

# Race, Breast Cancer Subtypes, and Survival in the Carolina Breast Cancer Study

Lisa A. Carey, MD

Charles M. Perou, PhD

Chad A. Livasy, MD

Lynn G. Dressler, PhD

David Cowan, BS

Kathleen Conway, PhD

Gamze Karaca, MSc

Melissa A. Troester, PhD

Chiu Kit Tse, MSPH

Sharon Edmiston, BS

Sandra L. Deming, PhD, MPH

Joseph Geradts, MD

Maggie C. U. Cheang, MMedSci

Torsten O. Nielsen, MD

Patricia G. Moorman, PhD

H. Shelton Earp, MD

Robert C. Millikan, DVM, PhD

**B**REAST CANCER IS A HETEROGENEOUS disease composed of a growing number of recognized biological subtypes. The prognostic and etiologic importance of this diversity is complicated by many factors, including the observation that differences in clinical outcomes often correlate with race. Age-adjusted mortality in the United States from breast cancer in white women is 28.3 deaths per 100 000 compared with 36.4 deaths per 100 000 in African American women.<sup>1</sup> This disparity is particularly pronounced among women younger than 50 years, in whom mortality is 77% higher among African American women compared with white women (11.0 vs 6.3 deaths per 100 000). Breast cancer in African American women has been characterized by higher grade,<sup>2,3</sup> later

**Context** Gene expression analysis has identified several breast cancer subtypes, including basal-like, human epidermal growth factor receptor-2 positive/estrogen receptor negative (HER2+/ER-), luminal A, and luminal B.

**Objectives** To determine population-based distributions and clinical associations for breast cancer subtypes.

**Design, Setting, and Participants** Immunohistochemical surrogates for each subtype were applied to 496 incident cases of invasive breast cancer from the Carolina Breast Cancer Study (ascertained between May 1993 and December 1996), a population-based, case-control study that oversampled premenopausal and African American women. Subtype definitions were as follows: luminal A (ER+ and/or progesterone receptor positive [PR+], HER2-), luminal B (ER+ and/or PR+, HER2+), basal-like (ER-, PR-, HER2-, cytokeratin 5/6 positive, and/or HER1+), HER2+/ER- (ER-, PR-, and HER2+), and unclassified (negative for all 5 markers).

**Main Outcome Measures** We examined the prevalence of breast cancer subtypes within racial and menopausal subsets and determined their associations with tumor size, axillary nodal status, mitotic index, nuclear pleomorphism, combined grade, p53 mutation status, and breast cancer-specific survival.

**Results** The basal-like breast cancer subtype was more prevalent among premenopausal African American women (39%) compared with postmenopausal African American women (14%) and non-African American women (16%) of any age ( $P < .001$ ), whereas the luminal A subtype was less prevalent (36% vs 59% and 54%, respectively). The HER2+/ER- subtype did not vary with race or menopausal status (6%-9%). Compared with luminal A, basal-like tumors had more *TP53* mutations (44% vs 15%,  $P < .001$ ), higher mitotic index (odds ratio [OR], 11.0; 95% confidence interval [CI], 5.6-21.7), more marked nuclear pleomorphism (OR, 9.7; 95% CI, 5.3-18.0), and higher combined grade (OR, 8.3; 95% CI, 4.4-15.6). Breast cancer-specific survival differed by subtype ( $P < .001$ ), with shortest survival among HER2+/ER- and basal-like subtypes.

**Conclusions** Basal-like breast tumors occurred at a higher prevalence among premenopausal African American patients compared with postmenopausal African American and non-African American patients in this population-based study. A higher prevalence of basal-like breast tumors and a lower prevalence of luminal A tumors could contribute to the poor prognosis of young African American women with breast cancer.

*JAMA*. 2006;295:2492-2502

www.jama.com

**Author Affiliations:** Division of Hematology/Oncology (Dr Carey), Departments of Medicine (Drs Dressler and Earp and Mr Cowan), Genetics (Dr Perou and Ms Karaca), and Pathology (Drs Perou and Livasy), School of Public Health, Department of Epidemiology (Drs Conway, Troester, Deming, and Millikan and Mss Tse and Edmiston), University of North Carolina-Lineberger Comprehensive Cancer Center, Chapel Hill; Department of Community and Family Medicine, Duke University

Medical Center, Durham, NC (Dr Moorman); Genetic Pathology Evaluation Centre, University of British Columbia, Vancouver (Dr Nielsen and Ms Cheang); and Roswell Park Cancer Institute, Buffalo, NY (Dr Geradts).

**Corresponding Author:** Lisa A. Carey, MD, Division of Hematology/Oncology, University of North Carolina-Lineberger Comprehensive Cancer Center, CB 7305, 3009 Old Clinic Bldg, Chapel Hill, NC 27599-7305 (Lisa\_Carey@med.unc.edu).

stage at diagnosis,<sup>2,4</sup> and worse survival even after controlling for stage at diagnosis.<sup>4-6</sup> The causes of this observed survival difference are likely multifactorial and include socioeconomic factors,<sup>4</sup> differences in access to screening<sup>7</sup> and treatment,<sup>6</sup> as well as potential biological differences among the cancers themselves.<sup>3,8,9</sup> Biological differences among breast cancers may reflect genetic influences, differences in lifestyle, or nutritional or environmental exposures. In addition, studies that include race as a characteristic must take into account that there is significant disagreement as to how race is measured and interpreted in medical research.<sup>10-12</sup>

Gene expression studies using DNA microarrays have identified several distinct breast cancer subtypes<sup>13</sup> based on an intrinsic gene list that includes 496 genes that differentiate breast cancers into separate groups based only on gene expression patterns. These subtypes differ markedly in prognosis<sup>14-16</sup> and in the repertoire of therapeutic targets they express.<sup>17</sup> The intrinsic subtypes include 2 main subtypes of estrogen receptor (ER)-negative tumors (basal-like and human epidermal growth factor receptor-2 positive/ER- [HER2+/ER-] subtype) and at least 2 types of ER+ tumors (luminal A and luminal B).<sup>14,15</sup> Basal-like tumors typically show low expression of HER2 and ER and exhibit high expression of genes characteristic of the basal epithelial cell layer, including expression of cytokeratins 5, 6, and 17.<sup>13</sup> The HER2+ (ie, gene amplified and/or highly overexpressed protein) tumors fall into at least 2 distinct expression groups: those that are ER- and typically cluster near the basal-like tumors (HER2+/ER- subtype), and those that are ER+ (and may also be progesterone receptor positive [PR+]) and cluster with tumors of luminal cell origins as part of the luminal B subtype.<sup>14,15</sup> The luminal subtype A and B tumors express ER, GATA3, and genes regulated by both ER and GATA3.<sup>18,19</sup> Compared with luminal B tumors, luminal A tumors express higher levels of ER and GATA3 and show more favorable patient outcomes,<sup>15</sup> whereas lu-

minal B tumors more often express human epidermal growth factor receptor-1 (HER1), HER2, and/or cyclin E1.<sup>14,15</sup>

Previous expression studies examined breast cancer subtypes in small data sets derived from frozen tumor banks.<sup>14-16,20,21</sup> The incidence of any of these molecular subtypes in a large population-based study and their relationship with demographic variables have not been systematically evaluated. The Carolina Breast Cancer Study (CBCS) is a population-based, case-control study of environmental and molecular determinants of breast cancer risk.<sup>22</sup> The CBCS is unique in that it oversampled African American and premenopausal women to allow better representation of these 2 subpopulations, making it well-suited for the examination of race- and age-related variables. We used immunohistochemical (IHC) surrogates to identify breast tumor intrinsic subtypes using formalin-fixed, paraffin-embedded tumor blocks collected for CBCS cases, and examined associations between tumor subtypes and race, menopausal status, tumor characteristics, and survival.

## METHODS

### Definition of Breast Cancer IHC Subtypes

Although breast cancer subtypes were originally identified by gene expression analysis using DNA microarrays, large-scale subtyping using gene expression profiling from formalin-fixed, paraffin-embedded samples is not currently feasible. For this reason, we used IHC markers that had been previously verified against gene expression profiles to estimate the prevalence of the intrinsic subtypes in a large population-based epidemiological study of African American and white women. The IHC profiles were developed previously by performing both microarray analysis and IHC for ER, HER2, HER1, and cytokeratin 5/6 on a single series of breast cancers; in that way, we identified combinations of these IHC markers that best matched the gene expression patterns, and then validated these IHC surrogates using a 930-case

tissue microarray from the University of British Columbia.<sup>17</sup> In that earlier study, the IHC-based definitions were luminal (ER+ and HER2-), HER2+ subtype, and basal-like (ER-, HER2-, cytokeratin 5/6+, and/or HER1+). We updated these IHC-based definitions in 2 ways: first, we included PR, which is another widely used breast tumor marker, in the definition of luminal because PR is an ER-regulated gene expressed in most ER+ tumors and is associated with response to hormonal therapy. Second, we recategorized HER2+ tumors into 2 groups based on their ER status since HER2+/ER- tumors cluster separately from HER2+/ER+ tumors in hierarchical clustering analyses.<sup>14,15</sup> In this way, we refined the previous IHC profiles for the breast cancer subtypes and created updated IHC subtype definitions: basal-like (ER-, PR-, HER2-, cytokeratin 5/6+, and/or HER1+), HER2+/ER- subtype (HER2+, ER-, PR-), luminal A (ER+ and/or PR+, HER2-), and luminal B (ER+ and/or PR+, HER2+). This definition for luminal B does not identify all luminal B tumors because only 30% to 50% are HER2+. The other luminal B tumors in this system would be classified with luminal A tumors. Tumors that were negative by IHC for all 5 markers (ER, PR, HER2, HER1, and cytokeratin 5/6) were considered unclassified. These refined IHC profiles are seen in FIGURE 1. In support of the updated profiles, the HER2+ and ER+ tumors (by gene expression) were found mostly within the ER+ tumor dendrogram branch and within the luminal B subtype, whereas the HER2+ and ER- tumors that represent the HER2+/ER- subtype gene expression pattern were seen within a distant ER- tumor dendrogram branch, which suggests that these 2 groups are different.

### Study Population

The CBCS is a population-based, case-control study conducted in 24 counties of eastern and central North Carolina.<sup>22</sup> The goal of the present analysis was to estimate the prevalence of breast cancer subtypes in a population-based

sample of breast cancer cases, and to examine correlations with clinico-pathologic variables and patient survival. The analysis was based on breast cancer cases ascertained between May 1993 and December 1996 (phase 1 of the CBCS) and excluded controls. Newly diagnosed (incident) cases of invasive breast cancer in women between the ages of 20 and 74 years were identified using a rapid ascertainment system developed in collaboration with the North Carolina Central Cancer Registry. Cases were selected by randomized recruitment with predetermined probabilities to increase enrollment of African American women and women younger than 50 years so that these otherwise underrepresented subpopulations would represent approximately 50% of the study population. The sampling strategy was intended to balance

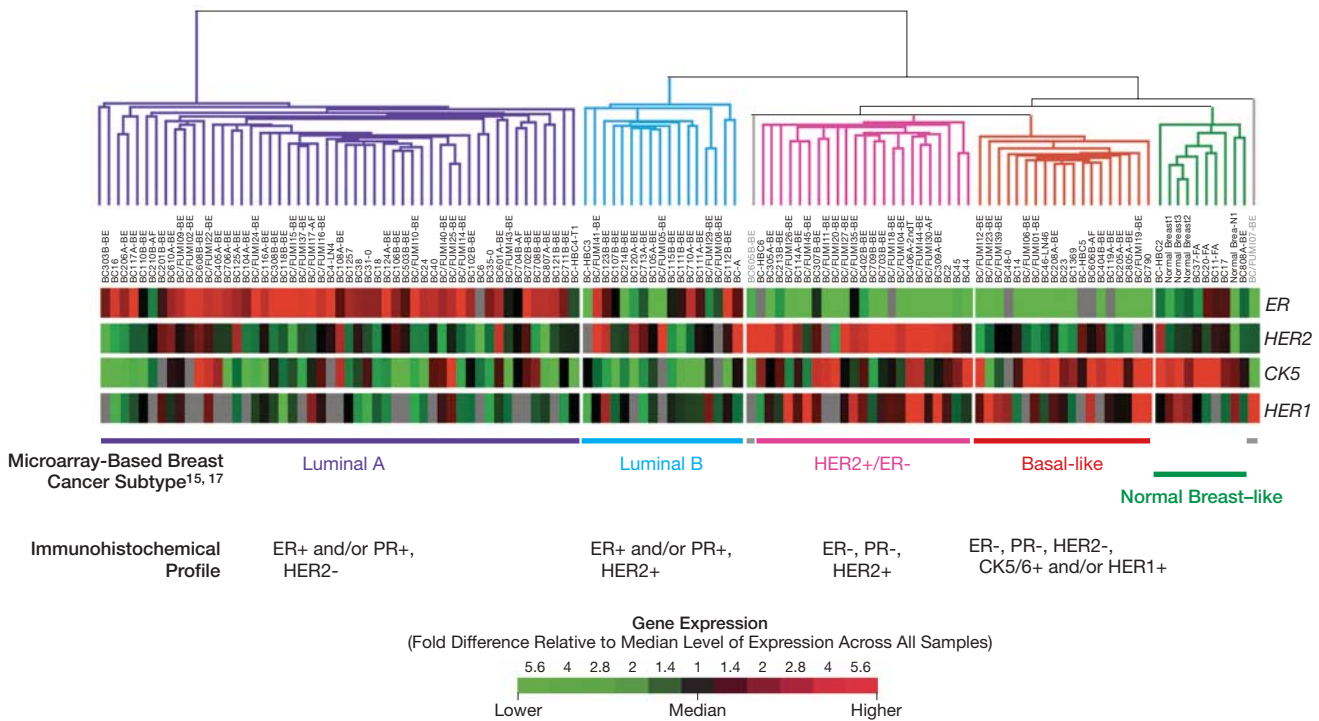
the 4 patient groups (younger African American, older African American, younger non-African American, older non-African American cases) so that statistically valid comparisons could be made for each of the 4 groups. To this end, the schema sampled 100% of African American cases younger than 50 years, 75% of African American cases at least 50 years old, 67% of non-African American cases younger than age 50 years, and 20% of non-African American cases at least 50 years old.<sup>22</sup> Other than the oversampling of younger and African American women by design, the CBCS population is representative of cases reported to the North Carolina Central Cancer Registry in that region of North Carolina during that time, except for a slightly lower proportion of African American cases aged 40 to 59 years with later-stage dis-

ease (2.4% vs 10.2%,  $P = .03$ ).<sup>2</sup> Contact rates in the CBCS were lowest among younger women and African American women, while participation rates were lowest among older women and African American women.<sup>23</sup> Compared with women who participated in the CBCS, nonparticipants were more likely to be of lower socioeconomic status, to have a lower educational level, and to have a recent history of unemployment.<sup>23</sup>

The study procedures for recruitment and enrollment were approved by the institutional review board of the University of North Carolina School of Medicine, and all study participants gave written informed consent.

Race was determined by self-identification and for analysis was categorized as African American or non-African American. Non-African

**Figure 1.** Immunohistochemical Identification of Breast Tumor Intrinsic Subtypes



The 115 patient/tumor sample dendrogram was taken from the hierarchical clustering analysis of the breast intrinsic gene list.<sup>15,17</sup> Tissue samples in gray indicate unknown subtype. The gene expression data for estrogen receptor (*ER*), human epidermal growth factor receptor-2 (*HER2*), *CK5* (cytokeratin), and *HER1* are shown with red squares representing the highest average expression, black representing average gene expression, and green representing the lowest below average expression. Gray indicates gene expression data not available. Note that *PR* (progesterone receptor) was not included in this gene expression analysis because it was not present on these early generation microarrays. Below the gene expression data are the revised immunohistochemical (IHC) classification schema used in this study. *PR* was added to the IHC profile since it is an *ER*-regulated gene expressed in most *ER*+ tumors.

American cases were predominantly white but also included 14 women who reported their race as Native American, Hispanic, Asian American, or multiracial. Information on race was obtained since a primary goal of the CBCS was to better understand breast cancer in African American women. Menopausal status was based on in-person interview data. Sampling was done according to age (since menopausal status was not obtained until interviews), but this did not affect the results (presenting by menopausal status rather than by age <50 and ≥50 years). Women younger than 50 years who had undergone natural menopause, bilateral oophorectomy, or irradiation to the ovaries were classified as postmenopausal and were considered together. In women aged 50 years or older, menopausal status was assigned based on cessation of menstruation.<sup>24</sup>

Centralized review of histology for all tumors was conducted by a single pathologist (J.G.),<sup>2</sup> who was blinded to patient demographics and other study variables. Based on histology, tumors were classified into 6 groups: A (invasive ductal carcinomas not otherwise specified, medullary, apocrine, neuroendocrine carcinomas), B (tubular, mucinous, papillary carcinoma, cribriform carcinomas), C (metaplastic, anaplastic, undifferentiated high-grade carcinomas), D (invasive lobular carcinomas), E (mixed ductal and lobular carcinomas), and unknown (unable to classify). Tumor size, lymph node status, and American Joint Committee on Cancer (AJCC, 5th edition) stage at diagnosis were abstracted from the medical records. Nuclear grade, histologic grade, and mitotic index were previously determined<sup>2</sup> according to the Nottingham modification of the Scarff-Bloom-Richardson criteria.<sup>25</sup> High mitotic index was defined as greater than 10 mitotic figures per 10 high-power fields.

Estrogen receptor and PR status were determined from medical records (80%) or by IHC performed at the University of North Carolina-Lineberger Comprehensive Cancer Center Immunohistochemistry Core Facility in Chapel Hill.<sup>26</sup>

For the cases in which ER and PR status was obtained from the medical record, various clinical laboratories determined the results. About half used IHC on paraffinized tissue with cutoffs for receptor positivity from more than 0% to more than 20%, and about half used biochemical assays on frozen tissue with cutoffs of 10 to 15 fmol/mg. For the remaining tumors, IHC was performed in the Core laboratory at the University of North Carolina.<sup>26</sup> Scoring for IHC was adapted from the method of the University of North Carolina Hospitals Department of Pathology with 5% invasive breast cancer nuclei-positive cells as the cutoff value for ER or PR status. In a 10% random sample of 23 cases that were ER+ and 24 cases that were ER- based on medical records, comparison of the medical record IHC result with IHC done by the Core Laboratory at the University of North Carolina revealed a  $\kappa$  statistic of 0.62, indicating substantial agreement beyond chance<sup>27</sup> with an overall concordance of 81%. The HER2 status was determined using the CB11 antibody (Biogenex, San Ramon, Calif) as previously defined.<sup>28</sup> HER2-positivity was defined as membrane or membrane plus cytoplasmic staining with weak or greater intensity in at least 10% of tumor cells. On a subset of 184 patients, a comparison of 2 independent scorers of the HER2 IHC assay, who were blinded to the other clinical variables, yielded a  $\kappa$  statistic of 0.58, indicating moderate agreement beyond chance<sup>27</sup> with an overall concordance of 82%. Staining for HER1 was categorized using a 0 to 3 scoring system,<sup>17</sup> and our assignment of HER1 positivity was defined as any HER1 staining. Cytokeratin 5/6 was scored positive if any cytoplasmic and/or membranous staining was seen.<sup>29</sup>

A TP53 mutational analysis was performed at the University of North Carolina-Lineberger Comprehensive Cancer Center Molecular Epidemiology Core Facility using single-strand conformational polymorphism analysis with direct sequencing of positive results as previously described.<sup>30</sup> Screening for germline mutations in *BRCA1* was accomplished using multiplex

single-strand conformation analysis as previously described on the first 211 cases in phase 1 of the CBCS.<sup>31</sup>

### Survival Data

The National Death Index provided vital status on CBCS cases as of May 11, 2004. These data were derived from death certificates and included all causes of death for overall survival and disease-specific cause of death for breast cancer-specific survival. In 1 large epidemiological study, the sensitivity of the National Death Index search was 98% and specificity was approximately 100% for breast cancer.<sup>32</sup> Breast cancer-specific survival was determined by the *International Classification of Diseases (ICD) breast cancer codes 174.9 (ICD-9) or C50.9 (ICD-10)* as the underlying cause of death on the death certificate.

### Statistical Analysis

To account for the sampling strategy that systematically overrepresented certain patient groups (eg, younger, African American), analyses are presented stratified by the 4 patient groups. Differences between breast cancer subtypes with regard to clinicopathologic characteristics were examined using 1-way analysis of variance (ANOVA) for age, and  $\chi^2$  tests for the remaining variables. The Fisher exact test was used when expected cell counts were less than 5 using the Monte Carlo method as implemented in SAS.<sup>33</sup> Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated to estimate magnitude and precision of association among breast cancer cases. Odds ratios represent prevalence and were calculated using logistic regression as implemented in SAS version 8.0 (SAS Institute Inc, Cary, NC). *P* values were similar when prevalence ratios were used as the measure of association, but several models did not converge. The reported *P* values reflect the  $\beta$  coefficient in the relevant logistic model. Variables were chosen based on clinical interest, and included age, race, and stage at diagnosis. Because of collinearity with stage, lymph node status was not included with stage in logistic mod-

els. To test for overfitting, we performed the Hosmer-Lemeshow goodness-of-fit test,<sup>34</sup> which did not reveal significant evidence for lack of fit. Likelihood ratio tests for interaction were conducted by comparing models with main effects to models with main effects plus an interaction term. *P* values were not corrected for multiple comparisons since the variables examined (clinicopathologic variables, definitions of breast cancer subtypes) were not independent and thus do not represent separate statistical tests. Survival curves were generated using the Kaplan-Meier method,<sup>35</sup> and the log-rank test<sup>36</sup> was used to compare mean survival across the IHC subtypes. To confirm that the assumptions of the log-rank test were fulfilled,<sup>36</sup> we determined that censoring due to non-breast cancer causes of death was unrelated to breast cancer subtype (*P* = .55), and the proportion of patients in each of the breast cancer subtypes did not differ across the years of enrollment in the study (*P* = .41). Censoring did not differ according to year of enrollment in the study for 5-year breast cancer-specific survival (*P* = .73) or overall survival (*P* = .33). Date and cause of death were obtained from the National Death Index and were thus assigned without knowledge of breast cancer subtype.

As a further test for differences in survival among breast cancer subgroups, we performed univariate Cox regression to estimate hazard ratios for basal-like breast cancer vs luminal A, and for HER2+/ER- breast cancer vs luminal A.<sup>37</sup> Power calculations were performed using a computer program developed by Dupont and Plummer,<sup>38</sup> and concluded that power was very good (70%-80%) or excellent (>80%) for the majority of comparisons in this analysis. Statistical analysis was performed by C.K.T. under the supervision of R.C.M.

## RESULTS

### Patient Population

A total of 1153 incident cases of invasive breast cancer were identified in

phase 1 of the CBCS. Successful contact was obtained in 861 cases (75%), and of these 807 (94%) had tumor blocks or tissue sections for centralized review and IHC. Of the 807 cases, 496 (61%) had both adequate tumor and interpretable IHC data for ER, PR, HER2, cytokeratin 5/6, and HER1, which was a requirement for inclusion in the subtype analysis. These cases included 196 African American and 300 non-African American women. Comparison of these 496 cases included those with the 365 excluded cases (on whom we did not have either adequate tumor tissue or complete IHC data) revealed the following differences: the included cases were more likely to be stage II (51% vs 39%) and less likely to be stage I (39% vs 48%), with little difference seen in stage III (8% vs 10%) or stage IV (3% vs 4%) percentages. The included cases also were more likely to have tumors with high mitotic indices (46% vs 34%, *P* < .001). These differences likely reflected the fact that tumor blocks from patients with smaller tumors were either unavailable or had insufficient tissue for subtype analysis. There were no differences between the included and excluded cases in age, race, menopausal status, lymph node status, nuclear grade, histologic grade, or survival.

### IHC Subtype Associations With Clinical and Demographic Data

Characteristics of the 496 CBCS cases with IHC data, overall and according to IHC subtypes, are presented in TABLE 1. The IHC subtypes differed significantly by age (*P* < .001), race (*P* = .03), menopausal status (*P* = .008), combined race and menopausal status (*P* < .001), axillary lymph node status at time of diagnosis (*P* = .04), histology group (*P* < .001), nuclear grade (*P* < .001), histologic grade (*P* < .001), and mitotic index (*P* < .001). Patients with luminal A and B tumors were older than the other patients, and patients with the HER2+/ER- subtype had the highest prevalence of positive lymph nodes. Patients with basal-like tumors were more likely to be African Ameri-

can, premenopausal, and to have tumors with high nuclear grade, high histologic grade, and high mitotic index. Basal-like tumors also showed the highest prevalence of unfavorable histologies (group C: metaplastic, anaplastic, and undifferentiated high-grade carcinomas).

In the overall study population, the prevalence of the basal-like subtype was 20% (100 cases total). The prevalence of basal-like breast cancer was significantly higher in African American breast cancer cases, comprising 52 of 196 African American women (26%) vs 48 of 300 non-African American cases (16%) (Table 1). Basal-like tumors were also more frequent in premenopausal cases, comprising 64 of 261 (24%) vs 36 of 235 (15%) postmenopausal cases. These prevalence estimates should be interpreted with caution, because they do not reflect the sampling probabilities used to define eligible cases in the CBCS. To account for the sampling strategy, separate estimates were derived for each of the 4 patient groups defined a priori (TABLE 2). The high prevalence of basal-like tumors in African American women was mostly seen in premenopausal women, in whom the prevalence was 39%. The prevalence of basal-like breast cancer in premenopausal African American women was significantly elevated compared with postmenopausal African American (14%) or non-African American women (16%) of any age (*P* < .001) (Table 2). The difference in prevalence of basal-like breast cancer between premenopausal and postmenopausal cases was statistically significant among African American cases (*P* < .001), but not among non-African American cases (*P* = .94). The luminal A subtype, conversely, was less frequent among premenopausal African American women (36%) compared with postmenopausal African American (59%) or non-African American (54%) women. The higher prevalence of basal-like breast cancers in younger African American patients was maintained when we stratified on stage at diagnosis. For example, among cases

**Table 1.** Characteristics of Carolina Breast Cancer Study Patients With Immunohistochemical Marker Data

Characteristic	No. (%)						P Value*
	All Cases (N = 496)	Basal-like (n = 100)	HER2+ /ER- (n = 33)	Luminal A (n = 255)	Luminal B (n = 77)	Unclassified (n = 31)	
Age, mean (SD), y	50 (12)	46 (10)	47 (9)	52 (12)	50 (12)	45 (11)	<.001
Race							
African American	196 (40)	52 (52)	16 (48)	93 (36)	25 (32)	10 (32)	.03
Non-African American	300 (60)	48 (48)	17 (52)	162 (64)	52 (68)	21 (68)	
Menopausal status							
Premenopausal	261 (53)	64 (64)	18 (55)	118 (46)	39 (51)	22 (71)	.008
Postmenopausal	235 (47)	36 (36)	15 (45)	137 (54)	38 (49)	9 (29)	
AJCC stage							
I	184 (39)	23 (24)	9 (28)	108 (44)	29 (39)	15 (48)	.06
II	242 (51)	59 (62)	17 (53)	114 (47)	40 (54)	12 (39)	
III	36 (8)	8 (8)	4 (13)	17 (7)	4 (5)	3 (10)	
IV	13 (3)	5 (5)	2 (6)	4 (2)	1 (1)	1 (3)	
Missing	21	5	1	12	3	0	
Lymph node status							
Positive	189 (39)	41 (41)	18 (56)	86 (34)	35 (47)	9 (29)	.04
Negative	298 (61)	58 (59)	14 (44)	165 (66)	39 (53)	22 (71)	
Missing	9	1	1	4	3	0	
ER status							
Positive	295 (60)	0	0	220 (86)	75 (97)	0	<.001
Negative	201 (40)	100 (100)	33 (100)	35 (14)	2 (3)	31 (100)	
PR status							
Positive	280 (56)	0	0	214 (84)	66 (86)	0	<.001
Negative	216 (44)	100 (100)	33 (100)	41 (16)	11 (14)	31 (100)	
Combined ER-PR status							
ER+ /PR+	243 (49)	0	0	179 (70)	64 (83)	0	<.001†
ER+ /PR-	52 (11)	0	0	41 (16)	11 (14)	0	
ER- /PR+	37 (7)	0	0	35 (14)	2 (3)	0	
ER- /PR-	164 (33)	100 (100)	33 (100)	0	0	31 (100)	
HER2 immunohistochemistry							
Positive	110 (22)	0	33 (100)	0	77 (100)	0	<.001
Negative	386 (78)	100 (100)	0	255 (100)	0	31 (100)	
Histology group‡							
A	375 (76)	84 (84)	31 (94)	178 (70)	61 (79)	21 (68)	<.001†
B	18 (4)	0	0	15 (6)	1 (1)	2 (7)	
C	20 (4)	10 (10)	0	8 (3)	1 (1)	1 (3)	
D	38 (8)	0	0	31 (12)	5 (7)	2 (7)	
E	45 (9)	6 (6)	2 (6)	23 (9)	9 (12)	5 (16)	
Unknown	0	0	0	0	0	0	
Nuclear grade							
Marked pleomorphism	212 (43)	80 (80)	25 (76)	66 (26)	21 (27)	20 (65)	<.001
Slight/moderate	283 (57)	19 (20)	8 (24)	189 (74)	56 (73)	11 (35)	
Missing	1	1	0	0	0	0	
Histologic grade							
Poorly differentiated	321 (65)	81 (82)	23 (70)	149 (58)	43 (56)	25 (81)	<.001
Well-/moderately differentiated	174 (35)	18 (18)	10 (30)	106 (42)	34 (44)	6 (19)	
Missing	1	1	0	0	0	0	
Combined grade (Nottingham)							
I	121 (25)	2 (2)	2 (6)	91 (36)	20 (26)	6 (19)	<.001†
II	144 (29)	14 (14)	6 (19)	85 (33)	33 (43)	6 (19)	
III	227 (46)	82 (84)	24 (75)	78 (31)	24 (31)	19 (62)	
Missing	4	2	1	1	0	0	
Mitotic index							
High, >10 per 10 hpf	226 (46)	85 (87)	22 (69)	78 (31)	25 (32)	16 (52)	<.001
Low, ≤10 per 10 hpf	267 (54)	13 (13)	10 (31)	177 (69)	52 (68)	15 (48)	
Missing	3	2	1	0	0	0	

Abbreviations: AJCC, American Joint Committee on Cancer; ER, estrogen receptor; hpf, high-power field; HER2, human epidermal growth factor receptor-2; PR, progesterone receptor.

\*Comparing 5 subgroups (basal-like, HER2+ /ER-, luminal A, luminal B, unclassified) using analysis of variance to test for differences in means, and  $\chi^2$  or Fisher exact test for the remaining characteristics.

†Fisher exact test.

‡Group A: invasive ductal carcinomas not otherwise specified, medullary, apocrine, neuroendocrine carcinomas; B: tubular, mucinous, papillary carcinoma, cribriform carcinomas; C: metaplastic, anaplastic, undifferentiated high-grade carcinomas; D: invasive lobular carcinomas; E: mixed ductal and lobular carcinomas; unknown: unable to classify.

with stage I disease, the prevalence of basal-like breast cancer was 40% in premenopausal African American women, 6% in postmenopausal African American women, 10% in premenopausal non-African American women, and 8% in postmenopausal non-African American women ( $P = .001$ ). This difference by race and menopausal status was not seen in the other ER- subtype (HER2+/ER-), which also was associated with high grade (Table 1).

Odds ratios for the association of breast cancer subtypes with lymph node status, histologic grade, and mitotic index are presented in TABLE 3, with the luminal A subtype (the most common IHC subtype representing 51% of the cases) serving as the referent group. Odds ratios were adjusted for age, stage, and race. Compared with the luminal A

subtype, patients with basal-like tumors were 2.1 times more likely to be African American ( $P = .004$ ). Likelihood ratio tests showed a significant interaction between race and menopausal status for developing the basal-like subtype ( $P = .02$ ), but not HER2+/ER- ( $P = .49$ ), luminal B ( $P = .62$ ), or unclassified tumors ( $P = .58$ ) compared with luminal A. In comparison with luminal A tumors and after adjustment for age, race, and stage, the basal-like subtype was 11 times more likely to have high mitotic index ( $P < .001$ ), 9.7 times more likely to have high nuclear grade ( $P < .001$ ), and 2.5 times more likely to have high histologic grade ( $P = .003$ ). The basal-like subtype was not associated with the presence of positive axillary lymph nodes at the time of diagnosis ( $P = .53$ ), whereas

both HER2+ subtypes (HER2+/ER- and luminal B) were significantly more likely to have positive lymph nodes at presentation ( $P = .04$ ). Notably, a strong association with high histologic, nuclear, and mitotic grade was seen for both subtypes of ER- tumors, namely the basal-like and HER2+/ER- tumors. However, the HER2+/ER- subtype was not significantly associated with race or menopausal status.

The TP53 sequence-based mutation analysis was performed on 330 of the 496 IHC classified breast cancer cases, of which 84 (25%) had TP53 mutations. The presence of TP53 mutations differed significantly with IHC subtype: 44% (28 of 63) of basal-like tumors and 43% (10 of 23) of HER2+/ER- subtype tumors contained TP53 mutations, whereas only 23% (12 of 52) of luminal B and 15% (25 of 175) of luminal A were mutation-positive ( $P < .001$ ). These findings were in agreement with previous comparisons of the breast tumor intrinsic subtypes and TP53 mutation status<sup>14</sup> as well as previous demonstration of a high proportion of p53-mutant tumors in BRCA1 and cytokeratin 5/6-positive tumors.<sup>39,40</sup>

A subset of CBCS patients were screened for BRCA1 germline mutations.<sup>31</sup> Of the 496 cases assayed, 211 were screened for mutations in BRCA1, with 4 carriers and 1 variant of un-

**Table 2.** Prevalence of Breast Cancer Subtypes According to Race and Menopausal Status

Tumor Status	All Cases	No. (%)			
		African American*		Non-African American†	
		Premenopausal (n = 97)	Postmenopausal (n = 99)	Premenopausal (n = 164)	Postmenopausal (n = 136)
Basal-like	100	38 (39)	14 (14)	26 (16)	22 (16)
HER2+/ER-	33	9 (9)	7 (7)	9 (6)	8 (6)
Luminal A	255	35 (36)	58 (59)	83 (51)	79 (58)
Luminal B	77	9 (9)	16 (16)	30 (18)	22 (16)
Unclassified	31	6 (6)	4 (4)	16 (10)	5 (4)

\* $P < .001$ ,  $\chi^2$  test for basal-like vs other tumor types in premenopausal vs postmenopausal African American women.  
 † $P = .94$ ,  $\chi^2$  test for basal-like vs other tumor types in premenopausal vs postmenopausal non-African American women.

**Table 3.** Adjusted Odds Ratios for Patient and Tumor Characteristics by Breast Cancer Subtype\*

Characteristic	Basal-like (n = 100)		HER2+/ER- (n = 33)		Luminal B (n = 77)		Unclassified (n = 31)	
	OR (95% CI)	P Value	OR (95% CI)	P Value	OR (95% CI)	P Value	OR (95% CI)	P Value
African American vs non-African American†	2.1 (1.3-3.4)	.004	1.8 (0.8-3.8)	.13	0.9 (0.5-1.5)	.63	0.9 (0.4-2.1)	.86
Premenopausal vs postmenopausal‡	0.9 (0.4-1.9)	.74	0.5 (0.2-1.5)	.21	0.9 (0.4-2.2)	.83	1.4 (0.4-6.0)	.61
Stage III-IV vs I-II§	1.4 (0.6-2.9)	.41	2.1 (0.8-5.7)	.16	0.8 (0.3-2.1)	.58	1.3 (0.4-4.3)	.62
Lymph node-positive vs negative§	1.2 (0.7-1.9)	.53	2.2 (1.0-4.7)	.04	1.7 (1.0-2.9)	.05	0.7 (0.3-1.7)	.45
Nuclear grade: marked pleomorphism vs slight/moderate‡	9.7 (5.3-18.0)	<.001	6.8 (2.8-16.3)	<.001	1.0 (0.5-1.8)	.92	4.6 (2.0-10.4)	<.001
Histologic grade: poorly differentiated vs well-/moderately‡	2.5 (1.4-4.6)	.003	1.2 (0.5-2.7)	.70	0.9 (0.5-1.5)	.67	2.6 (1.0-6.8)	.05
Combined grade (Nottingham) III vs I+II‡	8.3 (4.4-15.6)	<.001	6.2 (2.4-16.0)	<.001	1.0 (0.5-1.7)	.87	3.4 (1.5-7.9)	.004
Mitotic index >10 vs ≤10 per 10 hpff	11.0 (5.6-21.7)	<.001	4.3 (1.8-10.5)	.001	1.0 (0.6-1.8)	.95	2.3 (1.0-5.3)	.04

Abbreviations: CI, confidence interval; hpf, high-power field; OR, odds ratio.  
 \*The luminal A (ER+ and/or progesterone receptor positive, HER2-) subtype acts as the referent group.  
 †Adjusted for age (11-level ordinal variable) and stage (I, II, III+IV).  
 ‡Adjusted for age, race, and stage.  
 §Adjusted for age and race.

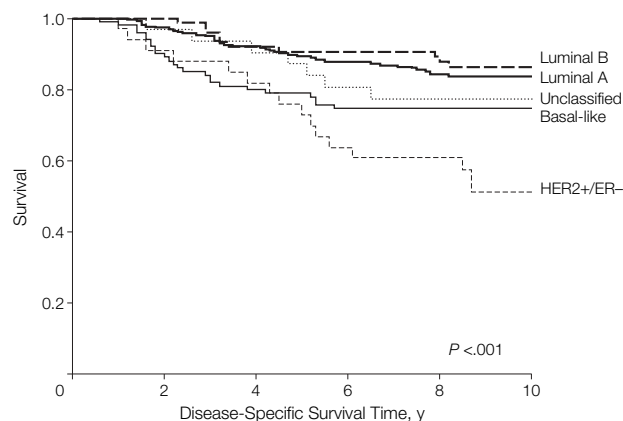
known effects being identified. The *BRCA1* mutation carriers comprised 1 luminal A tumor, 1 unclassified tumor, and 2 basal-like tumors. Although these numbers were very small, the data were consistent with earlier findings that most *BRCA1* mutant tumors show the basal-like phenotype<sup>15,41,42</sup> and that most *BRCA1* mutant tumors do not show HER2 positivity.<sup>43</sup>

### Survival by IHC Subtype

The maximum duration of follow-up for the CBCS phase 1 cases was 11.2 years (minimum of 8.1 years). During this period of observation, the study patients had 73% overall survival (232 deaths among 861 cases). Of the 232 deaths, 170 were considered breast cancer-specific, giving an overall disease-specific survival of 80% (691 of 861). African American cases had worse breast cancer-specific survival (74%) compared with non-African American cases (84%) ( $P < .001$ ). Age, race, menopausal status, stage, ER status, PR status, *TP53* mutation status, mitotic index, nuclear grade, and histologic grade were also significant predictors of breast cancer-specific survival ( $P < .001$  for each).

The breast cancer subtypes also differed significantly in breast cancer-specific survival ( $P < .001$ ): basal-like subtype (75%), HER2+/ER- subtype (52%), luminal A (84%), luminal B (87%), and unclassified (77%). Kaplan-Meier survival curves for breast cancer-specific survival are presented in FIGURE 2. A steep fall in breast cancer-specific survival was observed in the first 4 to 5 years for the basal-like and HER2+/ER- tumors, with particularly poor survival for the HER2+/ER- subtype. A similar early relapse pattern has been described for *BRCA1* tumors.<sup>44,45</sup> Over the entire observation period, breast cancer-specific survival was significantly worse among basal-like (hazard ratio, 1.8; 95% CI, 1.1-2.9;  $P = .03$ ) and HER2+/ER- breast cancer patients (hazard ratio, 3.5; 95% CI, 1.9-6.2;  $P < .001$ ) compared with luminal A as the referent group.

**Figure 2.** Survival Analysis of the Carolina Breast Cancer Study Cases Grouped Using the Refined Breast Tumor Immunohistochemical Intrinsic Subtypes



No. at Risk					
Luminal B	76	67	65	63	17
Luminal A	245	226	213	197	81
Unclassified	30	28	25	23	9
Basal-like	89	78	72	71	25
HER2+/ER-	30	27	21	20	7

Prevalence was luminal A=51%, luminal B=16%, basal-like=20%, HER2+/ER-=7%, and unclassified=6%.

The difference in survival by breast cancer subtype was seen both among lymph node-positive patients ( $P = .01$ ) and lymph node-negative patients ( $P = .03$ ). Data were sparse after stratifying on lymph node status and should be interpreted with caution. Breast cancer-specific survival within lymph node-positive patients by subtype was the following: basal-like (51%), HER2+/ER- (39%), luminal A (65%), luminal B (83%), and unclassified (44%). Within the lymph node-negative patients, breast cancer-specific survival was the following: basal-like (93%), HER2+/ER- (71%), luminal A (94%), luminal B (92%), and unclassified (91%).

The outcomes in premenopausal African American cases did not become more similar to the other groups when basal-like cases were removed. The breast cancer-specific survival by racial and menopausal subsets without basal-like breast cancers still differed significantly: premenopausal African American 64%, postmenopausal African American 81%, premenopausal non-African American 81%, and postmenopausal non-African American

91% ( $P < .001$ ). These data suggest that factors other than subtype, such as access to treatment, could also be influencing survival in younger African American women.

### COMMENT

Gene expression profiling has identified breast cancer intrinsic subtypes that predict distinct clinical outcomes<sup>14,15</sup> and which have been shown to be present in women of multiple ethnicities.<sup>46</sup> The basal-like subtype has been associated with poor clinical outcomes,<sup>15,16</sup> which likely reflect this subtype's high proliferative capacity<sup>14-16</sup> as well as the lack of directed therapies since basal-like tumors do not typically express ER- or overexpress HER2.<sup>17</sup> To facilitate investigation of the population-based frequencies of the basal-like breast cancer subtype, we refined an IHC-based assay to identify the main breast tumor intrinsic subtypes. We used the IHC method for categorization and determined for the first time the population-based prevalence of these subtypes. Although IHC-based assays do not provide as much biological insight into tumor biology as



mRNA-based assays containing thousands of genes, this IHC assay allowed classification of tumors into categories that have demonstrated associations between intrinsic subtypes and proliferation rates, overall survival, *TP53* status, and *BRCA1* mutation status.<sup>14,15,17,29,41,42</sup> The reproducible correlations across different studies and when using different assays (IHC and DNA microarray expression profiles) shows that we are tracking common tumor subtypes with similar biologic characteristics and clinical behaviors across distinct patient sets. The IHC-based classification system also allows analyses of subtypes to be conducted in patient populations where fresh tissue is not available.

In the population-based CBCS, the prevalence of the basal-like and luminal A breast cancer subtypes was strongly influenced by race and menopausal status; the highest prevalence of basal-like and lowest prevalence of luminal A tumors were observed among premenopausal African American breast cancer patients. Because the CBCS is a population-based sample, within defined race and age groups estimates of prevalence are likely to be representative of the underlying North Carolina population.<sup>2</sup> Differences between the CBCS and breast cancer patients reported to the North Carolina Central Cancer Registry include a lower proportion of African American women between the ages of 40 and 59 years with higher-stage tumors and lower participation among women from lower socioeconomic and educational strata.<sup>2,23</sup> Each of these factors could actually produce an underestimate of the prevalence of more aggressive breast cancer subtypes (basal-like and HER2+/ER-) among younger African American cases enrolled in the CBCS. However, this potential bias may have been partially offset by the fact that the analysis of IHC markers in the CBCS was based on patients with larger tumors.

A high frequency of basal-like tumors was observed in a study of breast cancer in Nigerian women, among whom ER-negative and HER2-nega-

tive tumors comprised 87 of 148 women, or 59% of total cases.<sup>47</sup> According to gene expression studies, ER-negative breast tumors fall into 1 of 2 categories,<sup>14,15</sup> namely basal-like tumors (ER-, PR-, and HER-) and the HER2+/ER- subtype (HER2+/ER-) (Figure 1). The HER2+/ER- group, which is also a high-grade and ER-negative tumor group, did not vary significantly with age or race. These findings suggest that associations between premenopausal breast cancer, race, and hormone status in the CBCS was driven by an excess of the basal-like subtype. Breast cancers that develop among *BRCA1* mutation carriers are generally basal-like.<sup>15,41,42</sup> However, very few *BRCA1* mutation carriers were present in the CBCS, with 2 out of the 4 known carriers falling into the basal-like category. No *BRCA1* carriers were identified among the African American cases tested in the CBCS and only a single variant of unknown biological significance was identified.<sup>31</sup> Thus *BRCA1* variants are unlikely to explain the high prevalence of basal-like breast cancer in younger African American patients in this study.

Basal-like breast cancers in the CBCS exhibited aggressive features, including high proliferative capacity (measured by mitotic index), high histologic grade, high nuclear grade, and frequent *TP53* mutations. Even after adjustment for age, race, and stage, the association of basal-like and HER2+/ER- subtypes with aggressive features remained significant. These findings were expected given the high expression of the proliferation cluster of genes in microarray analyses of basal-like and HER2+/ER- subtype tumors.<sup>13-15,48</sup> The association of race with high-grade breast tumors and ER negativity has been previously reported.<sup>2,3</sup> However, our study suggests that this association is driven by the increased prevalence of basal-like tumors and not by an increase in HER2+/ER- subtype.

The observation that the intrinsic breast cancer subtypes carry different prognoses was confirmed in the CBCS. Disease-specific survival was significantly lower among breast cancer cases

with basal-like and HER2+/ER- tumors, and more favorable among cases with luminal A tumors. The HER2+/ER- subtype appears particularly prone to early and frequent relapse, befitting the clinical experience with HER2 overexpressing tumors<sup>49</sup>; the CBCS cases in this study were diagnosed between 1993 and 1996 and were not treated with the anti-HER2 monoclonal antibody trastuzumab. Basal-like tumors were more frequent in younger African American women in the CBCS, and could contribute to their poor prognosis compared with other breast cancer patients. However, when cases of basal-like tumors were removed, the breast cancer-specific survival remained significantly worse among premenopausal African American cases. As noted previously, this may reflect the impact on prognosis of access to care, treatment, or other differences. In other words, while the high incidence of the poor-prognosis basal-like subtype may contribute to their relatively worse outcome, it does not entirely explain the poor outcomes seen in younger African Americans. We lacked treatment data in the CBCS, so we could not examine interactions between IHC subtypes and efficacy of cancer therapy. Examination of tumor microarray data using patients treated with surgery alone also suggests that these subtypes are prognostic and reflect the natural history of these tumors.<sup>15</sup> Interestingly, unlike HER2+/ER- and luminal B tumors, the basal-like subtype was not associated with involvement of positive axillary lymph nodes, a finding that was previously noted in a study of cytokeratin 5/6-positive tumors that oversampled *BRCA1* tumors.<sup>40</sup> Since basal-like breast cancers still carried a poor prognosis, it is possible, as suggested by others,<sup>40</sup> that this finding reflects a predominantly hematogenous, rather than lymphatic, pattern of dissemination. Further studies are needed to address this issue.

Further research is needed to confirm the finding that the basal-like breast tumor subtype shows a high prevalence in young African American breast cancer patients. In studies of

race and breast cancer, it is important that race be evaluated in the context of other variables such as stage at diagnosis and tumor histology. Information on breast cancer risk factors will help to determine whether basal-like tumors have a different underlying etiology compared with other types of breast cancer. Since *BRCA1* carriers tend to develop basal-like tumors, there may be other inherited genetic variants that predispose to developing specific subtypes of breast cancer.<sup>15,21</sup> The absence of *BRCA1* carriers among African American breast cancer patients in the CBCS suggests that genes other than *BRCA1* could predispose women to basal-like breast cancers; however, environmental and socioeconomic factors could also play a role in the observed distribution of breast cancer subtypes. Notably, in the CBCS, the prevalence of *BRCA1* mutations was 0 in African Americans and low (3.3%) in non-African Americans.<sup>31</sup> Most importantly, our data suggest that epidemiological studies of breast cancer in African American women should consider the joint distribution of ER, PR, and HER2 status (ie, subtypes), rather than rely on ER and PR status alone. Previous analyses typically group together HER2+/ER- tumors with basal-like tumors under the ER-negative designation; however, in the CBCS, HER2+/ER- tumors were not associated with race or menopausal status.

The high prevalence of basal-like tumors in younger African American women could contribute to their higher breast cancer mortality. Additional studies of long-term survival among patients with specific breast cancer subtypes are needed. Clinical trials aimed at identifying therapeutic approaches to the management of basal-like breast cancer are also needed, especially for young African American women.

**Author Contributions:** Dr Millikan had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

**Study concept and design:** Carey, Perou, Moorman, Millikan.

**Acquisition of data:** Livasy, Dressler, Conway, Edmiston, Deming, Geradts, Cheang, Nielsen, Moorman, Millikan.

**Analysis and interpretation of data:** Carey, Perou, Livasy, Dressler, Cowan, Conway, Karaca, Troester, Tse, Deming, Cheang, Nielsen, Earp, Millikan.

**Drafting of the manuscript:** Carey, Perou, Tse, Nielsen, Millikan.

**Critical revision of the manuscript for important intellectual content:** Carey, Perou, Livasy, Dressler, Cowan, Conway, Karaca, Troester, Edmiston, Deming, Geradts, Cheang, Nielsen, Moorman, Earp, Millikan.

**Statistical analysis:** Tse, Deming, Cheang, Millikan.

**Obtained funding:** Perou, Dressler, Conway, Earp, Millikan.

**Administrative, technical, or material support:** Carey, Livasy, Dressler, Cowan, Conway, Karaca, Troester, Edmiston, Nielsen, Moorman, Millikan.

**Study supervision:** Carey, Perou, Dressler, Conway, Edmiston, Nielsen, Millikan.

**Financial Disclosures:** None reported.

**Funding/Support:** This work was supported by an award to the University of North Carolina for a Breast Cancer Specialized Program of Research Excellence (SPORE) from the National Cancer Institute (NIH/NCI P50-CA58223), a grant from the General Clinical Research Centers Program of the Division of Research Resources/National Institutes of Health (M01RR00046 awarded to Dr Carey), and by the NCI (RO1-CA-101227-01 awarded to Dr Perou).

**Role of the Sponsor:** All study funding was from public grants for scientific research. The funding organizations had no role in the design and conduct of the study; the collection, analysis, and interpretation of the data; or the preparation, review, or approval of the manuscript.

**Previous Presentation:** This work was presented in part at the 40th Annual Meeting of the American Society of Clinical Oncology; New Orleans, La; June 2004.

**Acknowledgment:** For their critical review, we thank Barbara Rimer, PhD, School of Public Health; and Paul Godley, MD, PhD, and Matthew G. Ewend, MD, School of Medicine, University of North Carolina. They were not compensated for their time.

**REFERENCES**

1. Surveillance, Epidemiology, and End Results Program (SEER) SEER Stat Database: Mortality—All COD, Public-Use With State, Total U.S. for Expanded Races/Hispanics (1990-2001). National Cancer Institute, DC-CPS, Surveillance Research Program, Cancer Statistics Branch, released April 2004. <http://www.seer.cancer.gov>.
2. Furberg H, Millikan R, Dressler L, Newman B, Geradts J. Tumor characteristics in African American and white women. *Breast Cancer Res Treat*. 2001;68:33-43.
3. Porter PL, Lund MJ, Lin MG, et al. Racial differences in the expression of cell cycle-regulatory proteins in breast carcinoma. *Cancer*. 2004;100:2533-2542.
4. Eley JW, Hill HA, Chen VW, et al. Racial differences in survival from breast cancer: results of the National Cancer Institute Black/White Cancer Survival Study. *JAMA*. 1994;272:947-954.
5. Clegg LX, Li FP, Hankey BF, Chu K, Edwards BK. Cancer survival among US whites and minorities: a SEER (Surveillance, Epidemiology, and End Results) Program population-based study. *Arch Intern Med*. 2002;162:1985-1993.
6. Shavers VL, Brown ML. Racial and ethnic disparities in the receipt of cancer treatment. *J Natl Cancer Inst*. 2002;94:334-357.
7. del Carmen MG, Hughes KS, Halpern E, et al. Racial differences in mammographic breast density. *Cancer*. 2003;98:590-596.
8. Ademuyiwa FO, Olopade OI. Racial differences in genetic factors associated with breast cancer. *Cancer Metastasis Rev*. 2003;22:47-53.
9. Joslyn SA. Racial differences in treatment and survival from early-stage breast carcinoma. *Cancer*. 2002;95:1759-1766.

10. Mountain JL, Risch N. Assessing genetic contributions to phenotypic differences among "racial" and "ethnic" groups. *Nat Genet*. 2004;36(11 suppl):S48-S53.

11. Schwartz RS. Racial profiling in medical research. *N Engl J Med*. 2001;344:1392-1393.

12. Tate SK, Goldstein DB. Will tomorrow's medicines work for everyone? *Nat Genet*. 2004;36(11 suppl):S48-S42.

13. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406:747-752.

14. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001;98:10869-10874.

15. 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;100:8418-8423.

16. Sotiriou C, Neo SY, McShane LM, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci U S A*. 2003;100:10393-10398.

17. Nielsen TO, Hsu FD, Jensen K, et al. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin Cancer Res*. 2004;10:5367-5374.

18. Finlin BS, Gau CL, Murphy GA, et al. RERG is a novel RAS-related, estrogen-regulated and growth-inhibitory gene in breast cancer. *J Biol Chem*. 2001;276:42259-42267.

19. Usary J, Llaca V, Karaca G, et al. Mutation of GATA3 in human breast tumors. *Oncogene*. 2004;23:7669-7678.

20. Hedenfalk I, Duggan D, Chen Y, et al. Gene-expression profiles in hereditary breast cancer. *N Engl J Med*. 2001;344:539-548.

21. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002;415:530-536.

22. Newman B, Moorman PG, Millikan R, et al. The Carolina Breast Cancer Study: integrating population-based epidemiology and molecular biology. *Breast Cancer Res Treat*. 1995;35:51-60.

23. Moorman PG, Newman B, Millikan RC, Tse CK, Sandler DP. Participation rates in a case-control study: the impact of age, race, and race of interviewer. *Ann Epidemiol*. 1999;9:188-195.

24. Millikan RC, Pittman GS, Newman B, et al. Cigarette smoking, N-acetyltransferases 1 and 2, and breast cancer risk. *Cancer Epidemiol Biomarkers Prev*. 1998;7:371-378.

25. Genestie C, Zafrani B, Asselain B, et al. Comparison of the prognostic value of Scarff-Bloom-Richardson and Nottingham histological grades in a series of 825 cases of breast cancer: major importance of the mitotic count as a component of both grading systems. *Anticancer Res*. 1998;18(1B):571-576.

26. Huang WY, Newman B, Millikan RC, Schell MJ, Hulka BS, Moorman PG. Hormone-related factors and risk of breast cancer in relation to estrogen receptor and progesterone receptor status. *Am J Epidemiol*. 2000;151:703-714.

27. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.

28. Millikan R, Eaton A, Worley K, et al. HER2 codon 655 polymorphism and risk of breast cancer in African Americans and whites. *Breast Cancer Res Treat*. 2003;79:355-364.

29. van de Rijn M, Perou CM, Tibshirani R, et al. Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. *Am J Pathol*. 2002;161:1991-1996.

30. Conway K, Edmiston SN, Cui L, et al. Prevalence and spectrum of p53 mutations associated with smoking in breast cancer. *Cancer Res.* 2002;62:1987-1995.
31. Newman B, Mu H, Butler LM, Millikan RC, Moorman PG, King MC. Frequency of breast cancer attributable to BRCA1 in a population-based series of American women. *JAMA.* 1998;279:915-921.
32. Rich-Edwards JW, Corsano KA, Stampfer MJ. Test of the National Death Index and Equifax Nationwide Death Search. *Am J Epidemiol.* 1994;140:1016-1019.
33. SAS Support technical FAQ (354). What exact and Monte Carlo methods are available and in which procedures and releases do they appear? <http://support.sas.com/faq/003/FAQ00354.html>. Accessed February 8, 2006.
34. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. New York, NY: Wiley & Sons; 1989.
35. Bland JM, Altman DG. Survival probabilities (the Kaplan-Meier method). *BMJ.* 1998;317:1572.
36. Bland JM, Altman DG. The logrank test. *BMJ.* 2004;328:1073.
37. Cox DR, Oakes D. *Analysis of Survival Data*. London, England: Chapman & Hall; 1984.
38. Dupont WD, Plummer WD. Power and sample size calculations: a review and computer program. *Control Clin Trials.* 1990;11:116-128.
39. Crook T, Brooks LA, Crossland S, et al. p53 Mutation with frequent novel condons but not a mutator phenotype in BRCA1- and BRCA2-associated breast tumours. *Oncogene.* 1998;17:1681-1689.
40. Foulkes WD, Brunet JS, Stefansson IM, et al. The prognostic implication of the basal-like (cyclin E high/p27 low/p53+/glomeruloid-microvascular-proliferation+) phenotype of BRCA1-related breast cancer. *Cancer Res.* 2004;64:830-835.
41. Foulkes WD, Stefansson IM, Chappuis PO, et al. Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. *J Natl Cancer Inst.* 2003;95:1482-1485.
42. Olopade OI, Grushko T. Gene-expression profiles in hereditary breast cancer. *N Engl J Med.* 2001;344:2028-2029.
43. Grushko TA, Blackwood MA, Schumm PL, et al. Molecular-cytogenetic analysis of HER-2/neu gene in BRCA1-associated breast cancers. *Cancer Res.* 2002;62:1481-1488.
44. Foulkes WD, Wong N, Brunet JS, et al. Germline BRCA1 mutation is an adverse prognostic factor in Ashkenazi Jewish women with breast cancer. *Clin Cancer Res.* 1997;3:2465-2469.
45. Stoppa-Lyonnet D, Ansquer Y, Dreyfus H, et al. Familial invasive breast cancers: worse outcome related to BRCA1 mutations. *J Clin Oncol.* 2000;18:4053-4059.
46. Yu K, Lee CH, Tan PH, Tan P. Conservation of breast cancer molecular subtypes and transcriptional patterns of tumor progression across distinct ethnic populations. *Clin Cancer Res.* 2004;10:5508-5517.
47. Olopade OI, Ikpatt FO, Dignam JJ, et al. "Intrinsic gene expression" subtypes correlated with grade and morphometric parameters reveal a high proportion of aggressive basal-like tumors among black women of African ancestry. Paper presented at: American Society of Clinical Oncology Annual Meeting; June 2004; New Orleans, La.
48. Whitfield ML, Sherlock G, Saldanha AJ, et al. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell.* 2002;13:1977-2000.
49. Press MF, Bernstein L, Thomas PA, et al. HER-2/neu gene amplification characterized by fluorescence in situ hybridization: poor prognosis in node-negative breast carcinomas. *J Clin Oncol.* 1997;15:2894-2904.

I suppose every scholar has had the experience of reading something in a book which was significant to him, but which he could never find again. Sure he is that he read it there; but no one else ever read it, nor can he find it again.

—Ralph Waldo Emerson (1803-1882)