



Published in final edited form as:

Am J Health Behav. 2014 July ; 38(4): 492–500. doi:10.5993/AJHB.38.4.2.

Continuum of Mammography Use among US Women: Classification Tree Analysis

Annie Gjelsvik, PhD [Assistant Professor],

Brown University School of Public Health, Department of Epidemiology, Providence, RI.

Michelle L. Rogers, PhD [Senior Project Analyst],

Brown University School of Public Health, Center for Population Health and Clinical Epidemiology, Providence, RI.

Melissa A. Clark, PhD, MS [Professor],

Brown University School of Public Health, Department of Epidemiology, Providence, RI.

Hernando C. Ombao, PhD [Associate Professor], and

University of California at Irvine, Department of Statistics, Irvine, CA.

William Rakowski, PhD [Professor]

Brown University School of Public Health, Department of Behavioral and Social Sciences, Providence, RI.

Abstract

Objectives—To identify women with low mammography utilization. Methods: We used Classification Tree Analysis among women aged 42-80 from the 2008 Behavioral Risk Factor Surveillance System (N = 169,427) to identify sub-groups along a continuum of screening.

Results—Women with neither a primary care provider nor health insurance had the lowest utilization (33.9%) and were 2.8% of the sample. Non-smoking women aged 55-80, with a primary care provider, health insurance, and income of \$75,000 or more had the highest utilization (90.7%) and comprised 5% of the sample.

Conclusion—As access to primary care providers and health insurance increases with the Affordable Care act, classification tree analyses may help to identify women of high priority for intervention.

Keywords

Classification Tree Analysis (CTA); mammography; breast cancer screening

In order to achieve the morbidity and mortality reduction benefits of breast cancer screening, there must be high rates of mammography utilization across all sectors of the screening-

Correspondence Dr Gjelsvik; annie_gjelsvik@brown.edu.

Human Subjects Statement

This study only used publically available data.

Conflict of Interest Statement

The authors have no conflict of interest.

eligible population. The Healthy People 2020 goal is for 81.1% of women age 50-74 to have at least one mammography in the past 2 years (age adjusted to the year 2000).¹ Yet, even though population utilization rates have become relatively high in the past 20 years, the increase seems to have plateaued.²⁻⁴ The plateau has been attributed to a small decrease in the percent of recent mammography among women who were considered to be early adopters of mammography (women age 50-64, insured women, women with higher incomes and white women).^{4,5} In addition, pockets of low utilization remain, such as the uninsured, those who have no source of usual care or use an emergency department as their main source of care, women with less than a high school education, and women living in the United States for less than 10 years.⁶ Precise specification of groups with low utilization allows us to identify their salient barriers and facilitators, which better informs intervention audience targeting and the content of intervention messages and materials.

Surveillance of crude utilization rates and multivariable logistic regression have been typical ways to identify groups in the population with underutilization, as well for identifying the correlates of lower utilization. There are, however, constraints to this approach for population segmentation and intervention targeting. Crude-rate surveillance often uses only 1- or 2-variable classifications for specifying at-risk populations. Logistic regression produces refined (adjusted) estimates of the associations between independent variables and screening status, controlling for correlations among the independent variables. However, logistic regression output does not directly indicate how to combine variables, except additively, to specify low utilization groups using several variables concurrently; nor does the output specify which variables to combine. Non-linear associations with a dependent variable, such as U-shaped, inverted U-shaped (quadratic), or functions with asymptotes (those whose association plateaus after a “threshold” value on an independent variable), are also not routinely identified by logistic regression analysis.

In addition, main-effects logistic regression results are based on a logic of “ors.” That is, the main effects associations denote the risk of underutilization for persons with ‘characteristic #1,’ *or* ‘characteristic #2,’ *or* ‘characteristic #3’. In practice, however, interventions are conducted with priority populations who are defined by several characteristics simultaneously, which is a logic of “ands” (eg, urban, low-income women, and are members of one or more racial/ethnic minorities, and are without a regular source of care). For intervention planning, the specification of barriers and facilitators is most effective when it can be done for persons who have combinations of characteristics. It is therefore important to have a methodology that takes advantage of information available from crude screening rates and from identification of groups with low utilization based on a logic of “ands,” so that the resulting classification of priority populations is based on several characteristics in combination.

An analytic method that can be used to identify groups with low utilization is Classification Tree Analysis (CTA).^{7,8} Classification tree analysis is a non-parametric multivariable technique, sometimes referred to as signal detection, which is designed to identify combinations of variables efficiently. The results of a classification tree inherently show combinations among independent variables as a function of where they are most relevant in the process of tree creation. CTA also automatically combines categories of an independent

variable that have similar status on a dependent variable, which can account for non-linear relationships such as U-shaped, inverted-U, or threshold associations. In addition, unlike discriminate analysis, CTA does not require continuous-valued independent variables and allows for non-linearity in its segmentation scheme.

The objective of this study was to identify groups based on health status, health access, health behaviors and demographic characteristics in combination along the range of mammography utilization. In relation to breast cancer screening compliance, CTA has been used to identify subgroups of women age 50-75 who were not up-to-date on mammography screening,^{6,9} to prospectively identify those who did not adhere to mammography screening guidelines,¹⁰ to identify subgroups of the National Cancer Institute's Cancer Information Services users,⁹ and to predict mammography using population-based data in Portugal.¹¹ However, CTA has not been frequently used either with national-level mammography data specifically, or cancer screening behavior more broadly.

METHODS

Data Source and Sample

The analyses for this paper use the 2008 Behavioral Risk Factor Surveillance System (BRFSS). The BRFSS is a national-level dataset used to monitor cancer screening status. The BRFSS is a collaboration between the Centers for Disease Control and Prevention and each state, the District of Columbia, and affiliated U.S. territories. The analysis sample included women aged 42-80 (N = 169,427), exclusive of U.S. territories.

The BRFSS is an annual telephone survey using disproportionate random sampling methodology.¹² States are responsible for either conducting their own surveys or working through subcontractors. Each annual BRFSS is comprised of: (i) a "core" set of questions that are mandated to be in each state's survey; (ii) "topical modules," each having a standard set of questions, but asked by a state at its discretion; and (iii) items that are added by a state to address its own relevant issues. The BRFSS has traditionally been a survey of landline-only phone numbers, but in 2008, 18 states collected information via cell phone numbers as a pilot project. Those cell phone data are not part of the public-use dataset.

Tree Growth

The classification tree was grown using the Exhaustive Chi-Squared Automatic Interaction Detection (E-CHAID) routine within SPSS Answer Tree.¹³ The tree was grown by systematically splitting the sample from the root node through a succession of parent and child nodes, to the final set of terminal nodes. Our objective was to have a tree that could use several variables to define the terminal nodes, while still being of manageable size for population targeting. Therefore, our a-priori determined criteria were that the tree could have no more than 7 levels, that a split could occur only in a parent node with 5% or more of the total sample (ie, at least N = 4236), and each child node had to contain at least 2.5% of the total sample (ie, at least N = 2118). These levels were selected a priori to assure that the probabilities of on-schedule and off-schedule could be estimated with high precision and accuracy.

Variables for Tree Construction

Mammography status—The BRFSS asked 2 closed-ended questions. The first assessed ever having a mammogram. For women who ever had, the second question asked for time since most recent examination using the preset categories of: Less than 12 months; 12 months to less than 2 years, 2 years to less than 3 years, 3 to 5 years, and more than 5 years. Consistent with Healthy People 2020 objectives¹ and US Preventive Services Task Force guidelines¹⁴ we categorized the dependent variable as (1) less than or equal to 23 months, versus (0) greater than 23 months/Never/Don't Know (DK)/Refused.

Independent variables—The variables used to grow the tree were markers of access to care, health behaviors such as smoking, health status such as having limited activity, and sociodemographic characteristics: health insurance coverage (yes, no); whether or not needed health care could be afforded in the previous 12 months (yes, no); has a primary care provider (no personal physician, one or more personal physicians); smoking status (current, former, never); a combined measure of activity limitation and use of equipment (no activity limitation or equipment use, activity limitation but no equipment use, activity limitation and equipment use); age (42-54, 55-64, 65-80); number of people in the household (1, 2, 3 or more); education (less than high school, high school graduate/GED, some college, college graduate); marital status (never married, previously married, married/partnered); race/ethnicity (Hispanic; Black, non-Hispanic; White, non-Hispanic; Other, non-Hispanic); region of the country (West, Midwest, South, Northeast); income (\$0-<\$35,000; \$35,000-<\$50,000, \$50,000-<\$75,000, \$75,000 or more; missing/DK/refused); and Body Mass Index (not overweight, overweight, obese, missing/DK/refused) (Table 1). We retained DK/refused/missing data for income and Body Mass Index because of relatively large percentages of women in these categories, 13.8% and 5.8% respectively.

Similar to regression analysis, tree-based methods can be sensitive to differences in the covariate distributions of data among the independent variables across samples. Therefore, the final terminal nodes can differ even for 2 random-half samples from the same parent population, especially if the parent population is not large. This characteristic is consistent with the fact that CTA segments a sample into progressively more refined, smaller subsamples. As a result nodes created lower in the tree are more susceptible to showing variation due to smaller sample sizes, rather than staying at the level of algorithms that use the entire sample to estimate the association between the independent and dependent variables. Analysis samples are therefore sometimes split randomly into a “training sample” on which a tree is grown, and a “testing sample” to which results of the “training” tree are applied and compared.

Consistent with this methodology, we first ran the analyses with approximately 50% of respondents randomly assigned to each of the Training and Testing samples (N = 84,672 & N = 84,755 respectively; data not shown, available on request). Mammography rates were 77.9% (Training) and 77.9% (Testing). The 25% and 75% quartile points were 71.2% and 86.4% (Training); 69.8% and 85.6% (Testing). The CTA results were virtually identical to the full sample, the difference being an extra split at the terminal node level in the full sample, likely due to the larger sample size. We therefore present results for the full sample.

For ease of discussion, the nodes are shown in ascending order of mammography utilization. They are also shown by quartile (lowest 25%, 25%-75% interquartile range, and the highest 25%). A pictorial depiction of the classification tree is available from the corresponding author upon request.

RESULTS

Overall 77.9% (131,991/169,427) of women had a mammogram within the past 23 months. Women who were in compliance with mammogram screening were older, more likely to be college graduates, had higher incomes, and were more likely to live in the Northeast compared to women not in compliance. Women in compliance with mammogram screening were more likely to be married or partnered and more likely to live in a household with a total of 2 people. Women who had at least one mammogram in the past 23 months had better health care access; they were more likely to be insured, be able to afford needed health care, and to have a primary care provider. In addition, women who had a mammogram in the past 23 months were more likely to have no activity limitation and to have never smoked. The differences in race/ethnicity and obesity status were small (Table 2).

Overall results

For the full sample (Table 3), the overall mammography utilization rate was 77.9% with a 25%-75% interquartile range of 71.7% to 86.0%. There were a total of 26 terminal nodes, or distinct subgroups of the sample, along a large continuum of utilization. Node 5 (No primary care provider; Uninsured) had the lowest utilization, at 33.9%. On the other end, Nodes 30 and 38 (Primary care provider; Insured; Never or Former smokers; Income \$75K or more; Age 55-80) each reached utilization rates of 90%. The largest branch in the tree had 6 levels.

The bottom quartile of nodes had utilization rates ranging from 33.9% to 70.5%, and accounted for 21.2% of the population. The upper quartile had rates from 86.1% to 90.7% and accounted for 28.7% of the population. There was not as large a range of rates in the top quartile (4-5%), or even in the 25%-75% interquartile range (10.2%), as there was in the lowest quartile (36-37%). Among nodes with utilizations in the upper quartile there was clustering of middle and higher income women in nodes with 86% utilization.

Major splitting variables

The initial splitting variable was having a source of primary care (yes, no), followed by health insurance status (yes, no). In the lowest quartile, there were no further splitting variables for women who had *neither* a source of care nor health insurance (Node 5). Ability to afford care (yes, no) was the splitting variable for women with a source of care but who were uninsured (Nodes 7 and 8), the difference being 15% higher utilization for women who said they could afford care. However, even though there was higher utilization among women in Node 7, it was still only 64.4%. For Nodes 12 and 13, the difference between previously/never-married women and married/partnered women was 8%. Smoking status (current, former, never) was the splitting variable for women with both a source of care and health insurance; income (<\$35K, \$35-\$50K, \$50-\$75K, >\$75K, Missing) was the

subsequent splitting variable under all 3 smoking status nodes. Two of the 3 nodes that included current smokers were in the lowest quartile.

Less important splitting variables

Variables lower in the tree than income were activity limitation, age, marital status, number of persons in the household, and region of the country. Although meeting the eligibility criteria for inclusion in the tree, they made only small contributions in terms of discriminability within their nodes. The largest contribution was for activity limitation (yes; no) for Nodes 41 and 40 (a 5.3% difference), and for Nodes 36 and 37 (a 4.6% difference). Region of the country (West/Midwest; South/Northeast) showed a 3.8% difference between Nodes 43 and 42, while marital status (previously/never; married/partnered) had a 3.7% difference between Nodes 28 and 29. Similarly small differences were seen for number of people in the household (Nodes 33 and 32) and age (Nodes 35 and 34, Nodes 31 and 30, Nodes 39 and 38). Body Mass index did not enter the tree. The variable for activity limitation/equipment use always entered the tree with both equipment use categories combined for women reporting a limitation to activity.

DISCUSSION

In this study, the classification-tree results identified groups with very low and very high utilization. Using CTA has the benefit of not having to engage in exploratory, post-hoc combinations of individual independent variables to look for groups with lowest and highest utilization; the continuum is present in the terminal nodes of the tree. It is, however, still necessary to examine variables comprising the subgroups along the continuum of utilization rates in order to discern relative importance and trends in association.

The group reporting no source of primary care and no health insurance (Node 5) had startlingly low mammography utilization, even though they comprised only 2.8% (N = 4749) of the sample. The results for primary care provider and health insurance status are not surprising, as they are essential resources for service utilization. Any combination of primary care provider status and health insurance status where even one is a “no” should be a flag for follow-up to determine mammography status.

The 4 nodes with utilizations of less than 60% were 10.9% of the sample and the lowest quartile groups taken together were 21.2% of the sample. This result is therefore consistent with the notion that there are several relatively small “pockets” in the population, each with low utilization, but which add to a notable number of women, as was found by Rakowski and Clark⁶ in the 1992 Cancer Control Supplement to the National Health Interview Survey. In addition, there was wide variability around the overall rate of 77.9%, indicating that the national-level screening rate is simply a summary value that hides substantial variation across groups in the population.

Ability to afford care in the past 12 months and marital status helped to define several of the groups with lowest utilization. However even the groups with the relatively more favorable status on those variables (ie, could afford care; married/partnered) were among the 5 with lowest utilization.

The relatively high placement in the tree for smoking status is consistent with data showing that women who smoke have mammography rates up to 14% lower than former and never smokers.¹⁵⁻¹⁹ Similar to previous studies,^{6,10} income was an important splitting variable. Income was the splitting variable for each category of smoking, and there was somewhat of a gradient of income in conjunction with lower to higher screening rates, across the range of terminal nodes, denoting income's importance. These results suggest that even among populations of women who are insured and have a primary care provider, those who smoke should be identified and targeted. For instance, an integrated health care system could identify these women for additional intervention.²⁰

We did not prune the tree even though variables lower in the tree than income (activity limitation, age, marital status, number of persons in the household, and region of the country) were not associated with large differences in utilization. These variables are presented here to allow a comparison with results from future CTA analyses, similar to reporting results for variables that do not achieve standard levels of statistical significance in regression analyses. In fact, the tree was not complex, and the higher-placed nodes in branches of the tree would not have been affected by pruning. There is a useful balance to be gained between showing trees with only the clearly most important variables versus showing trees that include some variables of less importance but that may still be helpful for future comparisons.

Groups with missing income as the sole income category are challenging to interpret -- in this analysis there are 3 (Nodes 29, 20, and 28). In addition, missing income was combined with income of \$35K or less for Nodes 41 and 40. The prevalence of missing income in the 2008 BRFSS (13.8% in our sample) argues for keeping it in the analysis, as is commonly done also with logistic regression where it can be its own dummy variable. In classification tree analysis, however, where variables are combined to create the utilization groups, its presence can be frustrating. In the present study, the affected groups had utilization that was either lower (Nodes 41, 40) or higher (Nodes 29, 20, 28) than the overall utilization in the sample.

Similar to findings from previous classification tree and mammography screening behavior studies in the United States,^{6,10} race/ethnicity did not figure into the composition of the groups. There may be 2 reasons for this finding. First, in the 2008 BRFSS Black, White, and Hispanic women's mammography rates were fairly comparable (Black=80.5%, White=83.1%, Hispanic=76.6%). Second, for the first 2 splitting variables, racial/ethnic differences were also fairly comparable (primary care provider: Black=91.0%, White=92.5%, Hispanic=81.8%; has health insurance: Black=85.9%, White=93.1%, Hispanic=78.3%). It is therefore possible that race/ethnicity simply did not have the necessary discriminatory power to be selected, given CTA's "local" process of splitting based on the subsample in each respective node.

As with any multivariable procedure, the results of CTA are conditional on the independent variables that are used. The BRFSS core survey collects a range of information, although with a limited set of questions for each construct/variable. In addition, as a cross-sectional survey, in which receipt of mammography is assessed over the preceding period of time, it is

preferable that the independent variables have a likelihood of being relatively stable on the individual-level during the period of recall. As with other analytic procedures, therefore, consistencies of association, and non-association, across analyses and datasets are important in order to establish a variable's importance (or unimportance) for classification, as well as identifying the groups with lowest utilization. In the studies that had information on attitudes and knowledge regarding mammography^{10,11} or information on psychosocial correlates regarding susceptibility to breast cancer,^{6,10} these variables were important predictors of adherence. Unfortunately the BRFSS did not collect this information. Similarly while availability of mammography is associated with screening rates,²¹ the BRFSS does not collect area-level correlates of mammography such as availability of mammography facilities where a woman lives.

Results of CTA will also be influenced by the stopping rules for growing a tree, including maximum number of levels in the tree and minimum size of a node to allow splitting to occur. With samples as large as the BRFSS, trees can become enormous unless restrictions are placed. For example, Node 39 was just over 4500 women, but comprised only 2.7% of the sample. As was noted above, variables on the tree lower than income made small contributions to group differences in this analysis. Growing the trees further than we did would have added more groups along the continuum, but as Table 3 shows there was often not much difference between adjacent groups on the continuum.

Even with these considerations, the CTA analysis identified several groups of women with very low mammography utilization. Mammography utilization cannot be optimized until these group's rates are improved substantially. Primary care provider status and health insurance status are of highest priority to determine in women's interactions with health professionals. There is also opportunity to increase screening at the upper range of the continuum. For example, it could be beneficial to identify barriers to utilization among women in Node 31 (has primary care provider; has insurance; never smoked; income \$75K or more; age 42-54). Their mammography utilization was relatively high (86.1%), but reasons for non-screening in a seemingly high-resource group is a legitimate area for study. Given the nature of CTA, results can be useful for targeting policies and programs best suited for groups of individuals at their particular points along the continuum of screening utilization.

Acknowledgments

This study was supported by the NIH National Cancer Institute (5R21CA139179-02).

REFERENCES

1. Healthy People 2020. [March 26, 2012] U.S. Department of Health and Human Services. Office of Disease Prevention and Health Promotion (on-line). Available at: <http://www.healthypeople.gov/2020/topicsobjectives2020/objectiveslist.aspx?topicId=5>.
2. Cancer screening - United States, 2010. MMWR. Morb Mortal Wkly Rep. 2012; 61(3):41–45. [PubMed: 22278157]
3. Vital signs: breast cancer screening among women aged 50-74 years - United States, 2008. MMWR. Morb Mortal Wkly Rep. 2010; 59(26):813–816. [PubMed: 20613705]

4. Breen N, Gentleman JF, Schiller JS. Update on mammography trends. *Cancer*. 2011; 117(10):2209–2218. [PubMed: 21523735]
5. Breen N, Kessler L. Changes in the use of screening mammography: evidence from the 1987 and 1990 National Health Interview Surveys. *Am J Public Health*. 1994; 84(1):62–67. [PubMed: 8279613]
6. Rakowski W, Clark MA. Do groups of women aged 50 to 75 match the national average mammography rate? *Am J Prev Med*. 1998; 15(3):187–197. [PubMed: 9791636]
7. Breiman, L.; Friedman, JH.; Olshen, RS. C.J. *Classification and Regression Trees*. 2nd Edition. Wadsworth; Pacific Grove, CA: 1984.
8. Zhang, H.; Singer, B. *Recursive Partitioning in the Health Sciences*. Springer; New York, New York: 1999.
9. Sullivan HW, Finney Rutten LJ. Cancer prevention information seeking: a signal detection analysis of data from the cancer information service. *J Health Commun*. 2009; 14(8):785–796. [PubMed: 20029711]
10. Calvocoressi L, Stolar M, Kasl SV, et al. Applying recursive partitioning to a prospective study of factors associated with adherence to mammography screening guidelines. *Am J Epidemiol*. 2005; 162(12):1215–1224. [PubMed: 16221800]
11. Freitas C, Tura LF, Costa N, Duarte J. A population-based breast cancer screening programme: conducting a comprehensive survey to explore adherence determinants. *Eur J Cancer Care (Engl)*. 2012; 21(3):349–359. [PubMed: 22077789]
12. Behavioral Risk Factor Surveillance System Operational and Users Guide. Version 3.0. United States Department of Health and Human Services, Centers for Disease Control and Prevention; 2006.
13. AnswerTree 2.0 User's Guide. SPSS Inc; 1998.
14. Recommendations for Adults: Guide to Clinical Preventive Services, 2012. U.S. Preventive Services Task Force; Rockville, MD: 2011.
15. Fredman L, Sexton M, Cui Y, et al. Cigarette smoking, alcohol consumption, and screening mammography among women ages 50 and older. *Prev Med*. 1999; 28(4):407–417. [PubMed: 10090870]
16. Hall HI, Uhler RJ, Coughlin SS, Miller DS. Breast and cervical cancer screening among Appalachian women. *Cancer Epidemiol Biomarkers Prev*. 2002; 11(1):137–142. [PubMed: 11815412]
17. Rakowski W, Breen N, Meissner H, et al. Prevalence and correlates of repeat mammography among women aged 55-79 in the Year 2000 National Health Interview Survey. *Prev Med*. 2004; 39(1):1–10. [PubMed: 15207980]
18. Rakowski W, Clark MA, Truchil R, et al. Smoking status and mammography among women aged 50-75 in the 2002 Behavioral Risk Factor Surveillance System. *Women Health*. 2005; 41(4):1–21. [PubMed: 16260411]
19. Rakowski W, Meissner H, Vernon SW, et al. Correlates of repeat and recent mammography for women ages 45 to 75 in the 2002 to 2003 Health Information National Trends Survey (HINTS 2003). *Cancer Epidemiol Biomarkers Prev*. 2006; 15(11):2093–2101. [PubMed: 17119033]
20. Kempe KL, Larson RS, Shetterley S, Wilkinson A. Breast cancer screening in an insured population: whom are we missing? *Perm J*. 2013; 17(1):38–44. [PubMed: 23596367]
21. Elkin EB, Atoria CL, Leoce N, et al. Changes in the availability of screening mammography, 2000-2010. *Cancer*. 2013; 119(21):3847. [PubMed: 23943323]

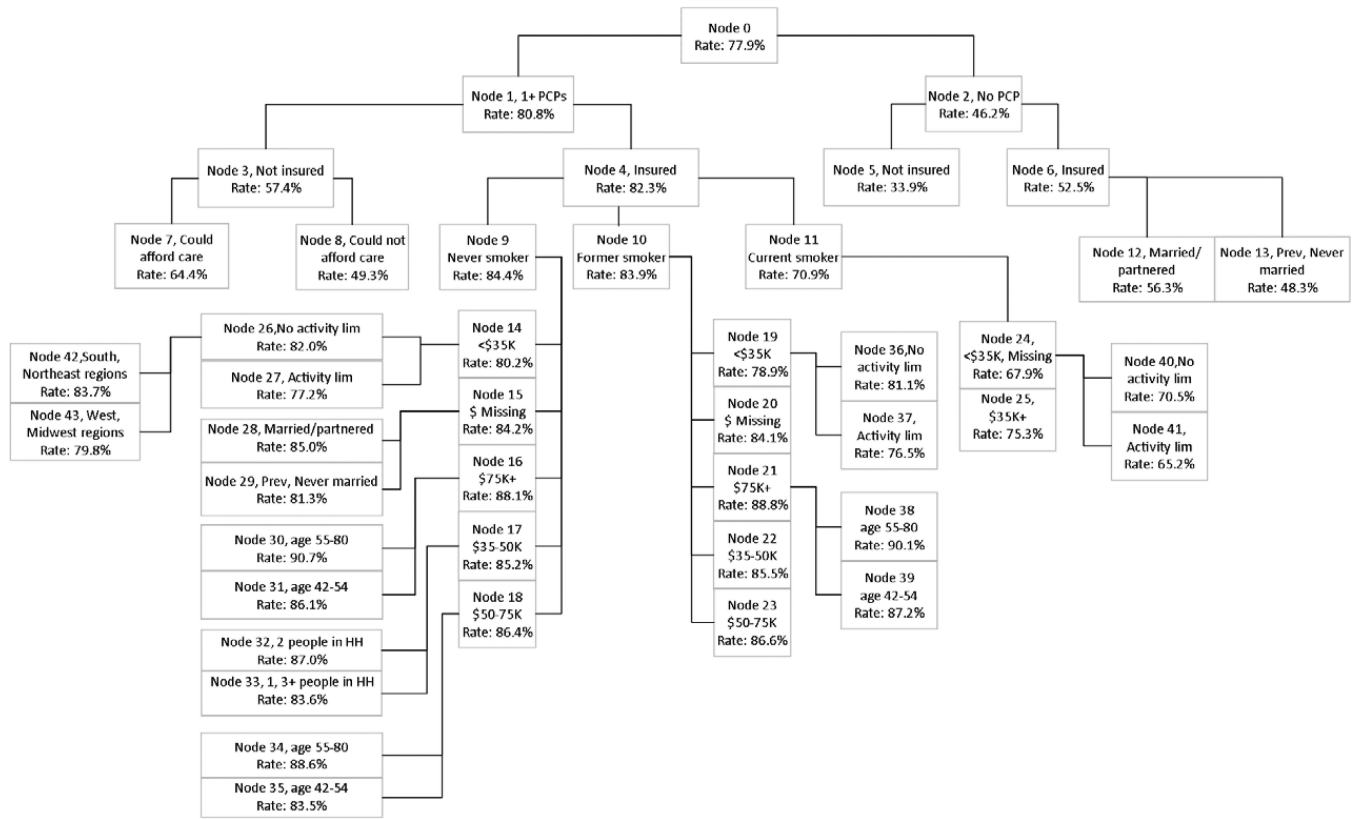


Figure 1. Classification Tree Diagram for the Dependent Variable of Having a Mammogram in the Past 23 Months, 169,427 Women age 42-80 in the 2008 Behavioral Risk Factor Surveillance System

Table 1

Independent Variable Categories

Variable	Categories
Health Insurance coverage	Yes No
Able to afford need health care in previous 12 months	Yes No
Primary care provider status	No personal physician One or more personal physician
Smoking status	Current smoker Former smoker Never smoker
Combined measure of activity limitation and use of equipment	No activity limitation or equipment use Activity limitation but not equipment use Activity limitation and equipment use
Age	42-54 years 55-64 years 65-80 years
Number people in household	1 2 3 or more
Education	Less than high school High school graduate or GED Some college College graduate
Marital status	Never married Previously married Married or partnered
Race/ethnicity	Hispanic Black non-Hispanic White non-Hispanic Other non-Hispanic
Region	West Midwest South Northeast
Income	\$0-<\$35,000 \$35,000-<\$50,000 \$50,000-<\$75,000 \$75,000 or more Missing/don't know/refused
Body Mass Index	Not overweight Overweight Obese Missing/don't know/refused

Table 2

Distribution of Independent Variables by Mammogram Status, 2008 Behavioral Risk Factor Surveillance System

Characteristic	Overall N = 169,427 %	Mammography in past 23 months*	
		Yes N = 131,991 %	No N = 37,436 %s
Age			
42-54	35.9	34.3	41.6
55-64	29.8	30.5	27.4
65-80	34.3	35.2	30.9
Education			
Less than HS	9.0	8.0	12.7
HS/GED	32.2	31.3	35.4
Some college	27.8	27.7	28.0
College	31.0	33.0	23.9
Race/Ethnicity			
Hispanic	4.6	4.5	4.8
White, Non-Hispanic	82.9	83.1	82.3
Black, Non-Hispanic	8.0	8.3	7.1
Other, Non-Hispanic	4.5	4.2	5.8
Income			
<\$35,000	37.7	34.5	49.2
\$35,000-<\$50,000	13.5	13.8	12.4
\$50,000-<\$75,000	14.3	15.1	11.5
\$75,000+	20.7	22.7	13.8
Missing/DK/Refused	13.8	14.0	13.1
Marital status			
Never married	6.8	6.4	8.5
Previously married	38.0	36.5	43.2
Married/partnered	55.2	57.1	48.3
Number of people in household			
1	34.0	33.4	36.3
2	42.6	44.1	37.3
3 or more	23.4	22.5	26.4
Health insurance coverage			
Not Insured	8.3	5.3	19.1
Insured	91.7	94.7	80.9
Region of country			
West	24.1	23.3	26.8
Midwest	22.5	22.3	23.3
South	32.9	32.6	33.9
Northeast	20.6	21.8	16.1

Characteristic	Mammography in past 23 months [*]		
	Overall N = 169,427 %	Yes N = 131,991 %	No N = 37,436 %
Activity limitation/equipment use			
No activity limitation or equipment use	69.7	71.1	65.1
Activity limitation with no equipment use	21.1	20.4	23.4
Activity limitation and equipment use	9.2	8.5	11.5
Able to afford needed care			
No	11.6	8.7	21.9
Yes	88.4	91.3	78.1
Has a primary care provider			
No personal physician	8.4	5.0	20.4
One or more personal physicians	91.6	95.0	79.6
Smoking status			
Current	16.3	13.2	27.0
Former	29.5	30.6	25.6
Never	54.3	56.2	47.4
Body Mass Index			
Not overweight	36.1	36.1	36.2
Overweight	27.6	27.3	28.6
Obese	30.9	31.4	29.0
Missing/DK/Refused	5.5	5.3	6.2

* All comparisons were statistically significant at $p < .0001$

Table 3

Results of Classification Tree Analysis of the 2008 Behavioral Risk Factor Surveillance System (N = 169,427; Women Aged 42-80)

Node #	Description and Quartile	Screening Rate	Node size	% of total sample
LOWEST QUARTILE				
5	PCP ^a =No, Insured=No	33.9%	4749	2.8%
13	PCP=No; Insured=Yes; Marital=Previously/Never Married	48.3%	4451	2.6%
8	PCP=Yes; Insured=No; Afford care=No	49.3%	4351	2.6%
12	PCP=No; Insured=Yes; Marital Status=Married/Partnered	56.3%	4982	2.9%
7	PCP=Yes; Insured=No; Afford care=Yes	64.4%	5036	3.0%
41	PCP=Yes; Insured=Yes; Smoking=Current; Income=Missing, <\$35K; Activity Limits=Yes	65.2%	6082	3.6%
40	PCP=Yes; Insured=Yes; Smoking=Current; Income=Missing, <\$35K; Activity Limits=No	70.5%	6204	3.7%
SEMI-INTERQUARTILE RANGE				
25	PCP=Yes; Insured=Yes; Smoking=Current; Income=\$35K+	75.3%	8724	5.2%
37	PCP=Yes; Insured=Yes; Smoking=Former; Income=<\$35K; Activity Limits=Yes	76.5%	7337	4.3%
27	PCP=Yes; Insured=Yes; Smoking=Never; Income=<\$35K; Activity Limits=Yes	77.2%	9147	5.4%
43	PCP=Yes; Insured=Yes; Smoking=Never; Income=<\$35K; Activity Limits=No; Region=West, Midwest	79.9%	7220	4.3%
36	PCP=Yes; Insured=Yes; Smoking=Former; Income=<\$35K; Activity Limits=No	81.1%	8325	4.9%
29	PCP=Yes; Insured=Yes; Smoking=Never; Income=Missing; Marital Status=Previously/Never Married	81.3%	4514	2.7%
35	PCP=Yes; Insured=Yes; Smoking=Never; Income=\$50-\$75K; Age=42-54	83.6%	5557	3.3%
33	PCP=Yes; Insured=Yes; Smoking=Never; Income=\$35-\$50K; # people in HH ^b =1, 3+	83.6%	5650	3.3%
42	PCP=Yes; Insured=Yes; Smoking=Never; Income=<\$35K; Activity Limits=No; Region=South, Northeast	83.7%	8791	5.2%
20	PCP=Yes; Insured=Yes; Smoking=Former; Income=Missing	84.1%	6143	3.6%
28	PCP=Yes; Insured=Yes; Smoking=Never; Income=Missing; Marital Status=Married/Partnered	85.0%	7398	4.4%
22	PCP=Yes; Insured=Yes; Smoking=Former; Income=\$35-\$50K	85.5%	6196	3.7%
HIGHEST QUARTILE				
31	PCP=Yes; Insured=Yes; Smoking=Never; Income=\$75K+; Age=42-54	86.1%	11210	6.6%
23	PCP=Yes; Insured=Yes; Smoking=Former; Income=\$50-\$75K	86.6%	6592	3.9%
32	PCP=Yes; Insured=Yes; Smoking=Never; Income=\$35-\$50K; # people in HH=2	87.0%	5274	3.1%
39	PCP=Yes; Insured=Yes; Smoking=Former; Income=\$75K+; Age=42-54	87.2%	4526	2.7%
34	PCP=Yes; Insured=Yes; Smoking=Never; Income=\$50-\$75K; Age=55-80	88.6%	7030	4.2%
38	PCP=Yes; Insured=Yes; Smoking=Former; Income=\$75K+; Age=55-80	90.1%	5443	3.2%

Node #	Description and Quartile	Screening Rate	Node size	% of total sample
30	PCP=Yes; Insured=Yes; Smoking=Never; Income=\$75K+; Age=55-80	90.7%	8495	5.0%

^aPCP = Has primary care provider

^b# people in HH = Number of persons in the household