UNIVERSITY OF NORTH TEXAS
HEALTH SCIENCE CENTER AT FORT WORTH
1970

Ylitalo, Kelly R., Urbanicity and Mammography Utilization: How Living in an Urban or Rural Area Affects Mammography Screening Utilization among Women in the United States. Master of Public Health (Epidemiology), May 2008, 67 pp., 7 tables, 1 figure, references.

Mammography is a widely used screening tool that can help prevent breast cancer mortality, yet utilization is not consistent. We utilized the 2006 Behavioral Risk Factor Surveillance System data to evaluate differences in mammography screening practices by urbanicity (urban vs. rural residence). Chi-square analyses, logistic regression, and propensity score matching were utilized to determine the association between urbanicity and mammography use. Women who lived in urban areas were more likely to utilize mammography compared to women who lived in rural areas, even after controlling for individual variables. Geographic access to health care as represented by individual urban or rural residence may contribute to mammography screening practices in the United States.

URBANICITY AND MAMMOGRAPHY UTILIZATION:

HOW LIVING IN AN URBAN OR RURAL AREA AFFECTS MAMMOGRAPHY

SCREENING UTILIZATION AMONG WOMEN IN THE UNITED STATES

Kelly R. Ylitalo, B.A.

APPROVED:

Major Professor

Committee Member

Committee Member

Department Chair

Dean, School of Public Health

URBANICITY AND MAMMOGRAPHY UTILIZATION: HOW LIVING IN AN URBAN OR RURAL AREA AFFECTS MAMMOGRAPHY SCREENING UTILIZATION AMONG WOMEN IN THE UNITED STATES

THESIS

Presented to the School of Public Health

University of North Texas

Health Science Center at Fort Worth

In Partial Fulfillment of the Requirements

For the Degree of

Master of Public Health

By

Kelly R. Ylitalo, B.A.

Fort Worth, Texas

May 2008

ACKNOWLEDGMENTS

I would like to thank Dr. Cardarelli for her guidance and mentorship not only throughout this thesis project but for my entire tenure at the UNTHSC SPH. Thank you to Dr. Zhang and Dr. Lackan as well for their encouragement, helpful comments, and time spent on this project. I appreciate their willingness to serve on my thesis committee.

Special thank you to John Adams for SAS tutoring and help, especially regarding propensity score matching. Also, thanks to John Adams and Rachael Jackson for always listening as I worked through my thought process out loud.

Thank you to Dr. Fischbach, Natalie Cole, and Eva Evans for allowing me to present propensity score matching in class and following up with insightful questions and comments that improved this investigation in its initial stages.

And thanks to Adam for his patience, always.

K.R.Y.

TABLE OF CONTENTS

TABLE OF CONTENTS, continued

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Breast cancer is a leading cause of cancer-related incidence and mortality for women in the United States (CDC, 2007). Early detection and treatment are the best methods to prevent breast cancer mortality, and mammography is widely accepted as an effective breast cancer screening modality for early detection. Screening mammography is reported to reduce breast cancer mortality by approximately 20% for women in their 40s and by approximately 30% for women in their 50s and 60s (Fletcher & Elmore, 2003). Despite mammography effectiveness, recent reports indicate that mammography utilization may be declining in the United States. Between the 1980s when it was introduced to community screening practice and the 1990s, the percentage of women over the age of 40 who self-reported mammography utilization increased appreciably (CDC, 2007). However, between the years 2000 and 2005, self-reported mammography decreased significantly (CDC, 2007).

Disparities in breast cancer incidence and mortality, as well as in mammography utilization, are widely documented. Income, education, race/ethnicity, and health insurance status are well-established predictors of self-reported mammography utilization (Meissner, Breen, Taubman, Vernon, & Graubard, 2007; Barrett & Legg, 2005; Calle, Flanders, Thun, & Martin, 1993). In addition, urbanicity can also be an important variable. Multiple studies have reported that residence in a rural area, as opposed to an urban area, predicts mammography underuse (Meissner, Breen, Taubman, Vernon, & Graubard, 2007; Calle, Flanders, Thun, & Martin, 1993; Mench & Mills, 2001).

Urbanicity may affect mammography utilization through a pathway of geographic health care access and provider availability. Compared to their urban counterparts, a greater percentage of individuals in rural areas lack health insurance and have poorer geographic access to a physician (Larson, Machlin, Nixon, & Zodet, 2004; Rosenthal, Zaslavsky, Newhouse, 2005).

Statement of Purpose

Urbanicity was conceptualized as a proxy for geographic access to health care. This study explored the relationship between individual urbanicity status (i.e., residence in an urban or rural area) and mammography utilization. Data from the 2006 Behavioral Risk Factor Surveillance System (BRFSS) survey was used to study this relationship among women aged 40 years and older in the United States. This study sought to answer the following question:

What is the effect of living in a rural area compared with living in an urban area on mammography utilization?

To determine the effect of urbanicity on mammography utilization, this study:

(1) Characterized a sample of women aged 40 years or older by individual sociodemographic variables of race, education, income, marital status, health insurance status, and urbanicity;

(2) Determined the proportion of these women aged 40 years or older who utilized mammography within the previous two years; and

(3) Explored the relationship between urbanicity and mammography utilization after accounting for covariates of race, education, income,

marital status, and health insurance status among a sample of women aged 40 years or older.

This study hypothesized that women living in urban areas are more likely to report utilization of screening mammography within the previous 2 years compared to women living in rural areas.

Public Health Importance

Mammography utilization among women in the United States is an important public health concern. *Healthy People 2010* aims to increase the proportion of women who have received a mammogram within the preceding two years (Department of Health and Human Services [DHHS], 2001a). Many studies have evaluated factors that serve as barriers or promoters of screening utilization. Baseline data from *Healthy People 2010* indicate that only 67% of all women aged 40 years and older received a mammogram within the preceding two years, and disparities in mammography utilization persisted among race/ethnicity, education, income, and urbanicity categories (see *Table 1*).

Table 1. Mammography Utilization Among Women Aged 40 Years and Older, United States, 1998[+].	Mammogram within previous 2 years (%)
TOTAL	67
Race/Ethnicity	
Non-Hispanic White	68
Non-Hispanic Black	66
Hispanic	61
American Indian or Alaska Native	45
Asian or Pacific Islander	61
Education level	
Less than high school	53
High school graduate	66
At least some college	73
Family Income Level	
Poor	50
Near poor	54
Middle/high income	73
Geographic location	
Urban	68
Rural	65

+ Age Adjusted to the year 2000 standard population.

Source: Healthy People 2010, from NHIS, NCHS, CDC.

While demographic variables like race/ethnicity, education, and income have been widely explored as predictors of utilization, urbanicity has received comparably less attention. Approximately twenty-five percent of all Americans live in rural areas, and residents of rural areas are less likely to utilize preventive screening services and report good health compared to residents of urban areas (DHHS, 2001b). The purpose of the present research study was to examine the effect of urbanicity, defined as urban or rural area of residence according to the U.S. Census Bureau guidelines, on mammography utilization. A nationwide sample from the BRFSS survey analyzed using logistic regression and propensity score matching was used to examine this effect. While the BRFSS is a widely used, publically-available survey data set, data from the 2006 BRFSS survey on mammography utilization have yet to be analyzed and publicly distributed at the time of this investigation. In addition, previous investigations have used logistic regression alone to analyze correlates of mammography utilization. These logistic regression models may be over-adjusted because collinear factors like race, education, and income are often included together as covariates in regression models yet likely influence screening utilization through the same pathway. The present study explored correlates of mammography utilization in the 2006 BRFSS data using propensity score matching, which mimicked randomization and provided a better method for controlling for confounding variables (Oakes & Johnson, 2006) . Ultimately, the evaluation of urbanicity and mammography utilization may assist public health officials to focus limited resources and efforts on the most appropriate areas to improve mammography screening rates. In summary, the present study is significant to public health and public health research because it:

- Examined the most recent, nationwide data available for mammography screening practices,
- Employed a novel statistical method for analysis, and
- Provided public health officials with an appropriate area of focus (urban or rural residence) for future breast cancer screening initiatives.

Delimitations

The guidelines given by the U.S. Preventive Services Task Force indicate that women aged 40 years or older should receive screening mammography at an interval of every one to two years. Therefore, women included in the present study were limited to those aged 40 years or older. While some organizations recommend that women who are considered "high-risk" for breast cancer receive a mammogram beginning at age 35, women who utilize mammography at an age younger than 40 years represent a distinct subset of individuals compared to the general population. Therefore, an age of 40 years and older was an eligibility criterion for this study.

The U.S. Census Bureau definitions for urbanicity provide distinctions of Metropolitan Statistical Areas, such that it was possible to characterize multiple degrees of urbanicity according to city centers and number of residents. However, creating a dichotomous urbanicity variable (urban/rural) accommodated the requirements for its use as the exposure of interest in propensity score matching.

Only women who provided complete information regarding age, race/ethnicity, income, education, martial status, and health insurance status were included in this study. Responses for all covariates of urbanicity were required to obtain a propensity score; therefore, excluding women without complete information accommodated the requirements of propensity score matching.

Limitations

The BRFSS is a nationally representative, cross-sectional telephone survey of the United States population. Thus, this study was limited by the cross-sectional design of the

BRFSS. While the present study allowed for measures of association to be calculated, causality between urbanicity and mammography utilization was not established.

This study was also limited by the sampling scheme. The BRFSS survey is a random-digit-dialed telephone survey of non-institutionalized individuals with landline telephone service in the United States (CDC, 2006a). Overall, 93.1% of households in the United States have landline telephone coverage; non-telephone coverage rates in 2006 varied from 3.1% to 12.1% (CDC, 2006b). Individuals who are from lower socioeconomic positions, live in rural areas, or have fewer opportunities to access medical care are less likely to have phone service, suggesting that these individuals are likely to be under-sampled (Frankel, Srinath, Hoaglin, Battaglia, Smith, Wright, & Khare, 2003). A non-coverage bias, especially considering that rural residence is an exposure of interest in this study, was probable. If women who did not have telephone service were more likely to live in a rural area and not utilize mammography screening, a negative bias toward the null may have occurred. This non-coverage bias may have led to an under-estimate of the effect of urbanicity on mammography utilization.

Wireless or cell phone usage is also an important consideration for random-digit-dialed landline telephone surveys like the BRFSS. Many cell phone users have chosen to live without a landline because of lifestyle preferences (Blumberg, Luke, & Cynamon, 2006), as opposed to socioeconomic forces, and thus may be underrepresented in telephone surveys. Blumberg, Luke, and Cynamon (2006) found that urban areas have higher proportions of adults who only use wireless service; in addition, compared to adults with landline service, adults who only used wireless service had lower proportions of having a usual place to go for medical care and higher proportions of uninsured. The magnitude of the noncoverage bias would be impacted by the differences between those with landline service and those without (Blumberg, Luke, & Cynamon, 2006). However, there is likely a smaller difference between individuals who have wireless only and those who have at least landline service, compared to the difference between individuals who have no telephone service and those who have at least landline service (Blumberg, Luke, & Cynamon, 2006). In this study, urban individuals who did not have landline telephone service may have been underrepresented in the study sample. If these individuals were also more likely to underutilize preventive services like mammography, then effect estimates were biased upward, away from the null value.

This study uses self-reports of mammography utilization. Several studies have indicated that when self-reports are compared to medical records, self-reports underestimate the time since the cancer screening test (Caplan, McQueen, Qualters, Leff, Garrett, & Calonge, 2003; Gordon, Hiatt, & Lampert, 1993) and thus may overestimate the rate of screening utilization in a population. In this study, it was likely that women underestimated the time since their last mammogram; however, it was unlikely that this report differed by urbanicity. The error may have resulted in nondifferential misclassification and thus biased the estimate downward, towards the null.

Women who did not have intact breasts and who were therefore unable to utilize mammography may have reported that they did not utilize mammography within the previous two years. While these women would not technically fall outside of screening guidelines, they would be counted as such in this study. However, it was not likely that this differed by urbanicity and therefore the error led to nondifferential misclassification with results likely biased towards the null.

Residual confounding is also possible. Covariates selected for the present study were obtained from the BRFSS survey and thus data were limited by the given BRFSS categorization. The BRFSS presents categorical, rather than continuous, income data. If income as a confounding variable or covariate was inappropriately categorized, residual confounding may have resulted. If the confounding variable was positively associated with both urbanicity and mammography utilization, then positive bias away from the null was likely to occur.

Designations of "urban" and "rural" populations are often problematic (Engelman, Hawley, Gazaway, Mosier, Ahluwalia, & Ellerbeck, 2002). In this study urbanicity was dichotomized to fulfill propensity score matching requirements, thus urbanicity was represented as either *urban* or *rural*. While the urbanicity variable was determined from a BRFSS variable calculated by the population size of the area, the present study conceptualized urbanicity as more than merely a population size. Instead, the study hypothesized that urbanicity represented geographic access to care. Important variations in geographic access to care may exist within urban or rural units (Stearns, Slifkin, and Edin, 2000). Individuals who resided on the outskirts of an urban area may have had poorer access to care than individuals who resided in a rural area located adjacent to an urban area. As such, the urban/rural dichotomization based on population may have actually contained an information bias with regard to the conceptualization of access to care. This information bias was nondifferential misclassification and likely biased the measure of association downward, towards the null.

There were several potential limitations for this investigation. Of particular importance were the self-reported nature of the mammography screening data and the nondifferential misclassification of urbanicity, both of which likely led to bias downward, towards the null. While the multitude of potential biases differently affected the measure of association, the net result of these biases was likely downward, towards the null, which suggested that the observed measure of association is an underestimate of the true measure of association.

Assumptions

In order to assume that the results of this study were valid, certain assumptions were made regarding self-reported data from the BRFSS survey.

- Participants responded honestly to questions regarding demographics and mammography utilization.
- Individuals who were contacted to participate in the BRFSS survey did not differ significantly from individuals who were not contacted to participate in the BRFSS survey.
- Individuals who chose to complete the survey did not differ significantly from individuals who chose not to complete the survey.
- All women who were aged 40 or older had intact breasts.
- Women did not change urbanicity residence since the time of their mammogram.

- Sociodemographic variables of race, income, education, and health insurance status were not results of urbanicity.
- Urban or rural area of residence had a fixed effect on individuals; that is, that area of residence had an equal effect on all individuals in the study.

Definitions of Terms

Terms were defined and collapsed for this investigation according to the 2006 BRFSS survey codebook (CDC, 2006a).

- Mammography – A mammogram is an x-ray of the breast.
- Gender – Participants were eligible for inclusion in this study if they were female. Gender was asked of all participants: "Indicate sex of respondent." Participants were included in this study if they indicated they were female.
- Age – Participants were eligible for inclusion in this study if they were aged 40 years or older. Age was asked of all participants: "What is your age?" Participants were included in this study if they indicated they were *Age 40-44* or older. Participants were excluded from this study if they indicated they were *Age 35-39* or younger, Don't Know/Not Sure, or Refused.
- Race – This variable was based on the CDC's calculated race term that combines all possible race and ethnicity responses. For the purposes of this study, race was categorized as *Non-Hispanic White, Non-Hispanic Black, Hispanic,* and *Other (including Asian, Native Hawaiian/Pacific Islander, American Indian/Alaskan Native,* or *Multiracial*). Participants were excluded from this study if race was indicated as Don't Know/Not Sure/Refused.

- Income – This variable measured of the total self-reported annual household income from all sources. Income was asked of all participants: "Is your annual household income from all sources…" Household income was categorized as *less than $10,000; $10,000 to $24,999; $25,000 to $49,999; $50,000 to $74,999;* and *$75,000 or greater*. Participants were excluded from this study if income was indicated as Don't Know/Not Sure/Refused.
- Education – This variable measured the highest level of education completed by the participant. Education was asked of all participants: "What was the highest grade or year of school you completed?" Education was categorized as *Less than high school, GED or High school graduate, Some college*, or *College graduate.* Participants were excluded from this study if education was indicated as Don't Know/Not Sure/Refused.
- Insurance Status – This variable measured the health insurance status of the participant. Insurance status was asked of all participants: "Do you have any kind of health care coverage, including health insurance, prepaid plans such as HMOs, or government plans such as Medicare?" Insurance status was categorized as either *Yes* or *No*, with participants excluded if they indicated insurance status was Don't Know/Not Sure/Refused.
- Marital Status – This variable measured an individual's self-reported marital status. Marital status was asked of all participants: "[Regarding marital status] are you…?" Marital status was categorized as *Currently Married* (including *Married* and *Separated*), *Never Married* (including *Never Married* and *Member of an*

Unmarried Couple), and *Previously Married* (including *Divorced* and *Widowed*). Participants were excluded if they indicated marital status was Don't Know/Not Sure/Refused.

- Urbanicity – This variable measured characteristics of the participants' areas of residence at the time of the survey. The BRFSS provided a Metropolitan Status Code for each participant. The U.S. Census Bureau (2007) defines a Metropolitan Statistical Area as an area with one or more counties of 50,000 or more individuals with a high degree of social and economic integration with an urban core. Urbanicity was dichotomized as *urban*, which included BRFSS data of *In the center city of an MSA*, *Outside the center city of an MSA but inside the county containing the center city*, *Inside a suburban county of the MSA*, and *In an MSA that has no center city.*; and *rural*, which included BRFSS data of *Not in an MSA.* Individuals residing outside of the United States were excluded from this study.
- Mammography utilization – This variable measured mammography screening utilization according to the U.S. Preventive Services Task Force guidelines. Two questions regarding mammography were asked of all participants: "A mammogram is an x-ray of each breast to look for breast cancer. Have you ever had a mammogram?" and "How long has it been since you had your last mammogram?" Participants who responded that they had ever had a mammogram and that their last mammogram was two years ago or less were considered as *utilizing mammography*. Participants who responded that they had never had a

mammogram or that their last mammogram was in excess of two years ago were considered as *not utilizing mammography*.

CHAPTER 2

LITERATURE REVIEW

Breast Cancer Incidence

Early detection and treatment are the best methods to prevent breast cancer mortality, and chances of survival are increased for breast cancer diagnosed at earlier rather than later stages (Roche, Skinner, & Weinstein, 2002). A review of cancer in rural versus urban populations indicated that urbanites have higher cancer incidence and mortality compared to their rural counterparts; however, results may be misleading if rural residents were undiagnosed with cancer due to unavailable screening facilities (Monroe, Ricketts, & Savitz, 1992). With regard to breast cancer, several investigations determined that the number of mammography facilities in a particular area was not associated with incidence of late stage breast cancer diagnosis (Roche, Skinner, & Weinstein, 2002; Marchick & Henson, 2005). Late stage breast cancer (Stage III or IV) can be diagnosed by self-examination or clinical breast exam without the use of mammography (Marchick & Henson, 2005). In the same study, however, incidence of ductal carcinoma in situ breast cancer, which can only be diagnosed with mammography, was significantly associated with the number of mammography facilities per 10,000 women in an area (Marchick & Henson, 2005). Likewise, higher percentages of in situ breast cancer were found in urban counties compared to nonurban counties in a California Cancer Registry study (Menck & Mills, 2001). Decreased access to mammography screening services increased the risk for late stage breast cancer diagnosis and resulted in poorer survival outcomes (Menck & Mills, 2001). It is generally accepted that increased ductal carcinoma in situ rates are the result of increased mammography use. Increased cancer incidence is of particular importance to urban and rural areas, as increased population density and thus increased number of mammography facilities will increase early diagnosis of breast cancer and improve future survival rates.

Socioeconomic Factors

Multiple research studies have evaluated if sociodemographic correlates such as race, education, and income impact mammography utilization. Studies have found that women with less than a high school education and incomes less than $20,000 are associated with lower breast cancer screening rates (Meissner, Breen, Taubman, Vernon, & Graubard, 2007; Barrett & Legg, 2005; Calle, Flanders, Thun, & Martin, 1993). White women are more likely to undergo screening compared to non-White women; however, some evidence indicates that this trend may be reversing as screening programs that target minority women have been successful (Jones, Caplan, & Davis, 2003). The urban-rural differences in cancer screening utilization may be more pronounced among minority women. One study reported that there is limited difference in urban-rural screening among Whites, yet there is a large difference between urban-rural screening among Blacks (Duelberg, 1992). The variables of race, education, and income are likely to impact screening utilization through pathways including socioeconomic position and financial health care access, such that women at lower levels of socioeconomic position have fewer financial resources to utilize in order to obtain a mammogram.

Urbanicity

Describing an individual's area of residence is a variable of interest in numerous cancer screening studies, perhaps because understanding the context in which one lives is important to an individual's health (Robert, 1998). As noted by Hall, Kaufman, and Ricketts (2006), there are many different ways to define *urban* and *rural* in epidemiological studies specific to the United States, and some measures are better suited than others for understanding geographic access to health care. While urbanization refers to a process over time, for example, an area becomes *more urban* over time, urbanicity describes characteristics of a particular area at a given point in time (Vlahov and Galea, 2002). It is the latter measure of *urban* and *rural* that may have the most impact on an individual's health (Vlahov and Galea, 2002). Urbanicity is valued as a proxy measure for a feature of a community that influences health (Hall, Kaufman, and Ricketts, 2006); in this investigation, urbanicity will serve as a proxy measure for access to health services. Rural areas are often conceptualized as the *counterfactual* urban experience in epidemiological investigations of urban-rural disparities (Hall, Kaufman, and Ricketts, 2006).

The urbanicity of a community may represent a contextual factor that is as or more important than the individual characteristics of the woman. The economic and social environment of the community can influence *predisposing characteristics* of the individual, such as age, race, and education, as well as *enabling characteristics*, such as health insurance, income, and employment (Litaker & Tomolo, 2007). Litaker and Tomolo (2007) maintain that

> [t]he economic and social attributes of a community, as well as the structure and function of its healthcare system, have the potential to influence the quality and availability of resources, including breast cancer screening. Previous work...suggests that communities with higher per capita income and employment rates, those providing greater support for safety net systems, and those with a greater supply of health facilities (e.g., community health centers) offer better overall access to medical services.[p.37]

There are multiple mechanisms that may explain how living in an urban or rural environment may affect health. The effect of urban stressors like air or water quality and noise levels on health may be buffered by the availability of health care and social service resources (Galea & Vlahov, 2005). On the other hand, while resources may be more available in urban areas compared to rural areas, disparities of socioeconomic position that are often found within cities suggest that all urban residents may not have the same access opportunities (Galea & Vlahov, 2005).

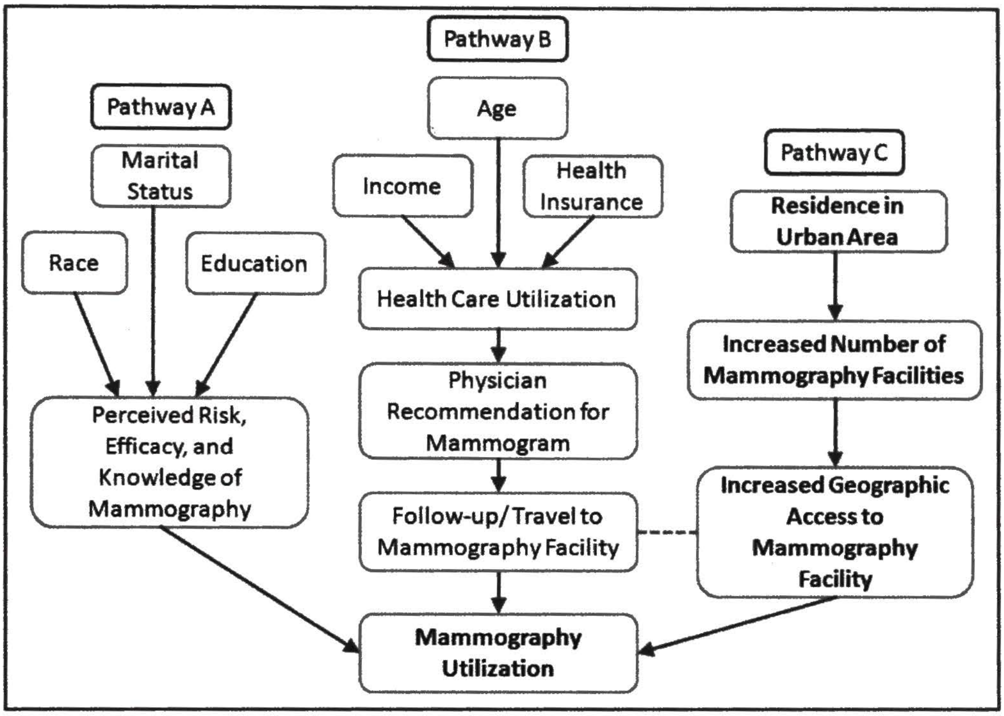
Characteristics of the built environments, such as roadway development and connectivity, may be of particular interest for an evaluation of health care utilization like mammography. Urban areas may have more public service clinics (Galea & Vlahov, 2005), like subsidized mammography screening facilities. Area of residence (urban vs. rural) therefore can represent a contextual factor like overall geographic access to care that impacts a women's ability to utilize mammography, independent of other sociodemographic influences. While it is well documented that a lower socioeconomic position can predict mammography underuse, one study found that women from disadvantaged areas who had no previous history of mammography screening were more likely to respond to a mammography screening invitation if they lived closer to the facility rather than further away from the facility (Hyndman, Holman, & Dawes, 2000). Likewise, among Kansas Medicare beneficiaries, the odds for utilizing mammography was lower for women who resided longer distances from the screening facility compared to women who lived closer even after adjusting for sociodemographic confounders (Engelman, Hawley, Gazaway, Mosier, Ahluwalia, & Ellerbeck, 2002). Carr et al. (1996) found similar results. After controlling for possible confounding variables, among both women who had ever received a mammogram and women who had received a mammogram in the previous year, women who were less than 15 miles from the screening facility were twice as likely to get a mammogram compared to women who lived greater than 15 miles from the facility (Carr et al., 1996). The findings suggest that perhaps distance to a mammography screening facility and issues like roadway development impact mammography utilization independently of socioeconomic position.

It is difficult to measure an individual's geographic access to mammography screening (Marchick & Henson, 2005). The number of mammography facilities is a given area is strongly correlated with population density, but is not found to be correlated with the geographic size of an area (Marchick & Henson, 2005). Therefore, women in sparsely populated areas may travel for long distances to obtain a mammogram. In a highly populated area, there are likely to be more mammography facilities and thus shorter travel distances to a facility. An increased number of mammography facilities in a given area indicate that women residing in populated areas have greater geographic access to screening, which is associated with higher rates of screening.

A mammography facility is a certified general radiology center (Susan G. Komen for the Cure, 2007). Permanent screening facilities can be located in a hospital, a physician's office, or a stand-alone center. There are also mobile mammography units, typically functioning as vans, which travel from site to site. Many of these mobile units serve to decrease geographic barriers to mammography utilization (Engelman, Hawley, Gazaway, Mosier, Ahluwalia, & Ellerbeck, 2002). Physician recommendation is an important predictor of mammography utilization (Meissner, Breen, Taubman, Vernon, & Graubard, 2007). After a physician recommends that a woman receive a mammogram, a woman must then follow-up with that recommendation. Because a physician's office may not house on-site mammography services, this means travelling to a permanent screening facility located outside of the physician's office or locating a mobile unit. Increased geographic access to a facility has been shown to increase the likelihood of screening. It is important to note that many mobile units accept self-referred women, which increases service to women who have not received a physician's referral (DeBruhl, Bassett, Jessop, & Mason, 1996). Therefore, a physician's referral is not always needed to utilize mammography. In a study of Medicare beneficiaries in Kansas, a primarily rural state, investigators found that 37% of counties contained only permanent mammography facilities, 23% of counties contained permanent and mobile mammography facilities, 29% of counties contained only mobile mammography facilities, and 11% of counties contained neither permanent nor mobile facilities (Engelman, Hawley, Gazaway, Mosier, Ahluwalia, & Ellerbeck, 2002).

A non-causal conceptual pathway is presented in *Figure 1*. Pathway A includes race, education, and martial status. These factors likely influence mammography utilization through the pathway of perceived risk, efficacy, and knowledge of mammography. In addition, and individual's race/ethnicity and marital status are important factors for cultural beliefs and social support, respectively, which may influence or modify an individual's willingness to obtain a mammogram. Pathway B includes income, age, and health insurance. These variables likely influence mammography utilization through the health care pathway, such that individuals who have more economic resources and have health insurance, either because of age criteria or because of income from employment, are more likely to visit a physician and thus receive a recommendation for a mammogram. While it is probable that all three pathways work together, Pathway C includes residence in an urban area and was the main pathway of interest for this investigation. Urban areas tend to have more mammography facilities compared to rural areas. This study hypothesized that an increased geographic access to health care services, like mammography facilities, had a positive effect on mammography utilization.

Figure 1. Conceptual Model for Mammography Utilization (Adapted from Carr et al., 1996).



Urbanicity and Preventive Service Utilization

Multiple studies have examined the association between urban/rural residents and preventive service utilization in the United States. Coughlin and Thompson (2004) hypothesized that the greater use of cancer screening by individuals in urban areas may be explained by the increased availability of health care and screening services. In contrast, individuals residing in a rural area may have limited access to health care services and practitioners. Because physician recommendation is an important predictor of screening utilization (Carr et al., 1996; Meissner, Breen, Taubman, Vernon, & Graubard, 2007), the decreased availability of practitioners and clinics in rural areas may impact screening rates. Compared to suburban and urban areas, residents of rural areas were more likely to be White, married, less educated, report fair or poor general health, have lower household income, be without health insurance, and live in a health professional shortage area.

Likewise, Zhang, Tao, and Irwin (2000) reported that residents of rural areas were more likely to be older, White, have lower education and household income, and be without health insurance compared to urban residents. However, the effect of urbanization on cancer screening was not statistically significant after adjusting for those socioeconomic characteristics. The results suggested that differences between urban and rural women with regard to mammography screening may be attributed to differences in education, income, and health insurance. The authors also noted that financial barriers may be an important consideration regarding access to health care for rural women. In some areas, rural residents have higher copayments and insurance premiums despite less coverage compared to urbanites.

Stearns, Slifkin, and Edin (2000) noted that if access problems faced by rural residents can be accounted for by socioeconomic characteristics like income, the access problems are likely to be shared by the poor in the inner city. The authors noted that among Medicare beneficiaries, utilization of physician services is lower among rural compared to urban residents. Individuals who resided in rural areas were more likely to travel at least an hour to their usual source of health care; however, rural women were also likely to report high levels of satisfaction with their care. Longer travel time to a health care facility may not be a problem in and of itself; longer travel time for residents in rural areas is only a problem if screening is compromised as a result.

Measurement of Urbanicity

Meissner, Breen, Taubman, Vernon, and Graubard (2007) investigated correlates of mammography utilization with a focus on women who reported not receiving a recent mammogram. Urbanicity was operationalized as *urban* or *rural*. The U.S. Census Bureau definition of the Metropolitan Statistical Area (MSA) was used to measure urbanicity, with *urban* including one or more counties with a population of at least 50,000 individuals. All other areas were classified as Non-MSA. The National Health Interview Survey data utilized by the investigators provided this distinction for each survey respondent. The study found that a larger proportion of women who reported no mammogram within the past two years lived in a Non-MSA area, compared to women who reported having a mammogram within the past two years (27.3% vs. 21.7%). Results persisted even after adjustment for sociodemographic covariates.

In a California study, Menck and Mills (2001) examined the influence of urbanicity, age, ethnicity, and income on the early diagnosis of breast cancer. Investigators purported that early diagnosis was a useful surrogate for mammography activity. All counties in California were included in the study. The classification of urbanicity was conceptualized as either *urban* or *non-urban* units, relative to the surrounding counties. Non-urban areas were measured based on large geographic area (on average 31,593 square miles) in combination with low population density (on average 14 inhabitants per square mile), while urban areas were measured as small geographic areas (on average 5,327 square miles) with high population density (on average 174 inhabitants per square mile). Investigators reported lower median family income in non-urban areas. Women who lived in urban counties, which were characterized by higher median household incomes, were diagnosed with breast cancer at an earlier stage.

Similarly, a Wisconsin study by Andersen, Remington, Trentham-Dietz, and Robert (2004) of trends in early detection of breast cancer found that areas characterized by lower levels of income, education, and urbanicity continue to be underserved and could benefit from screening programs. The study used early detection of breast cancer as documented by the state's cancer registry as a proxy for the unknown mammography screening rates. Urbanicity was measured as *urban*, which was characterized a zip code containing only urban census tracts; *rural*, which was characterized as containing only rural census tracts; and *mixed*, which was characterized as containing both urban and rural census tracts. The study found that between 1980 and 1998, breast cancer was diagnosed approximately one-third less frequently in rural areas compared to urban areas. The investigators concluded that screening rates varied by community socioeconomic status and urbanicity. It was hypothesized that the observed disparities in mammography utilization by urbanicity may be explained by limited access to healthcare practitioners and preventive services.

A study by the American Cancer Society evaluated demographic predictors of mammography screening among U.S. women (Calle, Flanders, Thun, & Martin, 1993). Urbanicity was classified as MSA, central city, and non-MSA. Presumably, investigators utilized the U.S. Census Bureau definition of MSA; however, operational measures of urbanicity were not reported. Overall, the study found that low income, Hispanic ethnicity, low education, and residence in a rural area were predictors of mammography underuse.

Available research supports the hypothesis that women living in rural areas are more likely to report mammography underuse because of geographic barriers in health care access. Independent of an individual's socioeconomic position and the impact that it has on mammography use, issues regarding geographic access to care like distance to a screening facility and roadway connectedness may also impact a woman's screening practices.

CHAPTER 3

METHODOLOGY

Study Overview

For this study, data were imported from the 2006 Behavioral Risk Factor Surveillance System (BRFSS), which is a publicly available data set from the CDC (CDC, 2006a; CDC, 2006b). The BRFSS is a cross-sectional, state-based telephone survey that was initiated by the CDC in 1984. The scope of the survey currently includes all fifty states, the District of Columbia, and Puerto Rico. The survey is developed jointly by the CDC's Behavioral Surveillance Branch (BSB) and the states. BSB provides guidelines and telephone numbers and the states conduct the survey. States submit data to the BSB, who then weights the data according to annual state-specific population estimates.

Sampling Procedures

Telephone numbers are garnered for eligible households (CDC, 2006b). An eligible household is considered a housing unit with a separate entrance for each group of occupants. The group of occupants must eat separately from other individuals on the property. The household must be the occupants' primary or secondary place of residence. Non-eligible households include vacation homes that are occupied for less than thirty days per year, group homes, or institutions.

Household members who are eligible to complete the survey include related or unrelated adults aged 18 or older who consider the household their home. The computer-assisted telephone interviewing (CATI) randomly selects survey respondents from eligible household members. A complete interview from each respondent must include data for age, race, and sex at a minimum.

The BRFSS uses a probability sample such that all households with telephones have a known, non-zero probability of inclusion in the sample. Disproportionate stratified random sampling was introduced in 2003 as the BRFSS sampling strategy. This method draws telephone numbers from two different strata which are based on the density of known household telephone numbers. The sampling achieves a statistically representative sample of households in the United States. Samples were obtained during every month and a total sample of 355,710 individuals was obtained for the 2006 BRFSS survey.

Data Collection Procedures

Data for the 2006 BRFSS survey were collected at the individual level. Demographic variables, including gender, race, age, income, education, and insurance status, as well as mammography utilization variables, are based on the respondents' self-reports. Information regarding urbanicity is provided by the CDC survey team because it corresponds to the residential telephone number of the respondent.

Instrumentation

The BRFSS questionnaire is comprised of three parts: the core component, which is given by all states; the optional modules, which are given by some states; and the state-added questions, which are developed by the individual states. Questions regarding women's health and specifically mammography utilization are included in the fixed core component section. From the BRFSS survey, a mammography utilization instrument was created to measure whether or not the participant subscribes to the current breast cancer screening guidelines. Responses were combined from the following questions: "A mammogram is an x-ray of each breast to look for breast cancer. Have you ever had a mammogram?" and "How long has it been since you had your last mammogram?" Results were dichotomized such that individuals who indicated that they had ever had a mammogram and that their last mammogram was two years ago or less will be considered as *utilizing mammography*, while participants who responded that they had never had a mammogram or that their last mammogram was in excess of two years ago were considered as *not utilizing mammography.* A detailed list of all survey items that were used from the 2006 BRFSS survey is provided in Appendix A.

Data Analysis Procedures

Propensity score matching has increased in use for observational study designs in social epidemiology. Oakes and Johnson (2006) maintain that in the absence of the true counterfactual design, propensity score matching methods allow the investigator the opportunity to more closely mimic a randomized experimental design and increase the transparency of the causal contrast inference. The use of the matched propensity score allows the investigator to evaluate the exchangeability between the exposed and unexposed groups, which allows for better comparisons and inference. Oakes and Johnson (2006) state that while propensity score matching may allow for better control of confounding, it does not account for unobserved or hidden bias. First, the investigator determines the exposure and the outcome of interest. Covariates, or potential confounding variables, are selected. These covariates are used in a logistic regression model to determine the probability (propensity score) that they predict the exposure of interest, which is the dependent variable in the logistic regression model. Once propensity scores are obtained, individuals are matched based on their propensity scores within a specific caliper and the effect of the exposure on the outcome of interest can be obtained.

In brief, propensity score methods have been used previously to examine the effect of a contextual variable, like neighborhood environment, on a particular outcome while controlling for individual socioeconomic indicators (Oakes & Johnson, 2006; Diez Roux, Borrell, Haan, Jackson, & Schultz, 2004). These effects were obtained without the use of multilevel models. For example, in a study of neighborhood environments and mortality in an elderly cohort, the propensity score methods reduced the number of participants who were included in the analysis but allowed for a balance of relevant individual characteristics across the lowest and highest (score) neighborhood groups (Diez Roux, Borrell, Haan, Jackson, & Schultz, 2004).

The present investigation was a cross-sectional study that utilized propensity score matching as a data analysis technique. Data management and analysis were conducted using the SAS 9.1 program. Variables of interest were imported from the 2006 BRFSS survey (see Appendix A) and participants who were eligible for this study were identified. Individuals were eligible if they were female, aged 40 years or older, and provided a response to all other questions of interest. Males, females younger than age 40, and individuals with missing data were excluded from this study.

The first level of analysis involved descriptive statistics, including frequencies and percents. Data were evaluated for missing responses and strata with small cell sizes were collapsed. A chi-square analysis was performed to determine the frequency and statistical association of age, race, education, income, marital status, health insurance status, and mammography utilization by urbanicity. P-values were reported to indicate statistically significant differences between individual characteristics and mammography utilization by urbanicity. In addition, separate logistic regression analyses were performed to determine the effect of each factor on mammography utilization. Odds ratios and 95% confidence intervals were reported to indicate the effect of each factor on mammography utilization.

Next, the probability that a woman belonged to either the urban or the rural group based on pre-selected covariates was obtained using logistic regression. A model was created to determine the probability of urbanicity, such that

$$\gamma = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6$$

where γ = urbanicity (*urban* or *rural*)

β_0 = *intercept*

β_1 = age

β_2 = race/ethnicity

β_3 = education

β_4 = income

β_5 = marital status

β_6 = health insurance status.

The regression model provided individual probabilities, or propensity scores, for each woman in the study sample. Collinearity was not a concern in this regression model because it estimated a probability, not an effect [J.M. Oakes (personal communication, February 26, 2008)]. Therefore, seemingly related variables like education and income were included together. These scores were output from the SAS program and saved as a new variable in the data set. Observations were matched by urbanicity such that after matching individuals who resided in urban areas had the same probability of age, race/ethnicity, income, education, health insurance status, and marital status as individuals who resided in rural areas. Using the Greedy matching technique, observations were ordered and sequentially matched to the nearest propensity score, such that the "best" match was made first and then the "next best" match was made until no more matches were possible (Parsons, 2008). After matching, distributions of the individual variables by urbanicity were evaluated using chi-square analysis to check for exchangeability between urban and rural groups. Finally, logistic regression was performed on the propensity score matched data set to examine the effect of urbanicity on mammography utilization after accounting for individual characteristics using propensity score matching. Odds ratios, 95% confidence intervals, and p-values were reported to examine the effect of urbanicity on mammography utilization.

CHAPTER 4

RESULTS

The purpose of this investigation was to (1) characterize a nationwide sample of women aged 40 years and older by individual sociodemographic variables of race, education, income, marital status, health insurance status, and urbanicity; (2) determine the proportion of these women aged 40 years or older who utilized mammography within the previous two years; and (3) explore the relationship between urbanicity and mammography utilization after accounting for covariates of race, education, income, marital status, and health insurance status among a nationwide sample of women aged 40 years or older.

The 2006 BRFSS contained 355,710 nationally representative participants. Of these, 165,311 participants were females aged 40 years and older and were contained in the study sample of this investigation. This study sample consisted of 41,270 women between the ages of 40 and 49 years (25.0%), 64,127 women between the ages of 50 and 64 years (38.8%), and 59,914 women aged 65 years or older (36.2%). The highest proportion of these women were non-Hispanic White (81.2%), had a household income of between $25,000 and $49,999 (29.1%), were high school graduates or GED recipients (32.6%), were currently married (50.4%), had some form of health insurance (91.3%), lived in an urban area (66.6%), and had received a mammogram within the previous two years (74.0%). Detailed results are presented in *Table 2*.

Table 2. Characteristics of women aged 40 years or older in the United States, BRFSS, 2006.

	n	%
Total	165,311	100.0
Age		
40-49 years	41,270	25.0
50-64 years	64,127	38.8
65 years or older	59,914	36.2
Race/Ethnicity		
White, non-Hispanic	133,044	81.2
Black, non-Hispanic	13,318	8.1
Hispanic	9,775	6.0
Other	7,622	4.7
Missing	1,552	
Income		
Less than $10,000	10,097	7.4
$10,000 to $24,999	37,842	27.7
$25,000 to $49,999	39,820	29.1
$50,000 to $74,999	20,757	15.2
$75,000 or more	28,349	20.7
Missing	28,446	
Education		
Less than high school degree	18,681	11.3
GED or high school graduate	53,671	32.6
Some college	44,583	27.0
College graduate	47,946	29.1
Missing	430	
Marital Status		
Currently married	82,971	50.4
Previously married	68,507	41.6
Never married	13,195	8.0
Missing	638	
Health Insurance		
Yes	150,641	91.3
No	14,385	8.7
Missing	285	
Urbanicity		
Urban	107,646	66.6
Rural	54,088	33.4
Missing	3,577	
Mammography within past 2 yrs		
Yes	122,252	74.0
No	43,059	26.0

Individual characteristics of women aged 40 years and older who were included in the study sample of this investigation differed by urbanicity. There were significant differences in the distributions and proportions of age, race/ethnicity, income, education, marital status, health insurance status, and mammography utilization between urban and rural areas ($p<0.0001$). Detailed results are provided in *Table 3*.

Table 3. Characteristics of women aged 40 years or older in the United States by urbanicity, BRFSS study sample, 2006.*

	Urbanicity		
	Urban	**Rural**	
Characteristic	n (%)	n (%)	*p-value*[+]
Age			
40-49 years	28,046 (26.1)	12,254 (22.7)	
50-64 years	41,503 (38.6)	21,169 (39.1)	<0.0001
65 years or older	38,097 (35.4)	20,665 (38.2)	
Race/Ethnicity			
White, non-Hispanic	86,706 (81.3)	46,089 (86.0)	
Black, non-Hispanic	9,936 (9.3)	2,626 (4.9)	<0.0001
Hispanic	5,433 (5.1)	1,907 (3.6)	
Other	4,565 (4.3)	2,961 (5.5)	
Income			
Less than $10,000	5,490 (6.2)	3,720 (8.3)	
$10,000 to $24,999	22,523 (25.3)	14,324 (32.0)	
$25,000 to $49,999	25,139 (28.2)	13,988 (31.3)	<0.0001
$50,000 to $74,999	14,128 (15.9)	6,365 (14.2)	
$75,000 or more	21,822 (24.5)	6,318 (14.1)	
Education			
Less than high school degree	10,478 (9.8)	7,107 (13.2)	
GED or high school graduate	33,200 (30.9)	19,526 (36.2)	<0.0001
Some college	29,205 (27.2)	14,788 (27.4)	
College graduate	34,464 (32.1)	12,549 (23.3)	
Marital Status			
Currently married	52,696 (49.2)	28,745 (53.3)	
Previously married	44,813 (41.8)	22,097 (41.0)	<0.0001
Never married	9,650 (9.0)	3,100 (5.7)	
Health Insurance			
Yes	99,061 (92.2)	48,373 (89.6)	<0.0001
No	8,412 (7.8)	5,608 (10.4)	
Mammography within past 2 yrs			
Yes	81,287 (75.5)	38,450 (71.1)	<0.0001
No	26,359 (24.5)	15,638 (28.9)	

*Urbanicity is defined as residence in an urban or rural area.
+ P-value corresponds to x^2 test across strata of urbanicity.

There were multiple factors that influenced mammography utilization among the study sample. Compared to women aged 40 to 49 years, women aged 50 to 64 years and women aged 65 years or older were 70% and 43% more likely to utilize mammography, respectively. Compared to non-Hispanic White women, non-Hispanic Black women were 2% more likely to utilize mammography, but results were not significant at the $\alpha = 0.05$ level [Odds ratio (OR)=1.02, 95% Confidence Interval (CI): 0.98, 1.06]. Hispanic women were 16% less likely to utilize mammography compared to non-Hispanic White women. A dose-response pattern was observed in the income variable, such that individuals who reported increasingly higher household incomes were also more likely to utilize mammography compared to individuals who reported a household income of less than $10,000. A similar pattern emerged for highest level of education, where individuals who reported increasingly higher levels of education were also increasingly more likely to utilize mammography compared to individuals who reported less than a high school education. Previously married or never married women were 30% and 38% less likely to utilize mammography, respectively, compared to individuals who were currently married. Women without any form of health insurance were 72% less likely to utilize mammography compared to women with some form of health insurance. Finally, women who resided in an urban area of greater than 50,000 residents were 25% more likely to utilize mammography compared to women who resided in a rural area of less than 50,000 residents. Detailed results are presented in *Table 4*.

Table 4. Logistic regression models for having a mammogram in the last 2 years among women aged 40 years or older in the United States, BRFSS study sample, 2006.

Characteristic	OR+ (95% CI)	*p-value**
Age		<0.0001
40-49 years	1.00	
50-64 years	1.70 (1.65, 1.74)	
65 years or older	1.43 (1.39, 1.47)	
Race/Ethnicity		<0.0001
White, non-Hispanic	1.00	
Black, non-Hispanic	1.02 (0.98, 1.06)	
Hispanic	0.84 (0.80, 0.87)	
Other	0.77 (0.73, 0.81)	
Income		<0.0001
Less than $10,000	1.00	
$10,000 to $24,999	1.22 (1.16, 1.27)	
$25,000 to $49,999	1.81 (1.73, 1.90)	
$50,000 to $74,999	2.30 (2.18, 2.43)	
$75,000 or more	2.71 (2.58, 2.85)	
Education		<0.0001
Less than high school degree	1.00	
GED or high school graduate	1.50 (1.45, 1.55)	
Some college	1.69 (1.63, 1.76)	
College graduate	2.24 (2.16, 2.33)	
Marital Status		<0.0001
Currently married	1.00	
Previously married	0.70 (0.68, 0.72)	
Never married	0.62 (0.60, 0.64)	
Health Insurance		<0.0001
Yes	1.00	
No	0.28 (0.27, 0.29)	
Urbanicity		<0.0001
Rural	1.00	
Urban	1.25 (1.23, 1.28)	

+ Unadjusted Odds Ratios, corresponding with 95% Confidence Intervals
* P-value is for trend within each factor

After propensity score matching, the study sample was reduced to 88,586 women aged 40 years and older. Women in this sample were matched based on the probability that they lived in an urban or rural area, such that 50% of the participants resided in urban areas and 50 % of the participants resided in rural areas. The propensity score matched sample consisted of 22,209 women between the ages of 40 and 49 years (25.1%), 36,202 women between the ages of 50 and 64 years (40.9%), and 30,175 women aged 65 years or older (34.1%). The highest proportion of these women were non-Hispanic White (85.7%), had a household income of between $10,000 and $24,999 (31.8%), were high school graduates or GED recipients (35.4%), were currently married (53.2%), had some form of health insurance (89.2%), and had received a mammogram within the previous two years (72.7%). Detailed results are presented in *Table 5*.

Table 5. Characteristics of women aged 40 years or older in the United States, BRFSS propensity score matched sample, 2006.

	n	%
Total	88,586	100.0
Age		
40-49 years	22,209	25.1
50-64 years	36,202	40.9
65 years or older	30,175	34.1
Race		
White, non-Hispanic	75,879	85.7
Black, non-Hispanic	4,543	5.1
Hispanic	3,262	3.7
Other	4,902	5.5
Income		
Less than $10,000	7,399	8.4
$10,000 to $24,999	28,197	31.8
$25,000 to $49,999	27,761	31.3
$50,000 to $74,999	12,661	14.3
$75,000 or more	12,568	14.2
Education		
Less than high school degree	10,504	11.9
GED or high school graduate	31,333	35.4
Some college	24,936	28.2
College graduate	21,813	24.6
Marital Status		
Currently married	47,148	53.2
Previously married	35,984	40.6
Never married	5,454	6.2
Health Insurance		
Yes	79,058	89.2
No	9,528	10.8
Urbanicity		
Urban	44,293	50.0
Rural	44,293	50.0
Mammography within past 2 yrs		
Yes	64,400	72.7
No	24,186	27.3

The differences in individual characteristics of women aged 40 years and older by urbanicity were reduced after propensity score matching. There were no statistically significant differences at the $\alpha = 0.05$ in the distributions and proportions of age (p=0.4103), income (p=0.6752), education (p=0.9559), marital status (p=0.7447), and health insurance status (p=0.3510). Significant differences in race/ethnicity across urbanicity did remain after propensity score matching (p=0.0018). Differences in mammography utilization, the outcome of interest in this investigation, remained significant between urban and rural areas (p<0.0001). Detailed results are provided in *Table 6.*

Table 6. Characteristics of women aged 40 years or older in the United States by urbanicity, BRFSS propensity score matched sample, 2006.*

	Urbanicity		
	Urban	**Rural**	
Characteristic	n (%)	n (%)	*p-value*+
Age			
40-49 years	11,181 (25.2)	11,028 (24.9)	
50-64 years	18,020 (40.7)	18,182 (41.0)	0.4103
65 years or older	15,092 (34.1)	15,083 (34.1)	
Race/Ethnicity			
White, non-Hispanic	37,881 (85.5)	37,998 (85.8)	
Black, non-Hispanic	2,351 (5.3)	2,192 (4.9)	0.0018
Hispanic	1,689 (3.8)	1,573 (3.6)	
Other	2,372 (5.4)	2,530 (5.7)	
Income			
Less than $10,000	3,760 (8.5)	3,639 (8.2)	
$10,000 to $24,999	14,050 (31.7)	14,147 (31.9)	
$25,000 to $49,999	13,873 (31.3)	13,888 (31.4)	0.6752
$50,000 to $74,999	6,325 (14.3)	6,336 (14.3)	
$75,000 or more	6,285 (14.2)	6,283 (14.2)	
Education			
Less than high school degree	5,279 (11.9)	5,225 (11.8)	
GED or high school graduate	15,652 (35.3)	15,681 (35.4)	0.9559
Some college	12,458 (28.1)	12,478 (28.2)	
College graduate	10,904 (24.6)	10,909 (24.6)	
Marital Status			
Currently married	23,517 (53.1)	23,631 (53.4)	
Previously married	18,041 (40.7)	17,943 (40.5)	0.7447
Never married	2,735 (6.2)	2,719 (6.1)	
Health Insurance			
Yes	39,572 (89.3)	39,486 (89.1)	0.3510
No	4,721 (10.7)	4,807 (10.9)	
Mammography within past 2 yrs			
Yes	32,721 (73.9)	31,679 (71.5)	<0.0001
No	11,572 (26.1)	12,614 (28.5)	

* Urbanicity is defined as residence in an urban or rural area.
+ P-value corresponds to x^2 test across strata of urbanicity.

Results persisted for the effect of urbanicity on mammography utilization even after adjustment for individual variables in logistic regression and using propensity score matching methods. Individuals who lived in urban areas were 25% more likely to report having a mammogram in the last two years compared to women who lived in rural areas, without adjustment for covariates (OR=1.25, 95% CI:1.23,1.28). After adjustment for age, race/ethnicity, education, marital status, and health insurance status, women who resided in urban areas were 21% more likely to utilize mammography compared to women who resided in rural areas (OR=1.21, 95% CI:1.18,1.24). Results were also obtained using propensity score matching, which accounted for individual variables by matching on a probability value. Using propensity score matching methods, women who resided in urban areas were still 13% more likely to utilize mammography compared to women who resided in rural areas (OR=1.13, 95% CI:1.09,1.16). Detailed results are provided in *Table 7*.

Table 7. The effect of urbanicity on having a mammogram in the last 2 years among women aged 40 years or older in the United States, BRFSS, 2006.

Statistical Analysis Method	Urbanicity	OR+ (95% CI)	p-value
Logistic Regression	Urban*	1.25 (1.23, 1.28)	<0.0001
	Urban ◊	1.21 (1.18, 1.24)	<0.0001
Propensity Score Matching	Urban	1.13 (1.09, 1.16)	<0.0001

+ Odds Ratios, corresponding with 95% Confidence Intervals

* Unadjusted

◊ Adjusted for Age, Race/Ethnicity, Education, Marital Status, Health Insurance Status

CHAPTER 5

DISCUSSION AND CONCLUSIONS

The purpose of this investigation was to characterize a nationwide sample of women aged 40 years and older by individual characteristics and then determine the effect of individual area of residence (urban vs. rural) on screening mammography utilization practices. Effect estimates were obtained using both logistic regression and propensity score matching methods.

Women between the ages of 50 and 64 years, women at the highest levels of education and household income, women who were currently married, and women with health insurance were all more likely to utilize mammography. Women who resided in urban areas were more likely to report receiving a mammogram in the previous two years compared with women who resided in rural areas. This association was persistent even after adjustment for individual variables of age, race/ethnicity, education, marital status, and health insurance status in logistic regression analysis and analysis using propensity score matching methods. At the time of this investigation, this is the first study to use propensity score matching to examine the effect of urbanicity on mammography utilization in the United States.

Propensity score matching allowed this study to more closely mimic a randomized experimental design and increase the transparency of the causal contrast inference (Oakes & Johnson, 2006). Differences in individual baseline characteristics were minimized using this approach. In this investigation, the counterfactual experience of *urban* residence was *rural* residence. After propensity score matching methods were employed, urban residents and rural residents had similar baseline characteristics. The only baseline characteristic that remained statistically significant between the urban and rural groups after propensity score matching was race/ethnicity. It is likely that differences among individuals who were non-Hispanic Black between urban and rural groups were too pronounced to mitigate with matching methods. The remaining differences in race/ethnicity between urban and rural groups may have led to residual confounding in the association between urbanicity and mammography utilization if non-Hispanic Blacks were more likely to utilize mammography screening compared to other race/ethnicity groups. However, the percentage of non-Hispanic Blacks in our study population was small and there was no clear evidence suggesting increased utilization of mammography screening in non-Hispanic Blacks compared to other groups. Residual confounding was unlikely to have a substantial impact on our study results.

There are several disadvantages to using propensity score matching methods. First of all, complete data were necessary for every participant. Observations with missing values were excluded because all values were needed to compute the probability or propensity score. Therefore, the study sample was reduced because of this requirement. It is important to note that propensity score matching did account for observed or hidden bias. Therefore, while the method mimicked a randomized design, it was possible that unknown baseline differences between urbanicity groups confounded the results.
Also, because a 1:1 matching scheme was employed, unequal frequencies between exposure groups means that unmatched observations were excluded. In this study the sample size was reduced from n=165,311 to n=88,586 by n=76,725 for propensity score

matching. The reduction in sample size was mostly due to a relatively small proportion of participants living in rural areas of the United States. Approximately two-thirds of the initial study sample was considered urban and one-third was considered rural, which meant that a large proportion of urban residents were excluded from the propensity score matched analysis in order to match urbanites to the smaller rural resident sample. Including different proportions of urban and rural residents may have introduced selection bias in the propensity score matched sample if participant inclusion differed by mammography utilization status. However, the benefits of propensity score matching may outweigh the potential risk for selection bias. Matching was not performed for statistical efficiency but rather to increase the transparency of the causal contrast and evaluate the rural counterfactual to the urban experience.

Although estimates of the effect of urbanicity on mammography utilization were presented together in *Table 7*, estimates obtained using propensity score matching should not be directly compared to estimates obtained using logistic regression (Oakes & Johnson, 2006). Propensity score matching excluded individuals with any missing variable values or unmatched probabilities of exposure, thus creating a different sample from the logistic regression methods.

Results from this study were consistent with other reports (Engelman, Hawley, Gazaway, Mosier, Ahluwalia, & Ellerbeck, 2002; Meissner, Breen, Taubman, Vernon, & Graubard, 2007), which suggests mammography use is higher in urban geographic areas. However, previous investigations have purported that the difference between urban and rural areas can be attributed to differences in socioeconomic disadvantage between the two areas (Smith, Humphreys, & Wilson, 2008). By using propensity score matching methods, this investigation was able to mimic randomization and attenuate the influence of socioeconomic differences between urban and rural areas.

This investigation adds to the growing body of literature that focuses on cancer screening practices and urbanicity as a proxy for geographic access to health care. However, although urbanicity was conceptualized as a proxy variable in this study, it did not elucidate the causal mechanism by which urbanicity impacts screening practices. Driving time, distance, roadway connectivity, and transportation are all important issues when evaluating geographic access to health care. Additional mixed method research approaches that include both qualitative and quantitative methods are needed to isolate the underlying reason for why residence in an urban area is positively associated with mammography utilization and why some women receive a mammogram and others do not. Related topics of culturally acceptable screening methods and promotion as well as affordability of services may also need investigation along with proximity of services (Coughlin, Leadbetter, Richards, & Sabatino, 2008).

Additional limitations included the cross-sectional nature of the study design which precluded causal inferences and the self-reported nature of the data. Individuals typically underestimate the time since their last screening (Caplan, McQueen, Qualters, Leff, Garrett, & Calonge, 2003; Gordon, Hiatt, & Lampert, 1993), which overestimates the rate of screening in a population. The designation of *urban* and *rural* populations is often problematic as well. This study was limited by the dichotomization of urbanicity, which prevented assessment based on degree of *urban* and degree of *rural*. The use of more urbanicity levels may have allowed for a within-strata approach that further elucidated individual barriers to mammography screening within a particular level of geographic access to care. An important strength of this study included the large sample size and thus power afforded by the BRFSS sample size. Weighting was not applied to the data, however, due to the specific subset of participants (females, aged 40 years and older) included in the study sample. Therefore, although this investigation used a nationwide sample of women, results cannot be generalized to the general population of the United States and thus external validity is limited.

Resources for public health initiatives are becoming increasingly scarce. Public health officials may wish to focus limited resources in rural areas to increase mammography screening utilization and early detection of breast cancer among women in the United States. However, with lower population density in rural areas compared to urban areas, officials must reconcile the investment of time and effort to reach dispersed populations with increased mammography rate returns. Mobile mammography vans have been shown to work effectively to bring screening services to many under-served areas of the United States (Brown & Fintor, 1995), including both rural and inner-city populations. Mobile units are more likely to have lower costs and more convenient operating hours (Brown & Fintor, 1995), which may also improve general access to healthcare instead of only geographic access to healthcare. Although more of a stop-gap measure rather than a long term solution, mammography vans could increase screening utilization among certain populations if van service can increase in overall frequency and regularity of site visits.

Long term solutions to mammography underuse may include more distal approaches to health equity, such as increased health literacy and decreased income and educational status disparities. Health literacy of the general population and communication between medical personnel and patients in clinical settings has much room for improvement. Physician recommendation is consistently reported as a strong predictor of mammography use (Meissner, Breen, Taubman, Vernon, & Graubard, 2007), which suggests that increased involvement in patient screening behaviors and follow-up by the clinician may improve mammography use. Decreasing financial barriers to mammography utilization may include the implementation of free or subsidized mammograms (Meissner, Breen, Taubman, Vernon, & Graubard, 2007). In addition, wider health insurance coverage may also ameliorate disparities of mammography utilization barriers.

In conclusion, mammography utilization is not consistent across all areas of the United States. In addition to other populations facing barriers to cancer screening, individuals residing in rural areas present lower rates of mammography use compared to individuals residing in urban areas. Future research and public health initiatives should focus on the reason for which urbanicity effects mammography use and how screening rates can be improved across the United States.

REFERENCES

Andersen, L.D., Remington, P.L., Trentham-Dietz, A., & Robert, S. (2004). Community Trends in the Early Detection of Breast Cancer in Wisconsin, 1980-1998. *American Journal of Preventive Medicine, 26*(1), 51-55.

Barrett, K. & Legg, J. (2005). Demographic and health factors associated with mammography utilization. *American Journal of Health Promotion, 19*(6), 401-5.

Blumberg, S.J., Luke, J.V., & Cynamon, M.L. (2006). Telephone Coverage and Health Survey Estimates: Evaluating the Need for Concern about Wireless Substitution. *American Journal of Public Health, 96*(5), 926-31.

Brown, M.L., & Fintor, L. (1995). U.S. screening mammography services with mobile units: Results from the National Survey of Mammography Facilities. *Radiology, 195*(2), 529-32.

Caliendo, M., & Kopeinig, S. (2005). Some practical guidance for the implementation of propensity score matching. IZA Discussion Paper No. 1588. Available at SSRN: http://ssrn.com/abstract=721907.

Calle, E.E., Flanders, W.D., Thun, M.J., & Martin, L.M. (1993). Demographic predictors of mammography and Pap smear screening in US women. *American Journal of Public Health, 83*(1), 53-60.

Caplan, L.S., McQueen, D.V., Qualters, J.R., Leff, M., Garrett, C., & Calonge, N. (2003). Validity of women's self-reports of cancer screening test utilization in a managed care population. *Cancer Epidemiology Biomarkers and Prevention, 12*(11), 1182-7.

Carr, W.P., Maldonado, G., Leonard, P.R., Halberg, J.U., Church, T.R., Mandel, J.H., Dowd, B., & Mandel, J.S. (1996). Mammogram Utilization Among Farm Women. *The Journal of Rural Health, 12*(4), 278-290.

Centers for Disease Control and Prevention. (2006a). BRFSS Annual Survey Data: Survey Data and Documentation. Retrieved February 13, 2008, from http://www.cdc.gov/brfss/technical_infodata/surveydata/2006.htm.

Center for Disease Control and Prevention. (2006b). Comparability of Data: BRFSS 2006. Retrieved January 18, 2007, from http://www.cdc.gov/brfss/technical_infodata/surveydata/2006/compare_06.rtf.

Center for Disease Control and Prevention. (2007). Breast Cancer: Statistics. Retrieved January 11, 2008, from http://www.cdc.gov/cancer/breast/statistics/.

Coughlin, S.S., Leadbetter, S., Richards, T., & Sabatino, S.A. (2008). Contextual analysis of breast and cervical cancer screening and factors associated with health care access among United States women, 2002. *Social Science and Medicine, 66*, 260-275.

DeBruhl, N.D., Bassett, L.W., Jessop, N.W., & Mason, A.M. (1996). Mobile mammography: Results of a national survey. *Radiology, 201*(2), 433-7.

Diez Roux, A.V., Borrell, L.N., Haan, M., Jackson, S.A., & Schultz, R. (2004). Neighbourbood environments and mortality in an elderly cohort: results from the cardiovascular health study. *Journal of Epidemiology and Community Health, 58*, 917-923.

Duelberg, S.I. (1992). Preventive health behavior among black and white women in urban and rural areas. *Social Science and Medicine, 34*(2), 191-8.

Engelman, K.K., Hawley, D.B., Gazaway, R., Mosier, M.C., Ahluwalia, J.S., Ellerbeck, E.F. (2002). Impact of geographic barriers on the utilization of mammograms by older rural women. *Journal of the American Geriatrics Society, 50*(1), 62-8.

Fletcher, S.W., & Elmore, J.G. (2003). Clinical practice: Mammographic screening for breast cancer. *New England Journal of Medicine, 348*(17):1672-80.

Frankel, M.R., Srinath, K.P, Hoaglin, D.C., Battaglia, M.P., Smith, P.J., Wright, R.A., & Khare, M. (2003). Adjustments for non-telephone bias in random-digit dialing surveys. *Statistics in Medicine, 22*(9):1611-26.

Gordon, N.P., Hiatt, R.A., & Lampert, D.I. (1993). Concordance of self-reported data and medical record audit for six cancer screening procedures. *Journal of the National Cancer Institute, 85*(7), 566-70.

Hall, S.A., Kaufman, J.S., & Ricketts, T.C. (2006). Defining urban and rural areas in U.S. epidemiologic studies. *Journal of Urban Health, 83*(2), 162-175.

Hyndman, J.C., Holman, C.D., & Dawes, V.P. (2000). Effect of distance and social disadvantage on the response to invitations to attend mammography screening. *Journal of Medical Screening, 7*(3), 141-5.

Jones, A.R., Caplan, L.S., & Davis, M.K. (2003). Racial/ethnic differences in the self-reported use of screening mammography. *Journal of Community Health, 28*(5), 303-16.

Larson, S.L., Machlin, S.R., Nixon, A., & Zodet, M. Health care in urban and rural areas, combines years 1998-2000. Rockville (MD): Agency for Healthcare Research and Quality; 2004. MEPS Chartbook No. 13. AHRQ Pub. No. 04-0050.

Litaker, D., & Tomolo, A. (2007). Association of Contextual Factors and Breast Cancer Screening: Finding New Targets to Promote Early Detection. *Journal of Women's Health, 16*(1), 36-45.

Marchick, J., & Henson, D.E. (2005). Correlations between Access to Mammography and Breast Cancer Stage at Diagnosis. *Cancer, 103*(8), 1571-80.

Meissner, H.I., Breen, N., Taubman, M.L., Vernon, S.W., & Graubard, B.I. (2007). Which women aren't getting memmograms and why? (United States). *Cancer Causes and Control, 18*(1), 61-70.

Menck, H.R. & Mills, P.K. (2001). The influence of urbanization, age, ethnicity, and income on the early diagnosis of breast carcinoma: Opportunity for screening improvement. *Cancer, 92*(5), 1299-1304.

Monroe, A.C., Ricketts, T.C., & Savitz, L.A. (1992). Cancer in Rural Versus Urban Populations: A Review. *The Journal of Rural Health, 8*(3), 212-220.

Oakes, J.M., & Johnson, P.J. (2006). Propensity Score Matching for Social Epidemiology. In: J.M. Oakes & J.S. Kaufman (Eds.), *Methods in Social Epidemiology* (370-392). San Francisco: Jossey-Bass.

Parsons, L. (2008). Proceedings of the Annual SAS Users Group International Conference 2008: *Reducing Bias in Propensity Score Matched-Pair Sample using Greedy Matching Techniques, Paper 214-26.* Austin, TX.

Robert, S.A. (1998). Community-level socioeconomic status effects on adult health. *Journal of Health and Social Behavior, 39*(1), 18-37.

Roche, L.M., Skinner, R., Weinstein, R.B. (2002). Use of a Geographic Information System to Identify and Characterize Areas with High Proportions of Distant Stage Breast Cancer. *Journal of Public Health Management and Practice, 8*(2), 26-32.

Rosenthal, M.B., Zaslavsky, A., & Newhouse, J.P. (2005). The geographic distribution of physicians revisited. *Health Services Research, 40*, 1931-52.

Smith, K.B., Humphreys, J.S., & Wilson, M.G.A. (2008). Addressing the health disadvantage of rural populations: How does epidemiological evidence inform rural health policies and research? *Australian Journal of Rural Health, 16*, 56-66.

Stearns, S.C., Slifkin, R.T., & Edin, H.M. (2000). Access to Care for Rural Medicare Beneficiaries. *The Journal of Rural Health, 16*(1), 31-42.

Susan G. Komen for the Cure. (2007). Facts for Life: Mammography. Item No. 806-375 9/07. Retrieved February 24, 2008, from http://cms.komen.org/komen/AboutBreastCancer/EarlyDetectionScreening/EDS3-3-1?ssSourceNodeId=292&ssSourceSiteId=Komen.

U.S. Census Bureau. (2007). Metropolitan and Micropolitan Statistical Areas: What are metropolitan and micropolitan statistical areas? Retrieved January 18, 2008, from http://www.census.gov/population/www/estimates/metroarea.html.

U.S. Department of Health and Human Services. (2001a). Healthy People 2010: Objectives for Improving Health (Part A). Retrieved February 6, 2008, from http://www.healthypeople.gov/document/pdf/Volume1/03Cancer.pdf.

U.S. Department of Health and Human Services. (2001b). Healthy People 2010: A Systematic Approach to Health Improvement. Retrieved February 6, 2008, from http://www.healthypeople.gov/document/html/uih/uih_bw/uih_2.htm#goals.

Vlahov, D., & Galea, S. (2002). Urbanization, urbanicity, and health. *Journal of Urban Health, 79*(4 Suppl 1), S1-S12.

Zhang, P., Tao, G., & Irwin, K.L. (2000). Utilization of Preventive Medical Services in the United States: A Comparison Between Rural and Urban Populations. *The Journal of Rural Health, 16*(4), 349-56.

APPENDIX A

BEHAVIORAL RISK FACTOR SURVEILLANCE SYSTEM CODEBOOK

STUDY VARIABLES

APPENDIX A

BEHAVIORAL RISK FACTOR SURVEILLANCE SYSTEM CODEBOOK

STUDY VARIABLES

Survey Column	Item Description
80	**Do you have any kind of health care coverage, including health insurance, prepaid plans such as HMOs, or government plans such as Medicare?**
	Yes
	No
	Don't Know/Not Sure
	Refused
109	**Are you: (marital status)**
	Married
	Divorced
	Widowed
	Separated
	Never married
	A member of an unmarried couple
	Refused
	Not asked or missing
112	**What is the highest grade or year of school you completed?**
	Never attended school or only kindergarten
	Grades 1-8 (Elementary)
	Grades 9-11 (Some high school)
	Grade 12 or GED (High school graduate)
	College 1 year to 3 years (Some college or technical school)
	College 4 years or more (College graduate)
	Refused
	Not asked or missing
114-115	**Is your annual household income from all sources:**
	Less than $10,000
	$10,000 to less than $15,000
	$15,000 to less than $20,000
	$20,000 to less than $25,000
	$25,000 to less than $35,000
	$35,000 to less than $50,000
	$50,000 to less than $75,000
	$75,000 or more
	Don't Know/Not Sure

	Refused Not asked or missing
135	**Indicate sex of respondent.** Male Female
172	**A mammogram is an x-ray of each breast to look for breast cancer. Have you ever had a mammogram?** Yes No Don't Know/Not Sure Refused Not asked or missing
173	**How long has it been since your last mammogram?** Within the past year (1 to 12 months ago) Within the past 2 years (1 to 2 years ago) Within the past 3 years (2 to 3 years ago) Within the past 5 years (3 to 5 years ago) 5 or more years ago Don't Know/Not Sure Refused Not asked or missing
993	**Metropolitan Status Code** In the center city of an MSA Outside the center city of an MSA but inside the county containing the center city Inside a suburban county of the MSA In an MSA that has no center city Not in an MSA GU, PR, VI (Outside the United States)
1202	**Race/ethnicity categories** White only, non-Hispanic Black only, non-Hispanic Asian only, non-Hispanic Native Hawaiian or other Pacific Islander only, non-Hispanic American Indian or Alaskan Native only, non-Hispanic Other race only, non-Hispanic Multiracial, non-Hispanic Hispanic Don't Know/Not Sure/Refused
1203-1204	**Fourteen-level age category** Age 18 to 24

Age 25 to 29
Age 30 to 34
Age 35 to 39
Age 40 to 44
Age 45 to 49
Age 50 to 54
Age 55 to 59
Age 60 to 64
Age 65 to 69
Age 70 to 74
Age 75 to 79
Age 80 or older
Don't Know/Refused/Missing

APPENDIX B

SAS PROGRAMMING CODE

APPENDIX B

SAS PROGRAMMING CODE

```
*************************************************************************************
***            THESIS: URBANICITY AND MAMMOGRAPHY UTILIZATION                    ***
***            AUTHOR: KELLY YLITALO                                             ***
***            DATE: MAY 2008                                                    ***
***            INPUT DATA: 2006 BRFSS FILES                                      ***
*************************************************************************************;

TITLE1 'KELLY YLITALO, THESIS, 2006 BRFSS SURVEY';

OPTIONS FORMDLIM='-';

DATA THESIS;
        INFILE 'C:\Documents and Settings\sph lab\Desktop\CDBRFS06.ASC' LRECL=1252;

*USE THE INPUT STATEMENT TO INPUT THE APPROPRIATE VARIABLES;
INPUT
SEX $135
_AGE_G $1206
EDUCA $112
MARITAL $109
INCOME2 114-115
HADMAM $172
HOWLONG     $173
HLTHPLAN $80
MSCODE $993
_RACEGR2 $1200
_AGEG5YR $1205-1206
;

IF MSCODE=' ' THEN URBAN=.;
IF MSCODE IN (1,2,3,4) THEN URBAN=1;
IF MSCODE IN (5) THEN URBAN=0;


IF _AGEG5YR IN (01,02,03, 04) THEN AGE=4;
IF _AGEG5YR IN (05, 06) THEN AGE=1;
IF _AGEG5YR IN (07, 08, 09) THEN AGE=2;
IF _AGEG5YR IN (10, 11, 12, 13) THEN AGE=3;
IF _AGEG5YR IN (14) THEN AGE=.;

IF MARITAL='1' THEN MARRIED=1;
IF MARITAL IN (2,3,4) THEN MARRIED=2;
IF MARITAL IN (5,6) THEN MARRIED=3;
IF MARITAL='9' OR MARITAL=' ' THEN MARRIED=.;

IF EDUCA IN (1,2,3) THEN EDUCAT=1;
IF EDUCA IN (4) THEN EDUCAT=2;
```

```
IF EDUCA IN (5) THEN EDUCAT=3;
IF EDUCA IN (6) THEN EDUCAT=4;
IF EDUCA='9' OR EDUCA=' ' THEN EDUCAT=.;

IF HADMAM IN (1) AND HOWLONG IN (1,2) THEN MAMM=1;
ELSE MAMM=0;

IF _RACEGR2 IN (1) THEN RACE=1;
IF _RACEGR2 IN (2) THEN RACE=2;
IF _RACEGR2 IN (5) THEN RACE=3;
IF _RACEGR2 IN (3,4) THEN RACE=4;
IF _RACEGR2=' ' THEN RACE=.;

IF HLTHPLAN IN (1) THEN INSURANCE=1;
IF HLTHPLAN IN (2) THEN INSURANCE=2;
IF HLTHPLAN IN (7,9) THEN INSURANCE=.;

IF INCOME2 IN (1) THEN INCOME=1;
IF INCOME2 IN (2,3,4) THEN INCOME=2;
IF INCOME2 IN (5,6) THEN INCOME=3;
IF INCOME2 IN (7) THEN INCOME=4;
IF INCOME2 IN (8) THEN INCOME=5;
IF INCOME2='77' OR INCOME2='99' OR INCOME2=' ' THEN INCOME=.;

RUN;

PROC PRINT DATA=THESIS (OBS=30);
RUN;

PROC FREQ DATA=THESIS;
        TABLES _AGEG5YR AGE;
RUN;

PROC FREQ DATA=THESIS;
        TABLES SEX*AGE/CHISQ;
RUN;

PROC FREQ DATA=THESIS;
        TABLES MAMM;
RUN;

PROC FREQ DATA=THESIS;
        TABLES _RACEGR2 RACE;
RUN;

*CREATE A DATA SET THAT CONTAINS ONLY ELIGIBLE PARTICIPANTS (FEMALE, AGED 40
YEARS OR OLDER);
DATA STUDY;
        SET THESIS;
        IF SEX=2 AND AGE IN (1,2,3);
        ID=_N_;
RUN;
```

```
PROC PRINT DATA=STUDY (OBS=100);
RUN;

PROC FREQ DATA=STUDY;
        TABLES URBAN RACE EDUCAT MARRIED INSURANCE INCOME MAMM AGE;
RUN;

PROC FREQ DATA=STUDY;
        TABLES (AGE RACE EDUCAT MARRIED INSURANCE INCOME
MAMM)*URBAN/CHISQ;
RUN;

PROC FREQ DATA=STUDY;
        TABLES (AGE RACE EDUCAT MARRIED INSURANCE INCOME
URBAN)*MAMM/CHISQ;
RUN;

PROC LOGISTIC DATA=STUDY DESCENDING;
        MODEL MAMM=URBAN;
RUN;

PROC LOGISTIC DATA=STUDY DESCENDING;
        CLASS URBAN (REF='0');
        MODEL MAMM=URBAN;
RUN;

PROC LOGISTIC DATA=STUDY DESCENDING;
        CLASS AGE (REF='1');
        MODEL MAMM=AGE;
RUN;

PROC LOGISTIC DATA=STUDY DESCENDING;
        CLASS RACE (REF='1');
        MODEL MAMM=RACE;
RUN;

PROC LOGISTIC DATA=STUDY DESCENDING;
        CLASS EDUCAT (REF='1');
        MODEL MAMM=EDUCAT;
RUN;

PROC LOGISTIC DATA=STUDY DESCENDING;
        CLASS MARRIED (REF='1');
        MODEL MAMM=MARRIED;
RUN;

PROC LOGISTIC DATA=STUDY DESCENDING;
        CLASS INSURANCE (REF='1');
        MODEL MAMM=INSURANCE;
RUN;

PROC LOGISTIC DATA=STUDY DESCENDING;
        CLASS INCOME (REF='1');
```

```
        MODEL MAMM=INCOME;
RUN;

PROC LOGISTIC DATA=STUDY DESCENDING;
        MODEL MAMM=URBAN AGE RACE EDUCAT INSURANCE MARRIED;
RUN;

PROC LOGISTIC DATA=STUDY DESCENDING;
        MODEL MAMM=URBAN AGE RACE EDUCAT INCOME INSURANCE MARRIED;
RUN;

*PROPENSITY SCORE MATCHING;
PROC LOGISTIC DATA=STUDY DESCENDING;
        MODEL URBAN=AGE RACE EDUCAT INCOME INSURANCE MARRIED;
        OUTPUT OUT=STUDY PROB=PROB;
RUN;

/* ******************************************** */
/* Call statement for Greedy Match Macro (Parsons, 2008)*/
/* ******************************************** */
%GREEDMTCH(WORK,STUDY,URBAN,MATCHED);

PROC FREQ DATA=WORK.MATCHED;
        TABLES URBAN RACE EDUCAT MARRIED INSURANCE INCOME MAMM AGE;
RUN;

PROC FREQ DATA=WORK.MATCHED;
        TABLES (AGE RACE EDUCAT MARRIED INSURANCE INCOME
MAMM)*URBAN/CHISQ MEASURES;
RUN;

PROC LOGISTIC DATA=WORK.MATCHED DESCENDING;
        MODEL MAMM=URBAN;
RUN;


/* ********************************************** */
/* Greedy 5->1 Digit Matching Macro (Parsons, 2008)*/
/* ********************************************** */
%MACRO GREEDMTCH
(
Lib, /* Library Name */
Dataset, /* Data set of all */
/* patients */
depend, /* Dependent variable */
/* that indicates */
/* Case or Control; */
/* Code 1 for Cases, */
/* 0 for Controls */
matches /* Output file of matched */
/* pairs */
);
/* Macro to sort the Cases and Controls
```

```
dataset */
%MACRO SORTCC;
proc sort data=tcases
out=&LIB..Scase;
by prob;
run;
proc sort data=tctrl
out=&LIB..Scontrol;
by prob randnum;
run;
%MEND SORTCC;

/* Macro to Create the initial Case and
Control Data Sets */
%MACRO INITCC(digits);
data tcases (drop=cprob)
tctrl (drop=aprob) ;
set &LIB..&dataset. ;
/* Create the data set of Controls*/
if &depend. = 0 and prob ne .
then do;
cprob = Round(prob,&digits.);
Cmatch = 0;
Length RandNum 8;
RandNum=ranuni(1234567);
Label RandNum=
'Uniform Randomization Score';
output tctrl;
end;
/* Create the data set of Cases */
else if &depend. = 1 and prob ne .
then do;
Cmatch = 0;
aprob =Round(prob,&digits.);
output tcases;
end;
run;
%SORTCC;
%MEND INITCC;
/* Macro to Perform the Match */
%MACRO MATCH (MATCHED,DIGITS);
data &lib..&matched. (drop=Cmatch randnum
aprob cprob start oldi curctrl matched);
/* select the cases data set */
set &lib..SCase ;
curob + 1;
matchto = curob;
if curob = 1 then do;
start = 1;
oldi = 1;
end;
/* select the controls data set */
DO i = start to n;
```

```
set &lib..Scontrol point= i nobs = n;
if i gt n then goto startovr;
if _Error_ = 1 then abort;
curctrl = i;
/* output control if match found */
if aprob = cprob then
do;
Cmatch = 1;
output &lib..&matched.;
matched = curctrl;
goto found;
end;
/* exit do loop if out of potential
matches */
else if cprob gt aprob then
goto nextcase;
startovr: if i gt n then
goto nextcase;
END; /* end of DO LOOP */
/* If no match was found, put pointer
back*/
nextcase:
if Cmatch=0 then start = oldi;
/* If a match was found, output case and
increment pointer */
found:
if Cmatch = 1 then do;
oldi = matched + 1;
start = matched + 1;
set &lib..SCase point = curob;
output &lib..&matched.;
end;
retain oldi start;
if _Error_=1 then _Error_=0;
run;
proc sort data=&lib..scase out=sumcase;
by ID;
run;
proc sort data=&lib..scontrol
out=sumcontrol;
by ID;
run;
proc sort data=&lib..&matched. out=smatched
(keep=ID matchto);
by ID;
run;
data tcases (drop=matchto);
merge sumcase(in=a) smatched;
by ID;
if a and matchto = . ;
cmatch = 0;
aprob =Round(prob,&digits.);
run;
```

```
data tctrl (drop=matchto);
merge sumcontrol(in=a) smatched;
by ID;
if a and matchto = . ;
cmatch = 0;
cprob = Round(prob,&digits.);
run;
%SORTCC
%MEND MATCH;
%INITCC(.00001);
/* Do a 5-digit match */
%MATCH(Match5,.0001);
/* Do a 4-digit match on remaining
unmatched */
%MATCH(Match4,.001);
/* Do a 3-digit match on remaining
unmatched */
%MATCH(Match3,.01);
/* Do a 2-digit match on remaining
unmatched */
%MATCH(Match2,.1);
/* Do a 1-digit match on remaining
unmatched */
%MATCH(Match1,.1);
data &lib..&matches.;
set &lib..match5(in=a)
&lib..match4(in=b) &lib..match3(in=c)
&lib..match2(in=d) &lib..match1(in=e);
if b then matchto=matchto + 100000;
if c then matchto=matchto + 10000000;
if d then matchto=matchto + 1000000000;
if e then matchto=matchto + 100000000000;
run;
/* Sort file -- Need sort for Univariate
analysis in tables
*/
proc sort data=&lib..&matches. out =
&lib..S&matches.;
by &depend.;
run;
%MEND GREEDMTCH;
```