Manuscript for MADA Project

Trend in mammogram screening for breast cancer and socio-economic status of hispanic and non-hispanic women during 2000, 2005, 2010, and 2015

Shiwani Sapkota

4/21/23

# 1. 1. Summary/Abstract

[\*\*To be added]

# 2. 2. Introduction

Breast cancer is one of the most common cancers and the second leading cause of cancer death among women in the US. Mammography screening is a screening process where an x-ray image of the breast is taken to detect irregularities in breast tissue which could be the sign of cancer (1). Mammography screening can help detect and treat breast cancer at early stages reducing the mortality rate (2,3,4,5). Hence, mammography screening can play a great role in early breast cancer detection and treatment, and in reduction in the number of deaths related to breast cancer. The U.S. Preventive Services Task Force (USPSTF) recommends a biennial (every other year) mammography screening for women aged 50 to 74 years and advises to determine the need for mammography screening on an individual basis for women aged 40 to 49 years (6).

There are well established racial differences in breast cancer incidence and mortality rates (7,8). The rate of breast cancer is higher among White women compared to Black and Hispanic women. But Black and Hispanic women are often diagnosed at later stages of the cancer and have lower survival rates compared to White women (7). Apart from the racial differences, there are some consistent findings regarding the positive association between breast cancer screening and SES factors such as income and having a usual/regular source of medical care but there are some conflicting results regarding the effects of SES factors such as education, age, and marital status on breast cancer screening (9).

Little is known about the predictors of using cancer screening services among minority women, particularly Hispanics. Very few studies have studied the influence of socioeconomic status (SES) and sociodemographic factors on utilizing mammography screening services among Hispanic/non-Hispanic groups of women (7). It is important to understand the influence of women’s SES and race/ethnicity on the use of available health care services for improving knowledge about service utilization pattern in women population and for designing appropriate interventions accordingly (10). Hence, for developing appropriate population-based interventions for increasing the use of mammography screening, it is essential to first identify and understand the association and interaction between the use of mammography screening and demographic/socio-economic characteristics of women (7).

This study aims to explore the relationship trend among socioeconomic status and other predictive factors influencing the likelihood of using mammography screening services within Hispanic and non-Hispanic women aged 40 to 74 years during 2000, 2005, 2010, and 2015.

# 3. 3. Methods

The study used the data from the National Health Interview Survey (NHIS). The NHIS survey collects information related to health, health care access, health behaviors of the civilian and non-institutionalized US population and makes digital data files available for years 1963 to present (11). Users can create their own costumed NHIS data extracts for analysis from the NHIS website. For this study, data with the variables of interest for years: 2000, 2005, 2010, and 2015 were directly extracted from the NHIS website. Those specific 4 years were selected for the study as there were most data points available for those years and 5-years difference comparison could give great insights at the trends of mammogram screening over years. The study was restricted among Hispanic and non-Hispanic women aged 40-74 years as per the mammography screening recommendation by USPSTF and those age groups women fall under high-risk groups for developing breast cancer.

Mammography screening status (yes or no) was defined as outcome variable of interest for the study. Similarly, the following variables were assessed as predictors of mammogram screening: (1) age (40-44, 45-49, 50-54, 55-59, 60-64, 65-69, or 70-74), (2) income (at or above poverty threshold, or below poverty threshold), (3) education (never attended/kindergarten only, less than high school, high school, some college or associate’s degree, or bachelor’s degree and higher), (4) marital status (never married, married, or single (separated/widowed/divorced)), (5) region of residence (Northeast, North Central/Midwest, South, or West), (6) insurance status (no coverage or coverage), (7) self-reported health status (excellent/very good/good or fair/poor), (8) having a usual place of medical care (no place or have usual place), (9) smoking status (never, current, or former), (10) alcohol status (never, current, or former), and (11) diabetic status (yes or no).

Data Cleaning:

Analysis:

The study was limited within 40-74 years Hispanic and non-Hispanic women with no report of cancer during the years 2000, 2005, 2010, and 2015. Exclusion of missing data points in 1 or more predictor variables yielded a final sample of 24848 subjects.

[\*\* Statistical analysis part to be added]

# 4. 4. Results

# 5. 4.1 Exploratory/Descriptive analysis

**Table 1: Demographic characteristics of study population**

|  | **2000** | | **2005** | | **2010** | | **2015** | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Characteristic** | **NH**, N = 4,8831 | **H**, N = 8391 | **NH**, N = 4,9531 | **H**, N = 8191 | **NH**, N = 4,8191 | **H**, N = 9791 | **NH**, N = 6,4011 | **H**, N = 1,1551 |
| mammogram\_status | 4,191 (86% | 646 (77% | 4,291 (87% | 633 (77% | 4,271 (89% | 821 (84% | 5,679 (89% | 934 (81% |
| age |  |  |  |  |  |  |  |  |
| 40-44 years | 1,101 (23% | 218 (26% | 969 (20% | 217 (26% | 798 (17% | 231 (24% | 871 (14% | 259 (22% |
| 45-49 years | 959 (20% | 176 (21% | 959 (19% | 173 (21% | 848 (18% | 209 (21% | 953 (15% | 214 (19% |
| 50-54 years | 857 (18% | 137 (16% | 846 (17% | 157 (19% | 829 (17% | 157 (16% | 967 (15% | 183 (16% |
| 55-59 years | 624 (13% | 94 (11% | 763 (15% | 103 (13% | 749 (16% | 131 (13% | 1,049 (16% | 155 (13% |
| 60-64 years | 507 (10% | 75 (8.9% | 579 (12% | 69 (8.4% | 707 (15% | 108 (11% | 981 (15% | 118 (10% |
| 65-69 years | 419 (8.6% | 70 (8.3% | 423 (8.5% | 60 (7.3% | 503 (10% | 81 (8.3% | 892 (14% | 141 (12% |
| 70-74 years | 416 (8.5% | 69 (8.2% | 414 (8.4% | 40 (4.9% | 385 (8.0% | 62 (6.3% | 688 (11% | 85 (7.4% |
| income |  |  |  |  |  |  |  |  |
| At or above poverty threshold | 4,341 (89% | 594 (71% | 4,431 (89% | 614 (75% | 4,177 (87% | 700 (72% | 5,596 (87% | 842 (73% |
| Below poverty threshold | 542 (11% | 245 (29% | 522 (11% | 205 (25% | 642 (13% | 279 (28% | 805 (13% | 313 (27% |
| education |  |  |  |  |  |  |  |  |
| Never attended/kindergarten only | 7 (0.1% | 23 (2.7% | 12 (0.2% | 15 (1.8% | 14 (0.3% | 24 (2.5% | 12 (0.2% | 19 (1.6% |
| Less than high school | 617 (13% | 386 (46% | 512 (10% | 331 (40% | 408 (8.5% | 357 (36% | 415 (6.5% | 402 (35% |
| High school | 1,686 (35% | 202 (24% | 1,549 (31% | 224 (27% | 1,378 (29% | 264 (27% | 1,653 (26% | 294 (25% |
| Some college or Associate degree | 1,399 (29% | 147 (18% | 1,511 (31% | 153 (19% | 1,575 (33% | 213 (22% | 2,117 (33% | 285 (25% |
| Bachelor degree and higher | 1,174 (24% | 81 (9.7% | 1,369 (28% | 96 (12% | 1,444 (30% | 121 (12% | 2,204 (34% | 155 (13% |
| marital\_status |  |  |  |  |  |  |  |  |
| Never married | 461 (9.4% | 90 (11% | 480 (9.7% | 90 (11% | 550 (11% | 126 (13% | 735 (11% | 150 (13% |
| Married | 2,446 (50% | 407 (49% | 2,486 (50% | 380 (46% | 2,366 (49% | 488 (50% | 3,086 (48% | 567 (49% |
| Single (separated/widowed/divorced) | 1,839 (38% | 313 (37% | 1,846 (37% | 315 (38% | 1,724 (36% | 334 (34% | 2,337 (37% | 395 (34% |
| Living with partner | 137 (2.8% | 29 (3.5% | 141 (2.8% | 34 (4.2% | 179 (3.7% | 31 (3.2% | 243 (3.8% | 43 (3.7% |
| region\_residence |  |  |  |  |  |  |  |  |
| Northeast | 937 (19% | 159 (19% | 855 (17% | 117 (14% | 830 (17% | 150 (15% | 1,138 (18% | 175 (15% |
| North Central/Midwest | 1,235 (25% | 37 (4.4% | 1,318 (27% | 52 (6.3% | 1,166 (24% | 54 (5.5% | 1,455 (23% | 77 (6.7% |
| South | 1,822 (37% | 321 (38% | 1,829 (37% | 342 (42% | 1,775 (37% | 363 (37% | 2,252 (35% | 399 (35% |
| West | 889 (18% | 322 (38% | 951 (19% | 308 (38% | 1,048 (22% | 412 (42% | 1,556 (24% | 504 (44% |
| insurance\_status |  |  |  |  |  |  |  |  |
| No Coverage | 4,370 (89% | 614 (73% | 4,441 (90% | 582 (71% | 4,224 (88% | 683 (70% | 6,002 (94% | 918 (79% |
| Coverage | 513 (11% | 225 (27% | 512 (10% | 237 (29% | 595 (12% | 296 (30% | 399 (6.2% | 237 (21% |
| health\_status |  |  |  |  |  |  |  |  |
| Excellent/very good/good | 4,161 (85% | 621 (74% | 4,166 (84% | 615 (75% | 4,017 (83% | 752 (77% | 5,391 (84% | 870 (75% |
| Fair/poor | 722 (15% | 218 (26% | 787 (16% | 204 (25% | 802 (17% | 227 (23% | 1,010 (16% | 285 (25% |
| usual\_medicalcare\_status |  |  |  |  |  |  |  |  |
| No place | 355 (7.3% | 120 (14% | 343 (6.9% | 147 (18% | 409 (8.5% | 178 (18% | 389 (6.1% | 138 (12% |
| Usual place | 4,528 (93% | 719 (86% | 4,610 (93% | 672 (82% | 4,410 (92% | 801 (82% | 6,012 (94% | 1,017 (88% |
| smoking\_status |  |  |  |  |  |  |  |  |
| Never | 2,511 (51% | 579 (69% | 2,740 (55% | 607 (74% | 2,775 (58% | 736 (75% | 3,758 (59% | 876 (76% |
| Current | 1,137 (23% | 118 (14% | 980 (20% | 110 (13% | 938 (19% | 112 (11% | 1,115 (17% | 115 (10.0% |
| Former | 1,235 (25% | 142 (17% | 1,233 (25% | 102 (12% | 1,106 (23% | 131 (13% | 1,528 (24% | 164 (14% |
| alcohol\_status |  |  |  |  |  |  |  |  |
| Never | 1,219 (25% | 377 (45% | 1,215 (25% | 352 (43% | 1,103 (23% | 411 (42% | 1,280 (20% | 449 (39% |
| Current | 2,834 (58% | 335 (40% | 2,901 (59% | 342 (42% | 2,906 (60% | 420 (43% | 4,005 (63% | 532 (46% |
| Former | 830 (17% | 127 (15% | 837 (17% | 125 (15% | 810 (17% | 148 (15% | 1,116 (17% | 174 (15% |
| diabetes\_status | 436 (8.9% | 108 (13% | 596 (12% | 133 (16% | 658 (14% | 209 (21% | 942 (15% | 239 (21% |
| 1n (% | | | | | | | | |

Table 1 shows the summary of mammography screening status and demographic characteristics of the study population.

**Figure 1: Summary of mammography screening status and the Hispanic status of study population in 2000, 2005, 2010, and 2015**

|  |
| --- |
| Summary of mammography screening status and the Hispanic status of study population in 2000, 2005, 2010, and 2015. |

Figure 1 shows the summary of mammography screening status and the Hispanic and non-Hispanic status of study population in different years (i.e. 2000, 2005, 2010, and 2015).

# 6. 4.2 Statistical Analysis

# 7. 4.2.1 Model Fitting

**Table 2: Full model fitting using mammogram\_status as outcome variable and other variables as predictor variables.**

# A tibble: 30 × 5  
 term estimate std.error statistic p.value  
 <chr> <dbl> <dbl> <dbl> <dbl>  
 1 (Intercept) -1.54 0.232 -6.67 2.57e- 11  
 2 as.factor(year)2005 -0.0216 0.0565 -0.383 7.01e- 1  
 3 as.factor(year)2010 0.221 0.0594 3.72 1.98e- 4  
 4 as.factor(year)2015 -0.0620 0.0556 -1.12 2.65e- 1  
 5 as.factor(age)2 1.05 0.0567 18.5 3.35e- 76  
 6 as.factor(age)3 1.48 0.0652 22.6 1.76e-113  
 7 as.factor(age)4 1.64 0.0723 22.7 8.15e-114  
 8 as.factor(age)5 1.77 0.0810 21.9 1.60e-106  
 9 as.factor(age)6 1.52 0.0848 18.0 3.80e- 72  
10 as.factor(age)7 1.50 0.0906 16.5 1.92e- 61  
11 as.factor(hispanic\_status)2 0.103 0.0578 1.77 7.59e- 2  
12 as.factor(income)2 -0.272 0.0555 -4.90 9.46e- 7  
13 as.factor(education)2 0.765 0.207 3.69 2.22e- 4  
14 as.factor(education)3 0.993 0.208 4.78 1.73e- 6  
15 as.factor(education)4 1.35 0.209 6.44 1.22e- 10  
16 as.factor(education)5 1.47 0.212 6.94 4.05e- 12  
17 as.factor(marital\_status)1 0.484 0.0610 7.94 2.06e- 15  
18 as.factor(marital\_status)2 0.362 0.0630 5.75 9.10e- 9  
19 as.factor(marital\_status)3 0.274 0.112 2.45 1.44e- 2  
20 as.factor(region\_residence)2 -0.209 0.0681 -3.06 2.21e- 3  
21 as.factor(region\_residence)3 -0.181 0.0613 -2.94 3.23e- 3  
22 as.factor(region\_residence)4 -0.256 0.0660 -3.88 1.06e- 4  
23 as.factor(insurance\_status)1 -0.706 0.0560 -12.6 2.37e- 36  
24 as.factor(health\_status)2 0.105 0.0578 1.83 6.80e- 2  
25 as.factor(usual\_medicalcare\_status)1 0.906 0.0604 15.0 6.20e- 51  
26 as.factor(smoking\_status)1 -0.344 0.0521 -6.60 3.99e- 11  
27 as.factor(smoking\_status)2 0.117 0.0575 2.04 4.12e- 2  
28 as.factor(alcohol\_status)1 0.517 0.0496 10.4 1.81e- 25  
29 as.factor(alcohol\_status)2 0.385 0.0635 6.06 1.37e- 9  
30 as.factor(diabetes\_status)1 0.173 0.0677 2.55 1.07e- 2

**Table 3: Model Performance Comparisons.**

Name Model AIC AIC\_wt AICc  
1 log\_fit\_year \_glm 19738.12 0.000000e+00 19738.12  
2 log\_fit\_hispanicstatus \_glm 19633.33 0.000000e+00 19633.33  
3 log\_fit\_age \_glm 18480.88 0.000000e+00 18480.89  
4 log\_fit\_income \_glm 19516.57 0.000000e+00 19516.58  
5 log\_fit\_education \_glm 19424.18 0.000000e+00 19424.19  
6 log\_fit\_maritalstatus \_glm 19616.95 0.000000e+00 19616.95  
7 log\_fit\_regionresidence \_glm 19728.81 0.000000e+00 19728.81  
8 log\_fit\_insurancestatus \_glm 18957.47 0.000000e+00 18957.47  
9 log\_fit\_healthstatus \_glm 19774.16 0.000000e+00 19774.16  
10 log\_fit\_usualmedicalcarestatus \_glm 19022.04 0.000000e+00 19022.04  
11 log\_fit\_smokingstatus \_glm 19549.05 0.000000e+00 19549.05  
12 log\_fit\_alcoholstatus \_glm 19610.43 0.000000e+00 19610.43  
13 log\_fit\_diabetesstatus \_glm 19732.16 0.000000e+00 19732.17  
14 log\_fit\_fullmodel \_glm 16980.97 1.000000e+00 16981.05  
15 log\_fit\_year\_age\_hispanic \_glm 18385.01 1.311485e-305 18385.02  
16 log\_fit\_demographic \_glm 17764.12 8.727922e-171 17764.15  
17 log\_fit\_medical \_glm 18638.94 0.000000e+00 18638.94  
18 log\_fit\_smoking\_alcohol \_glm 19380.92 0.000000e+00 19380.92  
 AICc\_wt BIC BIC\_wt R2\_Tjur RMSE Sigma  
1 0.000000e+00 19770.60 0.000000e+00 1.644080e-03 0.3426207 0.8911567  
2 0.000000e+00 19649.57 0.000000e+00 6.231817e-03 0.3418326 0.8888418  
3 0.000000e+00 18537.73 7.167617e-286 6.014278e-02 0.3324313 0.8622084  
4 0.000000e+00 19532.82 0.000000e+00 1.175688e-02 0.3408810 0.8861945  
5 0.000000e+00 19464.79 0.000000e+00 1.544778e-02 0.3402439 0.8840108  
6 0.000000e+00 19649.43 0.000000e+00 7.213888e-03 0.3416636 0.8884161  
7 0.000000e+00 19761.30 0.000000e+00 1.956404e-03 0.3425671 0.8909465  
8 0.000000e+00 18973.71 0.000000e+00 4.093044e-02 0.3358118 0.8734059  
9 0.000000e+00 19790.40 0.000000e+00 2.442188e-05 0.3428985 0.8920247  
10 0.000000e+00 19038.28 0.000000e+00 3.952442e-02 0.3360579 0.8748925  
11 0.000000e+00 19573.41 0.000000e+00 9.228383e-03 0.3413168 0.8869042  
12 0.000000e+00 19634.79 0.000000e+00 7.057120e-03 0.3416906 0.8882957  
13 0.000000e+00 19748.41 0.000000e+00 1.601106e-03 0.3426281 0.8910767  
14 1.000000e+00 17224.59 1.000000e+00 1.348906e-01 0.3202158 0.8257132  
15 1.354347e-305 18474.33 4.182509e-272 6.482649e-02 0.3316858 0.8598492  
16 8.923572e-171 17918.41 2.176579e-151 9.382133e-02 0.3269057 0.8449427  
17 0.000000e+00 18679.54 1.148866e-316 5.873770e-02 0.3329018 0.8659485  
18 0.000000e+00 19421.52 0.000000e+00 1.636665e-02 0.3401369 0.8830251  
 Log\_loss Score\_log PCP  
1 0.3970162 -Inf 0.7652221  
2 0.3949880 -Inf 0.7663010  
3 0.3715970 -Inf 0.7789789  
4 0.3926387 -Inf 0.7676003  
5 0.3906589 -Inf 0.7684682  
6 0.3945780 -Inf 0.7665319  
7 0.3968290 -Inf 0.7652955  
8 0.3813882 -Inf 0.7744608  
9 0.3978220 -Inf 0.7648412  
10 0.3826876 -Inf 0.7741302  
11 0.3932520 -Inf 0.7670056  
12 0.3944870 -Inf 0.7664950  
13 0.3969769 -Inf 0.7652120  
14 0.3404896 -Inf 0.7965569  
15 0.3695067 -Inf 0.7800803  
16 0.3566911 -Inf 0.7868989  
17 0.3748579 -Inf 0.7786485  
18 0.3897882 -Inf 0.7686843

# 8. 4.2.2 Model Evaluation

**Figure 2: ROC Curve for the full model**

|  |
| --- |
| ROC Curve for the full model. |

**Figure 3: ROC Curve for the alternative model**

|  |
| --- |
| ROC Curve for the alternative model. |

# 9. 4.2.3 Machine Learning

**Table 4: Model Selection using Null, Decision tree and Lasso methods.**

Model ROC\_AUC  
1 Null 0.5  
2 Decision Tree 0.7  
3 Lasso 0.8

# 10. 5 Discussion

[\*\*Summary, interpretation, strengths, and limitations of the study to be added]

# 11. 6 Conclusions

[\*\*Main take-home messages from the study to be added]

# 12. 7 References

1. Gorina Y, Elgaddal N. [Patterns of Mammography, Pap Smear, and Colorectal Cancer Screening Services Among Women Aged 45 and Over](https://www.ncbi.nlm.nih.gov/pubmed/34181518). *National Health Statistics Reports*. 2021;(157):1-18.

2. Tabár L, Gad A, Holmberg LH, et al. REDUCTION IN MORTALITY FROM BREAST CANCER AFTER MASS SCREENING WITH MAMMOGRAPHY. *The Lancet*. 1985;325(8433):829-832. doi:[10.1016/S0140-6736(85)92204-4](https://doi.org/10.1016/S0140-6736(85)92204-4)

3. Kerlikowske K, Grady D, Rubin SM, Sandrock C, Ernster VL. [Efficacy of screening mammography. A meta-analysis](https://www.ncbi.nlm.nih.gov/pubmed/7799496). *JAMA*. 1995;273(2):149-154.

4. Shapiro S, Venet W, Strax P, Venet L, Roeser R. [Ten- to fourteen-year effect of screening on breast cancer mortality](https://www.ncbi.nlm.nih.gov/pubmed/6955542). *Journal of the National Cancer Institute*. 1982;69(2):349-355.

5. Lăără E, Day NicholasE, Hakama M. TRENDS IN MORTALITY FROM CERVICAL CANCER IN THE NORDIC COUNTRIES: ASSOCIATION WITH ORGANISED SCREENING PROGRAMMES. *The Lancet*. 1987;329(8544):1247-1249. doi:[10.1016/S0140-6736(87)92695-X](https://doi.org/10.1016/S0140-6736(87)92695-X)

6. Siu AL, on behalf of the U.S. Preventive Services Task Force. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Annals of Internal Medicine*. 2016;164(4):279. doi:[10.7326/M15-2886](https://doi.org/10.7326/M15-2886)

7. Selvin E, Brett KM. Breast and cervical cancer screening: Sociodemographic predictors among White, Black, and Hispanic women. *American Journal of Public Health*. 2003;93(4):618-623. doi:[10.2105/ajph.93.4.618](https://doi.org/10.2105/ajph.93.4.618)

8. Aftab IB, Ahmed A, Mumu SK, Mouly TF, Commar DS. Trends and Disparities in Breast Cancer Incidence-Mortality Rates of Black-White Women in the U.S.: 2000-2016. *Advances in Breast Cancer Research*. 2021;10(04):200-217. doi:[10.4236/abcr.2021.104017](https://doi.org/10.4236/abcr.2021.104017)

9. Achat H, Close G, Taylor R. Who has regular mammograms? Effects of knowledge, beliefs, socioeconomic status, and health-related factors. *Preventive Medicine*. 2005;41(1):312-320. doi:[10.1016/j.ypmed.2004.11.016](https://doi.org/10.1016/j.ypmed.2004.11.016)

10. Monnat SM. Race/Ethnicity and the Socioeconomic Status Gradient in Women’s Cancer Screening Utilization: A Case of Diminishing Returns? *Journal of Health Care for the Poor and Underserved*. 2014;25(1):332-356. doi:[10.1353/hpu.2014.0050](https://doi.org/10.1353/hpu.2014.0050)

11. Blewett, Lynn A., Drew, Julia A. Rivera, Griffin, Risa, King, Miriam L., Williams, Kari C.W. IPUMS Health Surveys: National Health Interview Survey, Version 6.2. Published online 2016. doi:[10.18128/D070.V6.2](https://doi.org/10.18128/D070.V6.2)