

# Factors Associated with Surgery among South Asian American and Non-Hispanic White Women with Breast Cancer

Lydia Lo<sup>\*a</sup> & Jaya M. Satagopan<sup>b</sup>

<sup>a</sup>Department of Industrial and Systems Engineering, School of Engineering, Rutgers University, New Brunswick, NJ

<sup>b</sup>Department of Biostatistics and Epidemiology & Center for South Asian Quantitative Health and Education, School of Public Health, Rutgers University, New Brunswick, NJ

<https://doi.org/10.33697/ajur.2021.048>

Student: [lp133@scarletmail.rutgers.edu](mailto:lp133@scarletmail.rutgers.edu)\*

Mentor: [satagopj@sph.rutgers.edu](mailto:satagopj@sph.rutgers.edu)

## ABSTRACT

South Asian American (SA) women are diagnosed with more aggressive breast cancer than non-Hispanic White (NHW) women. Understanding the factors associated with the types of surgery received by these women sheds light on disease management in these culturally distinct populations. We used data on age at diagnosis, stage, grade, estrogen and progesterone receptors, and surgery from 4,590 SA and 429,030 NHW breast cancer cases in the Surveillance, Epidemiology and End Results (SEER) program. We used logistic regression with surgery as the binary outcome (subcutaneous, total, or radical mastectomy (STRM) versus partial mastectomy, no, unknown or other (PNUM)) and included additive effects of all the variables and interactions of age, stage, grade, and estrogen and progesterone receptors with race/ethnicity. Type I error of 5% was used to assess statistical significance of the effects. SA were significantly more likely than NHW cases to receive STRM relative to PNUM surgery among women diagnosed at or after age 50 years and having localized stage disease (Odds Ratio (OR) = 1.27, 95% Confidence Interval (CI) = 1.06 – 1.52). Further, SA were significantly less likely than NHW cases to receive STRM relative to PNUM surgery among those diagnosed before age 50 years and having regional or distant stage disease (OR = 0.75, 95% CI = 0.59 – 0.95 for age at diagnosis < 40 years; OR = 0.77, 95% CI = 0.62 – 0.95 for age at diagnosis 40-49 years). The type of surgery received by SA and NHW women differ according to age at diagnosis and disease stage.

## KEYWORDS

Breast Cancer; Surgery; Cancer Health Equity; Disease Characteristics; South Asian American; Non-Hispanic White; Logistic Regression; Interaction

## INTRODUCTION

South Asians (SA) are individuals with origins in Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, and Sri Lanka. SA are a fast-growing minority population in the U.S, increasing by about 40% from around 3.9 million in 2010 to around 5.4 million in 2017.<sup>1</sup> SA women in the U.S confront a wide range of chronic diseases, including breast cancer. They are more likely than non-Hispanic White (NHW) women to be diagnosed with aggressive breast cancers characterized by younger age, regional or distant stage tumors, higher grade tumors, and estrogen receptor (ER)-negative tumors at the time of diagnosis.<sup>2</sup> There are many different surgical treatments available for breast cancer. The most common types of surgery are lumpectomy and mastectomy.<sup>3,4</sup> Lumpectomy, also known as partial mastectomy, is a surgery to remove part of the breast that is affected by cancer.<sup>5</sup> There are various types of mastectomies – subcutaneous, total, or radical mastectomy – that involve removal of the entire breast, with or without the nipple, the areola, major muscles of the chest wall, and axillary or central lymph nodes.<sup>6</sup> In the U.S, early-stage breast cancers may be treated with lumpectomy, while large or later stage breast cancers may be removed by mastectomy.<sup>7</sup> In contrast, the rate of mastectomy relative to lumpectomy for breast cancers is high in Asian countries, regardless of disease stage.<sup>8</sup> Little is known about surgical treatment for breast cancer among SA women in the U.S, whether SA women are more likely than NHW women to undergo mastectomy, and whether any differences in surgery between SA and NHW persists after adjusting for tumor characteristics. To address this gap, we undertook this study to examine factors associated with type of breast cancer surgery in SA and NHW women using data from the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) program.<sup>9</sup> Understanding whether mastectomy rates are high among SA women in the US even after accounting for tumor characteristics such as stage can provide valuable insights into breast cancer-related treatment decisions in this culturally distinct minority population.

## METHODS AND PROCEDURES

### *Data Source*

We extracted breast cancer data from 18 population-based SEER registries that cover about 28% of the U.S. population<sup>9</sup>. We completed a data use agreement with SEER and downloaded a case listing of female breast cancers diagnosed between the years 2000 and 2016 via the SEER\*Stat software package.<sup>9</sup> Using the International Statistical Classification of Diseases (ICD), we abstracted data corresponding to ICD code “ICD-O-3: Breast from ages 0 to 85+”. This ICD classification includes the following breast cancers: 8500/3: Infiltrating duct carcinoma, NOS; 8521/3: Infiltrating ductular carcinoma; 8522/3: Infiltrating duct and lobular carcinoma; 8523/3: Infiltrating duct mixed with other types of carcinoma; and 8524/3: Infiltrating lobular mixed with other types of carcinoma, 8540/3: Paget disease, mammary; 8541/3: Paget disease and infiltrating ductal carcinoma of breast; and 8543/3: Paget disease and intraductal carcinoma.

### *Outcome Variable*

We obtained surgery information from the SEER variable “RX Summ—Surg Prim Site (1998)+” and used SEER surgery codes<sup>10</sup> to categorize this variable as subcutaneous mastectomy (surgery code 30), total mastectomy (surgery codes 40-49 and 75), radical mastectomy (surgery codes 50-74), partial mastectomy (surgery codes 20-24), no surgery (code 0), other surgery (codes 19 or 90), and unknown if surgery performed since patient information is from death certificate only (code 99). The binary outcome in our study was surgery, categorized as subcutaneous, total, or radical mastectomy (abbreviated as STRM) and partial mastectomy, no, other or unknown surgery (abbreviated as PNUM).

### *Explanatory Variables*

The explanatory variables were age at diagnosis, race/ethnicity, tumor stage, tumor grade, estrogen receptor (ER) status, and progesterone receptor (PR) status. We used the SEER variable “Age recode with < 1 year olds” to obtain age at diagnosis of breast cancer. We used the SEER variable “Race/ethnicity” to obtain data from non-Hispanic White (NHW), Asian Indian or Pakistani, NOS (1988+), Asian Indian (2010+), and Pakistani (2010+) women. The latter three race/ethnicity consisting of women of Asian Indian or Pakistani origin were grouped as South Asian (SA). We obtained tumor stage using SEER variables “SEER Summary Stage 1977”, “SEER Summary Stage 2000” and “SEER Combined Summary Stage 2000” for cases diagnosed in 2000, from 2001 to 2003, and from 2004 to 2016, respectively, and categorized stage as localized, regional, distant, and unknown. SEER variable “Grade” provides tumor grade as I, II, III, IV, or unknown. SEER variables “ER Status Recode Breast Cancer (1990+)” and “PR Status Recode Breast Cancer (1990+)” provide ER and PR statuses, respectively, as positive, negative, borderline, or unknown. Patients with unknown stage or grade and unknown or borderline ER or PR statuses were removed from analyses.

### *Statistical Analysis*

We calculated mean and standard deviation (SD) to examine baseline characteristics of age at diagnosis and calculated frequencies and proportions to examine baseline characteristics of race/ethnicity, stage, grade, ER status, PR status and surgery. We obtained boxplot of age at diagnosis of breast cancer of SA and NHW women receiving STRM and PNUM surgeries. We used a two-sample t-test to compare the mean age at diagnosis of breast cancer between SA and NHW women, and chi-squared tests to compare frequencies of race/ethnicity, stage, grade, ER status and PR status between SA and NHW women. Prior to conducting further analyses, we categorized age at diagnosis as “< 40 years”, “40 – 49 years”, and “≥ 50 years”. These age categories have been previously used in breast cancer studies and are taken to denote early-onset disease (age < 40 years), onset during perimenopausal period (age 40-49 years), and onset during post-menopausal period (age ≥ 50 years).<sup>2</sup> We also created binary variables for stage (localized versus regional or distant) and grade (I and II versus III and IV). Table 1 shows all the variables and their categories used in our analyses.

Using these categorized variables, we conducted logistic regression analyses to evaluate the association between age at diagnosis, stage, grade, ER and PR statuses and surgery.<sup>11</sup> First, we conducted univariable logistic regression analyses to gain insights into relationships between each explanatory variable and surgery type without adjusting for other explanatory variables. Next, we fitted two multivariable logistic regression models – first by including only additive effects of all the explanatory variables, and next by including all the additive effects and pairwise multiplicative interactions between race/ethnicity and age, stage, grade, ER status, and PR status. We calculated odds ratios (OR) and 95% confidence intervals (CI) and tested the statistical significance of the odds ratios using Wald tests. In the multivariable model with interaction terms, we derived odds ratios for race/ethnicity in distinct categories of age at diagnosis and stage as shown in the Appendix. All hypothesis tests were conducted at 5% type I error.

Variables in analysis	Binary Categories	
	Category 0 (or baseline category)	Category 1 (or comparison category)
<b>Outcome Variable</b>		
Surgery (mastectomy type)	Partial, None, Unknown, Other (PNUM)	Subcutaneous, Total, Radical (STRM)
<b>Explanatory Variables</b>		
Age		
Age 40 years	Age ≥ 50 years	Age < 40 years
Age 40-49 years	Age ≥ 50 years	Age 40-49 years
Race	Non-Hispanic White	South Asian
Grade	I and II	III and IV
Stage	Localized	Regional and Distant
ER	Positive	Negative
PR	Positive	Negative

**Table 1.** Variables used in the analysis and their binary categories. Note that since age has 3 categories, we created two dummy variables corresponding to the categories Age < 40 years and Age 40-49 years, keeping Age ≥ 50 years as the baseline category.

When the 95% CI for an OR corresponding to a variable included the value 1, the corresponding p-value for the Wald test also exceeded 0.05 as expected, suggesting no statistically significant association between the variable and surgery type or suggesting that individuals from distinct categories of that variable were equally likely to receive STRM relative to PNUM surgery at the 5% significance level. A statistically significant odds ratio for an interaction between race/ethnicity and another variable was taken as an indication that the effect of race/ethnicity on type of surgery differed across the categories of that variable.

The multivariable models fitted were evaluated using goodness-of-fit measures such as Akaike information criterion (AIC), Bayes information criterion (BIC), and area under the receiver operating characteristic curve (AUC). The AIC is defined as negative two times the maximum log-likelihood plus two times the number of parameters in the model. The BIC is defined as two times the maximum log-likelihood subtracted by the number of parameters in the model times the log of the sample size. The AUC is a measure of how well the model is able to distinguish between the binary surgery types. The best fitting logistic regression model among those considered will ideally have the smallest AIC and BIC and the largest AUC.

All statistical analyses were conducted using the R programming language (version 3.6.3).<sup>12</sup>

**RESULTS**

We obtained data for a total of 449,040 SA and NHW female breast cancer cases diagnosed between 2000 and 2016 from SEER. After removing cases with unknown stage, unknown grade, and unknown or borderline ER and PR statuses, the dataset for our analysis included 4,590 SA and 429,030 NHW cases. Table 2 shows the patients’ characteristics. Figure 1 shows boxplots of age at diagnosis of breast cancer in SA and NHW cases receiving STRM and PNUM surgery. SA cases were more likely to be younger than NHW cases among women receiving STRM surgery (mean (SD) age at diagnosis = 52.6 (13.6) years for SA and 59 (14.3) years for NHW;  $p < 0.05$ ) and among those receiving PNUM surgery (mean (SD) age at diagnosis = 55.2 (12.7) years for SA and 62.4 (13.0) years for NHW;  $p < 0.05$ ). Further, in both STRM and PNUM surgery groups, SA cases were significantly more likely to have regional or distant stage, higher grade, ER negative and PR negative tumors than NHW cases ( $p < 0.05$  for all comparisons).

Table 3 provides the results of univariable and multivariable logistic regression analyses. In univariable analyses (column 3 of Table 3), the characteristics of cases that were significantly more likely to receive STRM surgery relative to PNUM surgery included younger age, regional or distant stage, higher grade, ER negative, and PR negative tumors. Further, SA cases were significantly more likely than NHW cases to receive STRM relative to PNUM surgery (OR = 1.25, 95% CI = 1.18 – 1.33).

In all the multivariable analyses considered, cases with younger age at diagnosis, regional or distant stage, higher grade, ER negative status and PR negative status were significantly more likely to receive STRM relative to PNUM surgery. In multivariable analyses that included only additive effects (column 4 of Table 3), there was no statistically significant association between race/ethnicity and surgery type: both SA and NHW cases were equally likely to receive STRM relative to PNUM surgery (OR = 1.03, 95% CI = 0.97, 1.10). In contrast, in multivariable analyses that included additive and interaction effects (column 5 of Table 3), race/ethnicity had a statistically significant interaction with age at diagnosis and stage, thus highlighting the critical role of these variables on the relationship between race/ethnicity and surgery type. Table 4 gives the odds ratios for race/ethnicity in distinct categories of age at diagnosis and stage. Among cases with localized stage disease diagnosed before age 40 years and between age

40-49 years, SA and NHW women were equally likely to receive STRM relative to PNUM surgery (OR = 0.94, 95% CI = 0.77 – 1.14 for age at diagnosis < 40 years; OR = 0.96, 95% CI = 0.80 – 1.15 for age at diagnosis 40-49 years). However, among cases with localized disease diagnosed at age 50 years or later, SA women were significantly more likely than NHW women to receive STRM relative to PNUM surgery (OR = 1.33, 95% CI = 1.19 – 1.48). In contrast, among cases with regional or distant stage disease diagnosed before age 40 years and between age 40-49 years, SA women were significantly less likely than NHW women to receive STRM relative to PNUM surgery (OR = 0.79, 95% CI = 0.63 – 0.98 for age at diagnosis < 40 years; OR = 0.80, 95% CI = 0.69 – 0.93 for age at diagnosis 40-49 years). However, among cases with regional or distant stage disease diagnosed at age 50 years or later, SA women and NHW women were equally likely to receive STRM relative to PNUM surgery (OR = 1.12, 95% CI = 0.98 – 1.27).

Variable	STRM Surgery				PNUM Surgery			
	NHW (N = 159,533)		SA (N = 1,956)		NHW (N = 269,497)		SA (N = 2,634)	
Age	Mean: 59.03	SD: 14.26	Mean: 52.63	SD:13.53	Mean: 62.36	SD: 12.99	Mean: 55.17	SD: 12.74
Grade								
Grade I	23,944		182		66,523		455	
Grade II	65,713		744		114,995		1,036	
Grade III	62,959		929		76,800		1004	
Grade IV	1,513		17		1,684		16	
Stage								
Localized	82,152		930		195,244		1,708	
Regional	72,428		961		59,763		722	
Distant	4,626		60		12,608		174	
ER Status								
Positive	124,042		1,445		226,104		2,100	
Negative	35,491		511		43,393		534	
PR Status								
Positive	106,566		1,255		198,461		1,870	
Negative	52,967		701		71,036		764	

Table 2. Median and standard deviation (SD) of age at diagnosis of breast cancer and frequencies of grade, stage, ER status and PR status for SA and NHW breast cancer patients receiving STRM and PNUM surgery.

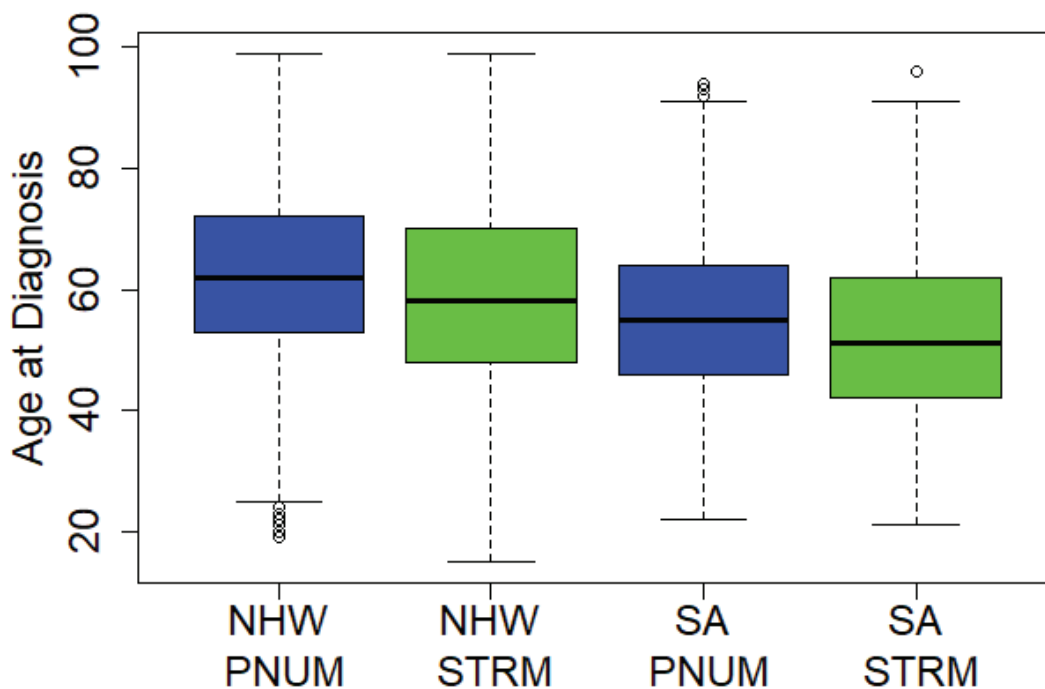


Figure 1. Boxplot of age at diagnosis of breast cancer of non-Hispanic White (NHW) and South Asian (SA) cases receiving PNUM and STRM surgery.

The multivariable model with additive and interaction terms had the smallest AIC but not the smallest BIC (Table 3). It had the same AUC as the multivariable additive model. As a benchmark for comparison, a multivariable additive model without race/ethnicity had AIC of 541585, BIC of 541661.8, and AUC of 0.6516 which are considerably worse than models containing race/ethnicity (detailed result of latter model is not shown).

Variable	Category	Odds Ratio (95% CI)		
		Univariable Analysis	Multivariable Analysis	
			Model with additive effects only	Model with Additive and interaction effects
<b>Additive effects</b>				
<b>Race/ethnicity</b>	SA	1.25 (1.18, 1.33)	1.03 (0.97, 1.10)	1.33 (1.19, 1.48)
	NHW	1.0	1.0	1.0
<b>Age</b>	<40 years	2.56 (2.49, 2.64)	2.34 (2.27, 2.41)	2.36 (2.28, 2.43)
	40-49 years	1.55 (1.52, 1.57)	1.53 (1.51, 1.56)	1.54 (1.51, 1.57)
	≥ 50 years	1.0	1.0	1.0
<b>Stage</b>	Regional or distant	2.51 (2.48, 2.54)	2.39 (2.35, 2.42)	2.39 (2.36, 2.42)
	Localized	1.0	1.0	1.0
<b>Grade</b>	III or IV	1.66 (1.64, 1.68)	1.28 (1.26, 1.30)	1.28 (1.26, 1.30)
	I or II	1.0	1.0	1.0
<b>ER status</b>	Negative	1.49 (1.47, 1.51)	1.07 (1.05, 1.10)	1.07 (1.05, 1.10)
	Positive	1.0	1.0	1.0
<b>PR status</b>	Negative	1.39 (1.37, 1.41)	1.15 (1.12, 1.17)	1.15 (1.12, 1.17)
	Positive	1.0	1.0	1.0
<b>Interaction effects</b>				
<b>Race/ethnicity * Age</b>	SA * Age < 40 years	-	-	0.70 (0.58, 0.85)
	SA * Age 40-49 years	-	-	0.72 (0.62, 0.83)
<b>Rage * Stage</b>	SA * regional or distant	-	-	0.84 (0.74, 0.95)
<b>Race * Grade</b>	SA * Grade III or IV	-	-	0.94 (0.82, 1.08)
<b>Race * ER</b>	SA * ER negative	-	-	0.96 (0.77, 1.19)
<b>Race * PR</b>	SA * PR negative	-	-	0.99 (0.82, 1.21)
<b>AIC</b>			521985.5	521955.6
<b>BIC</b>			522073	522108.7
<b>AUC</b>			0.6524	0.6524

Table 3. Results of univariable and multivariable logistic regression analyses.

Age	Stage	Odds Ratio (OR)	95% Confidence Interval (CI)
< 40 years	Localized	0.94	0.77, 1.14
	Regional / Distant	0.79	0.63, 0.98
40-49 years	Localized	0.96	0.80, 1.15
	Regional / Distant	0.80	0.69, 0.93
≥ 50 years	Localized	1.33	1.19, 1.48
	Regional / Distant	1.12	0.98, 1.27

Table 4. Odds ratio for receiving STRM relative to PNUM surgery for SA race/ethnicity compared to NHW race/ethnicity in distinct categories of age and stage, and 95% confidence interval (CI). These results are based on multivariable logistic regression model with additive and interaction terms,



## DISCUSSION

Surgery is one of the treatments available for breast cancer. Lumpectomy or partial mastectomy and mastectomy are the common surgical treatments for breast cancer. Mastectomy is more common than lumpectomy in Asian countries, regardless of disease stage. However, little is known about the use of mastectomy versus lumpectomy among breast cancer cases of SA race/ethnicity, who are among the fast-growing minority populations in the U.S, and how this compares to the use of surgeries among NHW breast cancer cases. Therefore, in this study, we examined factors associated with type of surgery (STRM versus PNUM) among female SA versus NHW breast cancer cases using data from SEER. Our analyses showed that: (i) there are statistically significant differences between SA and NHW breast cancer cases in the use of STRM relative to PNUM surgery; and (ii) these differences vary according to age at diagnosis and stage.

Our analyses showed the importance of interactions of race/ethnicity with age and stage. While the model with interactions had the smallest AIC, its BIC was larger than that of an additive model. This could be because BIC penalizes for having more terms in the model.

There can be several reasons why SA are more likely than NHW to have STRM relative to PNUM surgery when their disease is diagnosed at or after age 50 and they have localized stage disease but are less likely than NHW to have STRM relative to PNUM surgery when their disease is diagnosed before age 50 and they have regional or distant stage disease. Cultural factors are known to play a role in treatment preferences of female Chinese American breast cancer patients.<sup>13</sup> Health decisions of South Asian women living in Canada are strongly influenced by family and community responsibilities.<sup>14</sup> Further research is needed to examine cultural aspects of treatment preferences among South Asian American women, *i.e.*, South Asian women living in the United States. Whether other factors such as comorbidities play a role in surgery decisions of SA women is also an important question that requires further research.

Our work has several strengths and limitations. A strength of our study is the significance of our research question since SA are a rapidly growing minority population in the U.S and, hence, understanding their breast cancer health and treatment decisions is important. Another strength of our study is the availability of data on a large number of SA and NHW cases from SEER.

Several limitations must be considered when evaluating the results of our study. The registries included in the SEER 18 database cover only 28% of the entire US population. There is a large SA population in other major cities such as Chicago, Dallas, Houston, New York, and Washington DC that are not included in the databases that we have used in this study. The SA cases in our study are of Asian Indian and Pakistani origin. Cases with ancestry from other SA countries – Bangladesh, Bhutan, Maldives, Nepal, and Sri Lanka – are not included in our study since their race/ethnicity are not available through SEER. Our analyses are not adjusted for comorbidities since these data are not available through SEER. A well-designed study of SA and NHW cases will be required to address these limitations.

## CONCLUSIONS

Our analyses illustrated the important role of interaction terms in logistic regression model. An additive multivariable logistic regression model showed that SA and NHW cases were equally likely to receive STRM relative to PNUM surgery. Including interactions between race/ethnicity and age as well as stage revealed that differences in the type of surgery between SA and NHW varied according to age and stage categories. Disease characteristics play a key role on the type of surgery received by SA and NHW breast cancer patients. Further research is needed to gain insights into reasons underlying the choice of surgery made by SA women with specific disease characteristics.

## REFERENCES

1. South Asian Americans Leading Together (n.d) Resources and Fact Sheets. Retrieved January 6, 2021, from <https://saalt.org/resources/resources-factsheets/>
2. Satagopan, J. M., Stroup, A., Kinney, A. Y., Dharamdasani, T., Ganesan, S., & Bandera, E. V. (2021) Breast cancer among Asian Indian and Pakistani Americans: A surveillance, epidemiology and end results-based study. *International journal of cancer*, 148(7), 1598–1607. <https://doi.org/10.1002/ijc.33331>
3. Invasive Ductal Carcinoma: Diagnosis, Treatment, and More. (2019) Retrieved January 6, 2021, from <https://www.breastcancer.org/symptoms/types/idc>
4. Mayo Clinic. (2019) Paget's disease of the breast. Retrieved January 6, 2021, from <https://www.mayoclinic.org/diseases-conditions/pagets-disease-of-the-breast/diagnosis-treatment/drc-20351084>

5. Mayo Clinic. (2018) Lumpectomy. Retrieved January 6, 2021, from <https://www.mayoclinic.org/tests-procedures/lumpectomy/about/pac-20394650>
6. Oyelowo, T. (2009) Breast Conditions. Retrieved January 6, 2021, from <https://www.sciencedirect.com/science/article/pii/B9780323046015500083>
7. Mayo Foundation for Medical Education and Research. (2020) *Breast cancer surgery*. Mayo Clinic. <https://www.mayoclinic.org/tests-procedures/breast-cancer-surgery/about/pac-20385255>
8. Sinnadurai, S., Kwong, A., Hartman, M., Tan, E. Y., Bhoo-Pathy, N. T., Dahlui, M., See, M. H., Yip, C. H., Taib, N. A., & Bhoo-Pathy, N. (2018) Breast-conserving surgery *versus* mastectomy in young women with breast cancer in Asian settings. *BJS open*, 3(1), 48–55. <https://doi.org/10.1002/bjs.5.50111>
9. Surveillance, Epidemiology, and End Results (SEER) Program ([www.seer.cancer.gov](http://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.
10. Adamo MB, Johnson CH, Ruhl JL, Dickie, LA, (Eds.). 2013 SEER Program Coding and Staging Manual. National Cancer Institute, NIH Publication number 13-5581, Bethesda, MD
11. Agresti A. Categorical data analysis. John Wiley & Sons; 2003 Mar 31.
12. R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
13. Killoran, M., & Moyer, A. (2006) Surgical treatment preferences in Chinese-American women with early-stage breast cancer. *Psycho-Oncology: Journal of the Psychological, Social and Behavioral Dimensions of Cancer*, 15(11), 969–984. <https://doi.org/10.1002/pon.1032>
14. Bottorff, J. L., Johnson, J. L., Bhagat, R., Grewal, S., Balneaves, L. G., Clarke, H., & Hilton, B. A. (1998) Beliefs related to breast health practices: The perceptions of South Asian women living in Canada. *Social Science & Medicine*, 47(12), 2075-2085. [https://doi.org/10.1016/S0277-9536\(98\)00346-3](https://doi.org/10.1016/S0277-9536(98)00346-3)

#### ABOUT STUDENT AUTHOR

Lydia Lo graduated cum laude with a Bachelor of Science degree in Industrial Engineering from Rutgers University in May 2021. She currently works as an Operations Industrial Engineer at the United States Postal Service.

#### PRESS SUMMARY

South Asians (SA) are a fast-growing minority population in the U.S and confront a wide range of chronic disease, including breast cancer. SA women are diagnosed with more aggressive breast cancer than non-Hispanic White (NHW) women, characterized by younger age at diagnosis, regional or distant stage, higher grade, and ER negative tumors. Surgery – especially lumpectomy or mastectomy – is one of the treatments available for breast cancer. In the U.S, lumpectomy is used to treat early-stage breast cancers, while mastectomy is used for later stage breast cancers. In contrast, mastectomy is more commonly used in Asian countries, regardless of disease stage. It is important to understand the type of surgery and the factors that contribute to determining the type of surgery in SA breast cancer cases in the U.S relative to NHW cases. Our analyses reveal differences in the type of surgery received by SA and NHW breast cancer cases, and these differences vary according to age at diagnosis and disease stage. Further research is needed to gain deeper insights into these differences.

## APPENDIX

The Results section of the paper includes odds ratios corresponding to race/ethnicity in distinct subgroups of age and stage. The approach we used to obtain these odds ratios are given below.

Denote  $\pi$  as the probability that a person receives STRM surgery, given age, race/ethnicity, stage, grade, ER and PR. The logistic regression model consisting of interaction of race with age and stage is:

$$\log\left\{\frac{\pi}{1-\pi}\right\} = b_0 + b_1*x_1 + b_2*x_2 + b_3*race + b_4*stage + b_5*grade + b_6*er + b_7*pr + d_1*x_1*race + d_2*x_2*race + d_3*stage*race$$

The left hand of this model is known as the natural logarithm of odds or log-odds. The explanatory variables on the right hand side are as follows. First,  $x_1$  and  $x_2$  are dummy variables corresponding to age at diagnosis such that  $x_1 = 1$  if a person's age at diagnosis is  $< 40$  years and  $x_1 = 0$  otherwise, and  $x_2 = 1$  if a person's age at diagnosis is between 40 and 49 years and  $x_2 = 0$  otherwise. Thus,  $x_1 = 0$  and  $x_2 = 0$  for person whose age at diagnosis is  $\geq 50$  years. Further, stage, grade, er, pr, and race are binary variables coded as given in Table 1 of the manuscript. The model parameters are as follows:  $b_0$  is the y-intercept,  $b_i$  ( $b_1, b_2, b_3$ , etc.) are the additive effects of the explanatory variables and  $d_i$  ( $d_1, d_2, d_3$ ) are the effects of the interaction of race with age at diagnosis categories and stage, respectively. Although our multivariable analysis reported in the Results section includes interaction of race with grade, ER and PR, we have not included these interactions in writing the above model for simplicity of exposition. [Further, the estimated parameters for interactions of race grade, ER and PR are very small (near 0) and are not statistically significant in our analyses. Hence, including these interaction terms did not impact our estimated odds ratios corresponding to race in distinct categories of age and stage]

**Odds ratio corresponding to SA relative to NHW race among those diagnosed at age  $\geq 50$  years and having localized stage disease and 95% CI:**

Consider a person of age at diagnosis  $\geq 50$  years with localized stage disease and NHW race. For this person,  $x_1 = 0$ ,  $x_2 = 0$ , race = 0 and stage = 0. Plugging these values into the above model, the log-odds for this person will be:  $b_0 + b_5*grade + b_6*er + b_7*pr$ . Consider a person of age at diagnosis  $\geq 50$  years with localized stage disease and SA race. For this person,  $x_1 = 0$ ,  $x_2 = 0$ , race = 1 and stage = 0. The log-odds for this person will be:  $b_0 + b_3 + b_5*grade + b_6*er + b_7*pr$ .

To get the odds ratio of STRM versus PNUM surgery in SA race relative to NHW race among those of age at diagnosis  $\geq 50$  years and localized stage disease, take the difference between the two log odds and then the exponent of that. The difference between the two log odds is simply  $b_3$ . Hence, the odds ratio is  $\exp\{b_3\}$ .

The standard error corresponding to the estimated value of  $b_3$  can be obtained from the covariance matrix output of the logistic regression model in the R programming language. Suppose we denote this as  $SE_{b_3}$ . The 95% CI for the above odds ratio can be approximated as  $\exp\{b_3 \pm 1.96 * SE_{b_3}\}$ .

**Odds ratio corresponding to SA relative to NHW race among those diagnosed at age  $\geq 50$  years and having regional or distant stage disease and 95% CI:**

Consider a person of age at diagnosis  $\geq 50$  years with regional or distant stage disease and SA race. For this person,  $x_1 = 0$ ,  $x_2 = 0$ , race = 0 and stage = 1. The log-odds for this person will be:  $b_0 + b_4 + b_5*grade + b_6*er + b_7*pr$ . Consider a person of age at diagnosis  $\geq 50$  years with regional or distant stage disease and SA race. For this person,  $x_1 = 0$ ,  $x_2 = 0$ , race = 1 and stage = 1. The log-odds for this person will be:  $b_0 + b_3 + b_4 + b_5*grade + b_6*er + b_7*pr + d_3$ .

The difference between the two log odds is  $b_3 + d_3$ . Thus, the odds ratio for STRM relative to PNUM surgery in SA compared to NHW women with age at diagnosis  $\geq 50$  years and having regional or distant stage disease is  $\exp\{b_3 + d_3\}$ .



To calculate a 95% CI for this odds ratio, we need the standard error of  $b_3 + d_3$ . We shall denote this standard error as  $SE_{34}$ . It is the square root of the variance i.e.,  $SE_{34}$  is the square root of the variance  $\text{Var}(b_3 + d_3)$ , which is given by  $\text{Var}(b_3) + \text{Var}(d_3) + 2*\text{Cov}(b_3, d_3)$ . These variances and covariances can be obtained from the covariance matrix output of the logistic regression model in the R programming language. The 95% CI can then be approximated as:  $\exp\{ b_3 + d_3 \pm 1.96 * SE_{34} \}$ .

The odds ratios can be written in this manner for other categories of age and stage.