 (1.0 to 1.3) | 13 | 1.4 (1.2 to 1.6) |
| NH API | 8 | 6 | 0.8 (0.7 to 0.9) | 10 | 1.2 (1.1 to 1.4) | 12 | 1.8 (1.5 to 2.1) |
| Hispanic | 9 | 12 | 1.3 (1.2 to 1.5) | 11 | 1.1 (1.0 to 1.2) | 12 | 1.4 (1.2 to 1.6) |
| Age at diagnosis, y | | | | | | | |
| <50 | 19 | 26 | 1.0 (0.9 to1.1) | 30 | 1.3 (1.2 to 1.4) | 26 | 0.9 (0.8 to 1.0) |
| 50–64 (referent) | 37 | 41 | 1.0 | 39 | 1.0 | 44 | 1.0 |
| 65–74 | 23 | 19 | 0.9 (0.8 to 0.9) | 18 | 0.8 (0.7 to 0.9) | 17 | 0.7 (0.6 to 0.8) |
| ≥75 | 20 | 14 | 0.8 (0.7 to 0.9) | 13 | 0.7 (0.6 to 0.8) | 14 | 0.7 (0.6 to 0.8) |
| AJCC 7th stage at diagnosis | | | | | | | |
| I (referent) | 51 | 38 | 1.0 | 43 | 1.0 | 36 | 1.0 |
| II | 31 | 42 | 1.1 (1.0 to 1.2) | 36 | 1.1 (1.0 to 1.1) | 36 | 1.1 (1.0 to 1.2) |
| III | 10 | 16 | 1.0 (0.9 to 1.1) | 16 | 1.2 (1.1 to 1.4) | 20 | 1.6 (1.3 to 1.8) |
| IV | 3 | 5 | 1.0 (0.8 to 1.2) | 5 | 1.4 (1.2 to 1.7) | 8 | 2.1 (1.7 to 2.6) |
| Bloom–Richardson grade | | | | | | | |
| Low (referent) | 34 | 7 | 1.0 | 12 | 1.0 | 7 | 1.0 |
| Medium | 48 | 18 | 1.9 (1.7 to 2.1) | 42 | 2.3 (2.1 to 2.5) | 26 | 2.6 (2.1 to 3.2) |
| High | 17 | 75 | 20.0 (17.8 to 22.5) | 46 | 6.4 (5.8 to 7.1) | 67 | 16.8 (13.9 to 20.5) |

* AJCC = American Joint Committee on Cancer; API = Asian Pacific Islander; HER2, human epidermal growth factor receptor; HR = hormone receptor; NH = non-Hispanic; SEER = Surveillance, Epidemiology, and End Results.

† The $HR^+/HER2^-$ subtype (ie, the most common of all subtypes) serves as a reference group.

‡ All odds ratios are calculated after controlling for race/ethnicities, age, stage, tumor grade, and SEER registries. Analysis is based on complete cases. Polytomous logistic regression. All statistical tests were two-sided.

of being diagnosed with both stage III (OR = 1.2; 95% CI = 1.1 to 1.3) and stage IV (OR = 1.2; 95% CI = 1.1 to 1.4) disease. Analyses not adjusted for grade but adjusted for all of the other covariables showed that, compared with $HR^+/HER2-$ case patients, triple-negative case patients had an elevated risk of being diagnosed with either stage III (OR = 2.1; 95% CI = 1.9 to 2.3) or stage IV (OR = 2.0; 95% CI = 1.7 to 2.2) disease.

## Discussion

This study analyzed recently available data on HER2 status for breast cancer patients from SEER registries (based on 28% of the US population) to demonstrate differences in the occurrence of breast cancer subtypes, defined by ER, PR, and HER2 status. Previous studies carried out in observational studies (8,9,11,13,14) had limited ability to generalize results to the larger population, although data from California have been available and used for epidemiologic studies (6,10,11). The data presented here confirm the higher proportions of more aggressive breast cancer subtypes among younger, NH black, and Hispanic women and notable differences in clinical presentation across subtypes. Additional etiologic studies are recommended to better characterize contributors to age, racial, and ethnic differences in the occurrence of breast cancer subtypes.

Unlike the predominant subtype, $HR^+/HER2^-$, the proportion of women with the triple-negative, $HR^+/HER2^+$, and $HR^-/HER2^+$ subtypes decreased with advancing age such that, although these three comparison groups comprised 35% of case patients aged less than 50 years, they represented only 20% of case patients among women aged 75 years or older. This is consistent with the patterns seen in California (5,6,10,11). These patterns are directly relevant to individualized treatment decisions that influence clinical outcomes (27). Biological factors contributing to these differences are not completely understood. Among BRCA1 carriers, who commonly develop breast cancer at a young age, the vast majority are diagnosed with the triple-negative subtype (28). These mutations are rare, however, and account for a low attributable fraction of triple-negative case patients. Further etiologic studies are needed to more completely characterize contributors to these differences.

NH black women were twice as likely to be diagnosed with triple-negative breast cancer compared with NH whites, and Hispanics were 30% more likely to be diagnosed with triple-negative breast cancer than NH whites. This observation is consistent with existing literature indicating a disproportionate burden of triple-negative disease in these populations, with several studies having documented this among black women (29,30) and among Hispanic women (31). Similar to the unique age-specific pattern of triple-negative subtypes, the etiologic basis for different racial and ethnic patterns remains unclear. NH black, NH API, and Hispanic women also were more likely to be diagnosed with $HR^-/HER2^+$ breast cancer compared with NH white women, with NH API women having the highest risk. Little is known about the basis for these differences given the lower frequency of these $HR^-/HER2^+$ cancers, and studies that have explored their etiologies and risk factors have been hampered by small sample sizes. Looking carefully at individual risk factors such as reproductive history, lactation, weight, physical activity, mammography, postmenopausal hormone use, and longevity could explain the apparent differences in the

diagnosis of breast cancer subtypes by race and ethnicity in SEER areas (32).

These data also suggest some striking differences in stage and grade by breast cancer subtype. Using $HR^+/HER2^-$ as the comparison group in these analyses, little difference was found in the stage distribution of triple-negative case patients, unlike prior studies (29,33); however, triple-negative case patients were substantially more likely to have high-grade cancer (17% vs 75%) (Table 2). Although the difference in grade is well described (8,12,13) after controlling for stage, prior studies also found that triple-negative tumors were more likely to present at an advanced stage (2,6,11). The higher proportion of advanced stage and high-grade tumors among $HR^+/HER2^+$ and $HR^-/HER2^+$ case patients also has been reported previously and is consistent with the known aggressiveness of these tumor subtypes compared with $HR^+/HER2^-$ disease (4,8,13,14).

It is important to acknowledge the limitations of this study. The first limitation relates to missing data for ER, PR, and HER2 status. Although the proportion of case patients missing ER and PR status was low (5.4% and 6.1%, respectively), 8.8% of case patients had missing HER2 data (which led to an overall 12% of case patients missing molecular subtypes). The missing HER2 data were not entirely random but varied by age, stage, race/ethnicity, county-level SES, and registry. The magnitude and direction of potential biases introduced by the missing data are unknown. However, it is likely to differentially underestimate incidence rates by subtypes presented in this article and may also contribute to the observed lack of association between advanced-stage and triple-negative breast cancer. Multiple imputation methods have been used in previous studies (34,35) of SEER data to correct for missing ER status. However, we did not impute missing HER2 status for this analysis because we felt survival time would be an important predictor for missing HER2 observations, which is consequently not available to account for in the imputation model. The second limitation involves possible variations in laboratory techniques for testing biomarkers across multiple hospitals that might be expected in a population-based sample. Third, the data presented here are limited to a single diagnosis year, which may lend some inherent instability to the incidence rates observed, particularly for rarer subtypes. Thus, continued monitoring of subtypes is needed, both within population subgroups and over time. Finally, we acknowledge that there are different approaches to categorizing breast cancer case patients based on HR and HER2 status in the literature; we used the existing HR and HER2 information to best categorize breast cancers that approximate the subtypes of luminal A, luminal B, triple-negative, and HER2-overexpressing tumors (1).

In summary, this study provides large-scale, population-based estimates of incidence rates of breast cancer subtypes defined by ER, PR, and HER2 status in the United States. There were marked differences in the incidence of these subtypes by age and race/ethnicity. These findings have both clinical and public health implications given differences in available treatments and risks of recurrence and mortality by subtype. For example, $ER^-$ breast cancers are twice as likely to be missed by mammographic screening compared with $ER^+$ breast cancers (36). Furthermore, no targeted therapeutic agents currently are available for triple-negative breast cancer. Finally, triple-negative, $ER^+/HER2^+$, and $ER^-/HER2^+$ breast cancers carry a higher risk of mortality compared with $ER^+/HER2^-$ tumors. Understanding of the biological basis for differences in breast cancer subtype incidence and mortality rates across population groups is limited and warrants continued intensive study. SEER data can serve in the future to monitor clinical outcomes in women with different molecular subtypes of breast cancer.


## Funding

Surveillance Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health contracts with SEER registries.

## Acknowledgments

The authors thank SEER registries at the following locations: Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, San Francisco-Oakland, Seattle-Puget Sound, Utah, Los Angeles, San Jose-Monterey, Rural Georgia, Alaska, Greater California, Kentucky, Louisiana, New Jersey, and Greater Georgia. The authors would also like to thank Drs. William F. Anderson of the Division of Cancer Epidemiology and Genetics (DCEG) and Linda C. Harlan of the the Division of Cancer Control and Population Sciences (DCCPS) for providing a very helpful review of the manuscript.


## References


1. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747–752.
2. Prat A, Cheang MCU, Martín M, et al. Prognostic significance of progesterone receptor–positive tumor cells within immunohistochemically defined luminal A breast cancer. *J Clin Oncol*. 2013;31(2):203–209.
3. Fan C, Oh DS, Wessels L, et al. Concordance among gene-expression–based predictors for breast cancer. *New Engl J Med*. 2006;355(6):560–569.
4. Millikan R, Newman B, Tse C, et al. Epidemiology of basal-like breast cancer. *Breast Cancer Res Treat*. 2008;109:123–139.
5. Telli M, Chang E, Kurian A, et al. Asian ethnicity and breast cancer subtypes: a study from the California Cancer Registry. *Breast Cancer Res Treat*. 2011;127:471–478.
6. Clarke CA, Keegan TH, Yang J, et al. Age-specific incidence of breast cancer subtypes: understanding the black-white crossover. *J Natl Cancer Inst*. 2012;104(14):1094–1101.
7. Amirikia K, Mills P, Bush J, et al. Higher population-based incidence rates of triple-negative breast cancer among young African-American women: jmplications for breast cancer screening recommendations. *Cancer*. 2011;117:2747–2753.
8. Carey L, Perou C, Livasy C, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *J Am Med Assoc*. 2006;295:2492.
9. Anderson WF, Luo S, Chatterjee N, et al. Human epidermal growth factor receptor-2 and estrogen receptor expression, a demonstration project using the residual tissue repository of the Surveillance, Epidemiology, and End Results (SEER) program. *Breast Cancer Res Treat*. 2009;113(1):189–196.
10. Keegan T, DeRouen M, Press D, et al. Occurrence of breast cancer subtypes in adolescent and young adult women. *Breast Cancer Res*. 2012;14(2):R55.
11. Kurian A, Fish K, Shema S, et al. Lifetime risks of specific breast cancer subtypes among women in four racial/ethnic groups. *Breast Cancer Res*. 2010;12(6):R99.
12. Cronin KA, Harlan LC, Dodd KW, et al. Population-based estimate of the prevalence of HER-2 positive breast cancer tumors for early stage patients in the US. *Cancer Invest*. 2010;28(9):963–968.
13. Carey L, Dees E, Sawyer L, et al. The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. *Clin Cancer Res*. 2007;13:2329–2334.

14. Haque R, Ahmed SA, Inzhakova G, et al. Impact of breast cancer subtypes and treatment on survival: an analysis spanning two decades. *Cancer Epidemiol Biomarkers Prev*. 2012;21(10):1848–1855.
15. Howlader N, Noone AM, Krapcho M, et al. *SEER Cancer Statistics Review, 1975–2010*. Bethesda, MD: National Cancer Institute; 2013. http://seer.cancer.gov/csr/1975_2010/. Accessed June 17, 2013.
16. SEER*Stat Database: Incidence - SEER 20 Regs, Nov 2012 Sub (1973–2010 varying)—Linked To County Attributes—Total U.S., 1969–2011 Counties. National Cancer Institute, Division of Cancer Control and Population Sciences, Surveillance Research Program, Surveillance Systems Branch; released April 2013, based on the November 2012 submission. http://www.seer.cancer.gov/data/seerstat/nov2012/. Accessed June 17, 2013.
17. Krieger N, Chen JT, Waterman PD, et al. Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter?: the Public Health Disparities Geocoding Project. *Am J Epidemiol*. 2002;156(5):471–482.
18. Singh GK MB, Hankey BF, Edwards BK. Area Socioeconomic Variations in U.S. Cancer Incidence, Mortality, Stage, Treatment, and Survival, 1975–1999. NCI Cancer Surveillance Monograph Series, Number 4. NIH Publication No. 03-5417.Bethesda, MD: National Cancer Institute; 2003.
19. Collaborative Stage Data Collection System Web site. http://www.cancerstaging.org/cstage/manuals/index.html. Accessed June 17, 2013.
20. Hammond ME, Hayes DF, Dowsett M, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). *Arch Pathol Lab Med*. 2010;134(7):e48–e72.
21. Onitilo AA, Engel JM, Greenlee RT, et al. Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival. *Clin Med Res*. 2009;7(1–2):4–13.
22. SEER*Stat Software [computer program]. Version 7.0.4. Bethesda, MD: Surveillance Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute; 2010.
23. Census Population Estimates Web site. http://www.census.gov/popest/methodology/2000-2010_Intercensal_Estimates_Methodology.pdf. Accessed June 17, 2013.
24. Tiwari RC, Clegg LX, Zou Z. Efficient interval estimation for age-adjusted cancer rates. *Stat Methods Med Res*. 2006;15(6):547–569.
25. *AJCC Cancer Staging Manual*, 7th Edition: https://cancerstaging.org/referencestools/deskreferences/Pages/default.aspx. Accessed June 17, 2013.
26. SAS Institute. *Base SAS® 9.3 Utilities: Reference*. Cary, NC: These included age at diagnosis SAS Institute; 2011.
27. von Minckwitz G, Loibl S, Maisch A, et al. Lessons from the neoadjuvant setting on how best to choose adjuvant therapies. *Breast*. 2011;20(Suppl 3):S142–S145.
28. Peshkin BN, Alabek ML, Isaacs C. BRCA1/2 mutations and triple negative breast cancers. *Breast Dis*. 2010;32(1–2):25–33.
29. Ihemelandu CU, Naab TJ, Mezghebe HM, et al. Basal cell–like (triple-negative) breast cancer, a predictor of distant metastasis in African American women. *Am J Surg*. 2008;195(2):153–158,
30. Boyle P. Triple-negative breast cancer: epidemiological considerations and recommendations. *Ann Oncol*. 2012;23(Suppl 6):vi7–vi12.
31. Parise CA, Bauer KR, Brown MM, et al. Breast cancer subtypes as defined by the estrogen receptor (ER), progesterone receptor (PR), and the human epidermal growth factor receptor 2 (HER2) among women with invasive breast cancer in California, 1999–2004. Breast J. 2009;15(6):593–602.
32. Toriola A, Colditz G. Trends in breast cancer incidence and mortality in the United States: implications for prevention. *Breast Cancer Res Treat*. 2013;138(3):665–673.
33. Lin NU, Vanderplas A, Hughes ME, et al. Clinicopathologic features, patterns of recurrence, and survival among women with triple-negative breast cancer in the National Comprehensive Cancer Network. *Cancer*. 2012;118(22):5463–5472.
34. Anderson WF, Katki HA, Rosenberg PS. Incidence of breast cancer in the United States: current and future trends. *J Natl Cancer Inst*. 2011;103(18):1397–402.
35. Howlader N, Noone AM, Yu M, Cronin KA. Use of imputed population-based cancer registry data as a method of accounting for missing information: application to estrogen receptor status for breast cancer. *Am J Epidemiol*. 2012;176(4):347–356.
36. Porter PL, El-Bastawissi AY, Mandelson MT, et al. Breast tumor characteristics as predictors of mammographic detection: comparison of interval- and screen-detected cancers. *J Natl Cancer Inst*. 1999;91(23):2020–2028.

## Notes

Findings and conclusions are the authors' and do not necessarily represent the official positions of their affiliations, or those of the National Cancer Institute, the National Institutes of Health, or the US Department of Health and Human Services.

**Affiliations of authors:** Surveillance Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda MD (NH, SFA, LAGR, KAC); Department of Epidemiology, University of Washington, Seattle, WA (CIL); Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA (CIL); Louisiana Tumor Registry and Epidemiology Program, School of Public Health, Louisiana State University Health Sciences Center, New Orleans, LA (VWC); Cancer Prevention Institute of California, Fremont, CA (CAC); National Cancer Institute Contractor, RiesSearch, LLC, Rockville MD (LAGR).