Received: 16 August 2020 | Revised: 28 September 2020 | Accepted: 30 September 2020

DOI: 10.1002/ijc.33331

IJC International Journal of Cancer

CANCER EPIDEMIOLOGY

Breast cancer among Asian Indian and Pakistani Americans: A surveillance, epidemiology and end results-based study

Jaya M. Satagopan[1,2] | Antoinette Stroup[1,2,3] | Anita Y. Kinney[1,2] | Tina Dharamdasani[1] | Shridar Ganesan[4] | Elisa V. Bandera[1,2]

[1]Department of Biostatistics & Epidemiology, Rutgers School of Public Health, Piscataway, New Jersey

[2]Cancer Prevention and Control Program, Rutgers Cancer Institute of New Jersey, New Brunswick, New Jersey

[3]New Jersey State Cancer Registry, State of New Jersey Department of Health, New Brunswick, New Jersey

[4]Clinical Investigations and Precision Therapeutics Program, Rutgers Cancer Institute of New Jersey, New Brunswick, New Jersey

Correspondence
Jaya M. Satagopan, Department of Biostatistics and Epidemiology, Rutgers School of Public Health, 683 Hoes Lane West, Piscataway, NJ 08854.
Email: satagopj@sph.rutgers.edu

Funding information
Breast Cancer Research Foundation; National Institutes of Health, Grant/Award Numbers: R01CA197402, R01CA243547, R01CA202752, P30CA072720

Abstract

Breast cancer incidence is increasing among Asian Indian and Pakistani women living in the United States. We examined the characteristics of breast cancer in Asian Indian and Pakistani American (AIPA) and non-Hispanic white (NHW) women using data from the surveillance, epidemiology and end results (SEER) program. Breast cancer incidence rates were estimated via segmented Poisson regression using data between 1990 and 2014 from SEER 9 registries, including New Jersey and California. Disease characteristics, treatment and survival information between 2000 and 2016 for 4900 AIPA and 482 250 NHW cases diagnosed after age 18 were obtained from SEER 18 registries and compared using descriptive analyses and multivariable competing risk proportional hazards regression. Breast cancer incidence was lower in AIPA than NHW women, increased with age and the rate of increase declined after age of 46 years. AIPA women were diagnosed at significantly younger age (mean (SD) = 54.5 (13.3) years) than NHW women (mean (SD) = 62 (14) years, $P < .0001$) and were more likely than NHW cases ($P < .0001$) to have regional or distant stage, higher grade, estrogen receptor-negative, progesterone receptor-negative, triple-negative or human epidermal growth factor receptor 2-enriched tumors, subcutaneous or total mastectomy, and lower cumulative incidence of death due to breast cancer (hazard ratio = 0.79, 95% CI: 0.72-0.86, $P < .0001$). AIPA had shorter median follow-up (52 months) than NHW cases (77 months). Breast cancer in AIPA women has unique characteristics that need to be further studied along with a comprehensive evaluation of their follow-up patterns.

KEYWORDS
incidence, death, follow-up, segmented regression, competing risk

1 | INTRODUCTION

South Asians are the fastest growing major ethnic group in the United States, increasing by 40% between 2010 and 2017.[1] Of around 5.4 million South Asians in the United States, 92% are of Asian Indian or Pakistani origin.[1] Many Asian Indian and Pakistani Americans (AIPA) are employed in medical and technology fields, giving them a model minority stereotype. However, 13% to 22% of AIPA live under 125% of poverty line and lack health coverage.[2] More than 80% of AIPA live in male-led households,[2] and 40% live in multigenerational households.[3] Only 34% to 54% of AIPA women are in the workforce, of which 37% to 53% earn less than $12,500 annually.[2] Women

Abbreviations: AIPA, Asian Indian and Pakistani Americans; CI, confidence interval; ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; NHW, non-Hispanic white; PR, progesterone receptor; SA, South Asian; SEER, surveillance epidemiology and end results.

constitute 59% of AIPA adults without a high school diploma, but only 35% of adults with graduate degrees.[2] Given this diverse cultural background of AIPA, it is becoming increasingly critical to study the health of this rapidly growing population.

The AIPA population is at increased risk for cancer and other preventable diseases due to the effect of migration on lifestyle and sociocultural determinants of health.[4] The incidence of breast cancer, which is amenable to screening and early detection, is increasing among AIPA women.[5] Few studies have examined the characteristics of breast tumors in AIPA.[6,7] In these prior studies, AIPA women were younger when diagnosed and had more aggressive clinicopathologic features such as higher stage, higher grade and more estrogen receptor-negative (ER–) tumors than non-Hispanic white (NHW) cases. In analyses of a few discrete time points adjusted for tumor characteristics[6] and in unadjusted analyses,[7] these studies noted better, albeit nonsignificant, survival in AIPA than NHW cases. AIPA had significantly better survival than NHW cases in a study of early-stage breast cancers.[8]

Yet, for a fast-growing minority, surprisingly little is known about breast cancer in this socioculturally unique population. There is a dearth of nuanced assessment and interpretations of breast cancer incidence rates, correlates of breast cancer outcomes and follow-up patterns in AIPA, who are an understudied racial/ethnic minority in the United States for disproportionate burden of cancer.[9] These observations suggest a critical need for comprehensive investigations of age-specific breast cancer incidence, follow-up patterns and death due to breast cancer in the presence of other competing causes of death in AIPA women, and use the insights gained from this investigation to plan effective disease control strategies for this population. Therefore, we pursued these investigations using data from the surveillance, Epidemiology and End Results (SEER) program (https://seer.cancer.gov). As a benchmark for comparison, we also examined age-specific breast cancer incidence and death due to breast cancer and due to other causes in NHW women.

The contributions of our study include a segmented Poisson regression analysis to understand the rate of age-specific breast cancer incidence in AIPA and NHW women, and a proportional hazards regression analysis to understand death due to breast cancer in AIPA and NHW cases by accounting for death due to other causes as competing events. Our results provide insights into breast cancer burden for AIPA women, and suggest several hypotheses to guide future studies to better understand the factors influencing breast cancer etiology and prognosis in this population, with the ultimate goal of developing culturally relevant strategies to increase their utilization of preventive health care.

What's new?

While breast cancer incidence among Asian Indian and Pakistani American (AIPA) women is increasing, incidence patterns and breast cancer survival in this population remain understudied. Here, analyses of data for 1990-2014 in the SEER database show that breast cancer incidence among AIPA women increased with age, though the rate of increase declined after age 46. Relative to non-Hispanic white women, AIPA women were younger at diagnosis and exhibited more aggressive clinical-pathological characteristics but had shorter median follow-up time and lower disease-specific mortality. Comprehensive studies of breast cancer risk factors in AIPA women and their follow-up care are warranted.

2 | MATERIALS AND METHODS

2.1 | Data for age-specific incidence

We abstracted age-specific breast cancer incidence rates for AIPA and NHW women from the database of SEER 9 registries plus New Jersey, San Jose-Monterey and Los Angeles registries,[10] which includes diagnosis between 1990 and 2014. This database provides the total number of breast cancer cases and the size of the at-risk population, based on the US Census, in 5-year age groups for NHW and AIPA populations.

2.2 | Data for breast cancer survival

We examined the SEER 9 and SEER 18 databases to identify a data set for survival analysis. The SEER 9 database[11] includes cases diagnosed between 1975 and 2016 from nine registries (Supplementary Table S1). The SEER 18 database[12] includes cases diagnosed between 2000 and 2016 from the SEER 9 registries and nine additional registries, of which one—the Alaska Native Tumor Registry—does not include any South Asians. From each database, we obtained a case listing of age at diagnosis of breast cancer, year of diagnosis, tumor grade, ER status, progesterone receptor (PR) status, human epidermal growth factor receptor 2 (HER2) status, tumor subtype, chemotherapy, radiation therapy, type of surgery, cause of death and survival time for AIPA and NHW cases.

We obtained tumor stage using the variables "SEER Historic Stage A," "SEER Summary Stage 1977," "SEER Summary Stage 2000" and "SEER Combined Summary Stage 2000" for cases diagnosed from 1975 to 1994, 1995 to 2000, 2001 to 2003, and 2004 to 2016, respectively, and categorized stage as localized, regional and distant. SEER categorizes chemotherapy as "yes" or "no/unknown." Patients with "no/unknown" chemotherapy were deemed to have not received any chemotherapy. Cases with beam radiation, combination of beam with implants or isotopes, radiation with unspecified method, radioactive implants (including brachy therapy) and radioisotopes were deemed to have received radiation treatment. All others were deemed to have not received radiation. We used the variable "RX Summ—Surg Prim Site (1998)+" to categorize surgery as partial mastectomy/

lumpectomy, radical mastectomy, subcutaneous or total mastectomy, no surgery and unknown.

Patients were categorized as alive, died of breast cancer and died of other causes based on "cause of death to site recode variable." Survival time was measured as the number of months elapsed between the date of diagnosis and the date of death or, for alive cases, the date of last follow-up. The study end date was 31 December 2016.

2.3 | Statistical analysis

2.3.1 | Poisson regression model for age-specific breast cancer incidence rate

For each race/ethnicity E (= AIPA or NHW), denote Y_{AE} as the number of breast cancer cases and N_{AE} as the size of the at-risk population at age A (or in an age group with mid-point A). These data are available from the SEER age-specific cancer incidence database. First, we examined the age distribution of the at-risk population of AIPA and NHW women by plotting N_{AE} vs A and calculated the average and SD of the ages. Next, we fitted a Poisson model for Y_{AE}, with the mean or expected value of the number of breast cancer cases as $E\{Y_{AE}\} = N_{AE} \times R(A,E)$, where $R(A,E)$ is the breast cancer incidence rate at age A for race/ethnicity E.[13] Prior studies of carcinogenesis demonstrated a log-linear relationship between age and cancer rate, that is, a linear relationship between the natural logarithm of age and the natural logarithm of cancer rate.[14,15] Studies in NHW breast cancer cases found distinct log-linear relationships before and after a breakpoint age.[15] Therefore, we modeled the rate as a log-linear function of A, given by[13]:

$$\log\{R(A,E)\} = \alpha_E + \beta_E \times \log(A) + \theta_E \times [\log(A) - \log(B_E)] \times I[\log(A) > \log(B_E)],$$

where for race/ethnicity E, α_E is the intercept term, β_E is the slope of the log-linear relationship between age and breast cancer incidence rate, B_E is an unknown breakpoint age, θ_E is difference in the slope before and after the breakpoint age and the notation $I[C]$ is the indicator function taking value 1 if the condition C in the square parentheses is true and taking value 0 otherwise. We fitted this model using data for women of 30 years of age and above due to the low rate of breast cancer before 30 years of age. We estimated the breakpoint age, intercept, slopes, their standard errors and conducted hypothesis tests using previously described methods in the literature.[13,16,17] Briefly, for each race/ethnicity E, we estimated the breakpoint age B_E and obtained maximum-likelihood estimates of the intercept and the slopes using an iterative approach, and obtained the SEs using the delta method.[16,17] We tested the null hypothesis that $\beta_E = 0$ against a two-sided alternative using a Wald statistic. When $\theta_E = 0$, it means that there is no breakpoint age B_E. Since the parameter θ_E only appears in the model when B_E exists, we tested the null hypothesis $\theta_E = 0$ against a two-sided alternative using a score statistic in the presence of a nuisance parameter[16] (see Supplementary Material).

2.3.2 | Survival analysis

We examined the sample sizes, the number of deaths due to breast cancer and median follow-up times of AIPA and NHW cases in the SEER 9 and SEER 18 data sets and selected one data set for survival analysis. In the selected data set, we calculated descriptive statistics for age at diagnosis, stage, grade, ER, PR, HER2, tumor subtype, chemotherapy, radiation therapy and surgery separately for AIPA and NHW cases. For survival analysis, we categorized age at diagnosis as "<40 years," "40 to 49 years" and "≥50 years" and created binary variables for stage (localized vs regional or distant) and grade (I and II vs III and IV). We also categorized year of diagnosis as "2000 to 2004," "2005 to 2009" and "2010 to 2016." Cases with unknown stage, grade, surgery and borderline or unknown ER and PR statuses were removed from survival analyses. Since HER2 information and tumor subtype information were not available for cases diagnosed before 2010, these characteristics were not included in survival analyses.

We calculated median follow-up times separately for AIPA and NHW cases as the median of the follow-up time of cases without any event. We estimated the cumulative probability of death due to breast cancer by accounting for death due to other causes as competing risk factors[18-22] and compared the cumulative probability of death due to breast cancer between AIPA and NHW women.[21,22] We examined whether the cumulative probability of death due to breast cancer differed significantly between AIPA and NHW cases in distinct categories of age, stage, grade, ER and PR status, and examined whether this difference varied across the categories by evaluating the interaction between race/ethnicity and each variable.[22]

We estimated the hazard ratio of death due to breast cancer in AIPA (vs NHW) cases, adjusting for year of diagnosis, age, stage, grade, ER status, PR status, chemotherapy, radiation therapy and surgery, using two multivariable proportional hazards regression models.[22,23] In the first model, we used a sub-distribution hazard function, which allows us to estimate the effect of race/ethnicity on the cumulative incidence of death due to breast cancer over time and offers insights into prognosis.[23] Under this approach, the risk set at a given time includes all women who have not died of breast cancer by that time. In the second model, we used a cause-specific hazard function, which allows us to estimate the effect of race/ethnicity on the rate of death due to breast cancer. Under this approach, which offers insights into associations,[23] the risk set at a given time includes those subjects who are currently alive (ie, have not died of breast cancer or have not died of other causes by that time). Finally, to understand whether the effect of race/ethnicity on death due to breast cancer may be explained via its effect on death due to other causes, we also estimated the adjusted hazard ratio of death due to other causes in AIPA (vs NHW) cases using the two hazard functions.[21-23]

In all our hypothesis tests, comparisons with two-tailed P values <.05 were considered statistically significant.

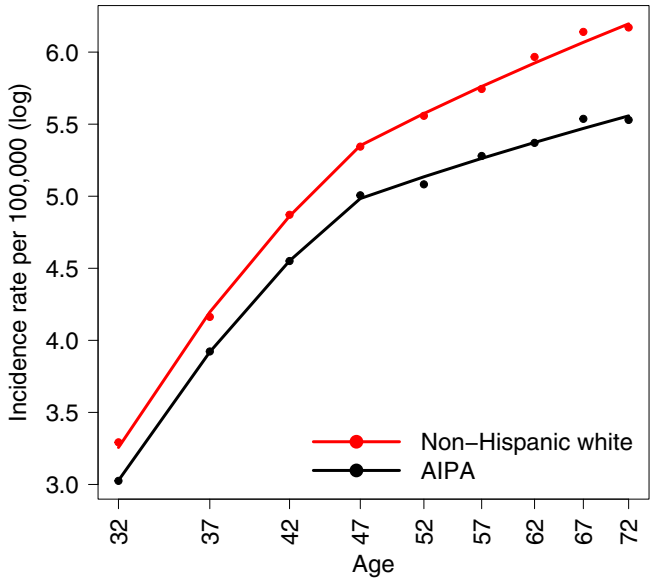

FIGURE 1 Age-specific breast cancer incidence rates for Asian Indian and Pakistani Americans (AIPA; black dots and curve) and non-Hispanic white (red dots and curve) women. The curves were obtained using segmented regression analysis relating logarithm of age to logarithm of incidence rates. The data used in this analysis come from the SEER 9 database including New Jersey and two California registries and contains cases diagnosed between 1990 and 2014

2.4 | Statistical software

All data analyses were conducted using the SEER*Stat software version 8.3.6,[24] SAS version 9.4[25] and the R programming language version 3.6.0[26] with libraries segmented,[17] cmprsk[27] and fastcmprsk.[28]

3 | RESULTS

3.1 | Age-specific breast cancer incidence rate

In the SEER cancer incidence database, the population of AIPA women were significantly younger than the population of NHW women (average age (SD): 31.1 (19.3) years for AIPA and 40.3 (23.3) years for NHW, $P < .0001$, Supplementary Figure 1).

AIPA women had lower age-specific incidence of breast cancer than NHW women (Figure 1). The estimated breakpoint ages from segmented Poisson regression models were 46 years (SE = 1.02 years) and 46.4 years (SE = 0.16 years) for AIPA and NHW women, respectively. Breast cancer rate increased significantly with slopes of 2.2 (SE = 0.11, $P < .0001$) and 2.3 (SE = 0.02, $P < .0001$) in AIPA and NHW women, respectively, before the breakpoint age. Breast cancer incidence continued to increase significantly, albeit at a slower rate, with slopes of 0.83 (SE = 0.06, $P < .0001$) and 1.22 (SE = 0.005, $P < .0001$) in AIPA and NHW women, respectively, after the breakpoint age (see also Table 1).

3.2 | Choice of SEER database for survival analysis

Since SEER 9 and SEER 18 databases had considerably more NHW than AIPA cases (Supplementary Table S1), our choice of a database for survival analysis was based on the characteristics of SA cases. The SEER 9 database had 2076 AIPA cases, of which 1753 (84%) were diagnosed between 2000 and 2016. The SEER 18 database had 5704 AIPA cases diagnosed between 2000 and 2016, of which 3710 were from New Jersey and three registries in California that are not part of SEER 9. The total number of AIPA events in the SEER 9 and SEER 18 databases were 223 and 502, respectively, and the median follow-up times were 57 months and 54 months, respectively. Thus, the SEER 18 database had 275% more AIPA cases, 225% more events and only 3 fewer months of follow-up than the SEER 9 database. Therefore, we selected the SEER 18 database for survival analysis.

3.3 | Patient characteristics

Table 2 shows the characteristics of AIPA and NHW cases. AIPA were diagnosed at a significantly younger age than NHW cases (mean age at diagnosis = 54.5 years for AIPA and 62 years for NHW cases; $P < .0001$; Figure 2), regardless of tumor characteristics at diagnosis (Supplementary Figure S2). AIPA were also significantly more likely than NHW cases to be diagnosed with regional or distant stage ($P < .0001$), higher grade ($P < .0001$), ER– ($P < .0001$) and PR– ($P < .0001$) tumors. Of the cases diagnosed after 2010, AIPA cases were more likely to have HER2+ ($P < .0001$) and triple-negative or HER2-enriched ($P < .0001$) tumors. AIPA and NHW cases were equally likely to receive radiation ($P = 0.08$). AIPA were more likely to receive chemotherapy ($P < .0001$) and mastectomy than NHW cases ($P = .002$). Among cases undergoing mastectomy, AIPA were more likely to receive subcutaneous or total mastectomy than NHW ($P < .0001$), regardless of stage.

Complete data were available for 4900 AIPA and 482 250 NHW cases. The remaining 804 AIPA cases and 88 667 NHW cases had missing data for at least one of the following variables: stage, grade, ER, PR and surgery. Cases with missing data were more likely to have distant stage and Grade IV disease, negative PR status and not receive chemotherapy, radiation therapy or surgery (Supplementary Table S2).

A total of 115 (2.3%) AIPA cases were diagnosed in 2000, and the annual number of new cases increased progressively to 497 (10%) in 2016, while the annual number of new NHW diagnosis was fairly uniformly distributed with around 28 500 (6%) new cases each year (Supplementary Figure S3). The median follow-up times based on AIPA and NHW cases having complete data were 52 and 77 months, respectively. A total of 376 (8%) AIPA and 47 981 (10%) NHW cases died of breast cancer, while 195 (4%) AIPA and 62 853 (13%) NHW cases died of other causes. The most common other causes of death were heart diseases (53 AIPA and 21 176 NHW cases), cerebrovascular disease (15 AIPA and 5380 NHW cases), lung cancer (12 AIPA and 4796 NHW cases), other malignancies (21 AIPA and 3217 NHW cases) and diabetes mellitus (9 AIPA and 1947 SA cases).

TABLE 1 Results of segmented Poisson regression analyses

Population	Estimated breakpoint (95% CI)	Estimated slope (95% CI) Before breakpoint	After breakpoint
AIPA	46.0 (44.0-48.0)	2.20 (1.99-2.41)	0.83 (0.71-0.95)
NHW	46.4 (46.1-46.7)	2.31 (2.29-2.35)	1.22 (1.21-1.23)

Note: Separate analyses were performed for South Asian (SA) and non-Hispanic White (NHW) women from the SEER 9 database including data from New Jersey and two California registries.
Abbreviations: AIPA, Asian Indian and Pakistani American; CI, confidence interval; NHW, non-Hispanic white; SEER, surveillance, epidemiology and end results.

3.4 | Survival analysis

The estimated 5-year cumulative probability of death due to breast cancer, our main outcome of interest, for the entire study sample was 8.2% (SE = 0.0004) (Figure 3). The characteristics of significantly high cumulative probability of death due to breast cancer were (Supplementary Figure S4): age at diagnosis before 40 years ($P < .0001$), regional or distant stage ($P < .0001$), Grade III or IV ($P < .0001$), ER– ($P < .0001$) and PR– ($P < .0001$) tumors. There was no statistically significant interaction between race/ethnicity and tumor characteristics (Supplementary Table S3).

In multivariable analysis based on the sub-distribution hazard function, adjusted for year of diagnosis, age, stage, grade, ER status, PR status, chemotherapy, radiation therapy and surgery, AIPA cases had significantly lower incidence of death due to breast cancer (hazard ratio = 0.79, 95% CI = 0.72-0.86, $P < .0001$) and lower incidence of death due to other causes (hazard ratio = 0.55, 95% CI = 0.47-0.64, $P < .0001$) than NHW cases. Similarly, AIPA also had significantly lower cause-specific hazard of death due to breast cancer (hazard ratio = 0.76, 95% CI = 0.69-0.84, $P < .0001$) and lower cause-specific hazard or death due to other causes (hazard ratio = 0.43, 95% CI = 0.46-0.61, $P < .0001$) than NHW cases. These results are summarized in Table 3.

4 | DISCUSSION

Asian Indian and Pakistani Americans are a fast-growing ethnic minority in the United States and face an increasing burden of breast cancer. Yet, the characteristics of breast cancer incidence and outcome in this population remain under studied. Our analyses attempt to address this issue by examining age-specific incidence rates, tumor characteristics and breast cancer-specific mortality in Asian Indian and Pakistani women living in the United States.

In our analyses, breast cancer incidence rate was lower in AIPA women than in NHW women. If, at any age, the rate of carcinogenesis process leading to cancer incidence is equivalent to the rate of breast tissue aging,[15] the lower rate of increase of breast cancer incidence in AIPA cases than in NHW cases after the breakpoint age suggests slower breast tissue aging in AIPA women than in NHW women after age 46 years. This might imply that AIPA and NHW women have different postmenopausal estrogen levels. Obesity, which leads to increased serum estrogen levels, is increasing in AIPA women.[29] AIPA women are at higher risk for developing weight-related diseases such as Type II diabetes, hypertension, dyslipidemia, coronary heart diseases and cancer even at lower body mass index levels compared to those in NHW women,[29] suggesting possible differences in body composition. Data on breast tissue, body mass index, body composition, estrogen levels and other obesity-related biomarkers in premenopausal and postmenopausal AIPA women are not available in our current study. A detailed study of these factors is needed to fill the knowledge gap regarding their role on breast incidence in the AIPA population.

In our analyses, as well as in prior works based on breast cancer cases diagnosed from 1988 to 2006[6] and 1988 to 2008,[7] AIPA cases were younger than NHW cases. In the SEER cancer incidence database, the at-risk population of AIPA women were younger than NHW women. Thus, we conjecture that the characteristics of breast cancer in AIPA women examined in our study very likely reflect disease occurring in this relatively young at-risk population. In our analyses, as also observed in a prior work,[8] AIPA women had more aggressive disease than NHW cases even after adjusting for age, ER status and PR status. In age-adjusted analyses, compared to NHW cases, AIPA cases had 16% lower odds of local stage tumors than regional or distant stage tumors, and 37% higher odds of Grade 3 or 4 tumors than Grade 1 or 2 tumors (see Supplementary Tables S4 and S5).

Therefore, studies of breast tissue, tissue aging, genetic, hormonal, clinical, reproductive and lifestyle factors are warranted to better understand potential drivers of aggressive disease in AIPA women. These data are not available from the SEER databases. Hence, independent studies of AIPA breast cancer cases that collect comprehensive data on these risk factors are needed. Studies of genetic risk factors for breast cancer in AIPA women are currently lacking,[30] and data on reproductive factors are sparse. In studies conducted in Asia, mean ages at menarche for Asian Indian and Pakistani women were 13.8 years and 11.7 years, respectively[31,32]; median ages at first child birth were 19.8 years and 22.2 years, respectively,[33] and the mean ages at menopause were 46.2 and 47.1 years, respectively.[34,35] In contrast, the median age at menarche for NHW women in the United States is 12.8 years,[36] the mean age at first child birth is 26.3 years[37,38] and the median age at menopause is 51.2 years.[38] Studies in AIPA breast cancer survivors are also needed to examine whether early age at first child birth contributes to overall lower incidence by lowering the risk of ER+ cancers but increased proportion of ER– tumors.

TABLE 2 Characteristics of South Asian and non-Hispanic white breast cancer patients from the SEER 18 database

Characteristics	NHW	AIPA	*P* value
Age at diagnosis[a]			
Mean (SD)	62.0 (13.8)	54.5 (13.3)	<.0001
Stage[b]			
Localized	361 543 (64.5%)	3261 (58.4%)	<.0001
Regional	169 968 (30.3%)	1998 (35.8%)	
Distant	29 063 (5.2%)	326 (5.8%)	
Grade[c]			
Grade I	123 500 (23.6%)	859 (16.5%)	<.0001
Grade II	229 513 (43.9%)	2125 (40.7%)	
Grade III	164 747 (31.5%)	2187 (41.9%)	
Grade IV	5028 (1.0%)	49 (0.9%)	
ER[c]			
Negative	90 761 (17.4%)	1160 (21.8%)	<.0001
Positive	432 168 (82.6%)	4172 (78.2%)	
PR[c]			
Negative	147 056 (28.5%)	1641 (31.0%)	<.0001
Positive	369 762 (71.5%)	3645 (69.0%)	
Chemotherapy			
No	353 380 (61.9%)	2760 (48.4%)	<.0001
Yes	217 537 (38.1%)	2944 (51.6%)	
Radiation			
No	281 457 (49.3%)	2746 (48.1%)	.08
Radiation	289 460 (50.7%)	2958 (51.9%)	
Surgery[c]			
None	42 861 (7.6%)	490 (8.7%)	<.0001
Partial mastectomy/lumpactomy	309 728 (54.7%)	2728 (48.3%)	
Radical mastectomy	118 989 (21%)	1243 (22%)	
Subcutaneous or total mastectomy	95 076 (17%)	1190 (21%)	
Surgery in localized stage cases[c]			
None	9698 (3%)	100 (3.1%)	<.0001
Partial mastectomy/lumpectomy	64 515 (67%)	1979 (61%)	
Radical mastectomy	46 284 (13%)	403 (12.4%)	
Subcutaneous or total mastectomy	64 515 (18%)	770 (24%)	
Surgery in regional of distant stage cases[c]			
None	33 163(16%)	390 (16%)	
Partial mastectomy/lumpectomy	69 992 (34%)	749 (31%)	.0008
Radical mastectomy	72 705 (42%)	840 (35%)	
Subcutaneous or total mastectomy	30 561 (15%)	420 (18%)	
HER2[b]			
Negative	187 683 (85.4%)	2452 (81.0%)	<.0001
Positive	32 120 (14.6%)	576 (19.0%)	
Subtype[b]			
HER2 enriched	9083 (4.1%)	188 (6.2%)	<.0001
Luminal A	165 713 (75.5%)	2101 (69.5%)	
Luminal B	22 941 (10.5%)	388 (12.8%)	
Triple negative	21 675 (9.9%)	346 (11.5%)	

Note: The sample size and proportion (in parenthesis) is shown for each characteristic. Mean and SD (in parentheses) are shown for age at diagnosis.
Abbreviations: AIPA, Asian Indian and Pakistani American; CI, confidence interval; ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; NHW, non-Hispanic white; PR, progesterone receptor; SEER, surveillance, epidemiology and end results.

[a]*P* value based on a two-sample *t* test. All other *P* values based on chi-squared tests.

[b]Available only for tumors diagnosed after 2010.

[c]Cases with unknown stage and grade, unknown or borderline ER and PR, unknown or other surgery (bilateral mastectomy, mastectomy and surgery not otherwise specified) are not included in the table.

The proportion of breast cancers diagnosed annually in AIPA women increased over the years, suggesting potentially increasing mammography screening in these women. However, a recent study found that mammography screening decreased in AIPA women between 2001 and 2009.[39] Recent studies of cancer screening have identified several sociocultural factors associated with delay in seeking health care as well as poor mammogram rates in AIPA women, including lack of family support, lack of transportation, modesty, fear, karmic beliefs that cancer is divine punishment for past deeds, having lived in the United States for less than 10 years, low English proficiency, and ethnocultural discordance in the health system.[40-42] Further work is needed to identify strategies for promoting nuanced interactions between healthcare workers and AIPA individuals by effectively understanding and addressing the myriad sociocultural factors associated with screening decisions in this population.

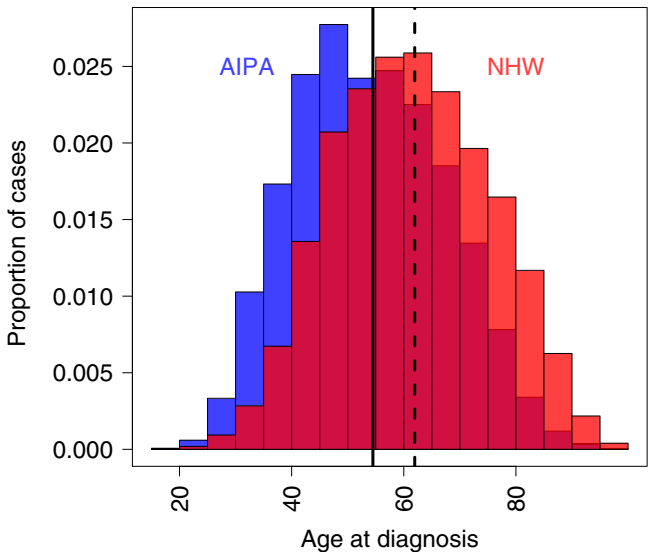

FIGURE 2 Distribution of age at diagnosis of breast cancer in Asian Indian and Pakistani Americans (AIPA; blue color) and non-Hispanic white (NHW; red color) cases from the SEER 18 database. The mean age at diagnosis in AIPA and NHW cases is shown as bold and dashed vertical lines, respectively

Despite having aggressive disease, relative to NHW cases, AIPA cases had significantly lower hazard for death due to breast cancer and death due to other causes under both sub-distribution and cause-specific hazard models. In survival analyses, when the cause-specific hazard ratios of a variable for competing events of interest are of opposite directions, the variable may be interpreted to have an indirect effect on the cumulative incidence of one of the events.[23,43] However, in our analyses, the cause-specific hazard ratios of AIPA (relative to NHW) race/ethnicity for death due to breast cancer and death due to other causes were both smaller than 1. Similar results emanated for the sub-distribution hazard ratios of AIPA race/ethnicity for the two events. Furthermore, for each event, analyses based on the sub-distribution and cause-specific hazard functions gave hazard ratios of similar magnitudes for AIPA race/ethnicity. These observations suggest that the lower hazard ratio of AIPA race/ethnicity for death due to breast cancer is unlikely to be due to an indirect effect of race/ethnicity on death due to competing causes. In other words, dying of other causes is not the reason why AIPA race/ethnicity has lower hazard ratio for death due to breast cancer. Further investigations are required to understand the reasons why AIPA cases have significantly lower hazard for death due to breast cancer than that of NHW cases. An important characteristic of these data is the short median follow-up of AIPA cases than NHW cases. Whether AIPA will continue to have lower cumulative incidence of death due to breast

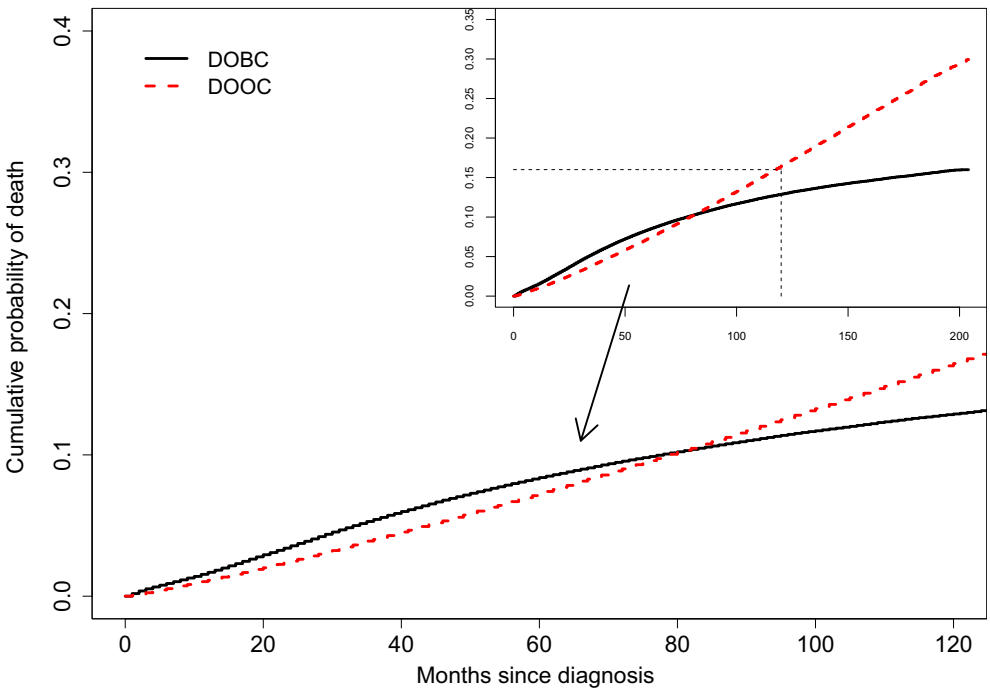

FIGURE 3 Estimated cumulative probability of breast cancer-specific mortality or dying of breast cancer (DOBC; black curve) and dying of other causes (DOOC; dashed red curve), shown for the combined cohorts of Asian Indian and Pakistani American and non-Hispanic white cases from the SEER 18 database. The inset in the top-right corner shows the estimated cumulative probabilities for the complete follow-up duration. The main figure shows the estimated cumulative probabilities for the first 10 years of follow-up, which is the portion within the dashed box in the inset. The arrow indicates that the main figure is the dashed box within the inset

TABLE 3 Hazard ratios, 95% confidence intervals (CI) and *P* values from multivariable proportional hazards regression analysis of two competing events—death due to breast cancer and death due to other causes—in the SEER 18 database, estimated using sub-distribution and cause-specific hazard function models

					Sub-distribution hazard model		Cause-specific hazard model	
Characteristic		N	N DOBC	N DOOC	Death due to breast cancer	Death due to other causes	Death due to breast cancer	Death due to other causes
Age	<40	22166	3141	443	Reference	Reference	Reference	Reference
	40 to 49	79589	7501	1976	0.870 (0.835-0.908)	1.096 (0.995-1.207)*	0.875 (0.839-0.913)	1.104 (0.996-1.224)*
	≥50	385395	37715	60629	1.087 (1.049-1.127)	6.498 (5.934-7.116)	1.188 (1.144-1.233)	6.912 (6.293-7.592)
Stage[a]	Regional and distant	169431	35389	19746	Reference	Reference	Reference	Reference
	Localized	317719	12968	43302	0.247 (0.241-0.253)	0.872 (0.854-0.890)	0.234 (0.228-0.239)	0.703 (0.690-0.716)
Grade[a]	≤2	329346	20909	44699	Reference	Reference	Reference	Reference
	≥3	157804	27448	18349	1.752 (1.714-1.790)	1.028 (1.009-1.048) **	1.789 (1.753-1.825)	1.127 (1.105-1.149)
Race/ethnicity	Non-Hispanic white	482250	47981	62853	Reference	Reference	Reference	Reference
	Asian Indian and Pakistani American	4900	376	195	0.787 (0.717-0.864)	0.553 (0.472-0.647)	0.759 (0.685-0.840)	0.532 (0.462-0.612)
ER[a]	Positive	402450	32177	53173	Reference	Reference	Reference	Reference
	Negative	84700	16180	9875	1.320 (1.279-1.362)	0.994 (0.967-1.022) ***	1.353 (1.317-1.391)	1.059 (1.029-1.090)
PR[a]	Positive	350207	25579	44976	Reference	Reference	Reference	Reference
	Negative	136943	22778	18072	1.427 (1.390-1.464)	1.061 (1.041-1.082)	1.451 (1.415-1.487)	1.115 (1.091-1.140)
Chemotherapy	Yes	195766	27470	12132	Reference	—	—	
	No/unknown	291384	20877	50916	1.105 (1.081-1.130)	2.606 (2.555-2.658)	1.258 (1.232-1.284)	2.823 (2.761-2.886)
Radiation	Yes	262738	21342	25445	Reference	Reference	Reference	Reference
	No	224412	27015	37603	1.060 (1.035-1.086)	1.758 (1.724-1.792)	1.096 (1.074-1.118)	1.795 (1.761-1.830)
Surgery	None	22312	9373	3273	Reference	Reference	Reference	Reference
	Partial mastectomy/ lumpectomy	278183	15192	34209	0.129 (0.125-0.133)	0.847 (0.812-0.885)	0.107 (0.104-0.110)	0.392 (0.377-0.407)
	Radical mastectomy	100958	17980	17438	0.212 (0.206-0.220)	0.873 (0.837-0.911)	0.181 (0.176-0.186)	0.413 (0.398-0.429)
	Subcutaneous or total mastectomy	85697	5812	8128	0.163 (0.156-0.170)	0.709 (0.677-0.741)	0.137 (0.132-0.141)	0.327 (0.314-0.341)
Year of diagnosis	2000 to 2004	126630	19687	30822	Reference	Reference	Reference	Reference
	2005 to 2009	141815	16831	20958	0.868 (0.847-0.890)	0.874 (0.859-0.888)	0.903 (0.884-0.923)	0.951 (0.933-0.969)
	2010 to 2016	218705	11839	9864	0.773 (0.753-0.794)	0.716 (0.700-0.732)	0.861 (0.839-0.883)	0.884 (0.861-0.908)

Abbreviations: N, total number of cases; N alive, number of cases alive; N DOBC, number died of breast cancer; N DOOC, number died of other causes.
[a]Cases with unknown stage, grade, borderline or unknown ER and PR, and unknown or other surgery (bilateral mastectomy, mastectomy or surgery not otherwise specified) were not included in analysis.
[a]**P* value of .06, ***P* value of .003, all other *P* values of <.0001; ****P* value of .69.

cancer than that of NHW cases even after comparable follow-up of the two populations is an important question that requires a comprehensive investigation of the characteristics of follow-up care in AIPA breast cancer survivors or putative barriers to follow-up in these individuals. We followed all the cases for *M* months since their time of diagnosis and set *M* = 36, 48 and 52 months. This gave comparable

follow-up of the two populations. However, AIPA cases continued to have significantly lower hazard for death due to breast cancer in multivariate analyses for all these choices of *M* (see Supplementary Tables S6-S8). Therefore, further investigations of a systematically assembled and followed cohort of AIPA breast cancer survivors are needed to better understand follow-up patterns of AIPA cases. Also needed are a detailed inquiry into unique cultural nuances that might influence follow-up, including spiritual needs, language support needs, fear of cancer and disease recurrence, social stigma, cultural and religious beliefs, and values.[44]

Several important limitations need to be considered when evaluating the results of our study. The AIPA population is younger than the NHW population, and there were considerably more NHW cases than AIPA cases. These are to be expected since Asian Indians and Pakistanis migrate to the United States primarily to seek education and a livelihood and, thus, are likely to be of the younger generation. AIPA cases had shorter median follow-up than NHW cases. Information about follow-up patterns of AIPA was not available. The registries included in the SEER 18 database cover only 28% of the entire US population.[45] There is a large AIPA population in other metropolitan regions such as Chicago, Dallas, Houston, New York and Washington DC (6) that are not part of SEER registries. Stage, grade, ER status and PR status were missing for some women. The reasons behind these missing data are, however, not known. SEER provides limited data on treatment, and we have used all available treatment information—namely, chemotherapy, radiation therapy and surgery (type of mastectomy)—in our analyses. Our analyses are not adjusted for comorbidities since comorbidity data are not available from SEER. Addressing these limitations will require a well-designed study of AIPA breast cancer survivors.

In summary, our study found that breast cancer incidence increases with age in AIPA women and the rate of increase reduces after age 46 years. The AIPA population of women is younger than NHW women. AIPA women have early age at onset, more aggressive breast cancer phenotypes, and shorter median follow-up compared to NHW women. Breast cancer in AIPA women has specific characteristics that need to be further studied for understanding disease etiology and outcomes in this understudied and growing population.

ACKNOWLEDGMENTS

This work was supported in part by research grants R01CA197402 (JMS), R01CA243547 (SG), R01CA202752 (SG) and P30CA072720 (SG) from the National Cancer Institute, USA, and the Breast Cancer Research Foundation (SG). The authors acknowledge the Office of Advanced Research Computing (OARC) at Rutgers, The State University of New Jersey, for providing access to the Amarel cluster and associated research computing resources that have contributed to the results reported here. URL: (http://oarc.rutgers.edu).

CONFLICT OF INTEREST

Dr. Shridar Ganesan has consulted for Merck, Foundation Medicine, Roche, Novartis Foghorn Therapeutics and Inspirata. The remaining authors do not have any conflict of interest.

DATA AVAILABILITY STATEMENT

SEER data are available to eligible individuals from the National Cancer Institute (NCI) upon completion of a data use agreement. We completed a data use agreement to access the data reported in this manuscript. Datasets used in the analysis are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

We signed a data use agreement per NCI/SEER policies in order to obtain approval to access de-identified SEER public-use data through SEER*Stat.

ORCID

Jaya M. Satagopan https://orcid.org/0000-0001-6612-4856

REFERENCES

1. Demographic snapshot of South Asians in the United States. South Asian Americans Leading Together. 2020 https://saalt.org/wp-content/uploads/2019/04/SAALT-Demographic-Snapshot-2019.pdf. Accessed January 10, 2020.
2. Demographic characteristics of South Asians in the United States. https://saalt.org/wp-content/uploads/2012/09/Demographic-Characteristics-of-SA-in-US-20001.pdf. Accessed September 14, 2020.
3. Lopez G, Ruiz NG, Patten E. Key facts about Asian Americans, a diverse and growing population. *Pew Res Center*. 2017. https://www.pewresearch.org/fact-tank/2017/09/08/key-facts-about-asian-americans/ Accessed September 14, 2020.
4. Davies AA, Blake C, Dhavan P. Social determinants and risk factors for non-communicable diseases (NCDs) in South Asian migrant populations in Europe. *Asia Eur J*. 2011;8:461-473.
5. Gomez SL, Noone AM, Lichetensztajn DY, et al. Cancer incidence trends among Asian American populations in the United States, 1990-2008. *J Natl Cancer Inst*. 2013;105(15):1096-1110. https://doi.org/10.1093/jnci/djt157.
6. Moran MS, Gonsalves L, Goss DM, Ma S. Breast cancers in U.S. residing Indian-Pakistani versus non-Hispanic White women: comparative analysis of clinical-pathological features, treatment, and survival. *Breast Cancer Res Treat*. 2011;128(2):543-551. https://doi.org/10.1007/s10549-011-1362-0.
7. Yi M, Liu P, Li X, et al. Comparative analysis of clinicopathologic features, treatment, and survival of Asian women with a breast cancer diagnosis residing in the United States. *Cancer*. 2012;118(17):4117-4125. https://doi.org/10.1002/cncr.27399.
8. Iqbal J, Ginsburg O, Rochon PA, Sun P, Narod SA. Differences in breast cancer stage at diagnosis and cancer-specific survival by race and ethnicity in the United States. *JAMA*. 2015;313(2):165-173. https://doi.org/10.1001/jama.2014.17322.
9. Mukherjea A, Ivey SL, Shariff-Marco S, Allen KN. Overcoming challenges in recruitment of South Asians for disparities research in the United States. *J Racial Ethn Health Disparities*. 2018;5(1):195-208. https://doi.org/10.1007/s40615-017-0357-x.
10. Surveillance, Epidemiology, and End Results (SEER) 2016 Program (www.seer.cancer.gov) SEER*Stat Database: Incidence—SEER 9, plus remainder of CA and NJ, Nov 2016 Sub (1990-2014) detailed API plus White Non-Hispanic—pops projected from populations, National Cancer Institute, DCCPS, Surveillance Research Program, released May 2017, based on the November 2016 submission.
11. Surveillance, Epidemiology, and End Results (SEER) 2018 Program (www.seer.cancer.gov) SEER*Stat Database: Incidence—SEER 9 Regs Custom Data (with additional treatment fields), Nov 2018 Sub

(1975-2016) <Katrina/Rita Population Adjustment>—Linked To County Attributes—Total U.S., 1969-2017 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2019, based on the November 2018 submission.
12. Surveillance, Epidemiology, and End Results (SEER) 2018 Program (www.seer.cancer.gov) SEER*Stat Database: Incidence—SEER 18 Regs Custom Data (with additional treatment fields), Nov 2018 Sub (2000-2016) <Katrina/Rita Population Adjustment>—Linked To County Attributes—Total U.S., 1969-2017 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2019, based on the November 2018 submission.
13. Frome EL, Checkoway H. Use of Poisson regression models in estimating incidence rates and ratios. *Am J Epidemiol*. 1985;121(2):309-323.
14. Doll R. The age distribution of cancer: implications for models for carcinogenesis. *J R Stat Soc A Stat*. 1971;134(2):133-155.
15. Pike MC, Pearce CL, Wu AH. Prevention of cancers of the breast, endometrium and ovary. *Oncogene*. 2004;23(38):6379-6391.
16. Muggeo V. Testing with a nuisance parameter present only under the alternative: a score-based approach with application to segmented modeling. *J Stat Comput Sim*. 2016;86(15):3059-3067.
17. Muggeo VMR. Package 'segmented'. http://nbcgib.uesc.br/mirrors/cran/web/packages/segmented/segmented.pdf. Accessed March 20, 2020.
18. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. New York: John Wiley & Sons; 2002.
19. Marubini E, Valsecchi MG. Analyzing survival data from clinical trials and observational studies. John Wiley & Sons; 1995.
20. Satagopan JM, Ben-Porat L, Berwick M, Robson M, Kutler D, Auerbach AD. A note on competing risks in survival analysis. *Br J Cancer*. 2004;91(7):1229-1235.
21. Gray RJ. A class of K-sample tests for comparing the cumulative incidence of a competing risk. *Ann Stat*. 1988;16(3):1141-1154.
22. Fine JP. Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94(446):496-509.
23. Austin PC, Lee DS, Fine JP. Introduction to the analysis of survival data in the presence of competing risks. *Circulation*. 2016;133:601-609. https://doi.org/10.1161/CIRCULATIONAHA.115.017719.
24. SEER Stat 2020: Surveillance Research Program, National Cancer Institute SEER*Stat software (www.seer.cancer.gov/seerstat) version 8.3.6.
25. SAS Institute Inc. *Base SAS® 9.3 Procedures Guide*. Cary, NC: SAS Institute Inc.; 2011.
26. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. 2013. http://www.r-project.org. Accessed January 10, 2020.
27. Gray RJ. Package 'cmprsk'. https://cran.r-project.org/web/packages/cmprsk/cmprsk.pdf. Accessed March 20, 2020.
28. Kawaguchi ES. Package 'fastcmprsk'. https://cran.r-project.org/web/packages/fastcmprsk/fastcmprsk.pdf. Accessed March 20, 2020.
29. Behl S, Misra A. Management of obesity in adult Asian Indians. *Indian Heart J*. 2017;69(4):539-544. https://doi.org/10.1016/j.ihj.2017.04.015.
30. Sundar S, Khetrapal-Singh P, Frampton J, et al. Harnessing genomics to improve outcomes for women with cancer in India: Key priorities for research. *Lancet Oncol*. 2018;19(2):e102-e112. https://doi.org/10.1016/S1470-2045(17)30726-X.
31. Pathak PK, Tripathi N, Subramanian SV. Secular trends in menarcheal age in India—evidence from the Indian human development survey. *PloS One*. 2014;9(11):e111027. https://doi.org/10.1371/journal.pone.0111027.
32. Ahmed SM, Waheed MA, Ahmad F, Danish SH. Factors contributing to early menarche in school girls. *J Pak Med Assoc*. 2016;66(5):629-633.
33. Marphatia AA, Ambale GS, Reid AM. Women's marriage age matters for public health: A review of the broader health and social implications in South Asia. *Front Public Health*. 2017;5:269.
34. Ahuja M. Age of menopause and determinants of menopause age: A PAN India survey by IMS. *J Midlife Health*. 2016;7(3):126-131.
35. Baig LA, Karim SA. Age at menopause, and knowledge of and attitudes to menopause, of women in Karachi, Pakistan. *J Br Menopause Soc*. 2006;12(2):71-74.
36. Koprowski C, Ross RK, Mack WJ, Henderson BE, Bernstein L. Diet, body size and menarche in a multiethnic cohort. *Br J Cancer*. 1999;79(11-12):1907-1911.
37. Butts SF, Seifer DB. Racial and ethnic differences in reproductive potential across the life cycle. *Fertil Steril*. 2010;93(3):681-690.
38. Mathews TJ, Hamilton BE. Mean age of mothers is on the rise: United States, 2000-2014. *NCHS Data Brief*. 2016;232:1-8.
39. Chawla N, Breen N, Liu B, Lee R, Kagawa-Singer M. Asian American women in California: A pooled analysis of predictors for breast and cervical cancer screening. *Am J Public Health*. 2015;105(2):e98-e109.
40. Boxwala FI, Bridgemohan A, Griffith DM. Factors associated with breast cancer screening in Asian Indian women In Metro-Detroit. *J Immigr Minor Health*. 2010;12(4):534-543.
41. Lobb R, Pinto AD, Lofters A. Using concept mapping in the knowledge-to-action process to compare stakeholder opinions on barriers to use of cancer screening among South Asians. *Implement Sci*. 2013;8:37.
42. Wu TY, Scheffer B, Lee C. Using focus groups to understand breast cancer screening beliefs and behaviors among Asian Indian women. *J Nurs Educ Pract*. 2012;2(4):42-51.
43. Latouche A, Allignol A, Beyersmann J, Labopin M, Fine JP. A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. *J Clin Epidemiol*. 2013;66(6):648-653.
44. Singh-Carlson S, Wong F, Martin L, Nguyen SK. Breast cancer survivorship and South Asian women: understanding about the follow-up care plan and perspectives and preferences for information post treatment. *Curr Oncol*. 2013;20(2):e63-e79. https://doi.org/10.3747/co.20.1066.
45. Surveillance, Epidemiology and End Results Program. https://www.epa.gov/sites/production/files/2015-06/documents/aceappendixb_surveillanceepidemiologyandendresults.pdf. Accessed May 9, 2020.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Satagopan JM, Stroup A, Kinney AY, Dharamdasani T, Ganesan S, Bandera EV. Breast cancer among Asian Indian and Pakistani Americans: A surveillance, epidemiology and end results-based study. *Int. J. Cancer*. 2020; 1–10. https://doi.org/10.1002/ijc.33331