UNIFORMED SERVICES UNIVERSITY OF THE HEALTH SCIENCES
4301 JONES BRIDGE ROAD
BETHESDA, MARYLAND 20814-4799

GRADUATE EDUCATION
(301) 295-3913
FAX (301) 295-6772

APPROVAL SHEET

Title of Dissertation: "Impact of Sociodemographic Factors on Racial/Ethnic Differences in Tumor Stage and Size for Cancer of the Female Breast"

Name of Candidate: CAPT Barry Miller
Doctor of Public Health
15 February 2000

Dissertation and Abstract Approved:

David Cruess, Ph.D. — 3/2/00 — Date
Department of Preventive Medicine and Biometrics
Committee Chairperson

Terry Thomas, Ph.D. — 15 Feb 2000 — Date
Department of Preventive Medicine and Biometrics
Committee Member

COL Gary Gackstetter, USAF — 15 Feb 00 — Date
Department of Preventive Medicine and Biometrics
Committee Member

Tracy Sbrocco, Ph.D. — 15 Feb 00 — Date
Department of Medical and Clinical Psychology
Committee Member

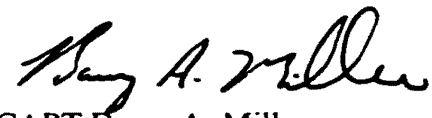
Ben Hankey, M.D. — 2/15/00 — Date
National Cancer Institute
Committee Member

COPYRIGHT STATEMENT

The author hereby certifies that the use of any copyrighted material in the thesis manuscript entitled:

"The Impact of Sociodemographic Factors on Racial/ethnic Differences in Tumor Stage and Tumor Size for Cancer of the Female Breast"

beyond brief excerpts is with the permission of the copyright owner, and will save and hold harmless the Uniformed Services University of the Health Sciences from any damage which may arise from such copyright violations.



CAPT Barry A. Miller
Department of Preventive Medicine and Biometrics
Uniformed Services University of the Health Sciences

ABSTRACT

Title of Thesis: The Impact of Sociodemographic Factors on Racial/Ethnic Differences in Tumor Stage and Tumor Size for Cancer of the Female Breast

Name, degree, year: Barry A. Miller, Doctor of Public Health, 2000

Thesis directed by: Terry L. Thomas, Ph.D., Associate Professor, Department of Preventive Medicine and Biometrics

A population-based, case-control study was conducted to determine the importance of sociodemographic factors in explaining racial/ethnic differences in tumor stage and size at the time of diagnosis among women with invasive, primary breast cancer. The study group included 106,607 women newly diagnosed with breast cancer during the years 1992 through 1996 while residing in any of the eleven reporting areas in the United States that comprise the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute (NCI).

Descriptive tabulations of the study variables indicated that Japanese and White women tended to be diagnosed at an earlier stage, with smaller diameter tumors, and at a lower tumor grade than other groups. Black and Hispanic women were more likely than other groups to be diagnosed with metastatic disease, with tumors 2 cm or larger in diameter, and with poorly differentiated tumors. In the regression analysis, elevated odds ratios among Black and Hispanic patients for later stage and larger size tumors were reduced by 50% to 60% when sociodemographic factors were added to a model already containing age and geographic area. Tumor grade and hormone receptor status only explained a small amount of the excess odds for distant stage disease among Black and Hispanic women, and did not explain any of the racial/ethnic differences in regional

stage disease or larger tumor size. In the analysis of tumor size, odds ratios for Black, Hispanic, Filipino, Chinese, and Korean women remained elevated relative to White women after adjustment for sociodemographic factors, tumor grade, and hormone receptor status. Japanese women, conversely, had consistently lower odds ratios (relative to White women) for every study outcome.

Results from this study suggest that sociodemographic factors account for a significant portion of the observed racial/ethnic differences in the stage of disease and tumor size at the time of diagnosis, but that unmeasured differences in socioeconomic or biological characteristics of breast tumors among some racial/ethnic groups may also exist. The special cancer data base created for this study may now be used to investigate the importance of sociodemographic factors in explaining population patterns for other types of cancer.

THE IMPACT OF SOCIODEMOGRAPHIC FACTORS ON RACIAL/ETHNIC DIFFERENCES IN TUMOR STAGE AND TUMOR SIZE FOR CANCER OF THE FEMALE BREAST

by

Barry A. Miller

Dissertation submitted to the Faculty of the
Department of Preventive Medicine and Biometrics of the
Uniformed Services University of the Health Sciences
in partial fulfillment of the requirements for the degree of
Doctor of Public Health 2000

ACKNOWLEDGMENTS

I thank my advisor Dr. Terry Thomas and the other members of my dissertation committee - Drs. David Cruess, Gary Gackstetter, Benjamin Hankey, and Tracy Sbrocco for their guidance and encouragement; my colleagues at the National Cancer Institute - Drs. Brenda Edwards, Benjamin Hankey, Barry Graubard and Donald Henson for their technical and administrative support; the principal investigators, data managers, and additional staff of the cancer registries participating in the Surveillance, Epidemiology and End Results Program for their technical assistance and data collection activities; and Scott Depuy, Todd Gibson and Steve Scoppa of Information Management Systems for preparation of the initial data base.

DEDICATION

To Allyson and both of our families for their support and understanding.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER I. BACKGROUND AND LITERATURE REVIEW

A. Background

1. Characterization of Breast Tumors

Breast cancer is the most common form of cancer diagnosed among women in the United States, accounting for about 29% of all malignancies [ACS 1999]. It is also the most common cancer in women worldwide [Parkin 1998]. About 16% of all cancer deaths among U.S. women are due to cancer of the breast, placing it second to cancer of the lung and bronchus [ACS 1999].

Over 90% of breast carcinomas arise as a neoplasm of the ductal epithelium, with the remainder developing as lower grade neoplasms from the lobular epithelium [Henderson 1996]. About 15% to 20% of breast cancers are diagnosed very early in their natural history and may be termed carcinoma in situ [PDQ 1999]. They have all of the characteristics of malignancy except invasion. An in situ cancer has not penetrated the basement membrane nor extended beyond the epithelial tissue. Some common synonyms are intraepithelial (confined to the epithelial tissue), non-invasive, and non-infiltrating. Once a cancer has invaded other tissues or spread to other parts of the body, it is termed invasive. Several histologic types of invasive breast cancer have been identified, but ductal carcinoma, not otherwise specified, is by far the most commonly recorded type. It comprises about 80% of all cases [Berg 1995]. A few specific variants of invasive ductal carcinoma have a better prognosis than other types. They include pure mucinous, pure tubular, pure medullary and pure papillary carcinoma [Fisher 1993, Donegan 1997].

These special types form a small group, however, representing less than 6% of all invasive breast cancers [Berg 1995].

The anatomic extent of a cancer, determined clinically or pathologically, is a classic and reliable indicator of prognosis [Simpson 1996, Donegan 1997]. The main components used in classifying the extent of disease are size of the tumor, extension of the tumor, evidence of metastasis, and lymph node involvement. General staging categories for invasive breast cancer include localized (confined to the breast tissue with no lymph node involvement), regional (direct invasion to extramammary tissues and/or metastasis to regional lymph nodes), and distant (metastasis beyond regional tissues) [Seiffert 1993]. These categories identify three general groups with distinctly different probabilities for survival after diagnosis and treatment. Five-year cumulative relative survival rates associated with this staging scheme for patients diagnosed between 1988-94 through the Surveillance, Epidemiology and End Results (SEER) population-based registry system are shown in **Table I-1**. This general staging scheme is useful for monitoring time trends in cancer rates for surveillance purposes since the stage definitions remain comparable over time. It differs from the more-detailed clinical staging scheme developed by the American Joint Committee on Cancer (AJCC), which makes use of tumor size in assigning the stage, but has changed its staging definitions over time. The AJCC staging is based on the size of the primary tumor (T), the absence or presence and extent of regional lymph node metastasis (N), and the absence or presence of distant metastasis (M). This scheme is referred to as the TNM system and is delineated in a manual published by the AJCC [AJCC 1997]. TNM-based stage groupings for invasive breast cancers are summarized in **Table I-2**. A limitation of all

cancer staging is that it provides a static picture of the disease. Within each stage are cases with differing biological potential and speed of progression [Donegan 1997].

Tumor size, measured as the largest dimension or diameter of the primary tumor, is second only to axillary lymph node status as an independent prognostic factor [Donegan 1997]. It is directly related to an increasing probability of regional metastasis, an increasing average number of involved axillary lymph nodes, and an increasing probability of recurrence and death. Studies indicate that tumors of equal size are prognostically similar whether they are palpable or not and independent of their method of detection [Tabar 1987, Pagana 1989]. Tumors 1.0 cm or less in diameter have an especially low risk of recurrence. Several studies have reported 5-year or 10-year disease-free survival exceeding 90 percent for node-negative patients with tumors 1.0 cm or less in diameter [O'Reilly 1990, Merkel 1993, Rosen 1993].

Another feature of invasive ductal and lobular breast carcinomas that has prognostic value is histologic grade. Histologic grade is a measure of intrinsic malignant characteristics of the tumor including the degree of tubule formation, number of mitoses, and nuclear pleomorphism in routine sections of breast tissue [Donegan 1997]. This information is used to assign a grade indicating the degree of tumor differentiation ranging from well differentiated (low grade), through moderately differentiated, to poorly differentiated (high grade). The degree of differentiation, in turn, is a morphologic indicator of tumor aggressiveness, with highly differentiated (i.e., low grade) tumors being less aggressive [Donegan 1997]. Histologic grade correlates with breast cancer patient survival, with high grade cancers having the lowest survival probabilities [Henson 1991]. This relationship persists in spite of interobserver and intraobserver variation

among pathologists grading breast cancer [Henson 1991], and even after the lymph node status of patients is taken into account [Fisher 1993, Garne 1994].

A variety of proteins involved in cellular differentiation, proliferation, and invasion are differentially expressed in neoplastic and normal breast epithelium. The most widely recognized among these are the estrogen (ER) and progesterone (PR) hormone receptors. Levels of ER and PR proteins in breast tumor tissue have undergone intensive study both as indicators of prognosis and as predictors of response to hormone and endocrine therapy [Donegan 1997, Osborne 1998]. These receptors are polypeptides that bind their respective hormones, translocate to the nucleus, and induce specific gene expression [ASCO 1996]. PR is expressed only after transcriptional activation of its gene by a functional ER-estrogen complex. ER positive or PR positive tumors are correlated with favorable prognostic features including evidence of tumor cell differentiation (i.e., low-grade histology) and a lower rate of cell proliferation [Mohla 1982, Pegoraro 1986, Dhingra 1996, Osborne 1998]. Tumors that are positive for ER generally have a low S-phase fraction, indicating that a low percentage of tumor cells are in the proliferation phases of the cell cycle [ASCO 1996, Donegan 1997, Beckmann 1997, Landberg 1997, Osborne 1998, Ravaioli 1998]. ER and PR levels have been widely used by oncologists to predict the likelihood of recurrent disease, although data supporting this use are inconsistent [ASCO 1996, Ferno 1998]. A recent review of published studies indicates that ER status and PR status are probably more reflective of tumor growth rate than of metastatic potential [Donegan 1997]. The measurement of ER and PR is most useful for predicting response to hormonal therapy [ASCO 1996, Dhingra 1996, Osborne 1998]. Tumors that express both ER and PR have the greatest benefit from hormonal therapy,

but those containing only ER or only PR still have significant responses.

Breast cancer is highly treatable by surgery, radiation therapy, chemotherapy, and hormonal therapy. Selection of therapy is influenced by the tumor stage; pathologic characteristics of the primary tumor, including ER and PR levels and lymph node involvement; menopausal status; patient age; and general health [PDQ 1999]. A summary of current treatment options by tumor stage, based on information from the National Cancer Institute's comprehensive cancer database [PDQ 1999], appears in **Table I-3a-d**.

2. Breast Cancer Etiology

Considerable experimental, clinical, and epidemiologic research aimed at clarifying the etiology of breast cancer indicates that hormones play a major role [Kelsey 1990, Le Marchand 1991, Habel 1993, Henderson 1996, Beckmann 1997]. The known risk factors can be thought of in terms of their influence on cumulative exposure of breast tissue to estrogen and perhaps progesterone [Pike 1993, Henderson 1996]. Endogenous and exogenous hormones appear to affect the expression of oncogenes and tumor-suppressor genes directly by altering promoter activity and indirectly by influencing the proliferation rate of breast epithelial cells [Beckmann 1997]. The activation of oncogenes and inactivation of tumor-suppressor genes produces a series of genetic changes that are believed to lead to malignancy [Pike 1993, Henderson 1996, Landberg 1997].

The most established risk factors for breast cancer include family history,

particularly among 1st degree relatives; early menarche; and late ages at first childbirth and menopause. These factors, however, are not readily modifiable for the purpose of disease prevention. There is evidence that menopausal estrogen replacement therapy increases breast cancer risk, but only to a small extent [Brinton 1993, Pike 1993, Henderson 1996, Colditz 1998]. The potential effect of oral contraceptives on risk is complex and seems to be limited to a subgroup of recent long-term users, though a confounding effect of increased medical surveillance in this group can not be ruled out [Malone 1993, Collaborative Group 1996]. The question of whether dietary intake of fat plays a role in the development of breast cancer has been the focus of many ecological, migrant, prospective cohort, case-control and experimental studies [Greenwald 1999, Hunter 1999]. This factor would be more amenable to change, but the analytic epidemiological studies generally do not support an association [Holmes 1999]. Obesity in post-menopausal women has been linked with mortality from breast cancer due, in part, to delayed diagnosis [Mohle-Boetani 1988, Hunter 1993, Hulka 1994, Yong 1996, Jones 1997] and to a worse prognosis that is independent of the stage of disease [Tretli 1990, Senie 1992]. Available data, however, suggest that obesity can account for only weak or moderate elevations in risk [Le Marchand 1991, Harris 1992, Henderson 1996]. Numerous studies have linked moderate to heavy alcohol intake with increases in breast cancer risk [Rosenberg 1993, Longnecker 1995, Henderson 1996], but the proportion of breast cancer cases attributable to alcohol consumption in the United States, assuming causality, is estimated to be quite small [Longnecker 1999]. Findings for light to moderate alcohol consumption are inconsistent and positive studies indicate only a very slight increase in risk [Longnecker 1995, Zhang 1999a, Zhang 1999b]. Several recent

epidemiologic studies have suggested that physical activity is related to a reduced risk for breast cancer, but the magnitude of the effect is unclear, the underlying biologic mechanisms remain unexplained, and confounding and effect modification by other factors can not be ruled out [Brinton 1998, Friedenreich 1998].

In summary, since we do not know how to effectively prevent this major cause of female cancer mortality, control strategies emphasize the early detection and treatment of breast tumors before they have reached an advanced stage. A high quality mammogram with a clinical breast exam is the most effective way to detect breast cancer early, when it is most treatable [Senie 1994]. There is not universal agreement, however, on the age at which screening mammography should begin. The National Cancer Institute [NCI 1997] and the American Cancer Society [ACS 1997] have accepted the March 1997 recommendations of the National Cancer Advisory Board stating (with one dissenting vote) that: 1) Women aged 40 and older should be screened every one to two years with mammography; and, 2) Women who are at higher than average risk of breast cancer should seek expert medical advice about whether they should begin screening before age 40 and about the frequency of screening. Members of a Consensus Development Panel on mammography sponsored by the National Institutes of Health in January 1997 [NIH 1997] concluded that "the available data did not warrant a single recommendation for all women in their forties" and that women in this age group should make their own decision in consultation with health professionals. The U.S. Preventive Services Task Force has also not recommended routine mammograms for average-risk women in their 40s [USPSTF 1996]. Early detection and treatment efforts in the United States have achieved only limited success, however, since recent mortality declines are modest and are not

comparable in all segments of the population [Ries 1998].

B. Significance of this Study

Survival from breast cancer among women in the United States varies by racial/ethnic group. These survival differences often persist after stage of disease at the time of diagnosis is taken into account. Proposed explanations for this disparity in survival include racial/ethnic differences in socioeconomic position and/or differences in the biological characteristics of breast tumors. Results from the few studies that have examined these factors concurrently are inconsistent.

Although several studies report that socioeconomic factors explain a large part of the racial/ethnic differences in breast cancer survival, evidence for an additional effect due to racial/ethnic differences in the biological characteristics of breast tumors is inconclusive. The three largest studies looking at racial/ethnic patterns of socioeconomic factors and tumor biology included only White and Black women [Chen 1994, Gordon 1995, Elmore 1998]. Two additional studies included Hispanic women [Weiss 1995] and Asian women [Krieger 1997a], but their populations were too small to draw reliable conclusions. Two of the three large studies were hospital-based, and one of these accrued its study group from patients participating in clinical trials. Since various selection factors may have influenced whether breast cancer patients were included in these two investigations, their findings may not be generalizable.

Thus, there is a need for population-based studies with larger study populations and more diverse racial/ethnic groups. A larger study size will provide additional power

for developing reliable estimates of the effect of racial/ethnic group and socioeconomic position on breast cancer outcomes. It will also improve our ability to detect differences in the patterns of various tumor characteristics (e.g., stage, size, grade, hormone receptor status) across racial/ethnic groups and socioeconomic levels.

C. Literature Review

1. Survival Differences by Race/Ethnicity

Survival rates among breast cancer patients are known to vary by racial/ethnic group. Data from population-based cancer registries in the United States have consistently reported that Black women have poorer survival than Whites [Axtell 1978, NIH 1980, Le Marchand 1984, Young 1984, Vernon 1985, Bain 1986, Baquet 1986, Ragland 1991, Elledge 1994, Simon 1996, Meng 1997, Ries 1998]. Two studies of the survival experience of women from five major racial/ethnic groups in Hawaii found that native Hawaiian and Filipino women had a higher risk of dying within five years following a breast cancer diagnosis than women from other racial/ethnic groups [Le Marchand 1984, Meng 1997]. Japanese and Chinese women had the highest five-year survival probabilities, followed by Whites among patients diagnosed between 1960 and 1979 [Le Marchand 1984]. Findings were similar for a later series of patients diagnosed between 1980 and 1988 [Meng 1997].

Differential proportions of more advanced disease at the time of diagnosis plays a role in the racial/ethnic disparities, though survival differences often persist after stage of

disease is taken into account [Le Marchand 1984, Vernon 1985, Le Marchand 1985, Bain 1986, Samet 1987, Ragland 1991, Elledge 1994, Simon 1996, Meng 1997, Ries 1998, Wojcik 1998]. A study of ten-year survival rates by race/ethnicity and stage included 1,983 breast cancer patients treated at M.D. Anderson Hospital and Tumor Institute in Houston, Texas between 1949 and 1968 [Vernon 1985]. Black women were found to have poorer survival than either White or Hispanic women, whose survival experience was similar. The racial/ethnic differences in survival remained after age, stage of disease at diagnosis, and delay in seeking treatment were taken into account. A more recent, multi-center, hospital-based study of breast cancer patients found that overall five-year survival among Black and Hispanic women was significantly worse than for Whites [Elledge 1994]. Minority women were more likely to present with clinically advanced disease. Within stage, however, Hispanic and White women had comparable five-year survival rates, while the prognosis for Black women remained worse than for the other groups.

Statistics reported by the National Cancer Institute for women diagnosed with breast cancer in the population-based cancer registries comprising the Surveillance, Epidemiology and End Results Program (SEER) indicate that Black women have poorer five-year relative survival rates than White women within every stage of disease [Ries 1998]. In metropolitan Atlanta, which is one of the SEER cancer registration areas, survival rates were compared among 2,322 White and 536 Black female residents with a diagnosis of primary breast cancer between January 1978 and December 1982 and followed through the end of 1983 [Bain 1986]. Black women in this study group were more likely to be diagnosed at an advanced stage and were less likely to receive surgical

treatment. However, even when the type of surgery and stage of disease were controlled in the analysis, racial/ethnic group remained as a significant prognostic indicator for survival. Similar studies conducted by the SEER registries in San Francisco/Oakland [Ragland 1991] and metropolitan Detroit [Simon 1996] reported that Black female breast cancer survival was poorer than that of White females at each stage of disease. Racial/ethnic differences were greatest for regional disease.

A study of Hispanic and non-Hispanic Whites residing in New Mexico and American Indians residing in New Mexico and Arizona compared survival rates in these groups for incident cancer cases diagnosed from 1969 through 1982 [Samet 1987]. American Indians were found to have significantly poorer one-year and five-year survival after a breast cancer diagnosis than non-Hispanic Whites, even after adjustment for stage and treatment. Hispanic Whites, initially showed lower survival than non-Hispanic Whites, but this difference disappeared after adjustment for stage and treatment.

In the studies of racial/ethnic differences in Hawaii, cited earlier, adjustment for stage of disease at diagnosis reduced breast cancer survival differences among Japanese, Chinese and White women to statistically non-significant levels [Le Marchand 1984, Meng 1997]. Five-year survival rates among Filipino and native Hawaiian women remained lower than the other groups, but were reduced after adjustment for stage. Similar results were found when follow-up for one of the study groups was extended to ten years [Le Marchand 1985]. Among cases diagnosed with localized disease, Filipino women in Hawaii had nearly a three-fold greater risk of dying within five years, while White women and Hawaiian women had an almost two-fold higher risk of dying than Japanese women [Meng 1997]. For advanced disease, defined as regional or distant

stage, Hawaiian women had a two-fold higher risk of dying than Japanese women. The combination of stage at diagnosis and marital status explained about 45% of the racial/ethnic differences in survival in their study group [Meng 1997]. Married patients in their study had the longest survival, a finding that has been previously reported among other cancer patients [Goodwin 1987]. Goodwin et al. studied over 27,000 epithelial cancers diagnosed from January 1969 through December 1982 among residents of the state of New Mexico and found that unmarried persons were more likely to be untreated for their cancer. After adjustment for stage distribution and treatment, unmarried persons still had poorer survival.

Breast cancer survival was recently examined by race/ethnicity and other factors in a review of records maintained by the Department of Defense Central Tumor Registry [Wojcik 1998]. The study group included 698 Black women and 6,577 White women diagnosed with breast cancer between 1975 and 1994 and treated in the U.S. military equal-access medical care system. After adjustment for age at diagnosis, the risk of death was 1.45 times greater for Black women than for White women. The risk only declined to 1.41 after adjustment for stage of disease at diagnosis and remained statistically significant. Additional covariates included waiting time between diagnosis and first treatment, marital status, alcohol usage, tobacco usage, and family history of cancer; but further adjustment for these factors had no effect on their findings. The authors concluded that potential differences in tumor biology, socioeconomic status, or sociocultural factors may be contributing to the survival differences they noted.

2. Survival Differences by Socioeconomic Position

In several studies, socioeconomic position has been found to partially or entirely explain racial/ethnic survival differences after other prognostic factors, such as stage of disease and age at diagnosis, have been considered [Dayal 1982, Bassett 1986, Cella 1991, Gordon 1992, Eley 1994, Greenwald 1996]. Women in lower socioeconomic groups tend to have poorer survival rates. Since race/ethnicity is usually confounded with socioeconomic position in analytic studies, it is important to examine the influence of both factors. Frequently, race/ethnicity acts as a surrogate marker for socioeconomic position in studies of risk factors [Gordon 1995], though its use in this manner is imprecise and potentially misleading [Harvard 1996].

In a study of survival patterns among breast cancer patients seen at the Medical College of Virginia between 1968 and 1977, socioeconomic information on the census tract of residence was available for a subset of the study group (117 White and 206 Black women) [Dayal 1982]. Each of the six socioeconomic indicators used in the study had a significant association with survival time. Age and stage at diagnosis did not explain survival differences between the two groups, but adjustment for socioeconomic position reduced the racial/ethnic disparity to a statistically non-significant level. In contrast to population-based studies, however, Dayal et al found that Black women in their study presented at an earlier stage than White women. This may be the result of selection bias with respect to the types of patients being treated at the study hospital.

In a larger study using a cancer surveillance system covering 13 counties in western Washington state, socioeconomic data for census block groups (subunits of

census tracts) was used to characterize the socioeconomic level of women diagnosed with breast cancer between January 1973 and December 1983 [Bassett 1986]. Survival patterns among 251 Black women and 1,255 White women were examined using a Cox regression model to adjust for Black-White differences in age, broad categories of stage (metastatic, non-metastatic), and tumor histology (ductal, lobular, other type). Black mortality was about 1.4 times that of Whites after adjustment for these factors. Following additional adjustment for socioeconomic level, Black mortality was only 1.1 times that of Whites (95% CI: 0.8, 1.5). In both groups, lower socioeconomic level was a strong predictor of shortened survival. The investigators suggested that studies of Black/White differences in breast cancer survival may be incomplete and potentially misleading if they do not jointly consider the role of socioeconomic position.

Another study examined survival data on patients diagnosed between 1977 and 1983 with one of six types of cancer and entered into the treatment protocols of a cooperative clinical trials group which included institutions in the United States and Canada [Cella 1991]. A strength of this study was that cancer patients admitted to the trials received the specified treatment regardless of income or insurance status. Race/ethnicity (White vs. Black) was not a significant predictor of survival time when data were adjusted for differences in general health status at entry, age, and protocol-specific prognostic factors (estrogen receptor status for breast cancer patients). Income and education, however, were important factors. Patients with lower annual incomes and those with lower educational level experienced significantly shorter survival times than those with higher income or education.

In a larger multi-center clinical trial of stage I and stage II breast cancer patients

based at Case Western Reserve University, those having less education and lower incomes were also found to have poorer disease-free survival and overall survival [Gordon 1992]. These differences remained after adjustment for estrogen receptor status, number of positive lymph nodes, and tumor size. Racial/ethnic group (White, Black) was not a significant determinant of survival once adjustment was made for socioeconomic status.

Another hospital-based study evaluated the importance of socioeconomic status and race/ethnicity in cancer survival by pooling information from 22 Comprehensive Cancer Centers in the United States [Greenwald 1996]. This data base, called the Centralized Cancer Patient Data System, included 6,896 breast cancer cases diagnosed between July 1977 and October 1981. Results from a Cox proportional hazards model which included age at diagnosis, racial/ethnic group (Black, White), and socioeconomic status (percentage of high school graduates in the postal code areas where patients resided), indicated that socioeconomic status and race/ethnicity were independent predictors of survival. A significant weakness of this study, however, was the lack of information on tumor characteristics or stage of disease.

Noting the well-documented disparity in cancer survival between Blacks and Whites, the National Cancer Institute, in 1983, planned and funded the Black/White Cancer Survival Study. Breast cancer survival differences were examined among 612 Black and 518 White women diagnosed in 1985 and 1986 in one of three population-based registry systems in Atlanta, GA, New Orleans, LA and San Francisco/Oakland, CA [Eley 1994]. Multivariable modeling using Cox proportional hazards regression were used to estimate the hazard ratio for Blacks compared to Whites, adjusting for stage of

disease, tumor characteristics (positive lymph nodes, histologic subtype, pathological grade, estrogen receptor status), treatment type, comorbid conditions, and sociodemographic factors (e.g., marital status, occupation, usual source of health care, health insurance status, an index of poverty). After controlling for geographic area of residence and age in the analysis, the risk of dying was 2.2 times (95% CI: 1.8, 2.8) greater for Blacks than Whites. Adjustment for stage of disease reduced the risk to 1.7 (95% CI: 1.4, 2.2) and further adjustment for tumor pathology, treatment, comorbidities, and sociodemographic variables resulted in a hazard ratio comparing Blacks to Whites of 1.3 (95% CI: 1.0, 1.8). Their results were similar, whether analyzing all-cause mortality or breast cancer-specific mortality. The authors concluded that about 40% of the racial/ethnic difference in survival was explained by more advanced stage of disease among Blacks, another 15% by histologic and pathologic differences, and a further 18% by the amount of comorbid illness and sociodemographic factors. They recommended that future efforts to reduce racial/ethnic differences in survival be aimed at early recognition of disease by means of community education, improved access to primary care and mammography, and increased compliance with screening recommendations.

3. Biological Tumor Markers, Race/Ethnicity, and Socioeconomic Position

Several studies have documented the pattern of a poorer breast cancer clinical stage distribution among persons in lower socioeconomic groups [Ownby 1985, Polednak 1986, Farley 1989, Mandelblatt 1991, Wells 1992, Weiss 1995, Bentley 1998,

Lannin 1998]. Since it is unclear, however, whether socioeconomic factors can entirely account for the racial/ethnic differences in breast cancer survival, potential differences in biological characteristics of the tumors have been studied by a number of investigators. Biological markers for breast tumors may have prognostic value: providing information on the expected clinical outcome of the malignancy; and/or have predictive value: indicating those patients likely to benefit from adjuvant systemic chemo- or hormonal therapy [Von Kleist 1996]. Studies of prognostic factors measure biological characteristics inherent to the breast tumor, such as tumor cell proliferation, tumor aggressiveness, and its potential for metastasis [Von Kleist 1996, Ferno 1998].

Several available tumor marker tests were recently evaluated for their utility in the prevention, screening, treatment and surveillance of breast cancers by a Tumor Marker Expert Panel convened by the American Society of Clinical Oncology. The Panel developed a set of clinical practice guidelines, based on their review of the published studies [ASCO 1996]. They determined that the receptor proteins for estrogen and progesterone should be measured on every primary breast cancer specimen. The Panel further concluded that the data were insufficient to recommend routine use of the other markers they considered in their review, namely: carcinoembryonic antigen (CEA), cancer antigen (CA 15-3), proliferative markers (DNA index or S-phase fraction), a marker of tumor invasion (cathepsin-D), a proto-oncogene (HER-2/*neu*), and a tumor suppressor gene (p53).

Racial/ethnic differences in the distribution of estrogen receptor (ER) status among breast cancer patients have been documented in several studies [Mohla 1982, Hulka 1984, Ownby 1985, Pegararo 1986, Beverly 1987, Stanford 1987, Stanford 1989,

Chen 1994, Ellege 1994, Gordon 1995, Gapstur 1996, Elmore 1998]. Most, but not all, of these studies were hospital-based or included only patients in clinical trials, so their findings may not be generalizable. Few adjusted for potential confounding variables such as age and socioeconomic position.

Information concerning the distribution of ER and PR by race/ethnicity and socioeconomic position is limited. Findings from three recent studies examining the relationship between Black/White differences in hormone receptor status and socioeconomic position are conflicting [Chen 1994, Gordon 1995, Elmore 1998]. In the cross-sectional study by Chen et al., data on tumor characteristics and socioeconomic variables were collected from medical records and in-person interviews with patients diagnosed in 1985 and 1986. Study subjects (n=506 Black and 457 White women) were identified from population-based cancer registries in three urban areas (Atlanta, New Orleans, San Francisco-Oakland). Black women in this study were more likely than Whites to have tumors that were ER-negative, poorly differentiated, with increased nuclear atypia, and more necrosis. With the exception of ER status, these associations remained statistically significant after controlling for age, geographic area, socioeconomic status, body mass index, use of alcohol and tobacco, reproductive experience, and health care access and utilization. Since the social and lifestyle factors of the study group did not entirely explain racial/ethnic differences in tumor characteristics associated with a poor prognosis, Chen et al. concluded that biological reasons for the racial/ethnic differences must be further explored.

The study by Gordon was based on newly-diagnosed breast cancer patients from northeastern Ohio who participated in one of three clinical trials during two time periods:

1974 to 1985 (n=164 Black and 723 White women) and 1986 to mid-1992 (n=167 Black and 437 White women). Since socioeconomic information was not obtained from individuals for this study, surrogate measures were used based on characteristics of the census tract of residence at the time of diagnosis. Gordon found that ER-negative tumors were associated with low socioeconomic level after controlling for race/ethnicity, age, and other patient characteristics. This relationship held for each of the time periods. Gordon concluded that the poorer prognosis of lower socioeconomic women might be explained by their less favorable ER status.

The hospital-based study by Elmore et al. included 100 Black and 300 White patients diagnosed with breast cancer at the Yale-New Haven Hospital from January 1985 through December 1993. Clinical and sociodemographic information was collected from each patient. In contrast to earlier studies, no racial/ethnic difference was noted for ER status in this study population. Black patients had increased age-adjusted odds ratios for several tumor characteristics that have been associated with a worse prognosis including, higher stage of disease, larger tumor size, positive lymph nodes, presence of necrosis, vascular/lymphatic invasion, and negative PR status. Further adjustment for income, medical insurance status, and method of detection, however, reduced the observed associations and only tumor size and necrosis remained statistically significant. The investigators concluded that the majority of histologic features of breast cancer measured in this study did not differ between Black and White patients. The significant differences in tumor size and necrosis suggested that a true biologic difference may exist, but confirmation is needed in larger studies.

Other investigators have reported no significant racial/ethnic differences [Weiss

1995, Krieger 1997a] or socioeconomic differences [Krieger 1997a] in the distribution of hormone receptors and other molecular biomarkers (e.g., oncogenes, cytoplasmic proteins, markers of cell growth) among breast cancer patients. The population sizes in these studies were too small, however, to draw reliable conclusions (Krieger: n=44 Black, 44 White, 43 Asian women; Weiss: n=32 Black, 172 White, 49 Hispanic women).

CHAPTER II. MATERIALS AND METHODS

A. General Study Design

A population-based, case-control design was chosen to evaluate the importance of socioeconomic position in explaining racial/ethnic differences in tumor characteristics among women newly diagnosed with invasive, primary breast cancer during the years 1992 through 1996 in any of the eleven reporting areas comprising the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute (NCI). Specific tumor characteristics at the time of diagnosis including stage, size, grade, estrogen receptor and progesterone receptor status, and a limited set of demographic variables were collected for individual study subjects. Socioeconomic variables were extracted from the 1990 decennial census data file and linked to individual patient records to provide census tract-level information.

The outcome variables in this study (tumor stage, tumor size) were coded as dichotomous variables. The analysis included descriptive tabulations of the study variables by racial/ethnic group and preliminary two-way comparisons between selected explanatory variables and outcome variables. Logistic regression models were used to estimate the relative importance of race/ethnicity and various socioeconomic measures in explaining the stage of disease and tumor size at the time of diagnosis.

B. Study Aims

The specific aims of this study were:

- To describe the racial/ethnic distribution of selected demographic, socioeconomic, and tumor characteristics (stage of disease, tumor size, tumor grade, estrogen/progesterone receptor status) that influence prognosis for cancer of the female breast.

- To assess the importance of sociodemographic factors in explaining racial/ethnic differences in tumor stage and size at the time of diagnosis.

C. Feasibility Assessment and Geocoding Improvements

Prior to conducting the study, I completed the following preliminary activities to evaluate the feasibility of the project and to improve the quality of the data:

- I evaluated the completeness and accuracy of geocoded information on residence at the time of diagnosis for cancer patients in the SEER program cancer data base. An edit check for valid census tract codes had never previously been conducted on the entire data base and was necessary to determine the feasibility of linking selected socioeconomic variables from the 1990 decennial census, at the census tract-level, with individual cancer records from all of the SEER reporting areas.

The percentage of valid census tract codes found in this edit ranged from 60% to 95% by registry. I reported these results to each of the 11 SEER registries.

- I created electronic files containing valid 1990 census tract codes and sent them to each of the registries. The registry staff then edited and recoded the census tract fields for their cancer patients and sent corrected data files to the NCI.

- I developed a survey form, with assistance from other NCI staff, to collect information from each of the SEER registries on their current geocoding procedures, any problems they encounter, and the associated costs (see data collection instrument, **Appendix II-1a-c**). I summarized the results of this survey and presented them at a special meeting with all of the registry data managers (see below).

- I planned and chaired a special section of the annual SEER data managers' meeting in Bethesda, MD in October 1998. The purpose was to exchange information on geocoding procedures and to find ways to improve the completeness and accuracy of geocoding in each of the registries.

- On the basis of what I found from the special data edits and learned from the data managers' meeting, I drafted new data reporting requirements aimed at improving the collection and geocoding of residence information. These requirements were incorporated as revisions in the SEER coding manual.

- I developed a new, global reporting rule that requires all SEER registries to provide a variable that indicates the completeness of address information used in assigning the census tract code. This variable will help data analysts to assess the validity of the census tract codes in future studies.

- The data collection changes described above are documented as revisions to the current SEER Program Code Manual **(Appendix II-2a-c)**.

As a result of these efforts, the percentage of cases that received a valid census tract code increased to 96%, overall.

D. Study Population

The targeted study population included all women with newly diagnosed primary breast cancers reported among women living in any of the eleven cancer registration areas in the SEER program of the National Cancer Institute during 1992 through 1996. The SEER registries were originally chosen for their ability to operate and maintain population-based cancer surveillance systems and for the characteristics and size of the population subgroups (e.g., racial/ethnic groups, urban/rural populations) within their reporting areas. The SEER geographic regions included in this study and the number of counties and census tracts they cover are identified in **Table II-1**. The locations include the States of Connecticut, Hawaii, Iowa, New Mexico, and Utah; and the metropolitan areas of Atlanta, Detroit, Los Angeles, San Francisco and Oakland, San Jose and

Monterey, and Seattle.

These areas cover about 14% of the total United States population, and include 78% of the Hawaiian population, 60% of the Japanese population, 49% of the Filipino population, 43% of the Chinese population, 34% of the Korean population, 31% of the Vietnamese population, 27% of the American Indian population, and 25% of the Hispanic population of this country. Selected demographic characteristics of the overall population covered by the eleven SEER registries are compared with those for the general United States population in **Figure II-1**. The population in the SEER coverage areas is similar to the general United States population with respect to the percentage of people living below the poverty level. The percentage of adults who graduated from high school is slightly higher in the SEER areas and a larger portion of the SEER population lives in urban areas. Finally, the percentage of foreign-born persons living in the SEER areas is nearly double that for the United States as a whole.

There were 126,400 women newly diagnosed with either in situ or invasive breast cancer among residents of the SEER coverage areas during the years 1992 through 1996 **(Figure II-2)**. Limiting the study to invasive cancers among the ten largest racial/ethnic groups results in a potential study group of 107,206 breast cancer patients among Hispanics; and non-Hispanic Whites, Blacks, American Indians in New Mexico, Chinese, Japanese, Filipinos, Hawaiians, Koreans, and Vietnamese. Cases that were identified only from an autopsy record or death certificate comprised less than one percent of the intended study population (n=599) and were excluded since they do not have useful information on tumor characteristics at the time of diagnosis. There were no notable differences between the excluded group and the study group with respect to racial/ethnic

category or registry.

The remaining study group (n=106,607) included over 84,000 invasive breast cancer cases among non-Hispanic White women, about 9,000 among non-Hispanic Black women, and over 7,000 among Hispanic women. There are over 1,800; 1,500; and 1,300 cases among non-Hispanic Japanese, Filipino, and Chinese women, respectively. Smaller numbers of cases occurred among native Hawaiian (n=508), Korean (n=301), Vietnamese (n=272), and American Indian (n=136) groups.

E. Evaluation of Sample Size

Sample size calculations were performed using Epi Info™ software version 6.04b [CDC 1994] to indicate the number of "unexposed" (White) and "exposed" (other specific racial/ethnic group) study subjects that would be required to detect a given range of odds ratios. The odds ratios reflect the odds of being in the "exposed" racial/ethnic group among "cases" (i.e., in this example, those with a diagnosis of distant stage disease) relative to that in the control group (i.e., localized stage disease). The specified level of power is 80% and the specified probability of making a Type I error, $\alpha = .05$. The expected proportion of the "unexposed" (White) group with "disease" (distant stage cancer) is 0.054, based on a preliminary examination of the data. Since additional planned case/control comparisons (e.g., regional vs. localized disease; and large tumor vs. small tumor) include larger numbers of study subjects than the distant vs. local comparison, these results represent a conservative assessment of sample size and power requirements for this study.

The calculated sample sizes for each value of the odds ratio **(Table II-2a,b)** may be compared to the available number of study subjects in the various racial/ethnic groups. This comparison indicates that there are sufficient numbers of Black and Hispanic study subjects to detect elevated odds ratios as low as 1.2 and reduced odds ratios as high as 0.8 at 80% power and an alpha level of .05. Among Japanese, Filipino, and Chinese women there are sufficient study subjects to detect elevated odds ratios as low as 1.4 and reduced odds ratios as high as 0.7 for Japanese and 0.6 for Filipino and Chinese. The number of study subjects available among Hawaiian, Korean and Vietnamese women will allow the detection of more moderate odds ratios (elevated OR as low as 1.7 for Hawaiian, 1.9 for Korean and 2.0 for Vietnamese; reduced OR as high as 0.3 for all groups). Among American Indian women, odds ratios equal to or greater than 2.4 or lower than 0.1 will be detectable with the same alpha level and power.

F. Data Linkage

Several variables describing the tumor and basic patient demographics were available for each breast cancer study subject **(Table II-3)**. Additional demographic variables relating to socioeconomic position are available from the 1990 decennial census [Census 1992] for small geographic areas (census tracts or block numbering areas) covering the SEER areas from which the study subjects are drawn. The census variables chosen for this study **(Table II-3)** were selected following a review of the published literature on the measurement of associations between socioeconomic position and health outcomes [Last 1987, Filkati 1995, Patrick 1995, Krieger 1997b].

The socioeconomic variables were linked to individual cases on the basis of their residence at the time of their cancer diagnosis. This linkage has never been attempted on the entire SEER program data base and is a unique feature of this study. Previous studies have been limited to one or a few of the registries located in metropolitan areas because of the lack of defined census tracts for many areas of the country in previous censuses and because of the poor quality of residence address information in rural areas.

The success of this linkage depended upon the completeness and accuracy of address information collected on cancer patients, thereby enabling the assignment of geocodes (i.e., census tract or block numbering area code) to individual records. These geocodes were then used to link patient records with socioeconomic information for census tracts or block numbering areas from the 1990 decennial census. The census tract data field, although collected since the beginning of the SEER program, has not undergone rigorous data editing prior to this study. Preliminary computer edits I conducted on the SEER data file indicated poor coding of census tract number by several of the SEER registry areas, with the percentage of valid codes ranging from 60% to 95% of cases. Most of the urban SEER areas had higher percentages of valid census tract codes than did registries covering entire states.

After I reported the edit results to each of the registries they reviewed their incorrectly coded cases and were able to assign new, valid geocodes to some of the study subjects. The percentage of valid census tract codes improved to 80% to 95% in their next data submission. This was still not sufficient, however, for the purposes of my proposed study. To help data managers verify the accuracy of their geocoding, I provided each of them with tabular results from the edits I had performed on their data

submission and with electronic files containing allowable county/census tract codes for their coverage areas. From this effort, I learned that many of the registries were providing outdated (1980) census tract codes instead of the current (1990) codes. I also learned that one of the registries was routinely failing to geocode cases from several of their counties due to a mistaken belief that census tracts had not yet been defined for the counties.

To obtain detailed information about the geocoding procedures, associated costs, and problems currently experienced by each of the registries, I developed a *Geocoding Update Instrument*, with assistance from other NCI staff **(Appendix II-1a-c)**. I summarized the results from this survey and presented them at a meeting I convened at the NCI in October 1998 that was attended by all of the SEER Registry Managers. The meeting facilitated an exchange of information between the registries and focused further attention on the need to improve the completeness and accuracy of geocoding.

Data managers from a largely rural state registry and an urban registry gave detailed presentations on the various techniques and data sources they use for geocoding. Issues of particular interest included the strengths and weaknesses of automated geocoding software and the use of rural route numbers, post office boxes, zip code centroids (for 5-digit, 7-digit and 9-digit zip codes), American Indian community codes, census tract maps, topologically integrated geographic encoding and referencing digital mapping system (i.e., TIGER files), crisscross directories, voting records, motor vehicle administration records and other sources to obtain necessary address and geocode information. During the meeting, all registry directors and managers became acquainted with a variety of available geocoding techniques and identified the strengths and

limitations of the methods. When complete street address and zip code information is available, the geocoding software is quick, relatively inexpensive and accurate. A further advantage is that the census block group (a subunit of the census tract) can be obtained from the software in addition to the census tract. Block group coding may not be feasible for cases with incomplete addresses that require manual geocoding.

Another important outcome from the meeting was the addition of a new variable to the SEER data base which I developed to indicate the level of certainty and the completeness of address information used to assign a geocode to each cancer case (e.g., high certainty = complete residence address available; low certainty = only rural route number or post office box and zip code available). This information was not available for subjects included in this study, but will be reported for all new cancer cases diagnosed on January 1, 1998 and thereafter. The new coding scheme and revisions to the SEER Program Code Manual are reproduced in **Appendix II-2a-c**.

As a result of the knowledge gained from this meeting additional efforts were undertaken by each of the registries to improve the geocoding of their cancer case records. In the next data file submission from the registries, the overall percentage of study subjects with valid census tract codes reached 96 percent. This seemed sufficient to conduct the proposed study and was used as the final data analysis file.

G. Variable Specification

1. Individual-level Variables

Cancer type

This study includes women newly diagnosed, between January 1, 1992 and December 31, 1996, with an invasive cancer of the breast. This includes codes C50.0 through C50.9 in the scheme of the International Classification of Diseases for Oncology, 2nd Edition [ICDO-2 1990].

Tumor stage

Descriptive information on the extent of disease at the time of diagnosis was collected on all cancer cases. This information is based on a combination of clinical, operative, and pathological assessments. If a discrepancy appears between pathology and operative reports concerning excised tissue, priority is given to the pathology report. The priority for using information to code the extent of disease is 1) pathologic, 2) operative and 3) clinical findings. The major components of the extent of disease are size of the tumor, extension of the tumor, evidence of metastasis, and lymph node involvement. This information allows the data to be collapsed into different staging schemes and provides flexibility in maintaining consistency over time, even if a staging scheme changes [Fritz 1998].

Cancer staging is a method for grouping patients based on the extent of the spread of the cancer from its site of origin. Detecting cancers at an early, more treatable stage is

a major goal of prevention and control efforts. Knowledge of the stage of disease at the time of diagnosis is essential for determining the choice of therapy and in assessing prognosis. Tumor stage is the strongest measure of the behavior of invasive breast cancer and forms the basis of prognostication [Simpson 1996]. The localized-regional-distant summary staging scheme has been found useful over the years for descriptive and statistical analysis of tumor registry data and is defined below [Seiffert 1993].

Localized: An invasive malignant neoplasm confined entirely to the organ of origin with no lymph node involvement.

Regional: A malignant neoplasm that 1) has extended beyond the limits of the organ of origin directly into surrounding organs or tissues; or 2) involves regional lymph nodes by way of the lymphatic system; or 3) has both regional extension and involvement of regional lymph nodes.

Distant: A malignant neoplasm that has spread to parts of the body remote from the primary tumor either by direct extension or by discontinuous metastasis (e.g., implantation or seeding) to distant organs, tissues, or via the lymphatic system to distant lymph nodes.

Unstaged: Information is not sufficient to assign a stage.

For this study, one of the main outcomes of interest was to differentiate breast

cancers diagnosed early enough in their natural history so as to afford a meaningful survival advantage. Comparisons of distant stage disease with localized disease and regional stage with localized stage are used because of the large difference in relative survival rates between these groups **(Table I-1)**. In addition, studies of the increased use of mammography in early detection and screening have demonstrated increases in the detection of localized lesions [Thomas 1977]. Thus, differences in the relative frequency of localized tumors versus more advanced stage tumors among different groups of individuals may reflect different levels of medical surveillance.

Tumor size

Tumor size was recorded in millimeters and refers to the exact size of the primary tumor at its largest dimension. If the patient has been pretreated with neoadjuvant chemotherapy, hormonal therapy, immunotherapy or radiation therapy, tumor size is not coded unless it was measured prior to the initiation of these therapies. In breast cancer, the size of the invasive component of the primary tumor reflects its natural history, its metastatic capacity, and is an independent predictor of survival [Simpson 1996]. The localized-regional-distant summary staging scheme does not explicitly use the size of the tumor in assigning a stage. Therefore, tumor size was also evaluated in relation to racial/ethnic and socioeconomic factors. Tumors 1.0 cm or less in diameter have an especially low risk of recurrence. Several studies have reported 5-year or 10-year disease-free survival exceeding 90 percent for node-negative patients with tumors 1.0 cm or less in diameter [O'Reilly 1990, Merkel 1993, Rosen 1993]. In addition, tumors smaller than 1.0 cm are more difficult to detect by clinical breast exam [Fletcher 1985,

Reintgen 1993, Helzlsouer 1995] and their increased identification in particular subgroups of the study population may reflect differential patterns of mammography screening. In the data analysis, patients with tumors 1.0 cm in diameter or greater at the time of diagnosis are compared with those diagnosed at smaller sizes.

Hormone receptor status

Results of testing for estrogen receptor and progesterone receptor were obtained from medical records and coded as positive, negative, borderline or unknown. This variable is included as an indicator of the tumor biology. Tumors that are positive for hormone receptors tend to be correlated with positive prognostic features such as a lower rate of cell proliferation and evidence of tumor cell differentiation.

Tumor grade

Tumors were classified into one of four grades or unknown. Grade 1 tumors are those considered to be well-differentiated (and the least aggressive); grade 2 corresponds to moderately differentiated; grade 3 tumors are poorly differentiated; and grade 4 includes undifferentiated tumors (or highly aggressive).

Age at diagnosis

The age of the patient at the time of their cancer diagnosis was available for all study subjects and was measured in completed years of life. It was included as a continuous variable in the analysis.

Race and ethnicity

Consistent with Office of Management and Budget federal data standards, race and ethnicity were treated as two independent variables [OMB 1978]. Race was coded into one of 36 categories on the basis of information in the medical records [Fritz 1998]. If a person's race was recorded in the medical record as a combination of White and any other specific race, they are routinely coded by SEER registries to the other specific race. Ethnicity is used in the SEER data base to denote persons of Hispanic (or Latino) origin. This group includes Spanish, Mexican, Puerto Rican, Cuban, and South or Central American (except Brazil). Persons of Hispanic origin may be of any race. Since information on the specific subgroup was available for less than half of the Hispanic cases, this group was analyzed in total. Other racial/ethnic groups were analyzed after Hispanics were removed, so that each group was mutually exclusive. A design variable with 10 levels was used to classify the ten racial/ethnic groups in the study.

Marital status at the time of cancer diagnosis

Many studies over the past century have shown that married individuals tend to be healthier and to live longer than non-married individuals [Last 1987, Smith 1997]. The mechanisms responsible for this association are hypothesized to relate to greater social and economic support and more healthy lifestyles among the married [Goodwin 1987, Umberson 1987, Corin 1995], as well as the possibility that healthier persons are more likely to be selected into marriage [Goldman 1993]. Marital status codes included single (never married), married (including common law), separated, divorced, widowed, and unknown. These categories were collapsed to not married (i.e., single, separated,

divorced, or widowed) and all other (i.e., married and unknown marital status) in the analysis. The unknowns likely included a mixture of married and not married persons, so grouping them with the married category probably diluted the effect of this variable somewhat.

SEER area

The study subjects were diagnosed in any of the geographic areas (State or cluster of contiguous counties) covered by the eleven SEER cancer registries. Since the SEER registries are located in different regions of the country which may have different patterns of medical practice or cancer risk factors, design variables representing the SEER areas were included in the analysis to control for potential confounding.

Census tract and county of residence at the time of diagnosis

The residence address of each cancer case at the time of diagnosis was used by registry personnel to assign a census tract or block numbering area and county code. Census tracts or block numbering areas are the smallest geographic areas currently recorded by SEER. They represent statistical subdivisions of a county and are established and maintained by local committees. Census tracts usually contain between 2,500 and 8,000 people, and average about 4,000 people. The geographic size of a census tract varies, therefore, depending on how densely an area is settled.

Census tracts are designed to be homogeneous with respect to population characteristics, economic status and living conditions at the time they are created [Census 1993]. At the time of the 1980 decennial census, not all counties in the U.S. were

covered by census tracts. The remaining untracted counties were subdivided into either census tracts or block numbering areas, however, by the time of the 1990 census. Block numbering areas are mutually exclusive of census tracts and are generally used in sparsely populated counties. As a result, the population size of a block numbering area is typically smaller than that of a census tract and, in some instances, a very thinly populated county may be covered by a single block numbering area. There are over 7,900 census tracts or block numbering areas within the geographic regions covered by the SEER Program **(Table II-1)**.

2. Census Tract-level Variables

The utility of community-level socioeconomic variables has been shown in studies assessing the impact of socioeconomic position on hospital admissions [Hofer 1998] and on selected health outcomes [Krieger 1992, Anderson 1997, Robert 1998]. Some investigators conclude that neighborhood-based measures of socioeconomic position merit greater use in public health research and surveillance because they characterize aspects of a person's living conditions that may not be evident from individual-level measures, particularly when studying diverse racial/ethnic groups [Kaplan 1996, Krieger 1997b]. For example, individuals coded as 'White' at each socioeconomic level may be more likely to live in more affluent, safer, and less polluted neighborhoods than individuals coded as 'non-White' [Massey 1990]. Neighborhood-based measures have the advantage of applicability across all age groups and both sexes. They also tend to provide a more stable estimate of the relevant economic situation of

individuals than do some of the more volatile individual measures such as personal income. Even when individual-level data are available, neighborhood-level measures enable the conduct of contextual analyses to determine how socioeconomic factors at multiple levels shape population patterns of health and disease [Krieger 1992, Anderson 1997, Krieger 1997b].

Working-class job

There are no census-derived data which explicitly measure socioeconomic level. Many socioeconomic measures are based on an occupational classification, however, since occupation is considered to be a reliable indicator of relative standing in industrial societies [Liberatos 1988]. Census occupational data can be used to create a measure of neighborhood socioeconomic level by selectively combining the census-defined occupational categories into a group that predominantly contains people in "working class" jobs **(Table II-3)**. This group largely consists of employees who do not own their own workplace, are not self-employed, and generally occupy subordinate positions at work [Wright 1982, Krieger 1992]. This scheme has been validated through comparisons with individual-level measures of social standing [Krieger 1991, Krieger 1992] and has been reported to be associated with breast cancer incidence and survival [Bassett 1986, Krieger 1990], prevalence of sexually transmitted diseases [Ellen 1995] and smoking status, parity, height and hypertension [Krieger 1991, Krieger 1992]. In the present study, this classification scheme was used to characterize census tracts by the percentage of employed persons in the tract, aged 16 and over, that are in "working-class" occupations. The census tract value for this variable (and all subsequent variables

in this section) was linked to individual study subjects on the basis of their residence at the time of their cancer diagnosis.

Income

Income derives from a variety of sources including wage earnings, interest, dividends, child support, alimony, transfer payments, and pensions. It has been found to be strongly associated with outcomes ranging from self-perceived health [DHHS 1991] to mortality [Backlund 1996]. Neighborhood-level gradients in income have also been linked to mortality [Smith 1996a, Smith 1996b], cancer incidence and survival [Devesa 1983, Greenwald 1996], and use of health services [Cherkin 1992]. A problem with the use of an income variable that is collected only for one point in time, is that it may fail to capture important information about income fluctuations. Another weakness is that, unlike the poverty variable described below, the family or household income variables are not adjusted for the number of persons supported by the income and will therefore have different meanings for different size households. The census tract-level measure of median family income was used in the present study.

Poverty

The poverty threshold set by the Bureau of the Census is an economic indicator of need. Unlike measures of median family income, poverty status takes into account the size and age structure of a family and is related to the ability to purchase a specific market basket of food [Census 1992]. In 1989, the average poverty threshold for a family of four persons was $12,674. Poverty thresholds are applied on a national basis,

without adjustment for regional, State or local variations in the cost of living. The poverty status variable represents the percentage of all persons in a census tract who are living below the poverty threshold for their given family size. Federally defined poverty areas are those in which 20% or more of the population lives below the poverty line [Census 1985]. This definition of poverty area, applied to census block group data, has been associated with several health outcomes [Krieger 1990, Krieger 1991, Krieger 1992, Ellen 1995]. Another poverty variable included for the analysis indicates the percentage of families headed by women with no husband at home, with one or more children, and who are living below the poverty level. It is possible that this variable captures additional factors, such as increased time demands or stress, that may help to explain patterns of health care utilization. The utilization patterns may, in turn, influence the severity of disease at the time of diagnosis.

Wealth

Privilege and wealth represent the opposite end of the socioeconomic spectrum from deprivation and poverty. Wealth encompasses accumulated assets, usually obtained through inheritance, investment or other forms of saving [Krieger 1997b]. Homes and cars represent the most commonly owned assets in the United States and information on ownership can usually be obtained easily and reliably. European studies have reported associations between car and home ownership and mortality rates [Filakti 1995] and cancer survival [Petridou 1994]. The percentage of households in a census tract that own their home and the percentage that do not own a car were calculated from the 1990 census data.

Education

Education is another widely used indicator of socioeconomic position in public health research and has been shown to be an important predictor of mortality and morbidity [Feldman 1989, Reis 1991, Heck 1997, Krieger 1997b]. The amount of education and knowledge attained by an individual influences lifestyle behaviors (e.g., exercise, diet) and may also provide qualifications for certain occupations and income [Liberatos 1988]. Its advantages include ease of measurement; relevance for persons not actively employed (e.g., house parent, unemployed, retired); and its stability over the adult lifespan, regardless of changes in health status. Education was selected as a practical measure for socioeconomic position in the 1989 revision of the U.S. standard death certificate [Tolson 1991]. Some investigators suggest that it is more meaningful to measure education in terms of certification or academic degrees achieved, rather than by the number of years of schooling, because the academic credentials have important implications for employment prospects [Faia 1981, Liberatos 1988]. When under-educated areas are defined as census tracts in which 25% or more of adults age 25 and older have not completed high school, associations between this ecological measure and selected health characteristics were found to be similar to associations based upon education data for individuals [Krieger 1992]. The percentage of persons aged 25 years and older who did not have at least a high school diploma and the percentage who had a bachelor's degree or higher was calculated for each census tract in the study and applied to all study subjects within each of the tracts.

Urban residence

Urban/rural designations are one of the most commonly used ecological variables in health research, however, the effect that an urban environment may be expected to have on health is unclear. Most urban environments have positive and negative qualities, and these qualities are not experienced equally by all residents [Verheij 1996]. Besides differences in exposure to physical risk factors (e.g., noise, pollution) or access to health care, people's values and attitudes about health may differ in urban versus rural areas. This can lead to differences in health behaviors (e.g., diet, exercise, smoking, care-seeking) and ultimately health status [Patrick 1995]. The percentage of the population in a census tract living in an urban area was calculated from the 1990 census data. About 77% of the study subjects were classified as living in a census tract that is considered to be 100% urban. Because of the extreme skewness of the study data, this variable was coded as a binary variable with census tracts classified as urban (100% of the population lives in an urban area) or not urban (<100% of the population lives in an urban area).

Unemployment

Most individuals in the United States obtain health insurance through an employer. The possession of health insurance, in turn, influences access to health care, including preventive care. Areas with high unemployment among persons in the labor force may therefore be related to tumor characteristics that influence prognosis. The percentage of the population that was unemployed among those in the labor force was calculated from the 1990 census data. Excluded from the labor force were: 1) persons under 16 years of age, and 2) among those over age 16 - students, housewives, retired

workers, seasonal workers enumerated in an "off" season who were not looking for work, institutionalized persons, and persons doing only incidental unpaid family work [Census 1992].

Foreign-born

The relationship between migration and health may vary by socioeconomic position and by the reason for migrating (e.g., political, economic). Migrants usually adopt at least some of the cultural characteristics of the community into which they move, although this may take as much time as a generation or more [Last 1987]. Affects on health may be mediated through changes in diet and levels of stress. The percentage of the population in each census tract that was born outside of the United States is calculated from the 1990 census data.

H. Data Quality

All data on cancer cases were collected by specially-trained medical records abstractors following well-documented, standardized procedures [Fritz 1998]. Collected data have passed extensive field and central office quality control edits. Overall completeness of case reporting by the SEER cancer registries has been measured to be about 98%, based on independent audits of a stratified random sample of hospitals in six of the coverage areas [Zippin 1995]. Cancers of the cervix (*in situ*), melanoma, unknown primary, and leukemia were disproportionately represented among the missing cases, based upon the overall distribution of cancers reported by the registries [Zippin 1995].

Since there is wide variability in the classification of *in situ* cancers of the cervix, these tumors are no longer routinely collected by SEER registries. The impact of this change is to further increase the overall completeness of case reporting. Cancers of the breast are not among the specific types of cancer reported to be missing in excess of expectation.

Ninety-nine percent of the 106,607 breast cancer cases in the study population had evidence in the medical record of microscopic confirmation of the diagnosis. A reabstracting study of breast cancer diagnoses in 1992 [Zippin unpublished], based on a stratified random sample (n=1,100) of cases from hospital facilities in all of the SEER reporting areas, found discrepancies in extent of disease codes that resulted in changing the summary stage of disease category for 3.8% of the cases. Among these, six cases (0.5%) were misclassified as invasive tumors when further investigation indicated that they were *in situ* lesions. The results from this reabstracting study were used to develop a new set of abstracting and coding guidelines which serve as training materials in annual workshops for data managers and were adapted for use in data editing software.

Special revisions (described in Section II.D. Data Linkage) were made to data collection manuals and procedures and new data edits were implemented in preparation for conducting the proposed study of socioeconomic factors and racial/ethnic differences in tumor characteristics for cancer of the female breast. These efforts resulted in improved assignment of valid census tract codes to cancer patient records used in this study. This enabled a more complete linkage of census demographic data to study records and thereby reduced the potential affect of a reporting bias on study findings. Since these changes are now incorporated into the standard data collection and

management operations of the SEER registries and the NCI, they will result in a permanent improvement in the overall quality and utility of the SEER Program data base.

I. Characterization of Study Subjects Excluded from Analysis

1. Analysis by Stage of Disease

The distribution of study subjects (n=106,607) by racial/ethnic group and availability of information on stage of disease at the time of diagnosis is shown in **Table II-4**. Ninety-seven percent (n=103,371) of the study subjects in these ten racial/ethnic groups had sufficient information from medical records to assign a tumor stage. The percentage of unstaged cases is slightly higher among Blacks and Koreans than among the other specific groups. The percentage of cases with staging information was fairly consistent across age groups with the exception of women diagnosed at age 90 and over **(Table II-5)**. Although this group had the highest percentage of cases with missing tumor stage, they accounted for only 8% of all cases without staging data.

Socioeconomic information could not be linked to 3.7% of the staged cases. Most of these were due to incomplete or non-specific residence address information (e.g., missing house number, post office box number, rural route number) which prevented the assignment of a valid census tract code. The age distribution of those missing socioeconomic data was similar to those with complete information. The Hawaii cancer registration area had the largest percentage of cases that could not be linked to census tract socioeconomic information (14%), while all other areas had fewer than 10% of their

cases that could not be linked **(Table II-6)**. As a result of the poorer match rate in Hawaii, the percentage of native Hawaiians missing socioeconomic data was higher than that for other racial/ethnic groups **(Table II-7)**.

2. Analysis by Tumor Size

The distribution of study subjects (n=106,607) by racial/ethnic group and availability of information on tumor size at the time of diagnosis is shown in **Table II-8**. Tumor size was available from medical records for 91 percent (n=96,871) of the study group. A small number of the cases (0.1%) had a diagnosis of Paget's disease, which refers to neoplastic eczematous changes around the nipple. These cases were not associated with an underlying invasive tumor mass, and therefore, had no tumor size measurement. Black women had a higher percentage of cases lacking tumor size information than the other racial/ethnic groups. Availability of tumor size data was fairly consistent by age group, with the exception of women aged 90 years and older. The oldest age women had the highest percentage of cases with missing tumor size information, but this group accounted for only 3% of all cases lacking tumor size data **(Table II-9)**.

Socioeconomic information could not be linked to 3.7% of the cases with tumor size information. The age distribution of those missing socioeconomic data was similar to those with complete information. The Hawaii cancer registration area had the largest percentage of cases that could not be linked to census tract socioeconomic information (14%), while all other areas had fewer than 9% of their cases that could not be linked

(**Table II-10**). As a result of the poorer match rate in Hawaii, the percentage of native Hawaiians missing socioeconomic data was higher than that for other racial/ethnic groups (**Table II-11**).

J. Analytic Methods

As stated earlier, a population-based case-control design was chosen to investigate the importance of socioeconomic position in explaining racial/ethnic differences in tumor characteristics among women newly diagnosed with invasive primary breast cancer during the years 1992 through 1996 in one of the eleven reporting areas comprising the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute. The results of preliminary two-way comparisons between outcome variables (tumor stage, tumor size) and explanatory variables were expressed as unadjusted odds ratios with 95% confidence intervals. The odds of being a case for specific racial/ethnic groups were compared to that for White women, primarily because they were the largest group available for study. The odds ratios reflect the odds of being diagnosed with late stage disease (or larger tumor size) in the "exposed" group (e.g., Black, Hispanic, Japanese, etc.) relative to the odds for a similar diagnosis among White women.

Multiple regression models were developed to evaluate the importance of socioeconomic variables and other demographic factors in explaining racial/ethnic differences in tumor stage and size at the time of diagnosis. The SAS statistical program LOGISTIC Procedure [SAS Logistic 1989] was used since the outcome variables for the

regression analysis are binary. The outcome variables included: 1) distant stage breast cancer vs. localized stage; 2) regional stage disease vs. localized stage; and 3) primary breast tumor size >1.0 cm vs. ≤1.0 cm. Several investigators have reported a diminishing of the effect of socioeconomic factors on mortality with increasing age. To determine whether a similar relationship might hold in the present study, where the health outcome is severity of disease, terms for the cross-product of age at diagnosis and socioeconomic variables are considered for inclusion in the models as potential confounders [Sorlie 1992, Backlund 1996, Kaufman 1998]. Finally, the significance of interaction terms between racial/ethnic group and sociodemographic factors were evaluated in the models. Thus, six types of factors were examined in the models: (1) age at diagnosis, (2) geographic area, (3) sociodemographic factors, (4) tumor biology, (5) cross-product terms of socioeconomic factors and age at diagnosis considered as potential confounders, and (6) interaction terms between racial/ethnic group and socioeconomic factors.

Odds ratios associated with the explanatory variables were computed by exponentiating the estimated coefficients of the fitted logistic model and are presented with their 95% profile likelihood confidence limits. In some instances, it is of greater interest to present the change in the odds ratio for something larger than a one-unit change in the explanatory variable (e.g., socioeconomic variables with values ranging from 0 to 100). The odds ratio may be customized in these cases by multiplying the estimated coefficient by a constant c (where c represents a change of say, 10 or 20 units) and then exponentiating the product. The goal of the modeling was to determine the degree to which racial/ethnic differences in tumor stage and size could be explained by socioeconomic factors. Therefore, changes in the magnitude of the odds ratios, after

socioeconomic variables and other control variables are added to logistic regression models containing race/ethnicity, are used to assess the importance of these additional factors. Since the main purpose of the regression analysis was to assess confounding by socioeconomic factors, model-building strategies such as stepwise or best subsets were not appropriate. The logistic regression models were also used to calculate the adjusted proportion with late stage disease or larger tumor size among selected racial/ethnic groups by education, tumor grade and hormone receptor status. This method standardized the proportions to the distribution of the remaining covariates in the full regression models for the entire study group [Graubard 1999]. Ninety-five percent confidence limits for these proportions were calculated as ± 1.96 times the standard error.

Since colinearities among the independent variables could have resulted in unrealistically large estimated coefficients and standard errors [Hosmer 1989], a correlation matrix was constructed to identify socioeconomic variables with strong linear relationships [SAS Corr 1990]. The goodness-of-fit of the models was assessed using the Hosmer and Lemeshow summary statistic [Hosmer 1989]. For this statistic, observed and expected numbers of observations were calculated for each of ten groups of approximately equal size based on the percentiles of the estimated probabilities of an event (an event is defined as: distant stage disease, regional stage disease, or tumor greater than 1 cm in diameter). Observations were sorted in increasing order of their estimated probability of having an event outcome. The discrepancies between the observed and expected number of observations in these groups were summarized by the Pearson chi-square statistic and compared to a chi-square distribution with degrees of freedom equal to the number of groups minus two.

K. Human Subjects and Confidentiality

I obtained Institutional Review Board (IRB) approval from the Uniformed Services University of the Health Sciences for this study involving human subjects under Project Number T087KO-01. All information collected on study subjects was treated as confidential. Study ID numbers replaced the subject's name on all data files submitted to the NCI for use in this study.

L. Roles as Study Investigator

My roles as the study investigator included:

1. Conducting the geocoding meeting and data linkage feasibility assessment;
2. Developing revised data coding instructions and creating a new variable to be reported by all SEER cancer registries and documented as changes in the SEER Coding Manual;
3. Conceiving the research questions;
4. Developing the study design, including selection of the analytic methods and obtaining IRB approval;
5. Designing and directing the creation of the data files needed for the analysis; and
6. Conducting all of the data analysis.

CHAPTER III. RESULTS

A. Descriptive Analysis

There were a number of racial/ethnic differences in tumor characteristics and sociodemograhic factors among the invasive breast cancer patients included in this study **(Table III-1a,b)**. Due to the large size of this study population, chi-squared tests for ordinal and nominal response variables indicated that there were statistically significant racial/ethnic differences for all study factors. Japanese and White women tended to be diagnosed at an earlier stage, with smaller diameter tumors and at a lower tumor grade than other groups. Black and Hispanic women were more likely than other groups to be diagnosed with metastatic disease, have tumors 2 cm or larger in diameter, and have poorly differentiated tumors. American Indian and Vietnamese patients also had a higher percentage of advanced disease than other groups. Relative to Japanese and White patients, a larger percentage of the tumors among all other racial/ethnic groups were 2.0 cm or greater at the time of diagnosis. Korean and Vietnamese women were also more likely to have poorly differentiated tumors. Black, Korean and American Indian women had the highest percentage of tumors that were negative for hormone receptors.

There were also notable differences in social and economic factors among the groups. Black women were less likely to be married at the time of diagnosis. American Indian, Hispanic and Black women were similar with regard to many of the census tract level indicators of socioeconomic position. They were much more likely to be living in less educated and poorer neighborhoods, as measured by the percentage of residents

without a high school diploma, median family income, and the percentage living below the poverty level. They also tended to live in areas where unemployment was higher; where residents held "working class" jobs; and where a high percentage of families were headed by women having one or more children, with no husband living at home, and whose income is below the poverty level. Korean, Vietnamese and Black patients lived more frequently in areas where home ownership was lowest, and Black and American Indian patients lived in areas where fewer households owned a car.

Tumor grade information was missing from hospital records for 25% of the study cases. Those missing tumor grade were more likely to have distant stage disease (8% versus 5% with distant stage among those having information on tumor grade) or to be missing staging information (7% versus 2%), and to be aged 80 years and over at the time of diagnosis (16% versus 11%). Black women had a higher percentage of tumors classified as unknown grade than other racial/ethnic groups. Since high grade tumors are associated with poorer survival [Henson 1991], survival rates were compared among patients by tumor grade as an aid to developing a coding scheme that would best utilize the available information. The following five-year cumulative relative survival probabilities by grade for patients in this study: grade 1: 99%; grade 2: 91%; grade 3: 74%; grade 4: 76%; unknown grade: 83%. Therefore, for use as an explanatory variable in the regression analysis, tumor grade was re-coded as a binary variable with poorly and undifferentiated tumors combined (grades 3 and 4 were coded as 1) versus all others (i.e., well differentiated, moderately differentiated, and unknown tumor grade were coded as 0). Since patients with unknown tumor grade probably include a mix of those with high and low tumor grades, a separate analysis was conducted excluding them from the

regression model to determine the impact on our study findings.

Information on hormone receptor status was not available for 23% of the intended study population (4% of these patients had a tumor marker assay performed but results were not included in the medical record, 6% did not have an assay performed, and 13% had no information). Patients lacking hormone receptor status information were more likely to have distant stage disease (10% versus 4% for those with information on hormone receptor status) or to be missing staging information (9% versus 1%), and to be aged 80 years or more at the time of diagnosis (16% versus 11%). Forty percent of the patients without information on hormone receptor status were also missing tumor grade information versus 21% of those with hormone receptor status. Hormone receptor status was missing more frequently among Black and Hispanic women, than for other racial ethnic groups. One percent of the study cases were classified as having a "borderline" test result for ER status. A similar percentage of cases were also reported to have a borderline test result for PR status. Since hormone receptor status reflects characteristics of the biology of breast tumors, survival rates were compared among study patients according to their hormone receptor status as an aid to developing a coding scheme that would best utilize the available ER and PR status information. Five-year cumulative relative survival probabilities were highest for patients in this study with either positive ER status or positive PR status (89%) and lowest for patients with both negative ER status and negative PR status (75%). Patients with all remaining combinations of codes ER status and PR status experienced a survival probability closest to that of the hormone receptor negative patients (78%). For the regression analysis, hormone receptor status is re-coded as a binary variable with tumors classified as being positive for either ER or PR

(coded as 0), versus all others (coded as 1).

The predominant histological type of breast cancer in the study group was ductal adenocarcinoma which accounted for over 73% of the cases **(Table III-2.)**. Other histological types included lobular (13%), adenocarcinoma not otherwise specified (4%), mucinous (2.5%), carcinoma not otherwise specified (2%), medullary carcinoma (1.3%) and inflammatory carcinoma (1.1%). Racial/ethnic variation in histological type was slight with White patients tending to have a higher percentage of lobular carcinomas (14%) and a lower percentage of ductal adenocarcinomas (73%) than other groups. Lobular carcinoma ranged from 6% of the cases among Chinese women to 11% among Hispanic women and ductal adenocarcinoma ranged from 73% of the cases among Hispanic women to 82% among Korean women.

The relationships between study outcomes (tumor stage or size) and potential explanatory variables are summarized (1) in plots of the log-odds of being a "case" for various levels of each explanatory factor **(Figures III-1a-c, III-2a-c, and III-3a-c)** and (2) in 2x2 tables for binary explanatory variables **(Tables III-3-5)**. With the exception of the oldest age group, age at diagnosis appears to be negatively associated with a diagnosis of distant or regional stage breast cancer and with tumors 1 cm or greater in diameter **(Figures III-1a, 2a, 3a)**. Measures of socioeconomic position are often treated as categorical variables in epidemiologic studies. In the present study, however, most of the socioeconomic variables show strong linear (on a log scale) associations with more advanced stage of disease and larger tumor size. Therefore, they are treated as continuous variables in the regression analysis. The log-odds of being a case is positively associated with the percentage of persons without a high school diploma; the

percentage living below the poverty level; the percentage of families headed by women having one or more children, with no husband living at home, and whose income is below the poverty level; the percentage of persons in "working class" jobs; the percentage of persons that are unemployed; and the percentage of households that do not own a car. The log-odds of being a case is negatively associated with median family income and with the percentage of households that own their own home. These patterns of association for the socioeconomic variables are consistent for each of the study outcomes. Since the percentage of foreign-born residents did not show an association with the outcome variables, either in the log-odds plots or when used as a single predictor in a logistic regression model, it was dropped from further analysis.

Tables III-3-5 show that having a high grade tumor is positively associated with late stage disease and larger tumor size in the univariate analysis (OR=2.2 for both distant and regional stage disease and OR=3.1 for tumors equal or greater than 1cm). Negative hormone receptor status is positively associated with distant stage disease but not with regional stage or with tumor size. Patients that are not married at the time of diagnosis show a positive association with distant stage disease and larger tumor size, but no association with regional stage disease. Living in an urban area is weakly associated with each of the study outcomes (95% CI does not include 1.0, but this is not apparent when the odds ratio is rounded to a single decimal place).

B. Regression Analysis

A matrix of Pearson correlation coefficients was constructed to identify socioeconomic variables having strong linear relationships **(Appendix III-1)**. Colinearities among the independent variables may produce inflated estimated coefficients and/or standard errors in the multiple logistic regression models. As might be expected, the percentage of unemployed persons was highly correlated with the two poverty measures (overall percent below poverty; the percentage of families headed by women having one or more children, with no husband living at home, and whose income is below the poverty level). The unemployment variable was therefore excluded from the regression analysis. Since the overall poverty variable showed strong linear relationships with the second poverty variable and with the percentage of residents who did not own a car, it was also excluded from the regression analysis.

The results from multiple logistic regression models used to assess the importance of selected study factors in explaining racial/ethnic differences in the severity of disease at the time of diagnosis are shown in **Tables III-6-8**. Regression models that included the socioeconomic x age product terms as potential confounders did not result in any meaningful change in the magnitude of the odds ratios for each of the racial/ethnic groups and there was no improvement in the precision of the odds ratios, so they were dropped from the analysis. Interactions between racial/ethnic group and socioeconomic variables were not statistically significant at the 5% level, so they were also excluded from the regression models.

The first model in **Table III-6** shows odds ratios (ORs) for being diagnosed with

distant stage disease versus localized disease among the racial/ethnic groups relative to White women with an adjustment only for age at diagnosis. Odds ratios for Black, Hispanic, and American Indian women were elevated while those for Japanese were significantly reduced. Odds ratios for other groups were not significantly different from 1.0. The addition of geographic area (registry where diagnosed, urban residence) to the model, slightly lowered the OR for American Indian women, while the OR for Hispanic women remained unchanged. When sociodemographic factors were incorporated in the model the excess odds for Hispanic women was reduced by 60% and the excess odds for American Indian and Black women were further lowered (by 62% and 50%, respectively). In the final model, which includes tumor grade and hormone receptor status, only the OR for Black women remained significantly elevated at 1.3. The OR for Hawaiian women was elevated even after adjustment for other study factors, but the 95% confidence interval included one.

Table III-7 shows a similar analysis of the importance of study factors in explaining racial/ethnic differences in the diagnosis of regional stage disease versus localized disease. Initial ORs for Black and Hispanic women are again elevated relative to White women, though at somewhat lower levels than those seen for distant stage disease. The OR for Japanese women is significantly lower than that for White women, while ORs for the other groups are not significantly different from 1.0. The addition of geographic area and sociodemographic factors reduced the ORs for Black and Hispanic women, but no further reduction occurred after the inclusion of biological characteristics of the tumors. The lower OR for Japanese women remained unchanged after the addition of each group of potential confounding factors.

Comparisons of the ORs for larger tumor size by racial/ethnic group are shown in **Table III-8**. In the initial model, ORs adjusted for age at diagnosis were significantly high for Black, Hispanic, Filipino, and Korean women relative to White patients. These elevated ORs were partially explained by sociodemographic factors and other tumor characteristics, but remained high for all groups. The OR for Japanese women was significantly reduced and remained unchanged after each set of study variables was added to the model.

Estimates of the strength of the associations of sociodemographic variables, tumor grade, and hormone receptor status (considered as confounders in this study) and the three study outcomes, while not the focus of this study, are shown in **Appendices III-2-4**. Among the sociodemographic factors, not being married and living in areas where a high percentage of persons do not have a high school diploma are consistently associated with a more advanced stage of disease at diagnosis and larger tumor size. Median family income is negatively associated with distant stage disease and larger tumor size, but is not a significant predictor of regional stage disease.

The estimated percentages of White, Black, Hispanic and Japanese patients with late stage disease or larger tumor size by education, hormone receptor status, and tumor grade are shown in **Tables III-9-11**. The percentages are adjusted for all other covariates in the full regression models and provide an alternative to the odds ratio in assessing the influence of these factors on the disease outcomes. The effects of education (as an indicator of socioeconomic position) and tumor biology (hormone receptor status, grade) are greater for tumor stage than for tumor size.

Results from the Hosmer-Lemshow goodness-of-fit test for the full logistic

regression model corresponding to each study outcome are shown in **Figures III-4-6**. The observed and expected number of observations appear to be fairly close in each model. Perhaps due to the extremely large size of the study populations, however, the Hosmer-Lemeshow test statistics indicate that the models for distant stage disease and regional stage disease (vs. localized stage) do not fit the observed data well. The model for tumor size does fit the observed data well, based upon the non-significant Hosmer-Lemeshow test statistic ($p = 0.12$).

The regression analysis was repeated after excluding patients that were missing information on tumor grade. The pattern of associations between the specific racial/ethnic groups and cancer outcomes remained unchanged. Only the odds ratios for tumor grade and hormone receptor status in the distant vs. localized tumor stage analysis changed noticeably. The odds ratio for high tumor grade increased from 2.0 to 3.6 (95% CI = 3.4-3.9) and the odds ratio for negative hormone receptor status changed from 2.1 to 1.5 (95% CI = 1.4-1.6). The Hosmer-Lemeshow goodness-of-fit statistic for the distant vs. localized tumor stage model also improved following the exclusion of patients with unknown tumor grade ($\chi^2 = 7.55$ with 8 DF, $p = 0.48$), indicating that the model fit the data well **(Figure III-7)**. When patients with ductal adenocarcinoma and those with other histological types combined were analyzed in separate regression models, the patterns of association between the study factors and outcomes were similar.

Individual logistic regression models were produced for White, Black, and Hispanic patients (the three largest groups) in order to examine the consistency across racial/ethnic group of associations between sociodemographic factors, tumor characteristics and each of the study outcomes. The odds ratios shown in **Tables III-12-**

14 are generally comparable in magnitude for each racial/ethnic group. One exception to this uniformity of effect is the lack of an association between marital status and larger tumor size in Black women. Another exception is the negative association between negative hormone receptor status and larger tumor size seen for White women, with no significant association in Black or Hispanic women.

CHAPTER IV. DISCUSSION

A. Study Aims Addressed

Findings from this population-based study of 106,607 female breast cancer patients addressed the following two main research aims: (1) To describe the racial/ethnic distribution of selected demographic, socioeconomic, and tumor characteristics (stage of disease, tumor size, tumor grade, estrogen/progesterone receptor status) that influence prognosis for cancer of the female breast; and (2) To assess the importance of sociodemographic factors in explaining racial/ethnic differences in the distribution of tumors by stage and size at the time of diagnosis.

Regarding the first study aim: several racial/ethnic differences in tumor characteristics and sociodemographic factors were noted in this study population. The tendency for White and Japanese women to be diagnosed at an earlier stage than other groups has been documented in Hawaii [LeMarchand 1984, Meng 1997]. The poorer stage distribution in Black [Ownby 1985, Polednak 1986, Bain 1986, Bassett 1986, Stanford 1989, Farley 1989, Ragland 1991, Wells 1992, Chen 1994, Eley 1994, Simon 1996, Jones 1998], Hispanic [Samet 1987, Bentley 1998], and American Indian women [Samet 1987] has also been noted by others. This is the only population-based study, to my knowledge, that has characterized tumor grade and hormone receptor status for breast cancer patients in specific racial/ethnic groups other than Whites or Blacks.

To address the second study aim, multiple logistic regression models were used to determine whether the observed racial/ethnic differences in tumor stage and tumor size at

the time of diagnosis persist after adjustment for sociodemographic factors and biological characteristics of the tumors. Elevated odds ratios for later stage or larger size tumors among Black patients and Hispanic patients were reduced by about 50%-60% after adjustment for sociodemographic factors. Evidence for the role of differential tumor biology in accounting for the racial/ethnic differences in tumor stage and size was not as compelling. When information on tumor grade and hormone receptor status were added to the regression models already containing sociodemographic variables, odds ratios for the Black and Hispanic women declined further for distant stage disease, but did not markedly change for regional stage disease or tumor size.

Odds ratios for Black and Hispanic women relative to White women remained slightly elevated after adjusting for sociodemographic and tumor biology characteristics for distant and regional stage disease. In the analysis of tumor size, odds ratios for Black, Hispanic, Filipino, and Korean women remained elevated relative to White women after adjustment for sociodemographic factors, tumor grade, and hormone receptor status. Japanese women, conversely, had a consistently lower odds than White women for each study outcome. These lower odds persisted even after adjusting for other study factors.

B. Study Strengths

Strengths of this study include the large patient population size and the fact that cases were identified through population-based cancer registries. SEER Program registries cover approximately 14% of the entire United States population and include geographic regions with diverse racial/ethnic groups. As a result, this study was able to

assess patterns of breast cancer severity in a larger number of racial/ethnic groups than prior investigations that were limited to one or a few central registries or a limited number of hospitals or clinical trial groups [Mohla 1982, Ownby 1985, Pegararo 1986, Polednak 1986, Beverly 1987, Stanford 1987, Stanford 1989, Farley 1989 Mandelblatt 1991, Wells 1992, Chen 1994, Ellege 1994, Hulka 1994, Gordon 1995, Weiss 1995, Gapstur 1996, Krieger 1997a, Bentley 1998, Elmore 1998, Lannin 1998]. Further, patients in the present study include all eligible breast cancer cases from the populations in a defined set of geographic areas and are not subject to the influence of referral patterns which may affect hospital-based or clinical trial-based case selection.

Other study strengths include the high percentage of patient diagnoses that were microscopically confirmed (99%) and our ability to assess the importance of selected biological characteristics of the tumors when evaluating racial/ethnic differences in tumor stage and size at diagnosis. Finally, the use of a geographic linkage enabled us to assess the role of neighborhood-level sociodemographic factors on the cancer outcomes.

C. Study Limitations

Only limited risk factor information is available from cancer registry records on the study subjects. Individual information on factors such as body mass, alcohol and tobacco use, reproductive history, medical insurance status, usual source of health care, and screening behavior would have been helpful in this analysis of tumor characteristics at the time of diagnosis. Although the utility and advantages of neighborhood-level measures of socioeconomic position are well documented [Hakama 1982, Massey 1990,

Kaplan 1996, Krieger 1997b], the addition of individual socioeconomic information would have allowed a multi-level assessment of the importance of these factors in our study population [Krieger 1992, Anderson 1997, Krieger 1997b]. The lack of individual socioeconomic data may also lead to residual confounding by socioeconomic position. This residual confounding and/or the influence of other important unmeasured factors could explain the persistence of slightly elevated odds ratios for some racial/ethnic groups in this study.

In spite of the large study population, the relatively small number of American Indians (n = 136) made it difficult to detect statistically meaningful differences for this group. Odds ratios associated with more advanced stage of disease were consistently elevated for American Indian women and were comparable to the excesses seen for Black women, but due to the small population size, 95% confidence limits always included one. Only one of the SEER registries in the present study, New Mexico, is currently able to accurately report cancer incidence data for American Indians. The Alaska Native tumor registry has recently entered the SEER Program and will provide useful data for future studies. Efforts to improve reporting are underway in other registries, but current misclassification of American Indians into other racial/ethnic groups leads to significant under-reporting for this group [Sugarman 1996]. Furthermore, the cancer patterns among American Indians are known to vary by region and tribe [IHS, 1997] , so small population sizes will continue to hinder epidemiologic research on these groups.

Another study limitation was the large percentage of patients with missing information on hormone receptor status and tumor grade. Given the widespread

recognition of the utility of determining hormone receptor status for predicting response to hormonal therapy and the importance of tumor grade in assessing prognosis, the lack of this information in patient records is troublesome. Henson reported that many physicians consider the assignment of tumor grade to be too subjective to be of much prognostic use and this may explain why it was missing for 25% of our study cases [Henson 1991]. Henson noted that a strong relationship between histologic grade and patient survival persists in spite of interobserver and intraobserver variability.

D. Interpretation

In spite of its limitations, this study represents the largest analysis of breast cancer among women in diverse racial/ethnic groups to date and clearly indicates that sociodemographic factors may play an important role in accounting for observed racial/ethnic differences in the stage of disease and tumor size at the time of diagnosis. This supports findings from several earlier studies [Ownby 1985, Polednak 1986, Farley 1989, Mandelblatt 1991, Wells 1992, Elmore 1995, Weiss 1995, Bentley 1998, Lannin 1998].

There are a number of ways that sociodemographic factors may be influencing the stage and size of breast tumors at the time of diagnosis. The association in this study between marital status and tumor stage and size supports the results from a study of several cancer types in New Mexico [Goodwin 1987]. It has been postulated that married persons may tend to have better health habits and less delay in seeking medical care after the occurrence of symptoms than unmarried persons. In addition, married

persons tend to have higher socioeconomic status and greater social support [Goodwin 1987]. Several investigators have emphasized the impact of socioeconomic factors on access to physician care or screening services [Gregorio 1983, Harper 1993, Hoffman-Goetz 1998]. Mammography use has been found to be positively associated with income, education, having health insurance coverage, having a usual source of care, and urban residence [Rakowski 1993, Horton 1992, Horton 1996, Breen 1994, Katz 1994, Anderson 1995, Coughlin 1999, Makuc 1999]. Therefore, programs to promote screening mammography that target primary care physicians and women with low incomes and education have been recommended [Breen 1994, Eley 1994]. Studies to evaluate the efficacy of this type of intervention would also be helpful.

Several surveys indicate that the use of mammography in the United States has risen over time, though the majority of breast cancers are still first discovered either by the patient through breast self-exam or as an incidental finding, or by a clinical breast exam [Norton 1992, Benedict 1996, McPherson 1997]. Self-reported data on women aged 40 years or more, who were interviewed in 38 states from 1989 to 1997 as a part of the Behavioral Risk Factor Surveillance System (BRFSS), showed that the largest increases in mammography usage (defined as having a mammogram within the previous two years) occurred in those with lower education and lower income [Blackman 1999]. Personal interview data from the National Health Interview Survey, spanning 1987 to 1994, indicated that recent increases in mammography screening were greatest for Black women with low family incomes and had stabilized for low-income White women and all women with higher family incomes [Makuc 1999]. National estimates from the Jacobs Institute of Women's Health (JIWH) Mammography Attitudes and Usage Study of 1995

also indicated that increases in mammography use occurred in recent years among women with lower family incomes [Horton 1996]. A surprising finding from this survey was the slight decline since 1992 in regular mammography screening among women with college degrees. The authors suggested that this may have been due to increased public concern about potential health risks associated with radiation exposure from mammography. Despite the general increase over time in the use of mammography for early detection, however, all of these surveys indicate that sociodemographic differentials persist with women in lower income and education groups having lower screening rates.

Current racial/ethnic patterns in the use of mammography has been reported from several large surveys [Breen 1994, Burns 1996, Horton 1996, Blackman 1999]. Personal interview data from the 1990 National Health Interview Survey indicated that White, Black and Hispanic women aged 40 years and older had comparable overall rates of screening mammograms within the previous year [Breen 1994]. A similar finding was reported in the JIWH national survey for Black and White women, although the investigators noted that the percentage of women in compliance with current American Cancer Society mammography screening guidelines is still less than optimal for every racial/ethnic group [Horton 1996]. Racial/ethnic patterns of mammography use were also reported from data collected through telephone interviews with a representative sample of the civilian, noninstitutionalized adult population of states participating in the Behavioral Risk Factor Surveillance System (BRFSS) [Blackman 1999]. The BRFSS study results indicated comparable rates of mammography within the past two years among White, Black, and Asian American or Pacific Islander groups, but a lower rate among the American Indian or Alaska Native population.

Tumor size greater than 1 cm was used in the present study as an indicator of delayed detection. Tumors smaller than 1 cm are primarily found by screening mammography, whereas larger tumors are often detected by other methods such as symptoms, clinical breast exam, or breast self-exam [Fletcher 1985, Reintgen 1993, Helzlsouer 1995]. A recently published analysis of tumor size and stage in Asian American women with breast cancer reported similar results to those from the present study; namely, that women in Chinese, Filipino and Korean American groups were more likely than White women to be diagnosed with a tumor size greater than 1 cm [Hedeen 1999]. Japanese women in both studies had a slightly lower odds than White women for a tumor size greater than 1 cm. The similar findings are not surprising since Hedeen et al also based their study on cases identified from SEER Program registries, though their study period (diagnoses between 1988 and 1994) differed somewhat from the present study and included cases only from the five registries with the greatest number of Asian Americans.

The findings from the current study and the study by Hedeen suggest that there may be a relative delay in the diagnosis of breast cancer among women in these ethnic groups. Survey data on health behaviors among women in California have indicated that Chinese, Filipino, Korean, and Vietnamese women are less likely to report ever having had a mammogram than are women in the general population [CDC 1992a, CDC 1992b, CDC 1994, Hiatt 1996, CDC 1997, Maxwell 1997]. This provides indirect evidence that lower utilization of mammography by these ethnic groups may be associated with the diagnosis of more advanced tumors.

An additional finding from the study by Hedeen et al was that the increased odds

ratio for larger tumor size was limited to Asian American women who were born in Asia. The investigators suggest that a woman's birthplace and level of acculturation or assimilation may influence her beliefs and behaviors with respect to medical care in general and mammography screening utilization in particular. Unfortunately, place of birth information was unavailable for nearly half of the patients in the current study, precluding an examination of this factor. A study of breast cancer screening and screening-related attitudes among Filipino-American women in California reported lower screening rates in this group than in Black or White women in the 1994 California Behavioral Risk Factor Study [Maxwell 1997]. Factors associated with lower screening rates in Filipino women in this survey included lack of a physician recommendation for a mammogram, concern over cost, belief that a mammogram is only needed in the presence of symptoms, perceived inconvenience or difficulties in getting to the mammography facility, and embarrassment [Maxwell 1997].

Many studies, in addition to the present one, have found that socioeconomic effects alone do not account for all of the racial/ethnic differences in tumor stage at diagnosis [Vernon 1985, Bain 1986, Mandelblatt 1991, Richardson 1992, Wells 1992, Hunter 1993]. Even in situations where universal access to medical care is provided, racial/ethnic disparities in breast cancer diagnosis or outcome persist [Trock 1993, Katz 1994, Wojcik 1998]. Cultural factors such as beliefs, attitudes and knowledge about cancer have been shown to vary by race/ethnicity and have been found to influence cancer screening and prevention behaviors [Michielutte 1982, Jepson 1991, Loehrer 1991, Perez-Stable 1992, Harper 1993, Pachter 1994, Maxwell 1997, Lannin 1998, Lobell 1998]. Results from a recent case-control study of breast cancer patients

diagnosed in a hospital primarily serving residents of two rural counties in eastern North Carolina indicated that psychosocial and cultural variables in conjunction with socioeconomic factors are sufficient to explain the difference in stage at diagnosis between Black and White women [Lannin 1998]. Because 30% of the cancers in Whites and 11% in Blacks were discovered by routine screening mammography in the study by Lannin et al, it would seem logical to conclude that cultural beliefs are associated with differential use of screening mammography. Another study conducted on women in the same community at the same time, however, found that a woman's knowledge and beliefs had little influence on her use of screening mammography [O'Malley 1997]. The most important factor was whether mammography was recommended to the patient by a physician. Since the majority of early and late stage cancers were found by the patient in the study by Lannin et al, the investigators concluded that the most important effect of the cultural beliefs is that they lead to delayed presentation once a woman has developed a palpable breast abnormality [Lannin 1998]. Another study of breast cancer patients identified within an HMO setting in North Carolina also reported that patient delay before reporting breast cancer symptoms to a physician was an important factor in explaining tumor stage at the time of diagnosis [Howard 1998].

Several studies have reported an inverse correlation between socioeconomic status and body mass [Allan 1993, Millar 1993]. Increased body mass or obesity, in turn, has been linked to later stage breast tumors [Mohle-Boetani 1988, Daniell 1988, Verreault 1989, Reeves 1996, Jones 1997] or larger size tumors [Senie 1992, Bastarrachea 1994]. However, at least one study has not found an association between obesity and tumor stage [Howson 1986]. The mechanism behind more advanced breast

cancer and obesity is unknown, but endocrinologic factors leading to increased levels of endogenous estrogen have been suggested [Morabia 1990, Schapira 1991, Bernstein 1993, Maggino 1993, Hulka 1994, Kuller 1994, Hankinson 1995]. Obesity may also (or alternatively) play a diagnostic role. Some studies have suggested that obesity makes early detection more difficult [Austin 1979, Zumoff 1983, Mohle-Boetani 1988, Ingram 1989] or that physician approach to patients that are obese may differ [Weiss 1995].

Findings from the present study indicate that differences in tumor grade and hormone receptor status play a role in explaining the increased diagnosis of distant stage cancers among Black women, even after sociodemographic factors are taken into account. This result supports earlier findings by Chen et al who compared tumor characteristics among Black and White breast cancer patients diagnosed in 1985 and 1986 in three urban SEER registries [Chen 1994]. Chen et al noted a racial/ethnic difference in estrogen receptor status after adjustment for socioeconomic position, body mass index, use of alcohol and tobacco, reproductive experience, health care access, and usual source of care. Our finding that Japanese patients tended to be diagnosed at an earlier stage and smaller tumor size than other groups after adjustment for all other study factors has been reported by others [Ward-Hinds 1982, LeMarchand 1984, Stemmermann 1985, Higuchi 1993, Hedeen 1999]. It has further been suggested that differences in histopathologic features between the breast cancers of Japanese and White women may indicate possible biological differences between the groups [Stemmermann 1991, Higuchi 1993].

E. Public Health Importance

Data base linkage, as was done for this study using the NCI cancer surveillance data file and the 1990 U.S. census data file, can provide an important means for enhancing the utility of routinely collected disease surveillance data. As additional risk factor information is added to a disease surveillance system, more analytic studies become feasible. These studies enable the surveillance system to be used, not only for routine monitoring of disease trends in the population, but also for improving our understanding of potential factors underlying and explaining the trends. Data base linkage also has the advantage of generally being less expensive and more quickly accomplished than having to conduct field studies requiring the collection of new data on individuals. Data base linkage does not replace the need for in-depth, epidemiologic investigations of specific public health questions, but does play a useful, complimentary role in efforts to better understand the patterns of disease in populations.

In this study, the addition of sociodemographic data to the cancer surveillance file provided a unique opportunity to evaluate the importance of these factors in explaining racial/ethnic differences in the stage and size of breast cancers at the time of diagnosis. This data base linkage will also enable the conduct of additional studies to evaluate the importance of sociodemographic factors in explaining population patterns for other types of cancer. It may be particularly useful in the study of specific cancers for which there have been recent advances in the methods of detection or treatment (e.g., prostate-specific antigen screening tests for cancer of the prostate) which might, in turn, be expected to differentially impact different socioeconomic groups.

F. Future Directions

The results from this study suggest that sociodemographic factors account for a significant portion of the observed racial/ethnic differences in the stage of disease and tumor size at the time of diagnosis, but that differences in biological characteristics of breast tumors, at least among Black women, can not be ruled out. It would be useful to confirm these findings in additional studies that include central histopathology review and that include patient-level socioeconomic data, as well as area-based measures. The identification of new, valid and reliable tumor markers would allow a more precise characterization of meaningful racial/ethnic or sociodemographic differences in breast tumor types for future studies. Further studies are also needed to determine whether differential exposure to carcinogens or genetic susceptibility are important in explaining the more aggressive forms of breast cancer in specific patient subgroups.

Additional studies should investigate the roles of recent immigration and culturally-linked health behavior patterns among breast cancer patients in explaining racial/ethnic patterns for late stage at diagnosis. Since a socioeconomic disparity in mammography screening levels has been documented in several population surveys, methods for increasing compliance with recommended guidelines should be identified, implemented, and then evaluated for their efficacy. Future studies could also focus on sociodemographic differences in the quality of mammography, whether mammography is received at regular intervals, and whether appropriate follow-up and treatment is given to identified cases.

REFERENCES

(ACS) American Cancer Society: Report of the Workshop on Guidelines for Breast Cancer Detection. Atlanta, GA, American Cancer Society, March 1997.

(ACS) American Cancer Society. Cancer facts and figures-1999. Atlanta:ACS, 1999.

(AJCC) American Joint Committee on Cancer. Cancer Staging Manual, 5th Edition. Philadelphia:Lippincott-Raven Publishers, 1997.

Allan JD, Mayo K, Michel Y. Body size values of white and black women. Res Nurs Health 1993;16:323-333.

Anderson LM, May DS. Has the use of cervical, breast, and colorectal cancer screening increased in the United States? Am J Publ Health 1995;85:840-842.

Anderson RT, Sorlie P, Backlund E, Johnson N, Kaplan GA. Mortality effects of community socioeconomic status. Epidemiology 1997;8:42-47.

(ASCO) American Society of Clinical Oncology. Clinical practice guidelines for the use of tumor markers in breast and colorectal cancer. J Clin Oncol 1996;14:2843-2877.

Austin H, Cole P, Wynder E. Breast cancer inblack American women. Int J Cancer 1979;24:541-544.

Axtell LM, Asire AJ, Myers MH. Cancer patient survival. A report from the Cancer Surveillance, Epidemiology and End Results (SEER) Program. Washington: DHHS publication no. (NIH) 77-992, 1976.

Backlund E, Sorlie PD, Johnson NJ. The shape of the relationship between income and mortality in the United States-evidence from the National Longitudinal Mortality Study. Ann Epidemiol 1996;24:12-20.

Bain RP, Greenberg RS, Whitaker JP. Racial differences in survival of women with breast cancer. J Chron Dis 1986;39:631-642.

Baquet CR, Ringen K. Cancer among blacks and other minorities: statistical profiles. National Cancer Institute, NIH Publication No. 86-2785, 1986.

Bassett M, Krieger N. Social class and black-white differences in breast cancer survival. Am J Public Health 1986;76:1400-1403.

Bastarrachea J, Hortobagyi Gn, Smith Tl. Obesity as an adverse prognostic factor for patients receiving adjuvant chemotherapy for breast cancer Ann Intern Med 1994;120:18-25.

Beckmann MW, Niederacher D, Schnurch H-G, Busterson BA, Bender HG. Multistep carcinogenesis of breast cancer and tumour heterogeneity. J Mol Med 1997;75:429-439.

Benedict S, Williams RD, Hoomani J. Method of discovery of breast cancer. Cancer Prac 1996;4:147-155.

Bentley JR, Delfino RJ, Taylor TH, Howe S, Anton-Culver H. Differences in breast cancer stage at diagnosis between non-Hispanic white and Hispanic populations, San Diego County 1988-93. Breast Cancer Res Treat 1998;50:1-9.

Berg JW, Hutter RV. Breast cancer. Cancer 1995;75:257-269.

Bernstein L, Ross RK. Endogenous hormones and breast cancer risk. Epidemiol Rev 1993;15:48-65.

Beverly LN, Flanders WD, Go RCP, Soong S-J. A comparison of estrogen and progesterone receptors in black and white breast cancer patients. Am J Publ Health 1987;77:351-353.

Blackman DK, Bennett EM, Miller DS. Trends in self-reported use of mammograms (1989-1997) and Papanicolaou tests (1991-1997) - behavioral risk factor surveillance system. In: CDC Surveillance Summaries, October 8, 1999. MMWR 1999;48:1-22.

Breen N, Kessler L. Changes in the use of screening mammography: evidence from the 1987 and 1990 national health interview surveys. Am J Public Health 1994;84:62-67.

Brinton LA, Schairer C. Estrogen replacement therapy and breast cancer risk. Epidiol Rev 1993;15:66-79.

Brinton LA, Bernstein L, Colditz GA. Summary of the workshop: Workshop on physical activity and breast cancer, November 13-14, 1997. Cancer 1998;83(suppl):595-599.

Burns RB, McCarthy EP, Freund KM, Marwill SL, Shwartz M; Ash A, Moskowitz MA. Black women receive less mammography even with similar use of primary care. Annals Internal Med 1996;125:173-182.

(CDC) Centers for Disease Control and Prevention. Behavioral risk factor survey of Chinese: California, 1989. MMWR 1992a;41:266-270.

(CDC) Centers for Disease Control and Prevention. Behavioral risk factor survey of Vietnamese: California, 1991. MMWR 1992b;41:69-72.

(CDC) Centers for Disease Control and Prevention. Behavioral risk factor survey of Korean Americans: Alameda County, California, 1994. MMWR 1997;46:774-777.

Cella DF, Osav EJ, Kornblith AB, Holland JC, Siberfarb PM, Lee KW, Comis RL, Perry M, Cooper R, Maurer LH, et al. Socioeconomic status and cancer survival. J Clin Oncol 1991;9:1500-1509.

(Census) U.S. Bureau of the Census. Poverty areas in large cities: 1980 census of the population, vol. 2, subject reports. Washington: The Bureau; 1985.

(Census) U.S. Bureau of the Census. Census of population and housing, 1990: summary tape file 3A on CD-ROM. Washington: The Bureau; 1992.

(Census) U.S. Bureau of the Census. A guide to state and local census geography. Washington: The Bureau; 1993.

Chen VW, Correa P, Kurman RJ, Wu X-C, Eley JW, Austin D, Muss H, Hunter CP, Redmond C, Coates R, Reynolds P, Herman AA, Edwards BK. Histological characteristics of breast carcinoma in blacks and whites. Cancer Epidemiol Biomarkers & Prev 1994;3:127-135.

Cherkin DC, Grothaus L, Wagner EH. Is magnitude of co-payment effect related to income? Using census data for health services research. Soc Sci Med 1992;34:33-41.

Colditz GA. Relationship between estrogen levels, use of hormone replacement therapy, and breast cancer. JNCI 1998;90:814-823.

Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormonal contraceptives: Further results. Contraception 1996;54(suppl):1S-106S.

Corin E. The cultural frame: context and meaning in the construction of health. In: Amick III BC, Levine S, Tarlov AR, Walsh DC (eds). Society and Health. New York: Oxford University Press, 1995.

Coughlin SS, Uhler RJ, Blackman DK. Breast and cervical cancer screening practices among American Indian and Alaska Native women in the United States, 1992-1997. Prev Med 1999;29:287-295.

Daniell HW. Increased lymph node metastases at mastectomy for breast cancer associated with host obesity, cigarette smoking, age, and large tumor size. Cancer 1988;62:429-435.

Dayal H, Power RN, Chiu C. Race and socioeconomic status in survival for breast cancer. J Chronic Dis 1982;35:675-683.

Devesa SS, Diamond EL. Socioeconomic and racial differences in lung cancer incidence. Am J Epidemiol 1983;118:818-831.

(DHHS) U.S. Department of Health and Human Services. Health status of minorities and low-income groups. Washington, DC: US GPO, 1991.

Dhingra K, Hortobagyi GN. Critical evaluation of prognostic factors. Sem Oncol 1996;23:436-445.

Donegan WL. Tumor-related prognostic factors for breast cancer. CA Cancer J Clin 1997;47:28-51.

Ellen JM, Kohn RP, Bolan Ga, Shiboski S, Krieger N. Socioeconomic differences in sexually transmitted disease rates among black and white adolescents, San Francisco, 1990 to 1992. Am J Public Health 1995;85:1546-1548.

Eley JW, Hill HA, Chen VW, Austin DF, Wesley MN, Muss HB, Greenberg RS, Coates RJ, Correa P, Redmond CK. Racial differences in survival from breast cancer. Results of the National Cancer Institute black/white cancer survival study. JAMA 1994;272:947-954.

Elledge RM, Clark GM, Chanmess GC, Osborne CK. Tumor biologic factors and breast cancer prognosis among white, Hispanic, and black women in the United States. J Natl Cancer Inst 1994;86:705-712.

Elmore JG, Moceri VM, Carter D, Larson EB. Breast carcinoma tumor characteristics in black and white women. Cancer 1998;83:2509-2515.

Faia MA. Selection by certification: a neglected variable in stratification research. Am J Sociol 1981;86:1093-1111.

Farley TA, Flannery JT. Late-stage diagnosis of breast cancer in women of lower socioeconomic status: public health implications. Am J Public Health 1989;79:1508-1512.

Feldman JJ, Makuc DM, Kleinman JC, Cornoni-Huntley J. National trends in educational differences in mortality. Am J Epidemiol 1989;129:919-933.

Ferno M. Prognostic factors in breast cancer: a brief review. Anticancer Res 1998;18:2167-2172.

Filakti H, Fox J. Differences in mortality by housing tenure and by car access from the POCS longitudinal study. Popul Trends 1995;81:27-30.

Fisher ER, Anderson S, Redmond C, Fisher B. Pathologic findings from the National Surgical Adjuvant Breast Project Protocol B-06:10-year pathologic and clinical prognostic discriminants. Cancer 1993;71:2507-2514.

Fletcher SW, O'Malley MS, Bunce LA. Physician's abilities to detect lumps in silicone breast models. JAMA 1985;253:2224-2228.

Friedenreich CM, Thune I, Brinton LA, Albanes D. Epidemiologic issues related to the association between physical activity and breast cancer. Cancer 1998;83(suppl):600-610.

FritzA, Ries L, editors. The SEER program code manual, 3rd edition. Bethesda, Md: National Cancer Institute; 1998. NIH Publication No. 98-2313.

Gapstur SM, Dupuis J, Gann P, Collila S, Winchester DP. Hormone receptor status of breast tumors in black, Hispanic, and non-Hispanic white women: an analysis of 13,239 cases. Cancer 1996;77:1465-1471.

Garne JP, Aspegren K, Linell F, Rank F, Ranstam J. Primary prognostic factors in invasive breast cancer with special reference to ductal carcinoma and histologic malignancy grade. Cancer 1994;73:1438-1448.

Goldman N. Marriage selection and mortality patterns: inferences and fallacies. Demography 1993;30:189-208.

Goodwin JS, Hunt WC, Key CR, Samet JM. The effect of marital status on stage, treatment, and survival of cancer patients. JAMA 1987;258:3125-3130.

Gordon NH, Crowe JP, Brumberg DJ, Berger NA. Socioeconomic factors and race in breast cancer recurrence and survival. Amer J Epidemiol 1992;135:609-618.

Gordon N. Association of education and income with estrogen receptor status in primary breast cancer. Am J Epidemiol 1995;142:796-803.

Graubard BI, Korn EL. Predictive margins with survey data. Biometrics 1999;55:652-659.

Greenwald HP, Polissar NY, Dayal HH. Race, soeioeconomic status and survival in three female cancers. Ethnicity Health 1996;1:65-75.

Greenwald P. Role of dietary fat in the causation of breast cancer: point. Cancer Epidemiol Biomarkers Prevention 1999;8:3-7.

Gregorio D, Cummings M, Michalek A. Delay, stage of disease and survival among White and Black women with breast cancer. Am J Public Health 1983;73:590-593.

Habel LA, Stanford JL. Hormone receptors and breast cancer. Epidemiol Rev 1993;15:209-219.

Hakama M, Hakulinen T, Pukkala E, Saxen E, Teppo L. Risk indicators of breast and cervical cancer on ecologic and individual levels. AM J Epidemiol 1982;116:990-1000.

Hankinson SE, Willett WC, Manson JE, Hunter DJ, Colditz GA, Stampfer MJ, Longcope C, Speizer FE. Alcohol, height, and adiposity in relation to estrogen and prolactin levels in postmenopausal women. J Natl Cancer Inst 1995;87:1297-302.

Harper AP. Mammography utilization in the poor and medically underserved. Cancer 1993;72:1478-1482.

Harris JR, Lippman ME, Veronesi U, Willett W. Breast cancer. N Eng J Med 1992;327:319-328.

(Harvard) Socioeconomic status. Harvard Report on Cancer Prevention, Volume 1: Causes of Human Cancer. In: Cancer Causes Control 1996;7:S33-S35.

Heck KE, Wabener DK, Schatzkin A, Devesa SS, Breen N. Socioeconomic status and breast cancer mortality, 1989 through 1993: an analysis of education data from death certificates. Am J Public Health 1997;87:1218-1222.

Hedeen AN, White E, Taylor V. Ethnicity and birthplace in relation to tumor size and stage in Asian American women with breast cancer. Am J Public Health 1999;89:1248-1252.

Helzlsouer KJ. Early detection and prevention of breast cancer. In: Cancer prevention and control. Greenwald P, Kramer BS, Weed DL (eds). New York:Marcel Dekker, Inc, pp 509-533.

Henderson BE, Pike MC, Bernstein L, Ross RK. Breast cancer. In: Cancer epidemiology and prevention, 2nd edition. Schottenfeld D and Fraumeni, Jr. JF (eds). New York:Oxford University Press, pp 1022-1039, 1996.

Henson DE, Ries L, Freedman LS, Carriaga M. Relationship among outcome, stage of disease, and histologic grade for 22,616 cases of breast cancer: The basis for a prognostic index. Cancer 1991;68:2142-2149.

Hiatt RA, Pasick RJ. Unresolved problems in early breast cancer detection: focus on the underserved. Breast Cancer Res Treat 1996;40:37-51.

Higuchi CM, Serxner SA, Nomura AMY, Stemmermann GN. Histopathological predictors of breast cancer death among Caucasians and Japanese in Hawaii. Cancer Epidemiol Biomarkers Prev 1993;2:201-205.

Hofer TP, Wolfe RA, Tedeschi PJ, McMahon LF, Griffith JR. Use of community versus individual socioeconomic data in predicting variation in hospital use. Health Serv Res 1998;33:243-259.

Hoffman-Goetz L, Breen NL, Meissner. The impact of social class on the use of cancer screening within three racial/ethnic groups in the United States. Ethnicity Dis 1998;8:43-51.

Holmes MD, Hunter DJ, Colditz GA, Stampfer MJ, Hankinson SE, Speizer FE, Rosner B, Willett WC. Association of dietary intake of fat and fatty acids with risk of breast cancer. JAMA 1999;281:914-920.

Horton JA, Romans MC, Cruess DF. Mammography attitudes and usage study, 1992. Womens Health Issues 1992;2:180-186.

Horton JA, Cruess DF, Romans MC. Compliance with mammography screening guidelines: 1995 mammography attitudes and usage study report. Womens Health Issues 1996;6:239-245.

Howard DL, Penchansky R, Brown MB. Disaggregating the effects of race on breast cancer survival. Fam Med 1998;30:228-235.

Howson CP, Kinne D, Wynder EL. Body weight, serum cholesterol, and stage of primary breast cancer. Cancer 1986;58:2372-2381.

Hulka BS, Liu ET, Lininger RA. Steroid hormones and risk of breast cancer. Cancer 1994;74:1111-1124.

Hunter CP, Redmond CK, Chen VW, Austin DF, Greenberg RS, Correa P, Muss HB, Forman MR, Wesley MN, Blacklow RS. Breast cancer: factors associated with stage at diagnosis in black and white women. J Natl Cancer Inst 1993;85:1129-1137.

Hunter DJ, Willett WC. Diet, body size, and breast cancer. Epidemiol Rev 1993;15:110-132.

Hunter DJ. Role of dietary fat in the causation of breast cancer: counterpoint. Cancer Epidemiol Biomarkers Prevention 1999;8:9-13.

(ICDO-2) International Classification of Diseases for Oncology, Second Edition. Geneva, World Health Organization, 1990.

(IHS) Department of Health and Human Services, Indian Health Service. Regional differences in Indian health. Rockville, MD: U.S. GPO, 1997.

Ingram DM, Huyang HY, Catchpole NM, Roberts A. Do big breasts disadvantage women with breast cancer? Aust N Z J Surg 1989;59:115-117.

Jepson C, Kessler LG, Portnoy B, Gibbs T. Black-white differences in cancer prevention knowledge and behavior. Am J Public Health 1991;81:501-504.

Jones BA, Kasl SV, Curnen MGM, Owens PH, Dubrow R. Severe obesity as an explanatory factor for the black/white difference in stage at diagnosis of breast cancer. Am J Epidemiol 1998;146:394-404.

Kaplan GA. People and places: contrasting perspectives on the association between social class and health. Intl J Health Serv 1996;26:507-519.

Katz SJ, Hofer TP. Socioeconomic disparities in prefentive care persist despite universal coverage: breast and cervical cancer screening in Ontario and the United States. JAMA 1994;272:530-534.

Kaufman JS, Long AE, Liao Y, Cooper RS, McGee DL. The relation between income and mortality in U.S. blacks and whites. Epidemiology 1998;9:147-155.

Kelsey JL, Gammon MD. Epidemiology of breast cancer. Epidemiol Rev 1990;12:228-240.

Krieger N. Social class and the black/white crossover in age-specific incidence of breast cancer: a study linking census-derived data to population-based registry records. Am J Epidemiol 1990;131:804-814.

Krieger N. Women and social class: a methodological study comparing individual, household, and census measures as predictors of black/white differences in reproductive history. J Epidemiol Community Health 1991;45:35-42.

Krieger N. Overcoming the absence of socioeconomic data in medical records: validation and application of a census-based methodology. Am J Public Health 1992;82:703-710.

Krieger N, Van Den Eeden SK, Zava D, Okamoto A. Race/ethnicity, social class, and prevalence of breast cancer prognostic biomarkers: a study of white, black, and Asian women in the San Francisco bay area. Ethnicity Dis 1997a;7:137-149.

Krieger N, Williams DR, Moss NE. Measuring social class in U.S. public health research: concepts, methodologies and guidelines. Annu Rev Public Health 1997b;18:341-378.

Kuller LH. Eating fat or being fat and risk of cardiovascular disease and cancer among women. Ann Epidemiol 1994;4:119-127.

Landberg G, Roos G. The cell cycle in breast cancer. APMIS 1997;105:575-589.

Lannin DR, Mathews HF, Mitchell J, Swanson MS, Swanson FH, Edwards MS. Influence of socioeconomic and cultural factors on racial differences in late-stage presentation of breast cancer. JAMA 1998;279:1801-1807.

Last JM. Social and behavioral determinants of health. In Public health and human ecology. East Norwalk: Appleton & Lange, 1987.

Le Marchand L, Kolonel LN, Nomura AMY. Relationship of ethnicity and other prognostic factors to breast cancer survival patterns in Hawaii. JNCI 1984;73:1259-1265.

Le Marchand L, Kolonel LN, Nomura AMY. Breast cancer survival among Hawaii Japanese and Caucasian women: Ten-year rates and survival by place of birth. Am J Epidemiol 1985;122:571-578.

Le Marchand L. Ethnic variation in breast cancer survival: a review. Breast Cancer Res Treat 1991;18(suppl):S119-S126.

Liberatos P, Link BG, Kelsey JL. The measurement of social class in epidemiology. Epidemiol Rev 1988;10:87-121.

Lobell M, Bay RC, Rhoads KVL, Keske B. Barriers to cancer screening in Mexican-American women. Mayo Clin Proc 1998;73:301-308.

Loehrer PJ, Greger HA, Weinberger M, Musick B, Miller M, Nichols C, Bryan J, Higgs D, Brock D. Knowledge and beliefs about cancer in a socioeconomically disadvantaged population. Cancer 1991;68:1665-1671.

Longnecker MP, Newcomb PA, Mittendorf R, Greenberg ER, Clapp RW, Bogdan GF, Baron J, MacMahon B, Willett WC. Risk of breast cancer in relation to lifetime alcohol consumption. JNCI 1995;87:923-929.

Longnecker MP. Invited commentary: The Framingham results on alcohol and breast cancer. Am J Epidemiol 1999;149:93-101.

Maggino T, Pirrone F, Velluti F, Bucciante G. The role of endocrine factors and obesity in hormone-dependent gynecological neoplasias. Eur J Gynaecol Oncol 1993;14:119-126.

Makuc DM, Breen N, Freid V. Low income, race, and the use of mammography. Health Services Res 1999;34:229-239.

Malone KE, Daling JR, Weiss NS. Oral contraceptives in relation to breast cancer. Epidemiol Rev 1993;15:80-97.

Mandelblatt J, Andrews H, Kerner J, Zauber A, Burnett W. Determinants of late stage diagnosis of breast and cervical cancer: the impact of age, race, social class, and hospital type. Am J Public Health 1991;81:646-649.

Massey DS, Eggers ML. The ecology of inequality: minorities and the concentration of poverty, 1970-1980. Am J Sociol 1990;95:1153-1188.

Maxwell AE, Bastani R, Warda US. Breast cancer screening and related attitudes among Filipino-American women. Cancer Epidemiol Biomarkers Prev 1997;6:719-726.

McPherson CP, Swenson KK, Jolitz G, Murray CL. Survival of women ages 40-49 years with breast carcinoma according to method of detection. Cancer 1997;79:1923-1932.

Meng L, Maskarinec G, Wilkens L. Ethnic differences and factors related to breast cancer survival in Hawaii. Intl J Epidemiol 1997;26:1151-1158.

Merkel DE, Winchester DJ, Gosdschmidt RA, August CZ, Wruck DM, Rademaker AW. DNA flow cytometry and pathologic grading as prognostic guides in axillary lymph node-negative breast cancer. Cancer 1993;72:1926-1932.

Michielutte R, Diseker RA. Racial differences in knowledge of cancer. Soc Sci Med 1982;16:245-253.

Millar WJ, Stephens T. Social status and health risks in Canadian adults:1985 and 1991. Health Rep 1993;5:143-156.

Mohla S, Sampson CC, Khan T, Enterline JP, Leffall, Jr. L, White JE, Gabriel BW, Hunter JB. Estrogen and progesterone receptors in breast cancer in black Americans: correlation of receptor data with tumor differentiation. Cancer 1982;50:552-559.

Mohle-Boetani JC, Grosser S, Whittemore AS, Malec M, Kampert JB, Paffengarger RS. Body size, reproductive factors, and breast cancer survival. Prev Med 1988;123:767-774.

Morabia A, Wynder EL. Epidemiology and natural history of breast cancer. Implications for the body weight-breast cancer controversy. Surg Clin North Am 1990;70:739-752.

(NCI) National Cancer Advisory Board Reccommendations for Women Aged 40-49. Rockville, MD, National Cancer Institute, March 1997.

(NIH) National Institutes of Health Consensus Development Conference Statement: Breast Cancer Screening for Women Ages 40-49. Bethesda, MD, January 21-23, 1997.

(NIH) National Institutes of Health. Comparison of survival for black and white patients. In Cancer Patient Survival Experience. NIH publication no. 80-2148, 1980.

Norton L, Crown JP. Malignancies of epithelial tissue: breast cancer. In Kelley WN (ed.). Textbook of Internal Medicine, 2nd Edition. Philadelphia:J.B. Lippincott Company, pp 1069-1072, 1992.

(OMB) Office of Management and Budget. Directive No. 15: Race and ethnic standards for federal statistics and administrative reporting. Statistical policy handbook. Washington, D.C.: U.S. Department of Commerce, Office of Federal Statistical Policy and Standards, 1978.

O'Malley MS, Earp JA, Harris RP. Race and mammography use in two North Carolina counties. Am J Public Health 1997;87:782-786.

O'Reilly SM, Camplejohn RS, Barnes DM, Millis RR, Rubens RD, Richards MA. Node-negative breast cancer: Prognostic subgroups defined by tumor size and flow cytometry. J Clin Oncol 1990;8:2040-2046.

Osborne CK. Steroid hormone receptors in breast cancer management. Breast Cancer Res Treat 1998;51:227-238.

Ownby HE, Frederick J, Russo J, Brooks SC, Swanson GM, Heppner GH, Brennan MJ, Breast Cancer Prognostic Study Associates. Racial differences in breast cancer patients. JNCI 1985;75:55-60.

Pachter LM. Culture and clinical care: folk illness beliefs and behaviors and their implications for health care delivery. JAMA 1994;271:690-694.

Pagana TJ, Lubbe WJ, Schwartz SM, Sprechini GD. A comparison of palpable and nonpalpable breast cancers. Arch Surg 1989;124:26-28.

Parkin DM. The global burden of cancer. Sem Cancer Biol 1998;8:219-235.

Patrick DL, Wickizer TM. Community and health. In Amick III BC, Levine S, Tarlov AR, Walsh DC, editors. Society and Health. New York: Oxford University Press, 1995.

PDQ Treatment information for health professionals: Breast cancer. CancerNet Cancer Information, National Cancer Institute, August 1999. Available from URL: http://cancernet.nci.nih.gov/.

Pegararo RJ, Karnan V, Nirmul D, Joubert SM. Estrogen and progesterone receptors in breast cancer among women of different racial groups. Cancer Res 1986;46:2117-2120.

Perez-Stable EJ, Sabogal F, Otero-Sabogal R, Hiatt RA, McPhee SJ. Misconceptions about cancer among Latinos and Anglos. JAMA 1992;268:3219-3223.

Petridou E, Kosmidis H, Haidas S, Tong D, Revinthi K, Flytzani V, Papaioannou D, Trichopoulos D. Survival from childhood leukemia depending on socioeconomic status in Athens. Oncology 1994;51:391-395.

Pike MC, Spicer DV, Dahmoush L, Press MF. Estrogens, progestogens, normal breast cell proliferation, and breast cancer risk. Epidemiol Rev 1993;15:17-35.

Polednak AP. Breast cancer in black and white women in New York State: case distribution and incidence rates by clinical stage at diagnosis. Cancer 1986;58:807-815.

Rackowski W, Rimer BK, Bryant SA. Integrating behavior and intention regarding mammography by respondents in the 1990 national health interview survey of health promotion and disease prevention. Public Health Rep 1993;108:605-624.

Ragland KE, Selvin S, Merrill DW. Black-white differences in stage-specific cancer survival: analysis of seven selected sites. Am J Epidemiol 1991;33:672-682.

Ravaioli A, Bagli L, Zucchini A, Monti F. Prognosis and prediction of response in breast cancer: the current role of the main biological markers. Cell Prof 1998;31:113-126.

Reeves MJ, Newcomb PA, Remington PL, Marcus PM, MacKenzie WR. Body mass and breast cancer: relationship between method of detection and stage of disease. Cancer 1996;77:301-307.

Reintgen D, Berman C, Cox C, Baekey P, Nicosia S, Greenberg H, Bush C, Lyman GH, Clark RA. The anatomy of missed breast cancers. Surg Oncol 1993;2:65-75.

Richardson JL, Langholz B, Bernstein L, Burciaga C, Danley K, Ross R. Stage and delay in breast cancer diagnosis by race, socioeconomic status, age, and year. Br J Cancer 1992;65:922-926.

Reis P. Educational differences in health status and health care. National Center for Health Statistics, DHHS Publication No. PHS 91-1507, 1991.

Ries LAG, Kosary CL, Hankey BF, Miller BA, Edwards BK, editors. SEER Cancer statistics review, 1973-1995. U.S. Department of Health and Human Services, National Cancer Institute, 1998. Available from URL: http://www-seer.ims.nci.nih.gov/Publications/.

Robert SA. Community-level socioeconomic status effects on adult health. J Health Soc Behav 1998;39:18-37.

Rosen PP, Groshen S, Kinne DW, Norton L. Factors influencing prognosis in node-negative breast carcinoma: Analysis of 767 T1N0M0/T2N0M0 patients with long-term follow-up. J Clin Oncol 1993;11:2090-2100.

Rosenberg L, Metzger LS, Palmer JR. Alcohol consumption and risk of breast cancer: A review of the epidemiologic evidence. Epidemiol Rev 1993;15:133-144.

Samet JM, Key CR, Hunt WC, Goodwin JS. Survival of American Indian and Hispanic cancer patients in New Mexico and Arizona, 1969-1982. JNCI 1987;79:457-463.

(SAS Corr) SAS Institute Inc. The Corr procedure. In SAS Procedures Guide, vers. 6, 3rd edition. Cary, NC: SAS Institute Inc., 1990.

(SAS Logistic) SAS Institute Inc. LOGISTIC procedure, release 6.12 for Windows. Cary, NC: SAS Institute Inc., 1989.

Schapira DV, Kumar NB, Lyman GH. Obesity, body fat distribution, and sex hormones in breast cancer patients. Cancer 1991;67:2215-2218.

Seiffert J, editor. SEER program comparative staging guide for cancer, version 1.1. Bethesda, MD: National Cancer Institute; 1993. NIH Publication No. 93-3640.

Senie RT, Rosen PP, Rhodes P, Lesser ML, Kinne DW. Obesity at diagnosis of breast carcinoma influences duration of disease-free survival. Ann Intern Med 1992;116:26-32.

Senie RT, Lesser M, Kinne DW, Rosen PP. Method of tumor detection influences disease-free survival of women with breast carcinoma. Cancer 1994;73:1666-1672.

Simon MS, Severson RK. Racial differences in survival of female breast cancer in the Detroit metropolitan area. Cancer 1996;77:308-314.

Simpson JF, Page DL. The role of pathology in premalignancy and as a guide for treatment and prognosis in breast cancer. Sem Oncol 1996;23:428-435.

Smith GD, Neaton JD, Wentworth D, Stamler R, Stamler J. Socioeconomic differentials in mortality among men screened for the Multiple Risk Factor Intervention Trial: I. White men. Am J Public Health 1996a;86:486-496.

Smith GD, Wentworth D, Neaton JD, Stamler R, Stamler J. Socioeconomic differentials in mortality among men screened for the Multiple Risk Factor Intervention Trial: II. Black men. Am J Public Health 1996b;86:497-504.

Smith KR, Waitzman NJ. Effects of marital status on the risk of mortality in poor and non-poor neighborhoods. Ann Epidemiol 1997;7:343-349.

Sorlie P, Rogot E, Anderson R, Johnson NJ, Backlund E. Black-white mortality differences by family income. Lancet 1992;340:346-350.

Stanford JL, Szklo M, Boring CC, Brinton LA, Diamond EA, Greenberg RS, Hoover RN. A case-control study of breast cancer stratified by estrogen receptor status. Am J Epidemiol 1987;125:184-194.

Stanford JL, Greenberg RS. Breast cancer incidence in young women by estrogen receptor status and race. Am J Publ Health 1989;79:71-73.

Stemmermann GN, Catts A, Fukunaga F, Horie A, Nomura AMY. Breast cancer in women of Japanese and Caucasian ancestry in Hawaii. Cancer 1985;56:206-209.

Stemmermann GN. The pathology of breast cancer in Japanese women compared to other ethnic groups: a review. Breast Cancer Res Treatment 1991;18:S67-S72.

Sugarman JR, Holliday M, Ross A, Castorina J, Hui Y. Improving American Indian cancer data in the Washington State cancer registry using linkages with the Indian Health Service and tribal records. Cancer 1996;78:1564-1568.

Tabar L, Duffy SW, Krusemo UB. Detection method, tumour size and node metastases in breast cancers diagnosed during a trial of breast cancer screening. Eur J Cancer Clin Oncol 1987;23:959-962.

Tolson GC, Barnes JM, Gay GA, Kowalexki JL. The 1989 revision of the US standard certificates and reports 1991; National Center for Health Statistics, DHHS publ. no. PHS 91-1465.

Tretli S, Haldorsen T, Ottestad L. The effect of pre-morbid height and weight on the survival of breast cancer patients. Br J Cancer 1990;62:299-303.

Trock B, Rimer BK, King E, Balshem A, Christinzio CS, Engstrom PF. Impact of an HMO-based intervention to increase mammography utilization. Cancer Epidemiol Biomarkers Prev 1993;2:151-156.

Umberson D. Family status and health behaviors: social control as a dimension of social integration. J Health Soc Behavior 1987;28:306-319.

(USPSTF) U.S. Preventive Services Task Force: Guide to Clinical Preventive Services, 2nd Ed. Screening for Breast Cancer. Philadelphia:Williams & Wilkins, 1996.

Verheij RA. Explaining urban-rural variations in health: a review of interactions between individual and environment. Soc Sci Med 1996;42:923-935.

Vernon SW, Tillery BC, Neale AV, Steinfeldt L. Ethnicity, survival, and delay in seeking treatment for sympotoms of breast cancer. Cancer 1985;55:1563-1571.

Verreault R, Brisson J, Deschenes L, Naud F. Body weight and prognostic indicators in breast cancer: modifying effect of estrogen receptors. Am J Epidemiol 1989;129:260-268.

Von Kleist S. Prognostic factors in breast cancer: theoretical and clinical aspects. Anticancer Res 1996;16:3907-3912.

Ward-Hinds M, Kolonel LN, Nomura AM, Lee J. Stage-specific breast cancer incidence rates by age among Japanese and Caucasian women in Hawaii, 1960-1979. Br J Cancer 1982;45:118-123.

Weiss SE, Tartter PI, Ahmed S, Brower ST, Brusco C, Bossolt K, Amberson JB, Bratton J. Ethnic differences in risk and prognostic factors for breast cancer. Cancer 1995;76:268-274.

Wells BL, Horm JW. Stage at diagnosis in breast cancer: race and socioeconomic factors. Am J Public Health 1992;82:1383-1385.

Wojcik BE, Spinks MK, Optenberg SA. Breast cancinoma survival analysis for African American and white women in an equal-acess health care system. Cancer 1998;82:1310-1318.

Wright EO, Costello C, Hachen D, Sprague J. The American class structure. Am Sociol Rev 1982;47:709-726.

Yong L-C, Brown CC, Schatzkin A, Schairer C. Prospective study of relative weight and risk of breast cancer: The Breast Cancer Detection Demonstration Project Follow-up Study, 1979 to 1987-1989. Am J Epidemiol 1996;143:985-995.

Young, Jr. JL, Ries LG, Pollack ES. Cancer patient survival among ethnic groups in the United States. J Natl Cancer Inst 1984;73:341-352.

Zhang Y, Kreger BE, Dorgan JF, Splansky GL, Cupples LA, Ellison RC. Alcohol consumption and risk of breast cancer: The Framingham study revisited. Am J Epidemiol 1999a;149:93-101.

Zhang Y, Kreger BE, Dorgan JF, Splansky GL, Cupples LA, Ellison RC. Authors' response to "The Framingham results on alcohol and cancer." Am J Epidemiol 1999b;149:105.

Zippin C, Lum D, Hankey BF. Completeness of hospital cancer case reporting from the SEER Program of the NCI. Cancer 1995;76:2343-2350.

Zumoff B, Dasbupta I. Relationship between body weight and the incidence of positive axillary nodes at mastectomy for breast cancer. J Surg Oncol 1983;22:217-220.

TABLE I-1. Five-year cumulative relative survival rates by stage of disease at diagnosis for female breast cancer cases, all ages, diagnosed 1989-94 [source: Ries 1998].

Localized	Regional	Distant	Local+Regional +Distant	Unstaged
0.96 (0.96, 0.97)[a]	0.77 (0.76, 0.77)	0.22 (0.20, 0.24)	0.86 (0.85, 0.86)	0.51 (0.48, 0.54)

[a] Confidence limits based on the survival rate +/- (2*standard error).

TABLE I-2. Summary of TNM-based stage groupings for invasive breast cancers.

Stage I:

The cancer is no larger than 2 centimeters and has not spread outside the breast.

Stage IIA is defined by either of the following:

The cancer is no larger than 2 centimeters but has spread to the lymph nodes under the arm (the axillary lymph nodes).

The cancer is between 2 and 5 centimeters but has not spread to the axillary lymph nodes.

Stage IIB is defined by either of the following:

The cancer is between 2 and 5 centimeters and has spread to the axillary lymph nodes.

The cancer is larger than 5 centimeters but has not spread to the axillary lymph nodes.

Stage IIIA is defined by either of the following:

The cancer is smaller than 5 centimeters and has spread to the lymph nodes under the arm, and the lymph nodes are attached to each other or to other structures.

The cancer is larger than 5 centimeters and has spread to the lymph nodes under the arm.

Stage IIIB is defined by either of the following:

The cancer has spread to tissues near the breast (skin or chest wall, including the ribs and the muscles in the chest).

The cancer has spread to lymph nodes inside the chest wall along the breast bone.

Stage IV:

The cancer has spread to other organs of the body (most often the bones, lungs, liver, or brain). Or, the tumor has spread locally to the skin and lymph nodes inside the neck, near the collarbone.

TABLE I-3a. Overview of treatment options for invasive cancer of the female breast by tumor stage [adapted from: PDQ 1999].

STAGE I BREAST CANCER

Treatment may be one of the following:

1. Breast-conserving surgery to remove only the cancer and some surrounding breast tissue (lumpectomy) or to remove part of the breast (partial or segmental mastectomy); both are followed by radiation therapy. Some of the axillary lymph nodes are also removed. This treatment provides identical long-term cure rates as those from mastectomy. A doctor's recommendation on which procedure to have is based on tumor size and location and its appearance on the mammogram.

2. Surgery to remove the whole breast (total mastectomy) or the whole breast and the lining over the chest muscles (modified radical mastectomy). Some of the axillary lymph nodes are also taken out.

Adjuvant therapy (given in addition to the treatments listed above):

1. Chemotherapy.

2. Hormone therapy.

3. Clinical trials of more aggressive adjuvant chemotherapy in certain patients.

4. Clinical trials of no adjuvant therapy for patients with a favorable prognosis.

5. Clinical trials of ovarian ablation or suppression.

TABLE I-3b. Overview of treatment options for invasive cancer of the female breast by tumor stage [adapted from: PDQ 1999].

STAGE II BREAST CANCER

Treatment may be one of the following:

1. Breast-conserving surgery to remove only the cancer and some surrounding breast tissue (lumpectomy) or to remove part of the breast (partial or segmental mastectomy); both are followed by radiation therapy. Some of the axillary lymph nodes are also removed. This treatment provides identical long-term cure rates as those from mastectomy. A doctor's recommendation on which procedure to have is based on tumor size and location and its appearance on the mammogram.

2. Surgery to remove the whole breast (total mastectomy) or the whole breast and the lining over the chest muscles (modified radical mastectomy). Some of the axillary lymph nodes are also taken out.

Adjuvant therapy (given in addition to the treatments listed above):

1. Chemotherapy with or without hormonal therapy.

2. Hormone therapy.

3. Clinical trial of chemotherapy before surgery (neoadjuvant therapy).

4. Clinical trials of high-dose chemotherapy with bone marrow transplantation for patients with cancer in more than three lymph nodes.

TABLE I-3c. Overview of treatment options for invasive cancer of the female breast by tumor stage [adapted from: PDQ 1999].

STAGE III BREAST CANCER

Stage IIIA cancer:

Treatment may be one of the following surgeries:

1. Surgery to remove the whole breast, the lining over the chest muscles, and many of the lymph nodes (modified radical mastectomy) or the whole breast, the chest muscles, and all of the lymph nodes (radical mastectomy).

2. Radiation therapy given after surgery.

3. Chemotherapy with or without hormone therapy given with surgery and radiation therapy.

4. Clinical trials are testing new chemotherapy with or without hormonal drugs; they are also testing chemotherapy before surgery (neoadjuvant therapy).

5. Clinical trials of high-dose chemotherapy with bone marrow or peripheral stem cell transplantation.

Stage IIIB cancer:

The patient will probably have a biopsy and then be given one or more of the following:

1. Surgery (radical or modified radical mastectomy) and/or radiation therapy to the breast and the lymph nodes.

2. Chemotherapy with or without hormones to shrink the tumor, followed by surgery and/or radiation therapy.

3. Hormonal therapy followed by additional therapy.

4. Clinical trials are testing new chemotherapy drugs and biological therapy, new drug combinations, and new ways of giving chemotherapy.

5. Clinical trials of high-dose chemotherapy with bone marrow or peripheral stem cell transplantation.

TABLE I-3d. Overview of treatment options for invasive cancer of the female breast by tumor stage [adapted from: PDQ 1999].

STAGE IV BREAST CANCER

The patient will probably have a biopsy and then be given one or more of the following:

1. Radiation therapy or, in some cases, a mastectomy to reduce the symptoms.
2. Hormonal therapy with or without surgery to remove the ovaries.
3. Combination chemotherapy.
4. Clinical trials are testing new chemotherapy and hormonal drugs and new combinations of drugs and biological therapy.
5. Clinical trials of high-dose chemotherapy with bone marrow or peripheral stem cell transplantation.

TABLE II-1. Counties and census tracts (or block numbering areas) included in SEER areas.

SEER Area	No. of Counties	No. of Census Tracts or BNA	Avg. Population Size per Tract or BNA
Los Angeles	1	1,652	5,365
Detroit	3	1,088	3,596
San Francisco & Oakland	5	843	4,373
Connecticut	8	834	3,941
Iowa	99	783	3,546
Seattle	3	754	4,465
San Jose & Monterey	4	546	3,882
Utah	29	400	4,307
New Mexico	33	390	3,885
Atlanta	5	367	5,933
Hawaii	5	265	4,182
All Areas	195	7,922	

SEER = Surveillance, Epidemiology and End Results program.
BNA = block numbering area.
Data source: Bureau of the Census, 1990 (STF-3A data file).

TABLE II-2a. Breast cancer study size requirements for detecting specified odds ratios (OR), where Case = distant stage cancer, Unexposed = white, Exposed = other specific racial/ethnic group, $\alpha = 0.95$, $1-\beta = 0.80$, P(D|Unexposed) = 0.054.

Available Study Size		Study Size Needed	
Unexposed : Exposed	OR	White	**Black**
= White : **Black**	1.5	8,093	861
= 84,446 : **9,031**	1.2	44,218	4,704
= 9.4 : 1	1.1	168,354	17,910
	0.9	155,222	16,513
	0.8	37,628	4,003
	0.6	8,789	935
Unexposed : Exposed	OR	White	**Hispanic**
= White : **Hispanic**	1.5	10,008	841
= 84,446 : **7,074**	1.2	54,776	4,603
= 11.9 : 1	1.1	208,690	17,537
	0.9	192,649	16,189
	0.8	46,731	3,927
	0.6	10,924	918
Unexposed : Exposed	OR	White	**Japanese**
= White : **Japanese**	1.5	35,403	785
= 84,446 : **1,871**	1.4	53,128	1,178
= 45.1 : 1	1.2	194,967	4,323
	0.8	167,637	3,717
	0.7	72,250	1,602
	0.6	39,327	872
Unexposed : Exposed	OR	White	**Filipino**
= White : **Filipino**	1.5	41,759	782
= 84,446 : **1,581**	1.4	62,692	1,174
= 53.4 : 1	1.2	230,047	4,308
	0.8	197,847	3,705
	0.7	85,280	1,597
	0.6	46,405	869
Unexposed : Exposed	OR	White	**Chinese**
= White : **Chinese**	1.5	47,502	780
= 84,446 : **1,387**	1.4	71,314	1,171
= 60.9 : 1	1.3	117,537	1,930
	0.8	225,147	3,697
	0.7	97,075	1,594
	0.6	52,800	867

TABLE II-2b. Breast cancer study size requirements for detecting specified odds ratios (OR), where Case = distant stage cancer, Unexposed = white, Exposed = other specific racial/ethnic group, $\alpha = 0.95$, $1-\beta = 0.80$, P(D|Unexposed) = 0.054.

Available Study Size		Study Size Needed	
Unexposed : Exposed	OR	White	**Hawaiian**
= White : **Hawaiian**	2.0	38,180	230
= 84,446 : **508**	1.7	70,218	423
= 166 : 1	1.5	127,820	770
	0.5	87,980	530
	0.4	58,598	353
	0.3	41,168	248
Unexposed : Exposed	OR	White	**Korean**
= White : **Korean**	2.0	64,349	229
= 84,446 : **301**	1.9	76,713	273
= 281 : 1	1.5	215,808	768
	0.5	148,649	529
	0.4	99,193	353
	0.3	69,407	247
Unexposed : Exposed	OR	White	**Vietnamese**
= White : **Vietnamese**	2.0	70,990	229
= 84,446 : **272**	1.9	84,630	273
= 310 : 1	1.5	238,080	768
	0.5	163,990	529
	0.4	109,430	353
	0.3	76,570	247
Unexposed : Exposed	OR	White	**American Indian (NM)**
= White : **American Indian (NM)**	2.4	81,351	131
= 84,446 : **136**	2.3	91,908	148
= 621 : 1	2.0	141,588	228
	0.3	153,387	247
	0.2	110,538	178
	0.1	79,488	128

TABLE II-3. Individual and area-based study variables.

INDIVIDUAL-LEVEL VARIABLES

1. Cancer type
2. Stage of disease at the time of cancer diagnosis
3. Tumor size at the time of cancer diagnosis
4. Tumor grade
5. Estrogen receptor status
6. Progesterone receptor status
7. Age in years at the time of cancer diagnosis
8. Race/ethnicity
9. Sex
10. Marital status at the time of cancer diagnosis
11. SEER registry
12. Census tract and county of residence at the time of cancer diagnosis

CENSUS TRACT-LEVEL VARIABLES

1. % Employed persons, age 16+, in "working-class" occupations (listed below)
   administrative support and clerical; sales; private household and other service occupations (excl. protective services); precision production, craft and repair; machine operators, assemblers and inspectors; transportation and material moving; handlers, equipment cleaners, helpers and laborers
2. Median family income
3. Median household income
4. % Persons with an income below the poverty line
5. % Families with an income below the poverty level, a female householder (no husband present), and related children <18 years old.
6. % Households owning their home
7. % Households owning no car
8. % Persons, age 25+, that have not completed high school
9. % Persons living in an urban area
10. % Unemployed among persons, age 16+, in labor force
11. % Persons born in foreign country

TABLE II-4. Distribution of staged and unstaged cancers of the female breast in study population by race/ethnicity.

Race/Ethnicity	Staged[a]	Unstaged[b]	Total
White	82,031 (97%)	2,415 (3%)	84,446 (100%)
Black	8,567 (95%)	464 (5%)	9,031 (100%)
Hispanic[c]	6,844 (97%)	230 (3%)	7,074 (100%)
Japanese	1,844 (99%)	27 (1%)	1,871 (100%)
Filipino	1,543 (98%)	38 (2%)	1,581 (100%)
Chinese	1,353 (98%)	34 (2%)	1,387 (100%)
Hawaiian	499 (98%)	9 (2%)	508 (100%)
Korean	287 (95%)	14 (5%)	301 (100%)
Vietnamese	268 (99%)	4 (1%)	272 (100%)
American Indian (NM)[d]	135 (99%)	1 (6%)	136 (100%)
Total	103,371 (97%)	3,236 (3%)	106,607 (100%)

[a] Staged = Extent of disease information was sufficient to assign one of the following stages: localized, regional, or distant disease.

[b] Unstaged = Extent of disease information was insufficient to assign a stage.

[c] Since persons of Hispanic ethnicity may be of any race, Hispanic cases were removed from all other racial/ethnic categories and combined to form this group.

[d] American Indians in New Mexico only.

TABLE II-5. Distribution of staged and unstaged cancers of the female breast in study population by age at diagnosis.

Age at Diagnosis	Staged[a]	Unstaged[b]	Total
<40	6,468 (97%)	223 (3%)	6,691 (100%)
40-49	18,199 (98%)	439 (2%)	18,638 (100%)
50-59	20,523 (98%)	471 (2%)	20,994 (100%)
60-69	23,391 (98%)	497 (2%)	23,888 (100%)
70-79	22,867 (97%)	635 (3%)	23,502 (100%)
80-89	10,481 (94%)	702 (6%)	11,183 (100%)
90+	1,442 (84%)	269 (16%)	1,711 (100%)
All Ages	103,371 (97%)	3,236 (3%)	106,607 (100%)

[a] Staged = Extent of disease information was sufficient to assign one of the following stages: localized, regional, or distant disease.

[b] Unstaged = Extent of disease information was insufficient to assign a stage.

TABLE II-6. Distribution of staged cancers of the female breast in study population by registry and availability of census tract-level socioeconomic information.

Registry	Socioeconomic Information Available? Yes	No	Total
San Francisco & Oakland	11,902 (96%)	535 (4%)	12,437 (100%)
Connecticut	11,705 (99%)	107 (1%)	11,812 (100%)
Detroit	12,652 (99%)	110 (1%)	12,762 (100%)
Hawaii	2,594 (86%)	425 (14%)	3,019 (100%)
Iowa	9,715 (99%)	84 (1%)	9,799 (100%)
New Mexico	3,875 (93%)	299 (7%)	4,174 (100%)
Seattle	10,420 (91%)	975 (9%)	11,395 (100%)
Utah	3,813 (>99%)	16 (<1%)	3,829 (100%)
Atlanta	5,675 (93%)	418 (7%)	6,093 (100%)
San Jose & Monterey	5,433 (93%)	402 (7%)	5,835 (100%)
Los Angeles	21,714 (98%)	502 (2%)	22,216 (100%)
Total	99,498 (96%)	3,873 (4%)	103,371 (100%)

TABLE II-7. Distribution of staged cancers of the female breast in study population by race/ethnicity and availability of census tract-level socioeconomic information.

Race/Ethnicity	Socioeconomic Information Available? Yes	No	Total
White	79,045 (96%)	2,986 (4%)	82,031 (100%)
Black	8,329 (97%)	238 (3%)	8,567 (100%)
Hispanic[a]	6,579 (96%)	265 (4%)	6,844 (100%)
Japanese	1,733 (94%)	111 (6%)	1,844 (100%)
Filipino	1,442 (93%)	101 (7%)	1,543 (100%)
Chinese	1,309 (97%)	44 (3%)	1,353 (100%)
Hawaiian	400 (80%)	99 (20%)	499 (100%)
Korean	277 (97%)	10 (3%)	287 (100%)
Vietnamese	257 (96%)	11 (4%)	268 (100%)
American Indian (NM)[b]	127 (94%)	8 (6%)	135 (100%)
Total	99,498 (96%)	3,873 (4%)	103,371 (100%)

[a] Since persons of Hispanic ethnicity may be of any race, Hispanic cases were removed from all other racial/ethnic categories and combined to form this group.

[b] American Indians in New Mexico only.

TABLE II-8. Distribution of female breast cancer cases in study population by race/ethnicity and availability of tumor size information.

Race/Ethnicity	Tumor Size Available? Yes	No	Paget's[a]	Total
White	76,843 (91.0%)	7,512 (8.9%)	91 (0.1%)	84,446 (100%)
Black	7,963 (88.2%)	1,062 (11.7%)	6 (0.1%)	9,031 (100%)
Hispanic[b]	6,463 (91.4%)	605 (8.5%)	6 (0.1%)	7,074 (100%)
Japanese	1,731 (92.5%)	137 (7.3%)	3 (0.2%)	1,871 (100%)
Filipino	1,467 (92.8%)	112 (7.1%)	2 (0.1%)	1,581 (100%)
Chinese	1,267 (91.4%)	34 (8.5%)	2 (0.1%)	1,387 (100%)
Hawaiian	470 (92.5%)	38 (7.5%)	0 (0.0%)	508 (100%)
Korean	280 (93.0%)	21 (7.0%)	0 (0.0%)	301 (100%)
Vietnamese	259 (95.2%)	13 (4.8%)	0 (0.0%)	272 (100%)
American Indian (NM)[c]	128 (94.1%)	8 (5.9%)	0 (0.0%)	136 (100%)
Total	96,871 (90.9%)	9,626 (9.0%)	110 (0.1%)	106,607 (100%)

[a] Paget's disease without an underlying tumor

[b] Since persons of Hispanic ethnicity may be of any race, Hispanic cases were removed from all other racial/ethnic categories and combined to form this group.

[c] American Indians in New Mexico only.

TABLE II-9. Distribution of female breast cancer cases in study population by age at diagnosis and availability of tumor size information.

Age at Diagnosis	Tumor Size Available? Yes	No	Paget's[a]	Total
<40	6,073 (90.8%)	612 (9.1%)	6 (0.1%)	6,691 (100%)
40-49	17,009 (91.3%)	1,619 (8.7%)	10 (<0.1%)	18,638 (100%)
50-59	19,172 (91.3%)	1,812 (8.6%)	10 (<0.1%)	20,994 (100%)
60-69	21,785 (91.2%)	2,066 (8.6%)	37 (0.2%)	23,888 (100%)
70-79	21,478 (91.4%)	1,999 (8.5%)	25 (0.1%)	23,502 (100%)
80-89	9,961 (89.1%)	1,201 (10.7%)	21 (0.2%)	11,183 (100%)
90+	1,393 (81.4%)	317 (18.5%)	1 (0.1%)	1,711 (100%)
All Ages	96,871 (90.8%)	9,626 (9.0%)	110 (0.1%)	106,607 (100%)

[a] Paget's disease without an underlying tumor

TABLE II-10. Distribution of female breast cancer cases in study population with tumor size information by registry and availability of census tract-level socioeconomic information.

Registry	Socioeconomic Information Available? Yes	No	Total
San Francisco & Oakland	11,294 (96%)	491 (4%)	11,785 (100%)
Connecticut	10,343 (99%)	91 (1%)	10,434 (100%)
Detroit	11,539 (99%)	102 (1%)	11,641 (100%)
Hawaii	2,440 (86%)	390 (14%)	2,830 (100%)
Iowa	9,280 (99%)	80 (1%)	9,360 (100%)
New Mexico	3,654 (93%)	281 (7%)	3,935 (100%)
Seattle	10,127 (91%)	944 (9%)	11,071 (100%)
Utah	3,597 (>99%)	14 (<1%)	3,611 (100%)
Atlanta	5,268 (93%)	387 (7%)	5,655 (100%)
San Jose & Monterey	4,974 (93%)	358 (7%)	5,332 (100%)
Los Angeles	20,743 (98%)	474 (2%)	21,217 (100%)
Total	93,259 (96%)	3,612 (4%)	96,871 (100%)

TABLE II-11. Distribution of female breast cancer cases in study population with tumor size information by race/ethnicity and availability of census tract-level socioeconomic information.

Race/Ethnicity	Socioeconomic Information Available? Yes	No	Total
White	74,053 (96%)	2,790 (4%)	76,843 (100%)
Black	7,744 (97%)	219 (3%)	7,963 (100%)
Hispanic[1]	6,215 (96%)	248 (4%)	6,463 (100%)
Japanese	1,630 (94%)	101 (6%)	1,731 (100%)
Filipino	1,369 (93%)	98 (7%)	1,467 (100%)
Chinese	1,230 (97%)	37 (3%)	1,267 (100%)
Hawaiian	379 (81%)	91 (19%)	470 (100%)
Korean	270 (96%)	10 (4%)	280 (100%)
Vietnamese	248 (96%)	11 (4%)	259 (100%)
American Indian (NM)[2]	121 (95%)	7 (5%)	128 (100%)
Total	93,259 (96%)	3,612 (4%)	96,871 (100%)

[1] Since persons of Hispanic ethnicity may be of any race, Hispanic cases were removed from all other racial/ethnic categories and combined to form this group.

[2] American Indians in New Mexico only.

TABLE III-1a. Distribution of selected characteristics among 106,607 female breast cancer patients, diagnosed 1992-1996.

Characteristics	White		Black		Hispanic		Japanese		Filipino		Chinese	
	%	(No.)	%	(No.)	%	(No.)	%	(No.)	%	(No.)	%	(No.)
Total	100.0	(84446)	100.0	(9031)	100.0	(7074)	100.0	(1871)	100.0	(1581)	100.0	(1387)
Age at diagnosis												
<35	1.7	(1431)	4.6	(419)	5.6	(397)	1.8	(33)	2.5	(39)	3.5	(49)
35-49	19.3	(16282)	28.8	(2602)	30.9	(2187)	19.3	(362)	36.4	(575)	34.6	(480)
50+	79.0	(66733)	66.6	(6010)	63.5	(4490)	78.9	(1476)	61.1	(967)	61.9	(858)
Stage of disease												
localized	64.9	(54780)	52.8	(4769)	56.6	(4006)	71.6	(1340)	62.2	(983)	63.2	(877)
regional	27.0	(22854)	33.2	(2994)	33.4	(2364)	22.9	(428)	30.5	(482)	29.1	(404)
distant	5.2	(4397)	8.9	(804)	6.7	(474)	4.1	(76)	4.9	(78)	5.2	(72)
unknown	2.9	(2415)	5.1	(464)	3.3	(230)	1.4	(27)	2.4	(38)	2.5	(34)
Tumor size												
< 1.0 cm	17.3	(14594)	9.9	(895)	11.5	(812)	21.2	(396)	12.0	(190)	14.0	(194)
1.0 - 1.9 cm	35.0	(29567)	26.9	(2425)	27.2	(1924)	38.2	(716)	28.9	(457)	33.4	(464)
2.0+ cm	38.7	(32682)	51.4	(4643)	52.7	(3727)	33.1	(619)	51.9	(820)	43.9	(609)
unknown or Paget's disease	9.0	(7603)	11.8	(1068)	8.6	(611)	7.5	(140)	7.2	(114)	8.7	(120)
Tumor grade												
1 (well differentiated)	12.3	(10348)	7.3	(664)	8.7	(617)	13.2	(247)	9.1	(144)	9.2	(128)
2 (moderately differentiated)	32.1	(27138)	22.6	(2038)	28.4	(2006)	33.7	(631)	32.6	(515)	33.0	(458)
3 or 4 (poorly or undifferentiated)	30.4	(25717)	40.2	(3630)	38.0	(2688)	30.1	(562)	34.4	(544)	35.8	(497)
unknown	25.2	(21243)	29.9	(2699)	24.9	(1763)	23.0	(431)	23.9	(378)	21.9	(304)
ER status												
positive	59.4	(50163)	41.2	(3723)	47.5	(3359)	63.8	(1194)	56.8	(898)	55.3	(767)
negative	17.6	(14851)	27.8	(2514)	21.7	(1537)	19.1	(358)	21.5	(340)	21.7	(301)
borderline	0.8	(676)	0.9	(81)	0.6	(44)	0.4	(7)	0.3	(4)	0.9	(12)
not done or unknown	22.2	(18756)	30.1	(2713)	30.1	(2134)	16.7	(312)	21.4	(339)	22.1	(307)
PR status												
positive	50.0	(42231)	35.4	(3201)	40.9	(2897)	56.0	(1047)	49.9	(789)	49.5	(687)
negative	24.2	(20437)	31.7	(2858)	26.3	(1861)	25.4	(475)	25.9	(409)	25.7	(356)
borderline	1.0	(856)	0.9	(80)	0.7	(47)	0.6	(12)	0.7	(12)	0.8	(11)
not done or unknown	24.8	(20922)	32.0	(2892)	32.1	(2269)	18.0	(337)	23.5	(371)	24.0	(333)
Marital status												
married	55.0	(46500)	35.9	(3244)	53.1	(3760)	61.6	(1152)	61.7	(975)	67.8	(940)
not married	42.3	(35696)	59.4	(5367)	44.4	(3138)	37.0	(693)	36.3	(574)	29.9	(415)
unknown	2.7	(2250)	4.6	(420)	2.5	(176)	1.4	(26)	2.0	(32)	2.3	(32)

TABLE III-1a, cont. Distribution of selected characteristics among 106,607 female breast cancer patients, diagnosed 1992-1996.

Characteristics	White		Black		Hispanic		Japanese		Filipino		Chinese	
	%	(No.)	%	(No.)	%	(No.)	%	(No.)	%	(No.)	%	(No.)
Urban residence												
urban area	70.9	(59883)	92.0	(8304)	83.6	(5918)	88.4	(1654)	87.1	(1377)	93.1	(1291)
not urban area	25.3	(21363)	5.0	(454)	12.3	(868)	5.6	(104)	6.0	(95)	3.5	(49)
unknown	3.8	(3200)	3.0	(273)	4.1	(288)	6.0	(113)	6.9	(109)	3.4	(47)
SEER area												
Atlanta	5.6	(4768)	18.0	(1628)	0.7	(49)	0.2	(3)	0.3	(5)	0.4	(5)
Connecticut	13.5	(11368)	6.9	(626)	4.5	(321)	0.5	(10)	0.1	(2)	0.6	(8)
Detroit	12.3	(10409)	29.1	(2626)	0.9	(61)	0.3	(6)	1.2	(19)	1.4	(20)
Hawaii	1.1	(951)	0.3	(26)	0.3	(24)	54.7	(1024)	18.9	(299)	14.8	(206)
Iowa	11.7	(9876)	1.2	(107)	0.7	(47)	0.1	(1)	0.2	(3)	0.3	(4)
Los Angeles	17.8	(14994)	27.8	(2506)	52.0	(3678)	23.2	(435)	38.1	(602)	29.0	(402)
New Mexico	3.5	(2977)	0.5	(48)	15.4	(1088)	0.3	(5)	0.3	(4)	0.5	(7)
San Jose & Monterey	5.6	(4712)	1.3	(119)	9.5	(675)	4.9	(91)	10.3	(163)	9.9	(138)
San Francisco Bay Area	11.3	(9535)	12.0	(1080)	12.8	(904)	9.3	(174)	24.4	(386)	38.8	(538)
Seattle	13.1	(11039)	2.8	(254)	1.7	(123)	5.8	(108)	6.1	(96)	4.0	(55)
Utah	4.5	(3817)	0.1	(11)	1.5	(104)	0.7	(14)	0.1	(5)	0.3	(4)
% Without high school diploma												
0-8	22.7	(19206)	4.5	(410)	7.7	(548)	14.6	(273)	8.0	(127)	22.3	(309)
9-28	61.8	(52205)	36.0	(3249)	39.2	(2771)	61.8	(1156)	50.0	(791)	50.5	(700)
29+	11.7	(9835)	56.5	(5099)	49.0	(3467)	17.6	(329)	35.1	(554)	23.4	(331)
unknown	3.8	(3200)	3.0	(273)	4.1	(288)	6.0	(113)	6.9	(109)	3.4	(47)
% Below poverty level												
0-2	16.5	(13924)	2.8	(255)	4.8	(342)	19.0	(355)	8.7	(138)	13.6	(189)
3-13	65.9	(55678)	31.7	(2860)	45.3	(3206)	62.7	(1173)	59.3	(938)	60.2	(835)
14+	13.8	(11633)	62.5	(5642)	45.8	(3238)	12.3	(230)	25.1	(396)	22.7	(315)
unknown	3.8	(3211)	3.0	(274)	4.1	(288)	6.0	(113)	6.9	(109)	3.5	(48)
% Working class												
0-50	21.1	(17836)	5.7	(514)	8.7	(616)	18.0	(336)	6.9	(109)	24.1	(334)
51-73	61.1	(51586)	42.3	(3821)	47.5	(3363)	63.3	(1185)	54.5	(861)	56.0	(777)
74+	14.0	(11811)	49.0	(4422)	39.7	(2807)	12.7	(237)	31.7	(502)	16.4	(228)
unknown	3.8	(3213)	3.0	(274)	4.1	(288)	6.0	(113)	6.9	(109)	3.5	(48)

TABLE III-1a, cont. Distribution of selected characteristics among 106,607 female breast cancer patients, diagnosed 1992-1996.

Characteristics	White %	White (No.)	Black %	Black (No.)	Hispanic %	Hispanic (No.)	Japanese %	Japanese (No.)	Filipino %	Filipino (No.)	Chinese %	Chinese (No.)
Median family income ($ thousands)												
<25	5.1	(4301)	37.9	(3424)	21.2	(1500)	2.9	(53)	8.5	(134)	8.2	(114)
25 - 49	54.6	(46119)	48.8	(4409)	57.5	(4069)	49.1	(919)	57.4	(907)	45.5	(632)
50+	36.5	(30862)	10.3	(925)	17.2	(1217)	42.0	(786)	27.2	(431)	42.8	(594)
unknown	3.8	(3200)	3.0	(273)	4.1	(288)	6.0	(113)	6.9	(109)	3.5	(47)
% Unemployed												
0-2	13.5	(11394)	2.0	(181)	4.5	(318)	31.8	(595)	8.0	(127)	18.3	(254)
3-7	70.4	(59488)	27.8	(2507)	47.4	(3352)	54.5	(1019)	61.5	(972)	63.7	(884)
8+	12.3	(10351)	67.2	(6069)	44.0	(3116)	7.7	(144)	23.6	(373)	14.5	(201)
unknown	3.8	(3213)	3.0	(274)	4.1	(288)	6.0	(113)	6.9	(109)	3.5	(48)
% Families with female head of household, no husband, 1 or more children <18 years old, and income below poverty level												
0	27.3	(23028)	5.8	(523)	11.8	(837)	33.1	(619)	18.1	(287)	29.6	(411)
1-4	54.1	(45708)	24.1	(2173)	43.8	(3100)	50.8	(950)	52.1	(823)	54.4	(755)
5+	14.8	(12496)	67.1	(6057)	40.3	(2849)	10.1	(189)	22.9	(362)	12.5	(173)
unknown	3.8	(3214)	3.1	(278)	4.1	(288)	6.0	(113)	6.9	(109)	3.5	(48)
% Own their home												
0-45	16.1	(13551)	39.3	(3553)	35.5	(2509)	25.2	(471)	29.1	(460)	32.6	(452)
46-85	59.7	(50423)	51.4	(4641)	52.5	(3715)	56.0	(1047)	54.7	(865)	47.3	(656)
86+	20.4	(17259)	6.2	(561)	7.9	(562)	12.8	(240)	9.3	(147)	16.7	(231)
unknown	3.8	(3213)	3.1	(276)	4.1	(288)	6.0	(113)	6.9	(109)	3.5	(48)
% Do not own a car												
0-2	27.2	(22929)	6.2	(562)	12.8	(907)	22.5	(421)	20.7	(327)	23.3	(323)
3-12	56.5	(47715)	28.1	(2542)	53.2	(3766)	49.4	(925)	46.6	(736)	41.7	(579)
13+	12.5	(10589)	62.6	(5651)	29.9	(2113)	22.0	(412)	25.9	(409)	31.5	(437)
unknown	3.8	(3213)	3.1	(276)	4.1	(288)	6.1	(113)	6.9	(109)	3.5	(48)
% Foreign-born												
0-3	23.7	(19999)	36.7	(3310)	7.7	(544)	1.7	(31)	1.3	(21)	0.6	(8)
4-21	58.8	(49707)	38.6	(3488)	35.6	(2519)	59.7	(1117)	31.1	(492)	40.1	(556)
22+	13.7	(11540)	21.7	(1960)	52.6	(3723)	32.6	(610)	60.7	(959)	55.9	(776)
unknown	3.8	(3200)	3.0	(273)	4.1	(288)	6.0	(113)	6.9	(109)	3.4	(41)

TABLE III-1b. Distribution of selected characteristics among 106,607 female breast cancer patients, diagnosed 1992-1996.

Characteristics	Hawaiian %	Hawaiian (No.)	Korean %	Korean (No.)	Vietnamese %	Vietnamese (No.)	American Indian %	American Indian (No.)
Total	100.0	(508)	100.0	(301)	100.0	(272)	100.0	(136)
Age at diagnosis								
<35	3.2	(16)	7.0	(21)	7.0	(19)	0.7	(1)
35-49	25.2	(128)	41.5	(125)	44.5	(121)	30.9	(42)
50+	71.6	(364)	51.5	(155)	48.5	(132)	68.4	(93)
Stage of disease								
localized	61.4	(312)	63.1	(190)	57.0	(155)	54.4	(74)
regional	29.9	(152)	28.6	(86)	37.5	(102)	36.8	(50)
distant	6.9	(35)	3.6	(11)	4.0	(11)	8.1	(11)
unknown	1.8	(9)	4.7	(14)	1.5	(4)	0.7	(1)
Tumor size								
<1.0 cm	13.0	(66)	10.6	(32)	13.2	(36)	16.2	(22)
1.0 - 1.9 cm	36.0	(183)	34.6	(104)	24.6	(67)	30.9	(42)
2.0+ cm	43.5	(221)	47.8	(144)	57.4	(156)	47.0	(64)
unknown or Paget's disease	7.5	(38)	7.0	(21)	4.8	(13)	5.9	(8)
Tumor grade								
1 (well differentiated)	8.1	(41)	7.6	(23)	9.2	(25)	5.2	(7)
2 (moderately differentiated)	29.1	(148)	28.6	(86)	29.4	(80)	36.8	(50)
3 or 4 (poorly or undifferentiated)	37.8	(192)	41.9	(126)	44.1	(120)	36.0	(49)
unknown	25.0	(127)	21.9	(66)	17.3	(47)	22.1	(30)
ER status								
positive	68.3	(347)	50.5	(152)	48.2	(131)	54.4	(74)
negative	19.3	(98)	26.6	(80)	22.8	(62)	26.5	(36)
borderline	1.4	(7)	0.6	(2)	0.7	(2)	1.5	(2)
not done or unknown	11.0	(56)	22.3	(67)	28.3	(77)	17.6	(24)
PR status								
positive	64.0	(325)	44.2	(137)	43.4	(118)	42.6	(58)
negative	22.8	(116)	31.9	(98)	27.2	(74)	37.5	(51)
borderline	0.8	(4)	0.7	(2)	0.0	(0)	1.5	(2)
not done or unknown	12.4	(63)	23.3	(106)	29.4	(80)	18.4	(25)
Marital status								
married	55.7	(283)	68.4	(206)	66.2	(180)	61.8	(84)
not married	43.7	(222)	29.6	(89)	32.7	(89)	34.5	(47)
unknown	0.6	(3)	2.0	(6)	1.1	(3)	3.7	(5)

TABLE III-1b, cont. Distribution of selected characteristics among 106,607 female breast cancer patients, diagnosed 1992-1996.

Characteristics	Hawaiian %	Hawaiian (No.)	Korean %	Korean (No.)	Vietnamese %	Vietnamese (No.)	American Indian %	American Indian (No.)
Urban residence								
urban area	59.6	(303)	90.7	(273)	93.8	(255)	7.3	(10)
not urban area	20.1	(102)	5.0	(15)	1.8	(5)	86.8	(118)
unknown	20.3	(103)	4.3	(13)	4.4	(12)	5.9	(8)
SEER area								
Atlanta	0.0	(0)	1.0	(3)	0.7	(2)	0.0	(0)
Connecticut	0.0	(0)	1.0	(3)	1.8	(5)	0.0	(0)
Detroit	0.2	(1)	1.0	(3)	1.1	(3)	0.0	(0)
Hawaii	93.1	(473)	20.3	(61)	1.4	(4)	0.0	(0)
Iowa	0.0	(0)	0.7	(2)	0.4	(1)	0.0	(0)
Los Angeles	3.1	(16)	45.5	(137)	44.5	(121)	0.0	(0)
New Mexico	0.2	(1)	1.7	(5)	0.4	(1)	100.0	(136)
San Jose & Monterey	0.8	(4)	7.6	(23)	22.8	(62)	0.0	(0)
San Francisco Bay Area	1.6	(8)	10.0	(30)	16.2	(44)	0.0	(0)
Seattle	0.6	(3)	9.3	(28)	9.6	(26)	0.0	(0)
Utah	0.4	(2)	2.0	(6)	1.1	(3)	0.0	(0)
% Without high school diploma								
0-8	5.1	(26)	15.6	(47)	8.8	(24)	0.7	(1)
9-28	47.8	(243)	59.5	(179)	46.3	(126)	30.9	(42)
29+	26.8	(136)	20.6	(62)	40.5	(110)	62.5	(85)
unknown	20.3	(103)	4.3	(13)	4.4	(12)	5.9	(8)
% Below poverty level								
0-2	12.2	(62)	10.7	(32)	8.1	(22)	0.7	(1)
3-13	47.4	(241)	62.1	(187)	48.2	(131)	6.6	(9)
14+	20.1	(102)	22.9	(69)	39.3	(107)	86.8	(118)
unknown	20.3	(103)	4.3	(13)	4.4	(12)	5.9	(8)
% Working class								
0-50	5.1	(26)	17.3	(52)	8.5	(23)	0.0	(0)
51-73	57.9	(294)	61.1	(184)	54.0	(147)	61.0	(83)
74+	16.7	(85)	17.3	(52)	33.1	(90)	33.1	(45)
unknown	20.3	(103)	4.3	(13)	4.4	(12)	5.9	(8)

TABLE III-1b, cont. Distribution of selected characteristics among 106,607 female breast cancer patients, diagnosed 1992-1996.

Characteristics	Hawaiian %	Hawaiian (No.)	Korean %	Korean (No.)	Vietnamese %	Vietnamese (No.)	American Indian %	American Indian (No.)
Median family income ($ thousands)								
<25	4.5	(23)	11.6	(35)	15.8	(43)	71.3	(97)
25 - 49	52.6	(267)	50.9	(153)	55.5	(151)	22.8	(31)
50+	22.6	(115)	33.2	(100)	24.3	(66)	0.0	(0)
unknown	20.3	(103)	4.3	(13)	4.4	(12)	5.9	(8)
% Unemployed								
0-2	23.4	(119)	17.9	(54)	6.3	(17)	0.0	(0)
3-7	46.3	(235)	56.5	(170)	59.9	(163)	17.6	(24)
8+	10.0	(51)	21.3	(64)	29.4	(80)	76.5	(104)
unknown	20.3	(103)	4.3	(13)	4.4	(12)	5.9	(8)
Families with female head of household, no husband, 1 or more children <18 years old, and income below poverty level								
0	17.5	(89)	23.9	(72)	15.4	(42)	1.5	(2)
1-4	44.1	(224)	58.5	(176)	44.5	(121)	12.5	(17)
5+	18.1	(92)	13.3	(40)	35.7	(97)	80.1	(109)
unknown	20.3	(103)	4.3	(13)	4.4	(12)	5.9	(8)
% Own their home								
0-45	16.5	(84)	40.5	(122)	39.0	(106)	3.7	(5)
46-85	58.3	(296)	43.5	(131)	49.6	(135)	61.7	(84)
86+	4.9	(25)	11.7	(35)	7.0	(19)	28.7	(39)
unknown	20.3	(103)	4.3	(13)	4.4	(12)	5.9	(8)
% Do not own a car								
0-2	15.5	(79)	25.6	(77)	17.3	(47)	5.9	(8)
3-12	45.7	(232)	40.5	(122)	49.3	(134)	29.4	(40)
13+	18.5	(94)	29.6	(89)	29.0	(79)	58.8	(80)
unknown	20.3	(103)	4.3	(13)	4.4	(12)	5.9	(8)
% Foreign-born								
0-3	8.1	(41)	3.6	(11)	2.2	(6)	80.9	(110)
4-21	52.7	(268)	37.9	(114)	30.2	(82)	13.2	(18)
22+	18.9	(96)	54.2	(163)	63.2	(172)	0.0	(0)
unknown	20.3	(103)	4.3	(13)	4.4	(12)	5.9	(8)

TABLE III-2. Percentage distribution of 106,607 invasive breast cancers by histological type and racial/ethnic group.

	Ductal Adenocarc.[a]	Lobular	Adenocarc.[a] NOS	Mucinous	Medullary	Inflammatory	All Other	TOTAL
White	73	14	4	3	1	1	4	100
Black	74	9	4	2	3	2	6	100
Hispanic	73	11	4	2	2	2	6	100
Japanese	79	8	4	3	1	1	4	100
Filipino	80	8	2	3	1	1	5	100
Chinese	79	6	3	4	2	1	5	100
Hawaiian	80	7	3	4	< 1	1	5	100
Korean	82	7	4	2	2	1	2	100
Vietnamese	76	6	4	4	2	1	7	100
Am. Indian	77	10	4	1	4	1	3	100

[a] Adenocarc. = adenocarcinoma.

TABLE III-3. Distribution of selected characteristics among invasive breast cancer patients by tumor stage.

	Distant (Total = 5,698)	Localized (Total = 65,018)		
	n	n	OR[a]	95% CI
Tumor grade				
1-2, unknown	3,198	47,952	1.0	
3-4	2,500	17,066	2.2	2.1, 2.3
Hormone receptor status				
ER or PR positive	2,344	40,703	1.0	
All other	3,354	24,315	2.4	2.3, 2.5
Marital status				
Married or unknown	2,683	37,267	1.0	
Not married	3,015	27,751	1.5	1.4, 1.6
Urban census tract				
Not urban	1,219	14,848	1.0	
Urban	4,479	50,170	1.1	1.0, 1.2

[a] OR = crude odds ratio.

TABLE III-4. Distribution of selected characteristics among invasive breast cancer patients by tumor stage.

	Regional (Total = 28,782)	Localized (Total = 65,018)		
	n	n	OR[a]	95% CI
Tumor grade				
1-2, unknown	16,249	47,952	1.0	
3-4	12,533	17,066	2.2	2.1, 2.2
Hormone receptor status				
ER or PR positive	18,040	40,703	1.0	
All other	10,742	24,315	1.0	1.0, 1.0
Marital status				
Married or unknown	16,617	37,267	1.0	
Not married	12,165	27,751	1.0	1.0, 1.0
Urban census tract				
Not urban	6,393	14,848	1.0	
Urban	22,389	50,170	1.0	1.0, 1.1

[a] OR = crude odds ratio.

TABLE III-5. Distribution of selected characteristics among invasive breast cancer patients by tumor size.

	1.0 cm+ (Total = 76,626)	<1.0 cm (Total = 16,633)		
	n	n	OR[a]	95% CI
Tumor grade				
1-2, unknown	48,711	14,008	1.0	
3-4	27,915	2,625	3.1	2.9, 3.2
Hormone receptor status				
ER or PR positive	48,248	10,302	1.0	
All other	28,378	6,331	1.0	0.9, 1.0
Marital status				
Married or unknown	42,817	10,057	1.0	
Not married	33,809	6,576	1.2	1.2, 1.3
Urban census tract				
Not urban	17,146	3,847	1.0	
Urban	59,480	12,786	1.0	1.0, 1.1

[a] OR = crude odds ratio.

Footnotes for **Table III-6.**

[a] OR = odds ratio is adjusted for other explanatory variables in the regression model.

[b] SDF = sociodemographic factors include % not married at time of diagnosis; % without high school diploma; % working class; % families headed by women with no husband at home, with one or more children, and who are living below the poverty level; median family income; % own their home; % having no car.

TABLE III-6. Effects of selected risk factors on the relative odds of a distant stage breast cancer diagnosis among specific racial/ethnic groups compared with whites.

Variables in model	Distant vs. Localized	
	OR[a]	(95% CI)
Age		
White	1.0	
Black	2.1	(1.9, 2.2)
Hispanic	1.5	(1.4, 1.7)
Japanese	0.7	(0.5, 0.9)
Filipino	1.0	(0.7, 1.2)
Chinese	1.1	(0.8, 1.3)
Hawaiian	1.4	(0.9, 2.0)
Korean	0.7	(0.3, 1.2)
Vietnamese	0.8	(0.4, 1.5)
Am. Indian	2.0	(1.0, 3.7)
Age, registry, urban area		
White	1.0	
Black	2.0	(1.9, 2.2)
Hispanic	1.5	(1.3, 1.6)
Japanese	0.7	(0.6, 1.0)
Filipino	1.0	(0.8, 1.3)
Chinese	1.2	(0.9, 1.5)
Hawaiian	1.5	(1.0, 2.3)
Korean	0.7	(0.4, 1.3)
Vietnamese	0.9	(0.4, 1.6)
Am. Indian	1.8	(0.9, 3.3)
Age, registry, urban area, SDF[b]		
White	1.0	
Black	1.5	(1.3, 1.6)
Hispanic	1.2	(1.0, 1.3)
Japanese	0.7	(0.5, 0.9)
Filipino	0.9	(0.7, 1.1)
Chinese	1.1	(0.8, 1.4)
Hawaiian	1.3	(0.8, 2.0)
Korean	0.6	(0.3, 1.2)
Vietnamese	0.7	(0.4, 1.3)
Am. Indian	1.3	(0.7, 2.5)
Age, registry, urban area, SDF[b], tumor grade, er/pr status		
White	1.0	
Black	1.3	(1.2, 1.5)
Hispanic	1.1	(1.0, 1.2)
Japanese	0.7	(0.6, 1.0)
Filipino	0.9	(0.7, 1.1)
Chinese	1.0	(0.8, 1.3)
Hawaiian	1.4	(0.9, 2.1)
Korean	0.6	(0.3, 1.0)
Vietnamese	0.7	(0.3, 1.2)
Am. Indian	1.4	(0.7, 2.6)

Footnotes for **Table III-7.**

[a] OR = odds ratio is adjusted for other explanatory variables in the regression model.

[b] SDF = sociodemographic factors include % not married at time of diagnosis; % without high school diploma; % working class; % families headed by women with no husband at home, with one or more children, and who are living below the poverty level; median family income; % own their home; % having no car.

TABLE III-7. Effects of selected risk factors on the relative odds of a regional stage breast cancer diagnosis among specific racial/ethnic groups compared with whites.

Variables in model	Regional vs. Localized	
	OR[a]	(95% CI)
Age		
White	1.0	
Black	1.4	(1.3, 1.4)
Hispanic	1.3	(1.2, 1.3)
Japanese	0.7	(0.7, 0.8)
Filipino	1.0	(0.9, 1.2)
Chinese	1.0	(0.9, 1.1)
Hawaiian	1.0	(0.8, 1.2)
Korean	0.9	(0.7, 1.2)
Vietnamese	1.3	(1.0, 1.6)
Am. Indian	1.5	(1.0, 2.2)
Age, registry, urban area		
White	1.0	
Black	1.3	(1.3, 1.4)
Hispanic	1.2	(1.2, 1.3)
Japanese	0.8	(0.7, 0.9)
Filipino	1.1	(0.9, 1.2)
Chinese	1.0	(0.9, 1.2)
Hawaiian	1.1	(0.9, 1.4)
Korean	0.9	(0.7, 1.2)
Vietnamese	1.2	(1.0, 1.6)
Am. Indian	1.4	(1.0, 2.0)
Age, registry, urban area, SDF[b]		
White	1.0	
Black	1.2	(1.2, 1.3)
Hispanic	1.1	(1.1, 1.2)
Japanese	0.8	(0.7, 0.9)
Filipino	1.0	(0.9, 1.1)
Chinese	1.0	(0.9, 1.2)
Hawaiian	1.1	(0.8, 1.4)
Korean	0.9	(0.7, 1.2)
Vietnamese	1.2	(0.9, 1.5)
Am. Indian	1.3	(0.9, 1.8)
Age, registry, urban area, SDF[b], tumor grade, er/pr status		
White	1.0	
Black	1.2	(1.1, 1.2)
Hispanic	1.1	(1.1, 1.2)
Japanese	0.8	(0.7, 0.9)
Filipino	1.0	(0.9, 1.1)
Chinese	1.0	(0.9, 1.1)
Hawaiian	1.0	(0.8, 1.3)
Korean	0.9	(0.7, 1.1)
Vietnamese	1.1	(0.9, 1.5)
Am. Indian	1.3	(0.9, 1.9)

Footnotes for **Table III-8.**

[a] OR = odds ratio is adjusted for other explanatory variables in the regression model.

[b] SDF = sociodemographic factors include % not married at time of diagnosis; % without high school diploma; % working class; % families headed by women with no husband at home, with one or more children, and who are living below the poverty level; median family income; % own their home; % having no car.

TABLE III-8. Effects of selected risk factors on the relative odds of a breast cancer diagnosis with tumor diameter greater or equal to 1 cm among specific racial/ethnic groups compared with whites.

Variables in model	Tumor size ge 1cm vs. <1 cm	
	OR[a]	(95% CI)
Age		
White	1.0	
Black	1.8	(1.7, 2.0)
Hispanic	1.6	(1.4, 1.7)
Japanese	0.8	(0.7, 0.9)
Filipino	1.5	(1.3, 1.8)
Chinese	1.2	(1.0, 1.4)
Hawaiian	1.2	(0.9, 1.7)
Korean	1.5	(1.0, 2.2)
Vietnamese	1.3	(0.9, 2.0)
Am. Indian	1.2	(0.7, 2.1)
Age, registry, urban area		
White	1.0	
Black	1.8	(1.7, 2.0)
Hispanic	1.5	(1.3, 1.6)
Japanese	0.9	(0.8, 1.0)
Filipino	1.5	(1.3, 1.8)
Chinese	1.2	(1.0, 1.4)
Hawaiian	1.4	(1.0, 2.0)
Korean	1.5	(1.0, 2.2)
Vietnamese	1.3	(0.9, 1.9)
Am. Indian	1.1	(0.6, 1.9)
Age, registry, urban area, SDF[b]		
White	1.0	
Black	1.5	(1.3, 1.6)
Hispanic	1.2	(1.1, 1.4)
Japanese	0.8	(0.7, 1.0)
Filipino	1.4	(1.2, 1.6)
Chinese	1.2	(1.0, 1.4)
Hawaiian	1.3	(0.9, 1.8)
Korean	1.4	(1.0, 2.1)
Vietnamese	1.1	(0.7, 1.7)
Am. Indian	0.9	(0.5, 1.5)
Age, registry, urban area, SDF[b], tumor grade, er/pr status		
White	1.0	
Black	1.4	(1.3, 1.5)
Hispanic	1.2	(1.1, 1.3)
Japanese	0.9	(0.7, 1.0)
Filipino	1.4	(1.1, 1.6)
Chinese	1.1	(1.0, 1.4)
Hawaiian	1.2	(0.9, 1.7)
Korean	1.4	(1.0, 2.1)
Vietnamese	1.1	(0.7, 1.7)
Am. Indian	0.9	(0.5, 1.6)

TABLE III-9. Estimated percentage of patients diagnosed with distant stage disease by education[a], hormone receptor status[b] and tumor grade; adjusted for all other factors[c] in full regression model.

	Percentage	95% CI
Japanese		
high education, positive hormone receptor, low tumor grade	3.1	(2.3, 3.9)
low education, negative/other hormone receptor, high tumor grade	15.0	(11.1, 18.9)
White		
high education, positive hormone receptor, low tumor grade	4.1	(3.8, 4.4)
low education, negative/other hormone receptor, high tumor grade	19.3	(17.1, 21.5)
Hispanic		
high education, positive hormone receptor, low tumor grade	4.4	(3.8, 5.0)
low education, negative/other hormone receptor, high tumor grade	20.3	(17.9, 22.9)
Black		
high education, positive hormone receptor, low tumor grade	5.3	(4.7, 5.9)
low education, negative/other hormone receptor, high tumor grade	23.8	(20.7, 27.1)

[a] High education is defined as the midpoint of the lower two quintiles in the distribution of the percentages of persons within census tracts with no high school diploma (equal to 6.5% in this study population). Low education is defined as the midpoint of the highest two quintiles in the distribution of the proportions of persons within census tract with no high school diploma (equal to 52% in this study population).

[b] Hormone receptor status and tumor grade are the same binary variables used in regression models (Tables III-5-7).

[c] All other factors in full model as specified in Table III-5.

TABLE III-10. Estimated percentage of patients diagnosed with regional stage disease by education[a], hormone receptor status[b] and tumor grade; adjusted for all other factors[c] in full regression model.

	Percentage	95% CI
Japanese		
high education, positive hormone receptor, low tumor grade	21.7	(19.6, 23.8)
low education, negative/other hormone receptor, high tumor grade	36.0	(32.8, 39.4)
White		
high education, positive hormone receptor, low tumor grade	25.7	(25.0, 26.3)
low education, negative/other hormone receptor, high tumor grade	41.2	(39.3, 43.1)
Hispanic		
high education, positive hormone receptor, low tumor grade	27.9	(26.4, 29.3)
low education, negative/other hormone receptor, high tumor grade	43.9	(41.8, 46.0)
Black		
high education, positive hormone receptor, low tumor grade	28.8	(27.5, 30.1)
low education, negative/other hormone receptor, high tumor grade	45.2	(42.7, 47.5)

[a] High education is defined as the midpoint of the lower two quintiles in the distribution of the percentages of persons within census tracts with no high school diploma (equal to 6.5% in this study population). Low education is defined as the midpoint of the highest two quintiles in the distribution of the proportions of persons within census tract with no high school diploma (equal to 52% in this study population).

[b] Hormone receptor status and tumor grade are the same binary variables used in regression models (Tables III-5-7).

[c] All other factors in full model as specified in Table III-5.

TABLE III-11. Estimated percentage of patients diagnosed with tumors 1 cm or greater in diameter by education[a], hormone receptor status[b] and tumor grade; adjusted for all other factors[c] in full regression model.

	Percentage	95% CI
Japanese		
high education, positive hormone receptor, low tumor grade	74.0	(71.2, 76.8)
low education, negative/other hormone receptor, high tumor grade	89.8	(88.2, 91.4)
White		
high education, positive hormone receptor, low tumor grade	77.5	(76.8, 78.2)
low education, negative/other hormone receptor, high tumor grade	91.4	(90.6, 92.2)
Hispanic		
high education, positive hormone receptor, low tumor grade	80.6	(79.1, 82.1)
low education, negative/other hormone receptor, high tumor grade	92.8	(92.0, 93.6)
Black		
high education, positive hormone receptor, low tumor grade	82.9	(81.6, 84.2)
low education, negative/other hormone receptor, high tumor grade	93.7	(92.9, 94.5)

[a] High education is defined as the midpoint of the lower two quintiles in the distribution of the percentages of persons within census tracts with no high school diploma (equal to 6.5% in this study population). Low education is defined as the midpoint of the highest two quintiles in the distribution of the proportions of persons within census tract with no high school diploma (equal to 52% in this study population).

[b] Hormone receptor status and tumor grade are the same binary variables used in regression models (Tables III-5-7).

[c] All other factors in full model as specified in Table III-5.

TABLE III-12. Odds ratios for selected explanatory variables in full multiple logistic regression model comparing female breast cancer patients with distant stage disease to those with localized stage disease.

Variables in model	Distant vs. Localized					
	White (n = 57,033)		Black (n = 5,418)		Hispanic (n = 4,308)	
	OR[a]	(95% CI)	OR[a]	(95% CI)	OR[a]	(95% CI)
Urban	1.0	(0.9, 1.1)	1.0	(0.7, 1.6)	0.6	(0.4, 0.9)
Not married	1.4	(1.3, 1.5)	1.3	(1.1, 1.6)	1.3	(1.1, 1.6)
No high school diploma (10% increments)	1.1	(1.0, 1.1)	1.1	(1.0, 1.2)	1.1	(1.0, 1.2)
Working class job (10% increments)	1.0	(1.0, 1.1)	1.1	(0.9, 1.2)	0.9	(0.8, 1.1)
Median family income ($ 20 thousands)	0.9	(0.9, 1.0)	1.2	(0.9, 1.5)	0.8	(0.6, 1.2)
Families <poverty, female head of house (10% inc.)	1.1	(1.0, 1.2)	1.0	(0.9, 1.1)	1.3	(1.0, 1.7)
Home ownership (10% increments)	1.0	(1.0, 1.0)	0.9	(0.9, 1.0)	1.0	(0.9, 1.0)
No car (10% increments)	1.0	(1.0, 1.1)	1.0	(0.9, 1.1)	0.9	(0.8, 1.1)
Negative for hormone receptors	2.1	(2.0, 2.3)	1.9	(1.6, 2.2)	2.0	(1.6, 2.5)
Advanced grade tumor (grades 3 or 4)	2.0	(1.9, 2.1)	1.6	(1.4, 1.9)	1.9	(1.5, 2.3)

[a] OR = odds ratio adjusted for all other explanatory variables in the full model.

TABLE III-13. Odds ratios for selected explanatory variables in full multiple logistic regression model comparing female breast cancer patients with regional stage disease to those with localized stage disease.

Variables in model	Regional vs. Localized					
	White (n = 74,842)		Black (n = 7,563)		Hispanic (n = 6,119)	
	OR[a]	(95% CI)	OR[a]	(95% CI)	OR[a]	(95% CI)
Urban	1.0	(0.9, 1.0)	1.0	(0.8, 1.2)	0.9	(0.8, 1.1)
Not married	1.1	(1.1, 1.1)	1.0	(0.9, 1.1)	1.0	(0.9, 1.1)
No high school diploma (10% increments)	1.1	(1.0, 1.1)	1.0	(0.9, 1.0)	1.0	(1.0, 1.1)
Working class job (10% increments)	1.0	(1.0, 1.0)	1.0	(1.0, 1.1)	1.0	(0.9, 1.1)
Median family income ($ 20 thousands)	1.0	(1.0, 1.0)	1.0	(0.9, 1.2)	1.0	(0.8, 1.1)
Families <poverty, female head of house (10% inc.)	1.0	(1.0, 1.1)	1.0	(0.9, 1.1)	1.1	(0.9, 1.2)
Home ownership (10% increments)	1.0	(1.0, 1.0)	1.0	(1.0, 1.0)	1.0	(1.0, 1.1)
No car (10% increments)	1.0	(1.0, 1.0)	1.0	(0.9, 1.1)	1.0	(0.9, 1.1)
Negative for hormone receptors	0.8	(0.8, 0.8)	0.8	(0.7, 0.9)	0.8	(0.7, 0.9)
Advanced grade tumor (grades 3 or 4)	2.1	(2.0, 2.2)	2.1	(1.9, 2.3)	1.9	(1.7, 2.1)

[a] OR = odds ratio adjusted for all other explanatory variables in the full model.

TABLE III-14. Odds ratios for selected explanatory variables in full multiple logistic regression model comparing female breast cancer patients with tumor diameter of 1 cm or greater to those with tumors <1 cm in diameter.

Variables in model	Tumor Size 1 cm+ vs. <1 cm					
	White (n = 74,053)		Black (n = 7,744)		Hispanic (n = 6,215)	
	OR[a]	(95% CI)	OR[a]	(95% CI)	OR[a]	(95% CI)
Urban	0.9	(0.9, 1.0)	1.0	(0.7, 1.4)	0.8	(0.6, 1.1)
Not married	1.2	(1.2, 1.3)	1.0	(0.9, 1.2)	1.3	(1.1, 1.5)
No high school diploma (10% increments)	1.1	(1.0, 1.1)	1.1	(1.0, 1.2)	1.1	(1.0, 1.2)
Working class job (10% increments)	1.0	(1.0, 1.1)	1.0	(0.9, 1.1)	0.9	(0.8, 1.0)
Median family income ($ 20 thousands)	1.0	(0.9, 1.0)	1.0	(0.8, 1.3)	1.0	(0.8, 1.2)
Families <poverty, female head of house (10% inc.)	1.1	(1.0, 1.2)	1.0	(0.9, 1.1)	1.2	(1.0, 1.6)
Home ownership (10% increments)	1.0	(1.0, 1.0)	1.0	(0.9, 1.0)	1.0	(0.9, 1.0)
No car (10% increments)	1.0	(0.9, 1.0)	0.9	(0.9, 1.0)	0.9	(0.8, 1.1)
Negative for hormone receptors	0.7	(0.7, 0.8)	1.1	(0.9, 1.2)	0.9	(0.8, 1.1)
Advanced grade tumor (grades 3 or 4)	3.0	(2.9, 3.2)	3.5	(2.9, 4.2)	3.3	(2.7, 4.0)

[a] OR = odds ratio adjusted for all other explanatory variables in the full model.

FIGURE II-1. Selected demographic characteristics of the SEER population compared with those for the total U.S. population.

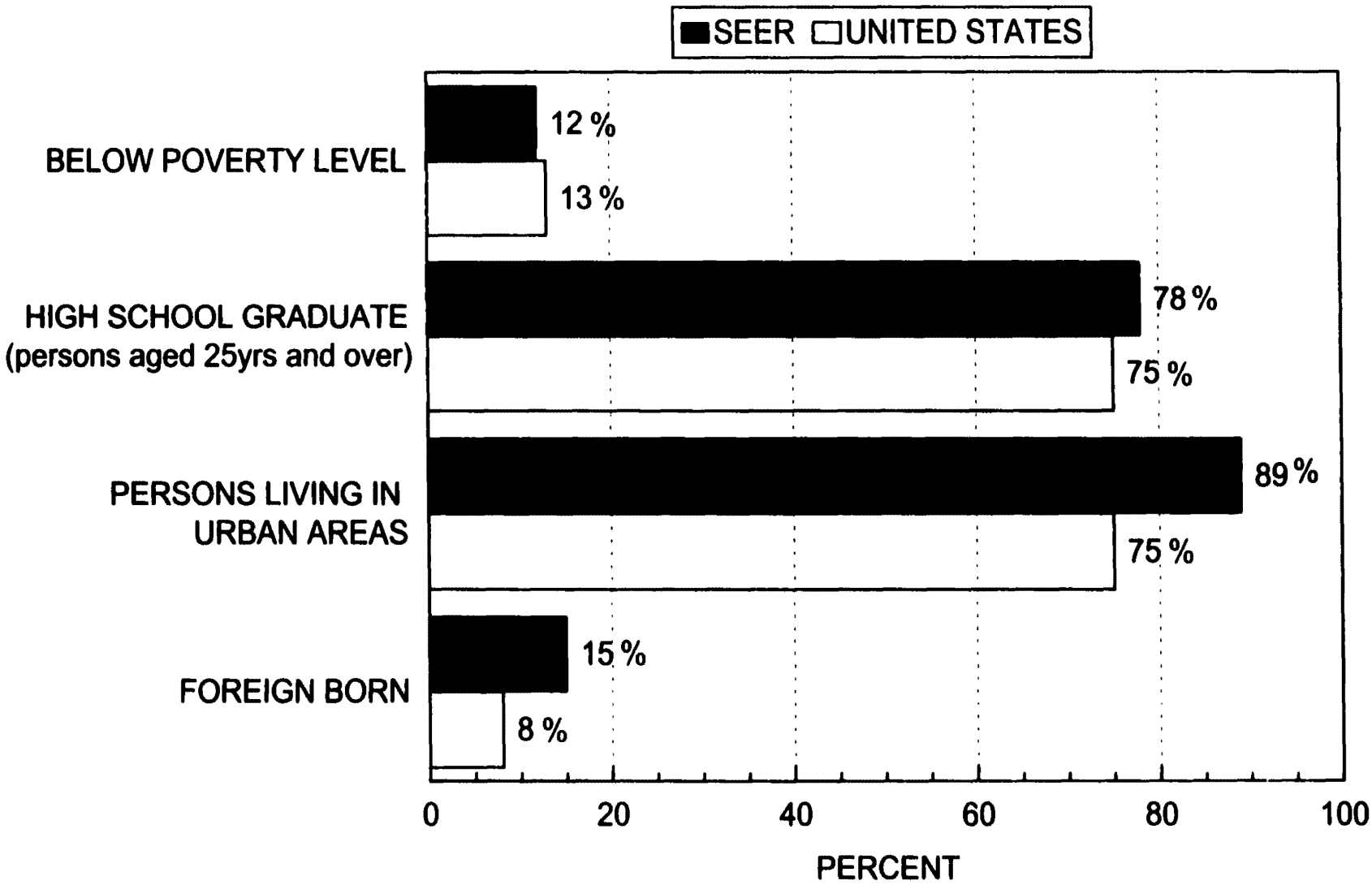


FIGURE II-2. Selection of female breast cancer study group.

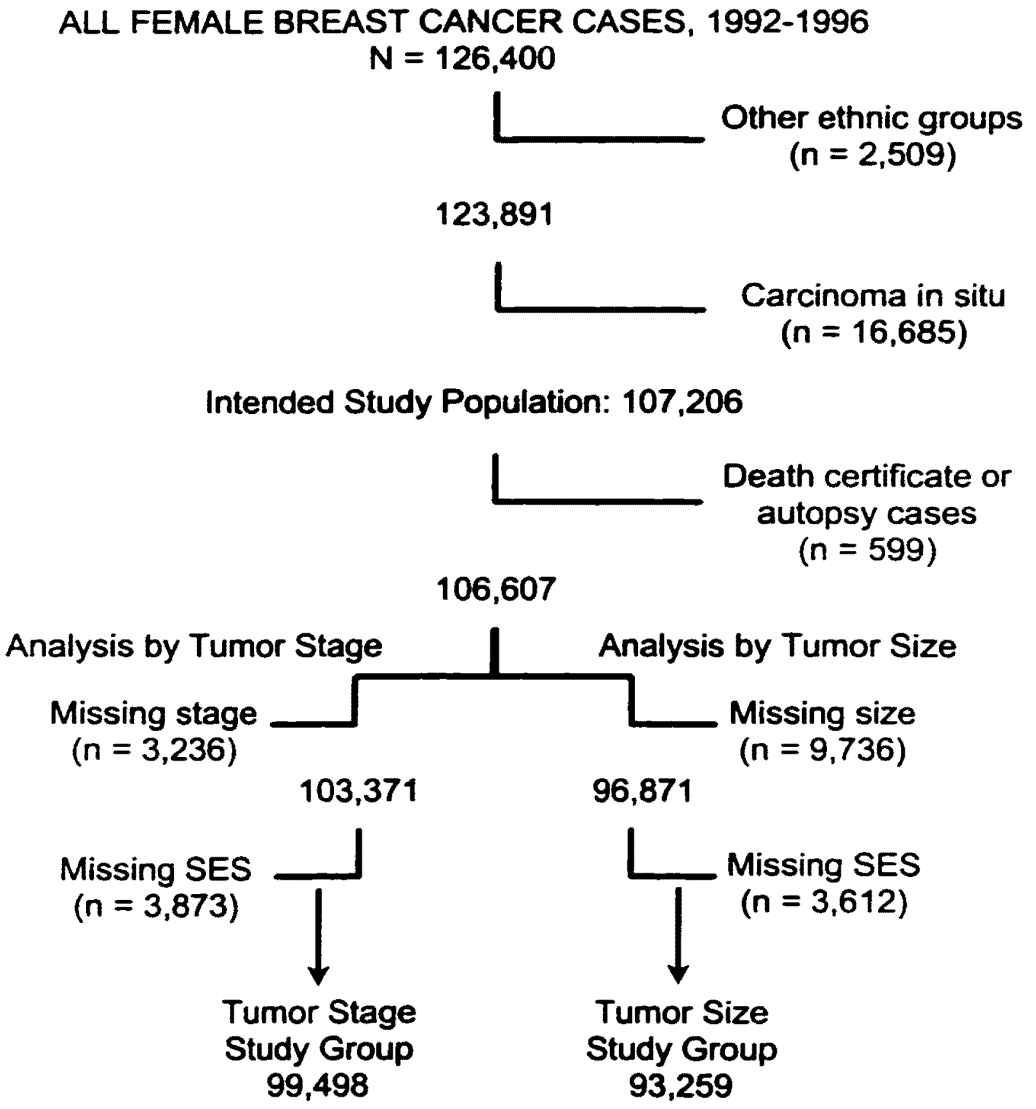
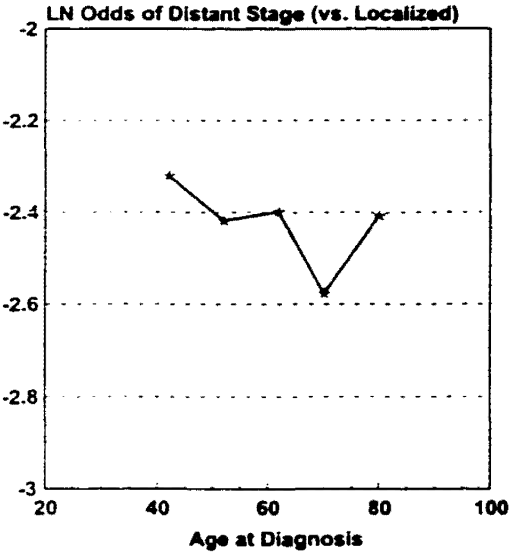
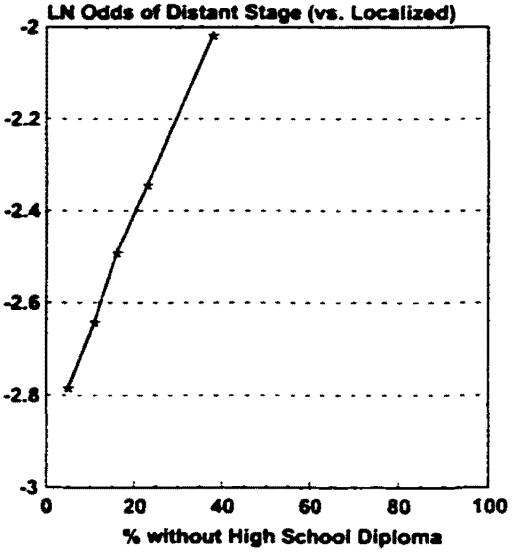
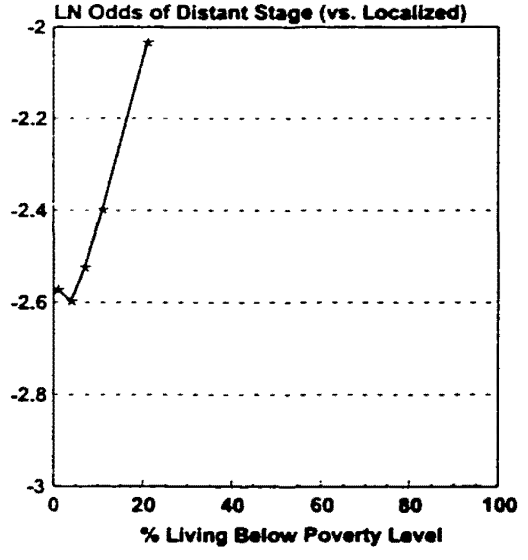
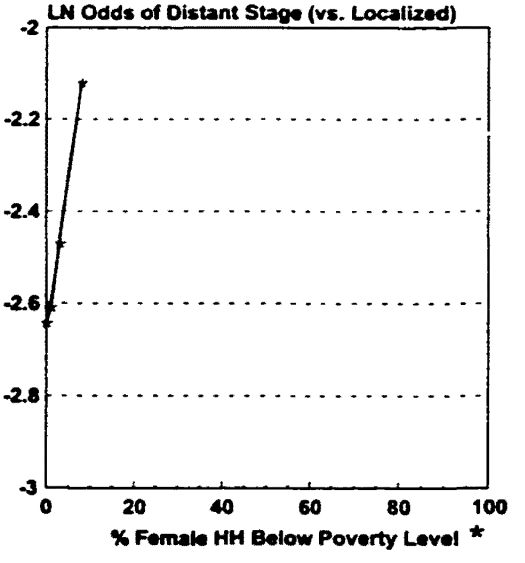


FIGURE III-1a. Ln-Odds plots of distant stage disease by explanatory variables.



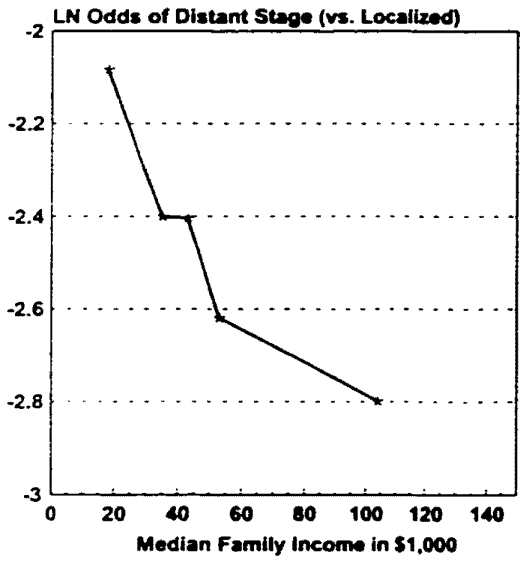


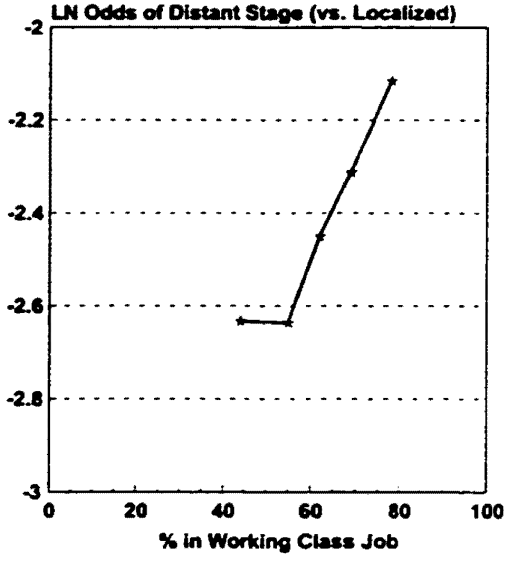


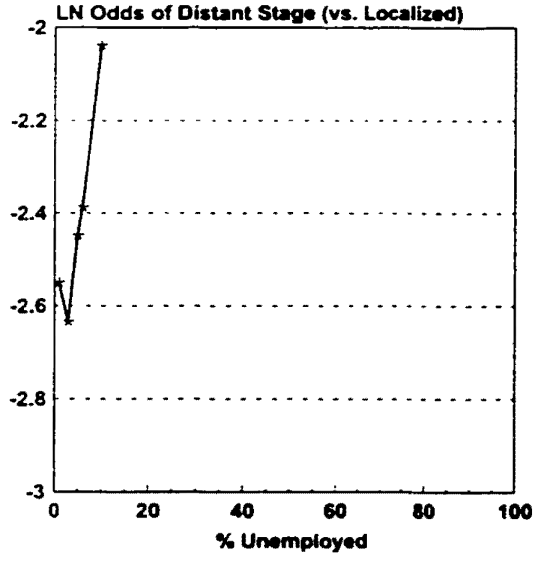


* Percentage of families headed by women with no husband at home, with one or more children, and who are living below the poverty level.

FIGURE III-1b. Ln-Odds plots of distant stage disease by explanatory variables.







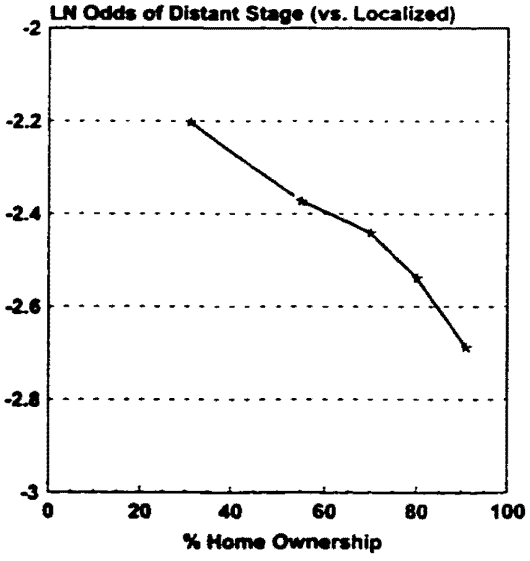
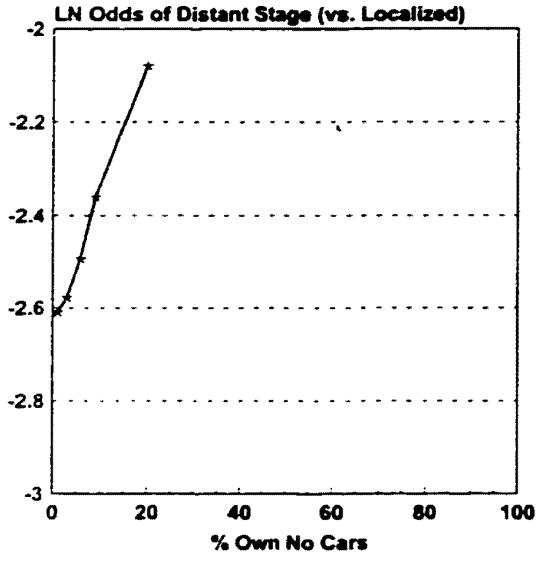


FIGURE III-1c. Ln-Odds plots of distant stage disease by explanatory variables.



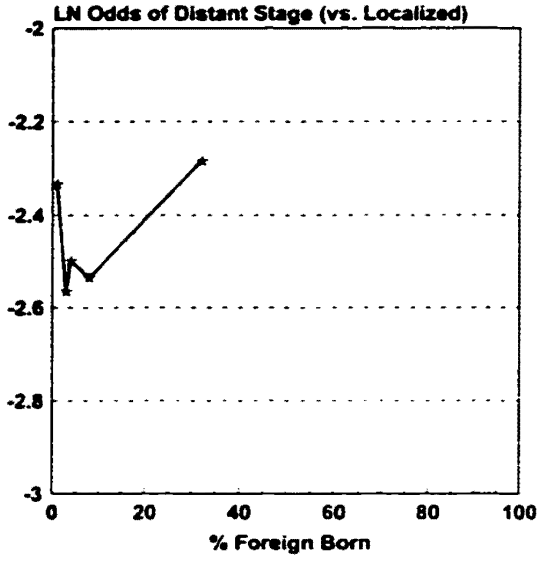
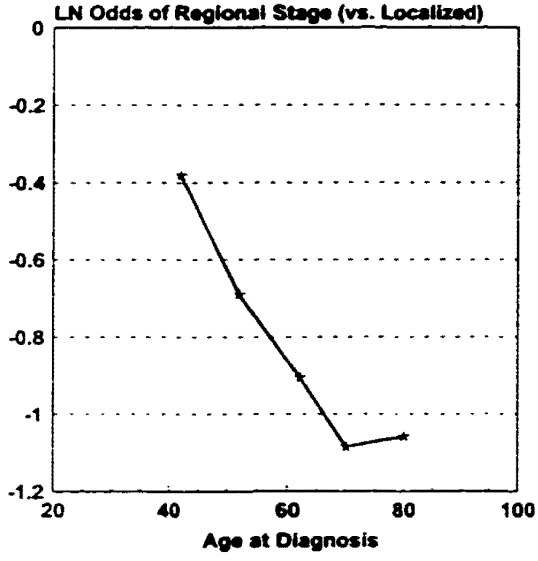
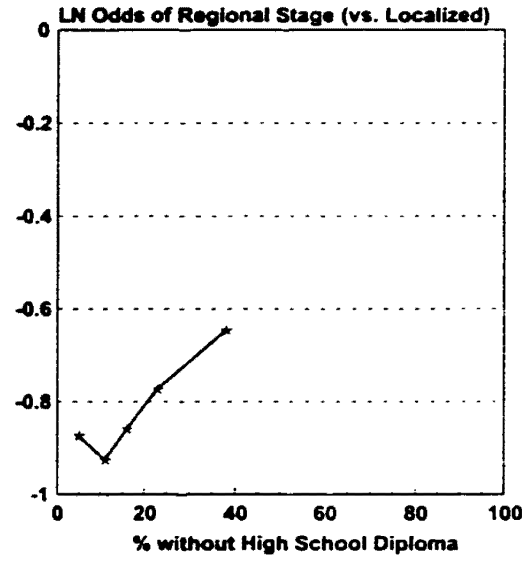
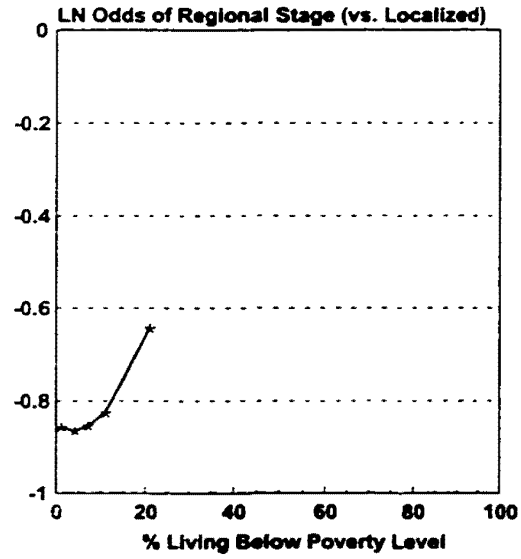
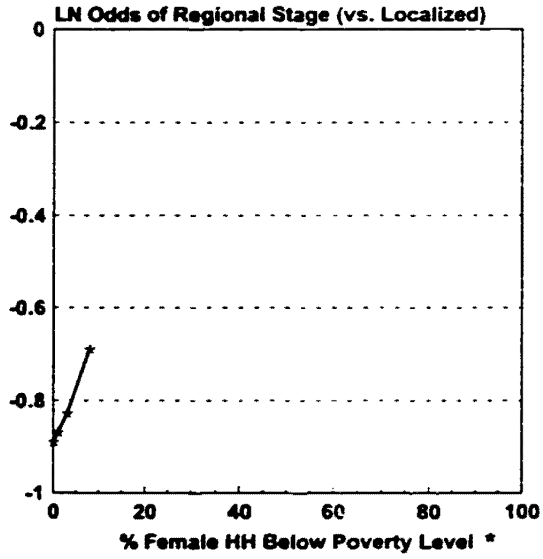


FIGURE III-2a. Ln-Odds plots of regional stage disease by explanatory variables.



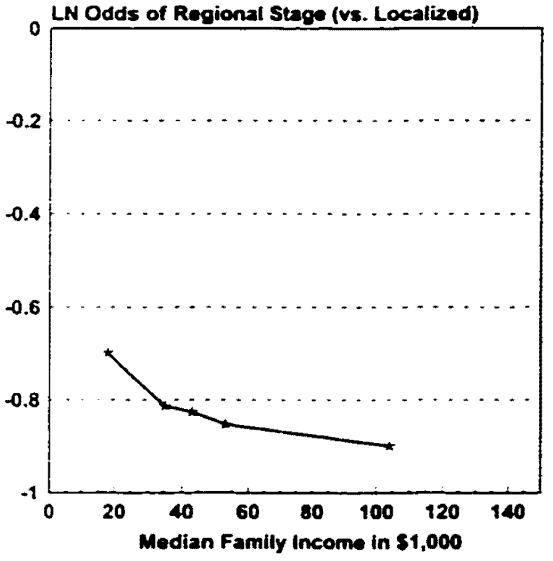


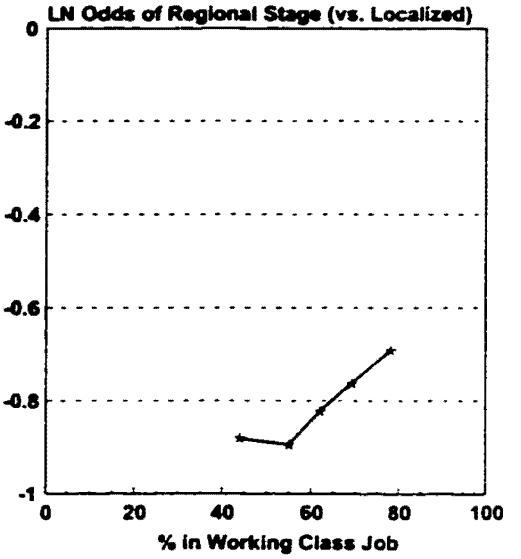


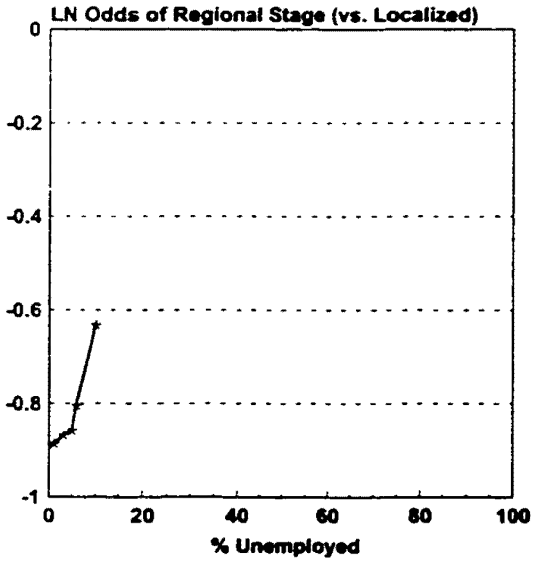


* Percentage of families headed by women with no husband at home, with one or more children, and who are living below the poverty level.

FIGURE III-2b. Ln-Odds plots of regional stage disease by explanatory variables.







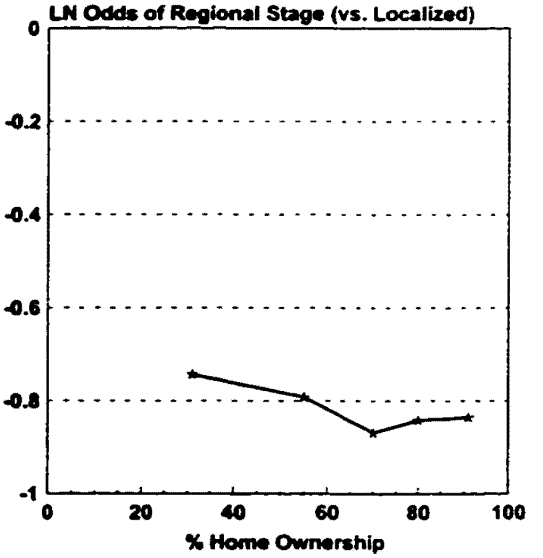
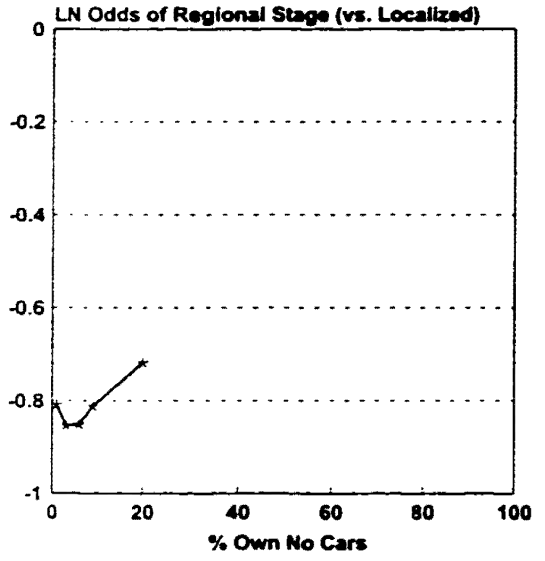


FIGURE III-2c. Ln-Odds plots of regional stage disease by explanatory variables.



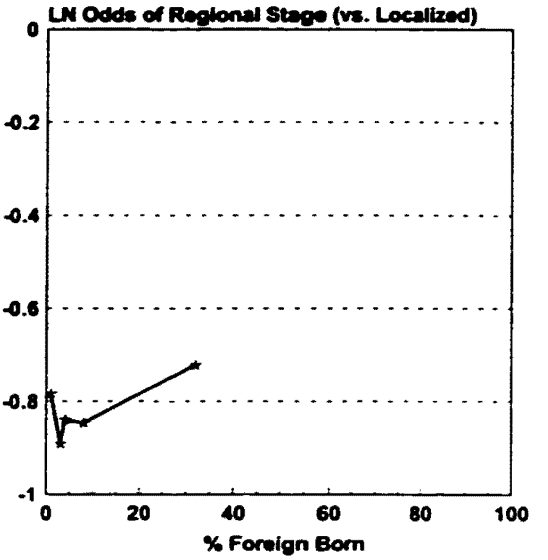
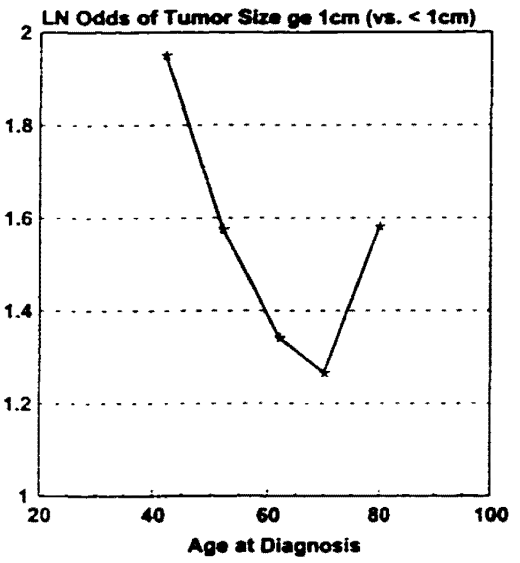
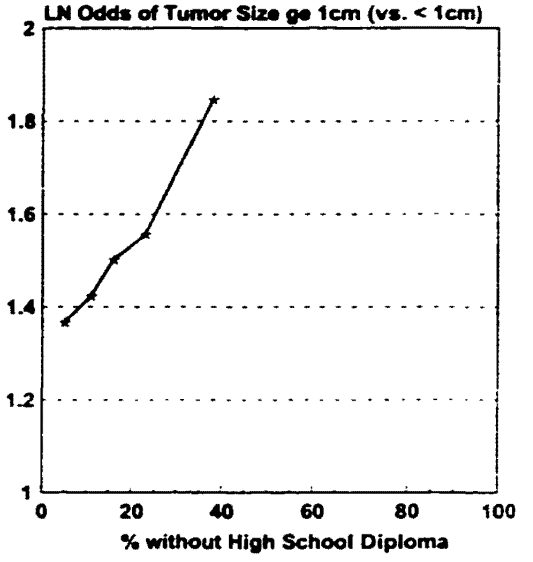
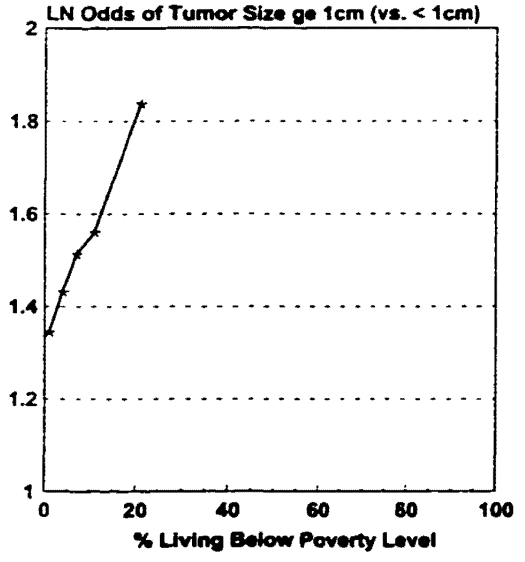
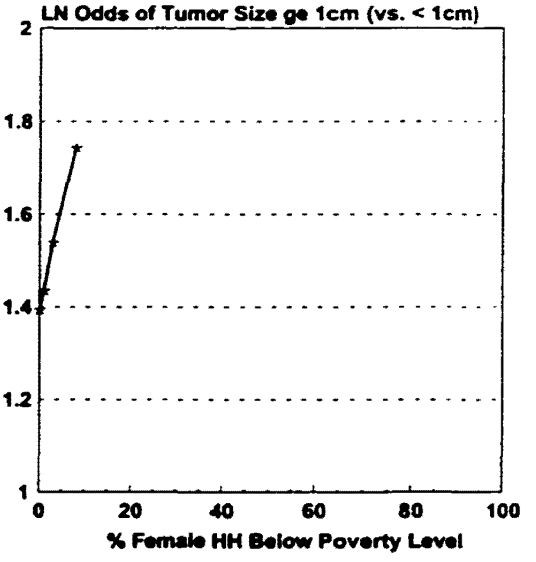


FIGURE III-3a. Ln-Odds plots of tumor size greater than or equal to 1 cm by explanatory variables.



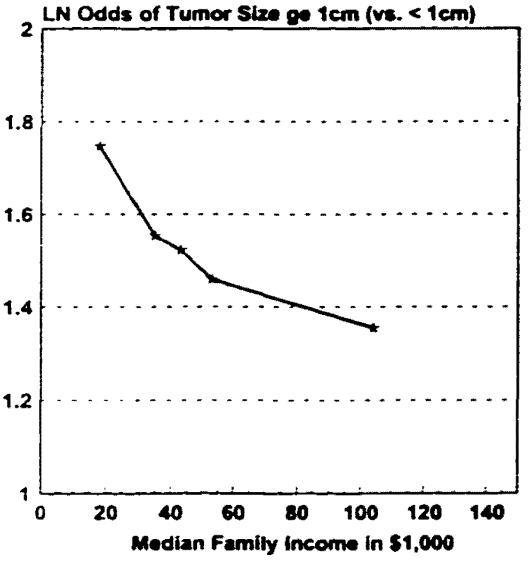


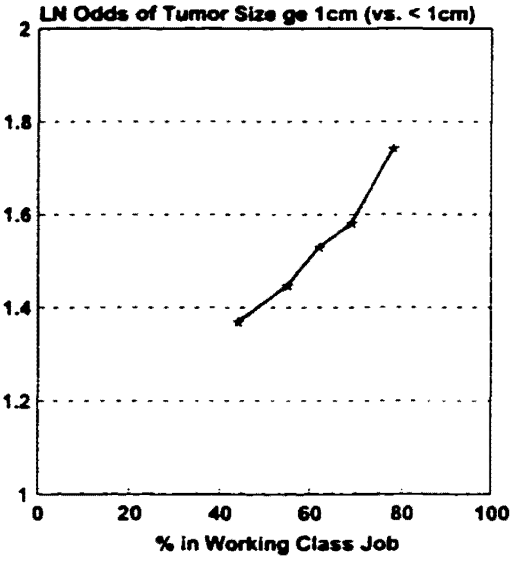


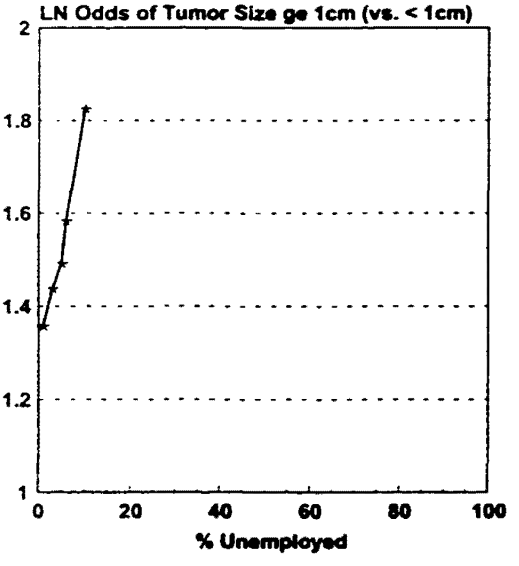


* Percentage of families headed by women with no husband at home, with one or more children, and who are living below the poverty level.

FIGURE III-3b. Ln-Odds plots of tumor size greater than or equal to 1 cm by explanatory variables.







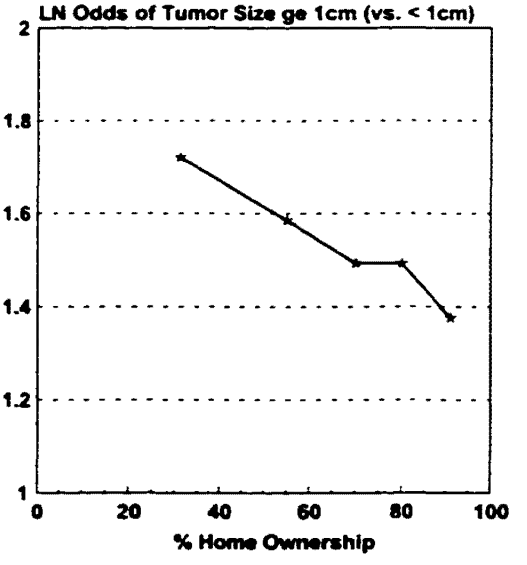
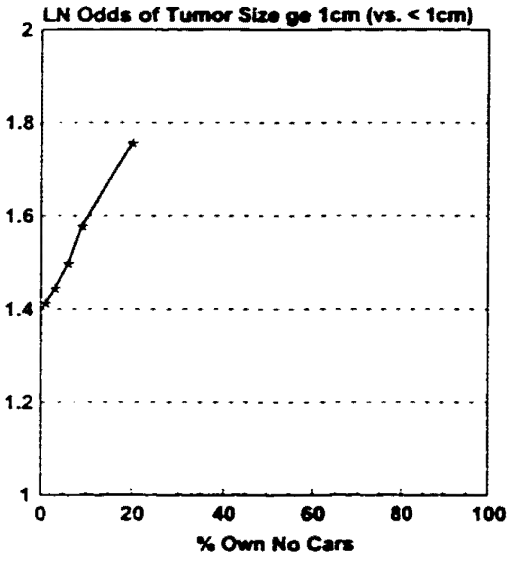


FIGURE III-3c. Ln-Odds plots of tumor size greater than or equal to 1 cm by explanatory variables.



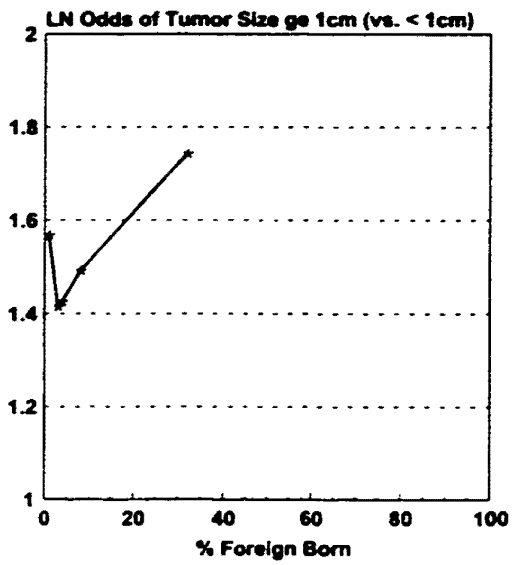


FIGURE III-4. Plot of observed and expected number of observations for Hosmer-Lemeshow test of goodness-of-fit of model for distant stage disease vs. localized stage disease.

Hosmer and Lemeshow GOF Statistic = 56.228 with 8 DF (p = 0.0001)

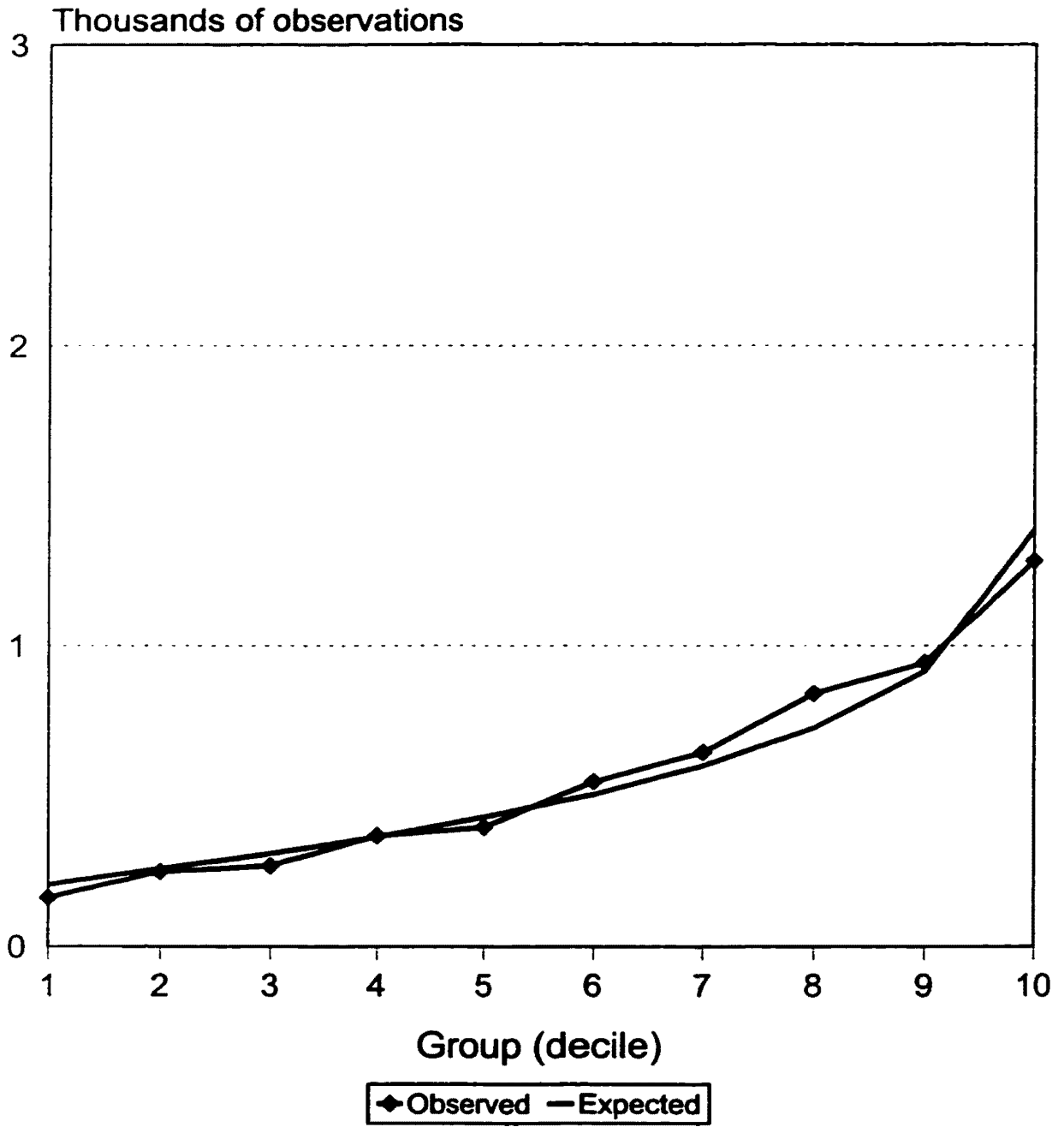


FIGURE III-5. Plot of observed and expected number of observations for Hosmer-Lemeshow test of goodness-of-fit of model for regional stage disease vs. localized stage disease.

Hosmer and Lemeshow GOF Statistic = 49.709 with 8 DF (p = 0.0001)

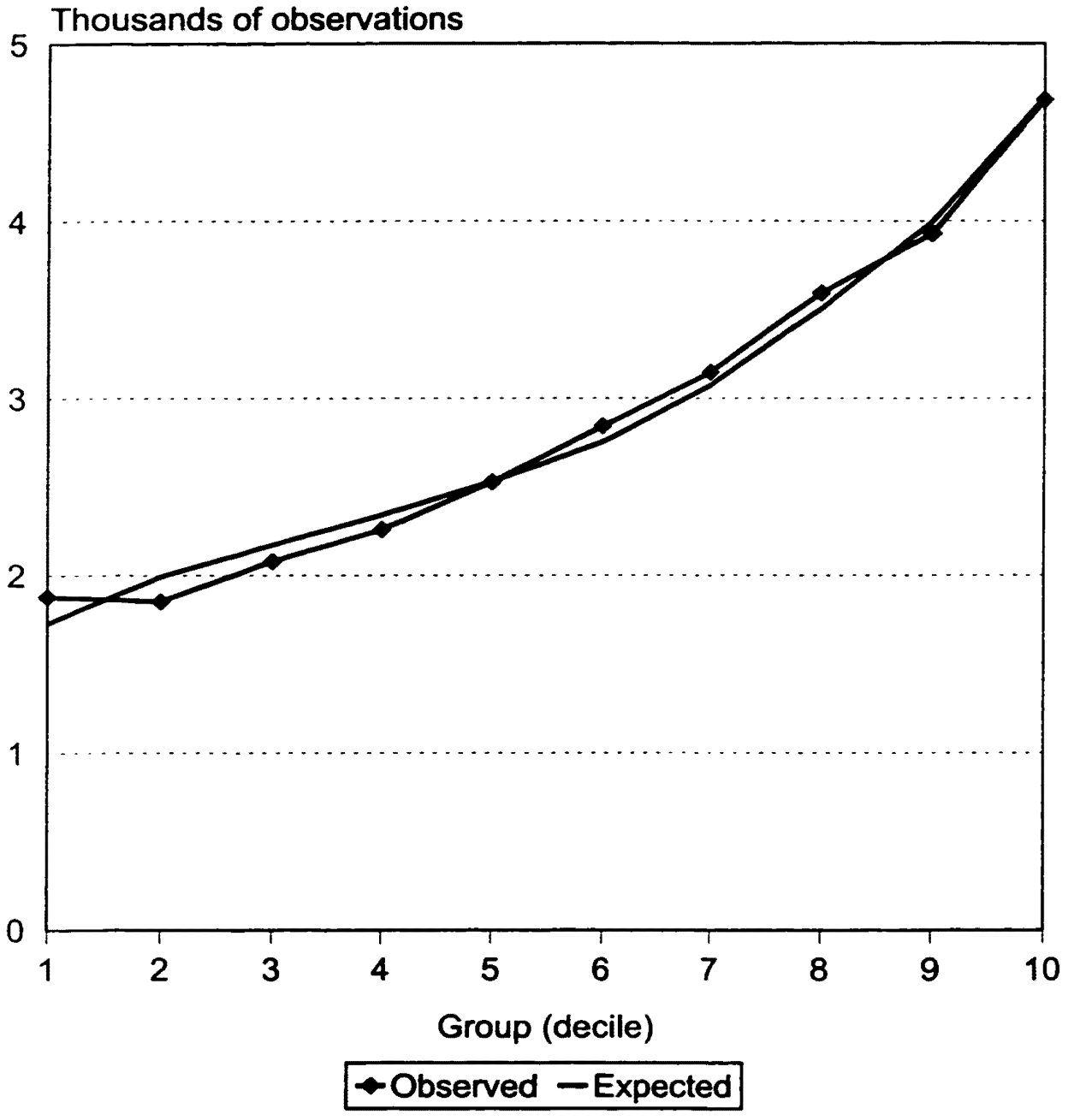


FIGURE III-6. Plot of observed and expected number of observations for Hosmer- Lemeshow test of goodness-of-fit of model for tumor size greater than or equal to 1.0 cm vs. tumor size less than 1.0 cm.

Hosmer and Lemeshow GOF Statistic = 6.7037 with 8 DF (p = 0.5689)

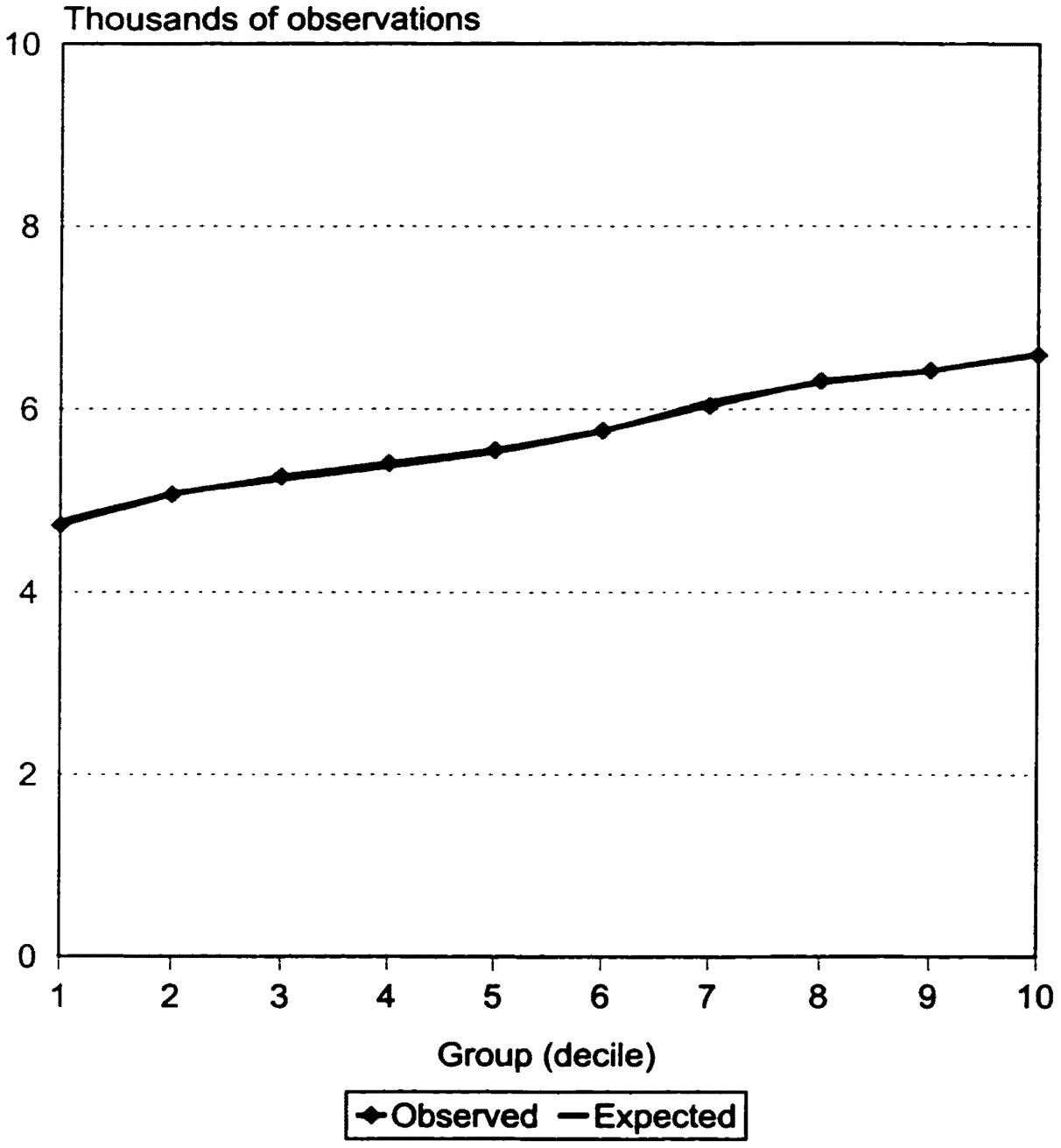
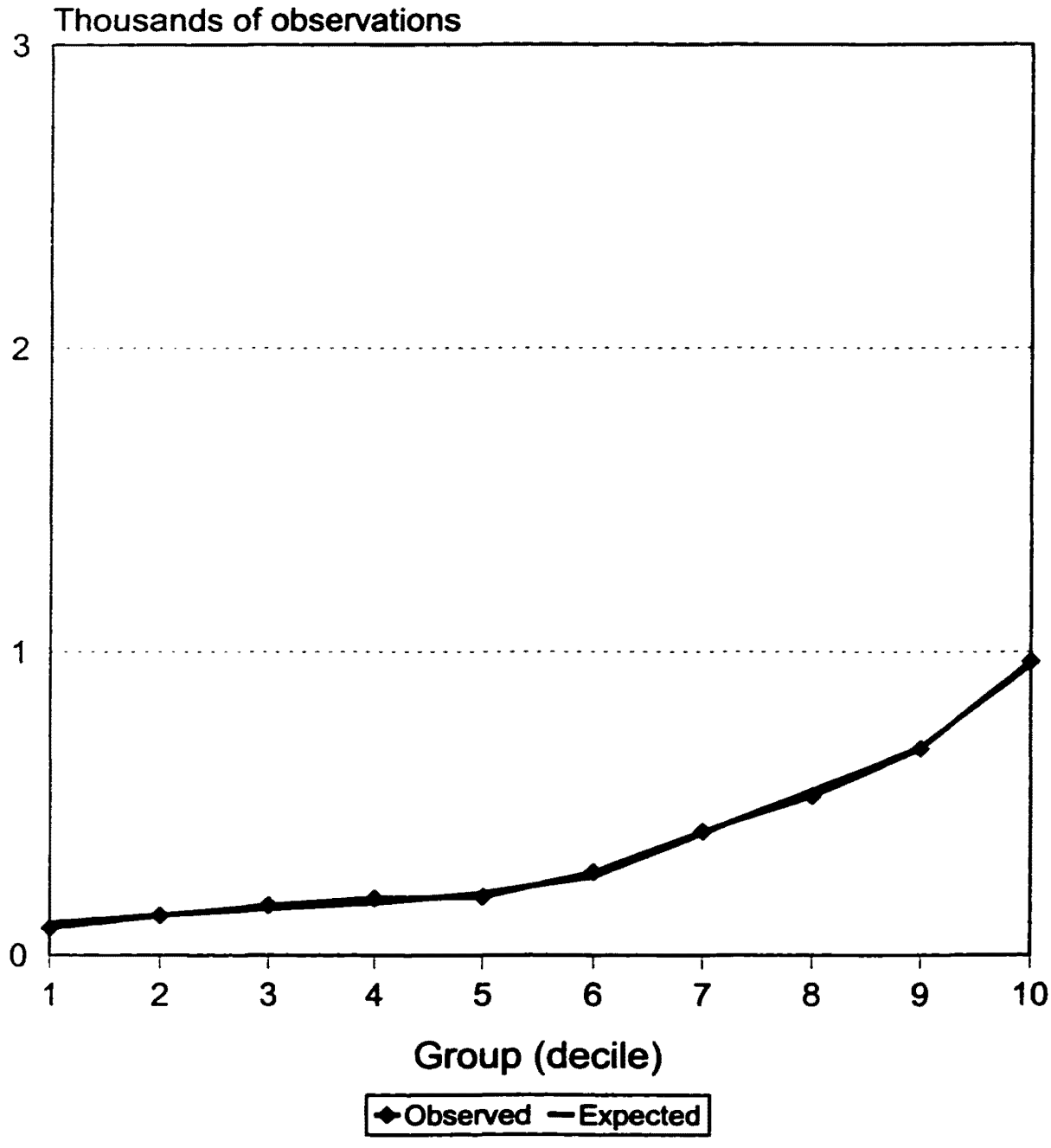


FIGURE III-7. Plot of observed and expected number of observations for Hosmer-Lemeshow test of goodness-of-fit of model for distant stage disease vs. localized stage disease after excluding patients with unknown tumor grade.

Hosmer and Lemeshow GOF Statistic = 7.5475 with 8 DF (p = 0.4789)



APPENDIX II-1a. Geocoding Update Instrument.

Geocoding Update Instrument — September 1997, Version 3

Reporting Site Name: _____

Address: _____

_____

Phone: _____ Fax: _____

The most knowledgeable expert on geocoding at your registry should complete this form:

Name: _____

Phone: _____ Fax: _____ E-mail: _____

1. **What year did you begin geocoding?**
Census Tract? ☐ _____Year Block group? ☐ _____Year

2. **Current Practices:**

- What software is used?_____
- How do you currently obtain geocoded information at the Census tract level?

☐ In-house ☐ Offside, Skip to #4 ☐ Both, Complete #3 & #4

The following questions pertain to equipment and personnel costs incurred by your registry in obtaining geocoded information.

3. **If Inhouse:**

Software costs to the registry:

- Initial purchase price _____
- Ongoing maintenance and support costs
(Specify annual, quarterly, etc.)_____

Personnel costs to the registry:

FTE Equivalent	Job Description (skill level, degree required, etc.)	Annual Salary

Specify other personnel costs you incur in performing geocoding.
Other costs (e.g. supplies, equipment, transportation):

Category	Description	Annual Costs

APPENDIX II-1b. Geocoding Update Instrument

4. **If Offside: (Specify private contractor, state department of health, etc.)**

- Facility Name: _____
- Individual Contract Name:

_____

- Phone: _____ Fax: _____ E-mail: _____
- Charge to registry (specify annual, quarterly, etc.) $ _____
- List any "no cost" services provided to the registry:

_____

5. **Do you test whether Census tract codes generated are valid?**

☐ Yes ☐ No

Describe method used to test for validity or verify accuracy and year implemented:

Year	Method

6. **What additional effort do you take to obtain information on unmatched cases?**

☐ None ☐ Specify _____

7. **Is a probability method used to assign uncertain matches?**

☐ None ☐ Specify _____

8. **Do you geocode at the block group level?**

☐ Yes ☐ No
Specify level: _____

9. **If you don't currently geocode at the block group level, how would you have to change your operations in order to do so? Describe necessary procedures, costs, and time involved in modifying and upgrading your system to achieve this.**

_____

_____

_____

_____

_____

APPENDIX II-1c. Geocoding Update Instrument

10. **Specify any structural problems that make geocoding difficult in your area.**

11. **Do you collect information on Socio-economic status (SES, measured in terms of income, education or employment/ occupation) on your SEER cases? Please specify:**

12. **Please list concerns your registry has pertaining to geocoding that you wish to discuss at the SEER meeting.**

Please return this form NO LATER THAN SEPTEMBER 26 to:

Kathleen C. Barry, Applied Research Program
Executive Plaza North, Room 313
6130 Executive Boulevard, MSC 7344
Bethesda, MD 20892-7339
Phone: 301-496-5410
Fax: 301-435-3710
E-mail: barryk@dcpcepn.nci.nih.gov

APPENDIX II-2a. Changes to SEER Data Coding Manual as a Result of Pilot Study.

CENSUS TRACT/BLOCK NUMBERING AREA (BNA)

Section III, Field 01.B

Census Tract/Block Numbering Area (BNA)

The census tract/block numbering area is assigned to the patient's residence at the time of diagnosis. For cases diagnosed 1988 and forward, 1990 definitions of census tract and block numbering area must be used.

If an area is assigned a census tract/block numbering area (BNA) code and the code is not available, code as '999999.'

If an area is not assigned a census tract or BNA (1980 or prior censuses only), code as '000000.'

Census tract numbers should be right justified and zero filled so that all six positions have a code entered. For purposes of coding census tract, assume that the decimal point is located ***between*** the fourth and fifth positions of this field. Thus, census tract '409.6 ' would be coded '040960'and census tract '516.21' would be coded '051621.'

BNA codes are 6 digits in length, as are census tract codes. They can be distinguished by their range. Census tract codes range from 0001.00 to 9499.99, while BNAs range from 9501.00 to 9989.99. The decimal point is ignored. For example, BNA code 9607.23 would be coded '960723.'

A census tract is a small statistical subdivision of a county with (generally) between 2,500 and 8,000 residents. The boundaries of census tracts are established cooperatively by local committees and the Census Bureau. An attempt is made to keep the same boundaries from census to census so that historical comparability will be maintained. This goal is not always achieved; old tracts may be subdivided due to population growth, disappear entirely, or have their boundaries changed. Between 1970 and 1980 the number of tracts increased by over 20 percent. Thus it is important to know which definitions were used for the coding of the census tracts: the 1970 definitions, the 1980 definitions, or starting with 1988 diagnoses, the 1990 definitions.

Some parts of the country identify areas with block numbering areas (BNAs) codes. These are the geographic equivalent of a census tract. BNAs were implemented in the 1990 census. The BNA is always a subunit of a county and census tracts/BNAs are mutually exclusive; that is, a given county is subdivided into either census tracts or BNAs, but not both. There may be as few as one or two BNAs per county, or more than 20 BNAs per county.

Note that Block Group coding is different than block numbering area coding and is not currently collected by SEER.

APPENDIX II-2b. Changes to SEER Data Coding Manual as a Result of Pilot Study.

CODING SYSTEM FOR CENSUS TRACT

Section III, Field 01.C

Coding System for Census Tract

Code	
0	Not tracted
1	1970 Census Tract Definitions (1973-77)
2	1980 Census Tract Definitions (1978-87)
3	1990 Census Tract Definitions (1988+)
4	2000 Census Tract Definitions

Note: Do not implement code '4' until instructed by SEER.

APPENDIX II-2c. Changes to SEER Data Coding Manual as a Result of Pilot Study.

CENSUS TRACT CERTAINTY

Section III, Field 01.D

Census Tract Certainty

Code

1 Census tract/BNA based on complete and valid street address of residence
2 Census tract/BNA based on residence ZIP+4
3 Census tract/BNA based on residence ZIP+2
4 Census tract/BNA based on residence ZIP only
5 Census tract/BNA based on ZIP of post office box
9 Unable to assign census tract or block numbering based on available information/unknown

This field is a code indicating the basis of assignment of census tract or block numbering area (BNA) for an individual record. It is helpful in identifying cases census tracted/BNA'd from incomplete information or a post office box address. Most of the time, this information is provided by a geocoding vendor service. Alternatively, the code is manually assigned by central registry staff. Codes are hierarchical, with lower numbers having priority.

Use code '1' when census tract or block numbering area is assigned with certainty. This can result either from a computer match using geocoding software or from manual searches.

Example 1 Complete and valid street address used.

Example 2 Rural route or incomplete street address is used, but is known to lie entirely within one census tract.

Use codes '2' through '5' when census tract or block numbering area is assigned with some uncertainty.

Example 3 Street address is incomplete or invalid, or only rural route number is available, but ZIP code of residence is known. The case may be geocoded manually or geocoded using software. The case is placed at the geographic center of the ZIP code area, i.e., the ZIP code "centroid." *Use code '4.'*

Example 4 Post office box number and ZIP code used. *Use code '5.'*

Note: Avoid using P.O. box mailing address, when possible, as this is not the true residence of the patient.

Use code '9' when the ZIP code is missing, when the complete address of the patient cannot be determined or when there is insufficient information to assign census tract or BNA.

Use of this code is required effective with cases diagnosed 1/1/98 and after. It is strongly suggested that this information be obtained from the geocode vendor for cases back to 1988.

APPENDIX III-1. Pearson Correlation Coefficients; Prob > |R| under Ho: Rho=0; N = 102,419.

	RTEDUNOH	RTUNEMPL	RTWRKCLA	RTPOVBEL	RTPOVFAM	RTMEDFIN	RTOWNERS	RTNONEOR	RTFORGBO
RTEDUNOH	1.00000	0.68137	0.77064	0.70971	0.59310	-0.64547	-0.38929	0.57959	0.38187
	0.0	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
RTUNEMPL	0.68137	1.00000	0.52612	0.79655	0.80151	-0.51955	-0.36103	0.67352	0.11998
	0.0001	0.0	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
RTWRKCLA	0.77064	0.52612	1.00000	0.51850	0.48217	-0.75377	-0.25725	0.37870	0.10833
	0.0001	0.0001	0.0	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
RTPOVBEL	0.70971	0.79655	0.51850	1.00000	0.85976	-0.62956	-0.54601	0.75089	0.19579
	0.0001	0.0001	0.0001	0.0	0.0001	0.0001	0.0001	0.0001	0.0001
RTPOVFAM	0.59310	0.80151	0.48217	0.85976	1.00000	-0.52020	-0.42012	0.70372	0.00634
	0.0001	0.0001	0.0001	0.0001	0.0	0.0001	0.0001	0.0001	0.0425
RTMEDFIN	-0.64547	-0.51955	-0.75377	-0.62956	-0.52020	1.00000	0.47063	-0.47564	-0.02846
	0.0001	0.0001	0.0001	0.0001	0.0001	0.0	0.0001	0.0001	0.0001
RTOWNERS	-0.38929	-0.36103	-0.25725	-0.54601	-0.42012	0.47063	1.00000	-0.64435	-0.43924
	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0	0.0001	0.0001
RTNONEOR	0.57959	0.67352	0.37870	0.75089	0.70372	-0.47564	-0.64435	1.00000	0.24847
	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0	0.0001
RTFORGBO	0.38187	0.11998	0.10833	0.19579	0.00634	-0.02846	-0.43924	0.24847	1.00000
	0.0001	0.0001	0.0001	0.0001	0.0425	0.0001	0.0001	0.0001	0.0

RTEDUNOH = % without high school diploma; RTUNEMPL = % unemployed; RTWRKCLA = % working class; RTPOVBEL= % below poverty level; RTPOVFAM = % families headed by women with no husband at home, with one or more children, and who are living below the poverty level; RTMEDFIN = median family income; RTOWNERS = % own their home; RTNONEOR = % having no car; RTFORGBO = % foreign-born.

APPENDIX III-2. Odds ratios for explanatory variables in full model comparing female breast cancer patients with distant stage disease to those with localized disease.

Variables in model	Distant vs. Localized		
	OR[a]	(95% CI)	p[b]
White	1.0		
Black	1.3	(1.2, 1.5)	.0001
Hispanic	1.1	(1.0, 1.2)	.2452
Japanese	0.7	(0.6, 1.0)	.0316
Filipino	0.9	(0.7, 1.1)	.2610
Chinese	1.0	(0.8, 1.3)	.7159
Hawaiian	1.4	(0.9, 2.1)	.1741
Korean	0.6	(0.3, 1.0)	.0951
Vietnamese	0.7	(0.3, 1.2)	.2032
American Indian	1.4	(0.7, 2.6)	.3343
Age (10 yr increments)	1.0	(1.0, 1.0)	.6218
Registry 0	1.0		
Registry 1	0.9	(0.9, 1.1)	.7405
Registry 2	1.1	(1.0, 1.3)	.0183
Registry 3	1.1	(0.9, 1.2)	.4187
Registry 4	1.0	(0.8, 1.3)	.8857
Registry 5	1.1	(1.0, 1.2)	.2799
Registry 6	1.2	(1.0, 1.4)	.0386
Registry 7	1.0	(0.9, 1.1)	.5184
Registry 8	0.9	(0.7, 1.1)	.1864
Registry 9	1.0	(0.9, 1.2)	.6217
Registry 10	1.0	(0.8, 1.1)	.5121
Urban	1.0	(0.9, 1.1)	.9467
Not married	1.4	(1.3, 1.5)	.0001
No high school diploma (10% increments)	1.1	(1.0, 1.1)	.0008
Working class job (10% increments)	1.0	(1.0, 1.1)	.3355
Median family income ($ 20 thousands)	0.9	(0.9, 1.0)	.0107
Families <poverty, female head of house (10% increments)	1.0	(0.9, 1.1)	.8715
Home ownership (10% increments)	1.0	(1.0, 1.0)	.2417
No car (10% increments)	1.0	(1.0, 1,0)	.8409
Negative for hormone receptors	2.1	(2.0, 2.2)	.0001
Advanced grade tumor (grades 3 or 4)	2.0	(1.8, 2.1)	.0001

[a] OR = odds ratio adjusted for all other explanatory variables in the model.

[b] *p*-value for the Wald chi-square statistic with 1 degree of freedom.

APPENDIX III-3. Odds ratios for explanatory variables in full model comparing female breast cancer patients with regional stage disease to those with localized disease.

Variables in model	Regional vs. Localized		
	OR[a]	(95% CI)	p[b]
White	1.0		
Black	1.2	(1.1, 1.2)	.0001
Hispanic	1.1	(1.1, 1.2)	.0003
Japanese	0.8	(0.7, 0.9)	.0008
Filipino	1.0	(0.9, 1.1)	.9928
Chinese	1.0	(0.9, 1.1)	.9968
Hawaiian	1.0	(0.8, 1.3)	.8213
Korean	0.9	(0.7, 1.1)	.3344
Vietnamese	1.1	(0.9, 1.5)	.3967
American Indian	1.3	(0.9, 1.9)	.2256
Age (10 yr increments)	0.9	(0.9, 0.9)	.0001
Registry 0	1.0		
Registry 1	1.0	(0.9, 1.0)	.0726
Registry 2	0.9	(0.8, 0.9)	.0002
Registry 3	1.0	(1.0, 1.1)	.3796
Registry 4	0.8	(0.7, 0.9)	.0001
Registry 5	0.8	(0.8, 0.9)	.0001
Registry 6	1.0	(0.9, 1.1)	.9880
Registry 7	0.8	(0.8, 0.9)	.0001
Registry 8	1.1	(1.0, 1.2)	.1491
Registry 9	1.0	(0.9, 1.1)	.8744
Registry 10	1.0	(0.9, 1.0)	.1922
Urban	1.0	(0.9, 1.0)	.0304
Not married	1.1	(1.0, 1.1)	.0001
No high school diploma (10% increments)	1.0	(1.0, 1.1)	.0008
Working class job (10% increments)	1.0	(1.0, 1.0)	.2931
Median family income ($ 20 thousands)	1.0	(1.0, 1.0)	.9581
Families <poverty, female head of house (10% increments)	1.0	(1.0, 1.1)	.1763
Home ownership (10% increments)	1.0	(1.0, 1.0)	.4512
No car (10% increments)	1.0	(1.0, 1.0)	.1686
Negative for hormone receptors	0.8	(0.8, 0.8)	.0001
Advanced grade tumor (grades 3 or 4)	2.1	(2.0, 2.2)	.0001

[a] OR = odds ratio adjusted for all other explanatory variables in the model.

[b] *p*-value for the Wald chi-square statistic with 1 degree of freedom.

APPENDIX III-4. Odds ratios for explanatory variables in full model comparing female breast cancer patients with tumors greater than or equal to 1 cm in diameter to those with tumors less than 1 cm in diameter.

Variables in model	1 cm + vs. <1 cm		
	OR[a]	(95% CI)	p[b]
White	1.0		
Black	1.4	(1.3, 1.5)	.0001
Hispanic	1.2	(1.1, 1.3)	.0001
Japanese	0.9	(0.7, 1.0)	.0660
Filipino	1.4	(1.1, 1.6)	.0009
Chinese	1.1	(1.0, 1.4)	.1469
Hawaiian	1.2	(0.9, 1.7)	.3202
Korean	1.4	(1.0, 2.1)	.0956
Vietnamese	1.1	(0.7, 1.7)	.7326
American Indian	0.9	(0.5, 1.6)	.6709
Age (10 yr increments)	0.9	(0.9, 1.0)	.0001
Registry 0	1.0		
Registry 1	0.9	(0.9, 0.9)	.0166
Registry 2	0.9	(0.8, 1.0)	.0036
Registry 3	0.8	(0.7, 0.9)	.0001
Registry 4	0.7	(0.6, 0.9)	.0001
Registry 5	0.7	(0.6, 0.8)	.0001
Registry 6	0.9	(0.8, 1.0)	.0506
Registry 7	0.7	(0.6, 0.8)	.0001
Registry 8	0.9	(0.8, 1.0)	.4035
Registry 9	1.0	(0.9, 1.1)	.6223
Registry 10	1.0	(0.9, 1.1)	.9709
Urban	0.9	(0.9, 1.0)	.0003
Not married	1.2	(1.2, 1.3)	.0001
No high school diploma (10% increments)	1.1	(1.0, 1.1)	.0030
Working class job (10% increments)	1.0	(1.0, 1.1)	.0579
Median family income ($ 20 thousands)	<1.0	(0.9, 1.0)	.0250
Families <poverty, female head of house (10% increments)	1.0	(1.0, 1.1)	.2954
Home ownership (10% increments)	1.0	(1.0, 1.0)	.2285
No car (10% increments)	1.0	(0.9, 1.0)	.3461
Negative for hormone receptors	0.7	(0.7, 0.8)	.0001
Advanced grade tumor (grades 3 or 4)	3.3	(3.1, 3.4)	.0001

[a] OR = odds ratio adjusted for all other explanatory variables in the model.

[b] *p*-value for the Wald chi-square statistic with 1 degree of freedom.