AD_____

GRANT NUMBER: DAMD17-94-J-4509

TITLE: Massachusetts Cancer Control Evaluation Project

PRINCIPAL INVESTIGATOR: Susan Gershman, Ph.D.

CONTRACTING ORGANIZATION: Massachusetts Health Research Institute, Inc.
Boston, Massachusetts 02108

REPORT DATE: October 1995

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1995	Annual 23 Sep 94 - 22 Sep 95

4. TITLE AND SUBTITLE
Massachusetts Cancer Control Evaluation Project

5. FUNDING NUMBERS
DAMD17-94-J-4509

6. AUTHOR(S)
Susan Gershman, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Massachusetts Health Research Institute, Inc.
Boston, Massachusetts 02108

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The purpose of the Massachusetts Cancer Control Evaluation Project is to develop a model for evaluating the effectiveness of cancer control programs and to integrate this model within a computerized cancer registry management system. The evaluation components of this model include incidence, mortality, staging shifts, health behavior regarding cancer screening, location of and access to mammography, and socioeconomic factors.

Year 1 activities have focused on examining the distribution of breast cancer in MA and throughout the US. Using data from MA, SEER, CT and CA, we have explored trends in cancer incidence, staging, mortality and mammography screening, and have begun integration of these data sources. We have also analyzed census data, prepared population data for multiple geographic units of analysis and time periods, and examined correlations between various socioeconomic factors. Additionally, we have compiled a master file of data sources in preparation for developmental modeling,and have begun this statistical modeling.

Year 2 activities will focus upon completion of the statistical model, and integration of this model into the Mass. Cancer Registry's database (MCR=CIMS). In addition to the socioeconomic variables created from our measurement modeling of the census tract measures, other known covariates will be analyzed in Year 2.

14. SUBJECT TERMS
Breast cancer, Cancer control, program evaluation, multivariate modeling

15. NUMBER OF PAGES
48

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements***.

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	**PR**	- Project
G	- Grant	**TA**	- Task
PE	- Program Element	**WU**	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

- **DOD** - See DoDD 5230.24, "Distribution Statements on Technical Documents."
- **DOE** - See authorities.
- **NASA** - See Handbook NHB 2200.2.
- **NTIS** - Leave blank.

Block 12b. Distribution Code.

- **DOD** - Leave blank.
- **DOE** - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
- **NASA** - Leave blank.
- **NTIS** - Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

*U.S.GPO:1993-0-358-779

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

PI - Signature

10/19/95
Date

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION:

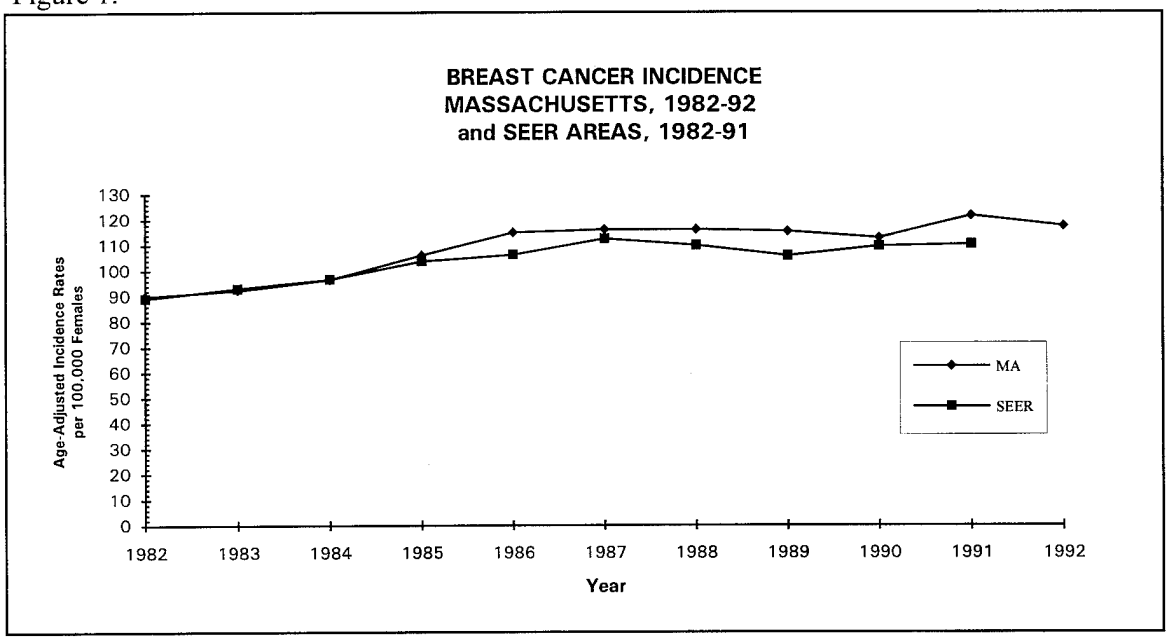
Nature of the Problem

Breast Cancer Incidence and Mortality:

Breast cancer is the most common cancer among women in the United States (excluding cancers of the skin), and is second only to lung cancer in causing cancer deaths in women (American Cancer Society, 1995a). According to the American Cancer Society (ACS), the average woman has approximately a 12.6% lifetime risk of developing invasive breast cancer, or about a one in eight chance (American Cancer Society, 1995b). The ACS estimates that 184,300 new breast cancer cases will be diagnosed among women in the United States during 1996 (American Cancer Society, 1995a). The incidence of breast cancer has risen dramatically over the past twenty years. According to the National Cancer Institute's Surveillance, Epidemiology and End Results (SEER) Program -- currently the best information available on national cancer incidence -- the incidence of breast cancer increased 24% between 1982 and 1991, from 89.1 per 100,000 in 1982 to 110.2 per 100,000 in 1990 (Ries et al., 1994) (Figure 1).

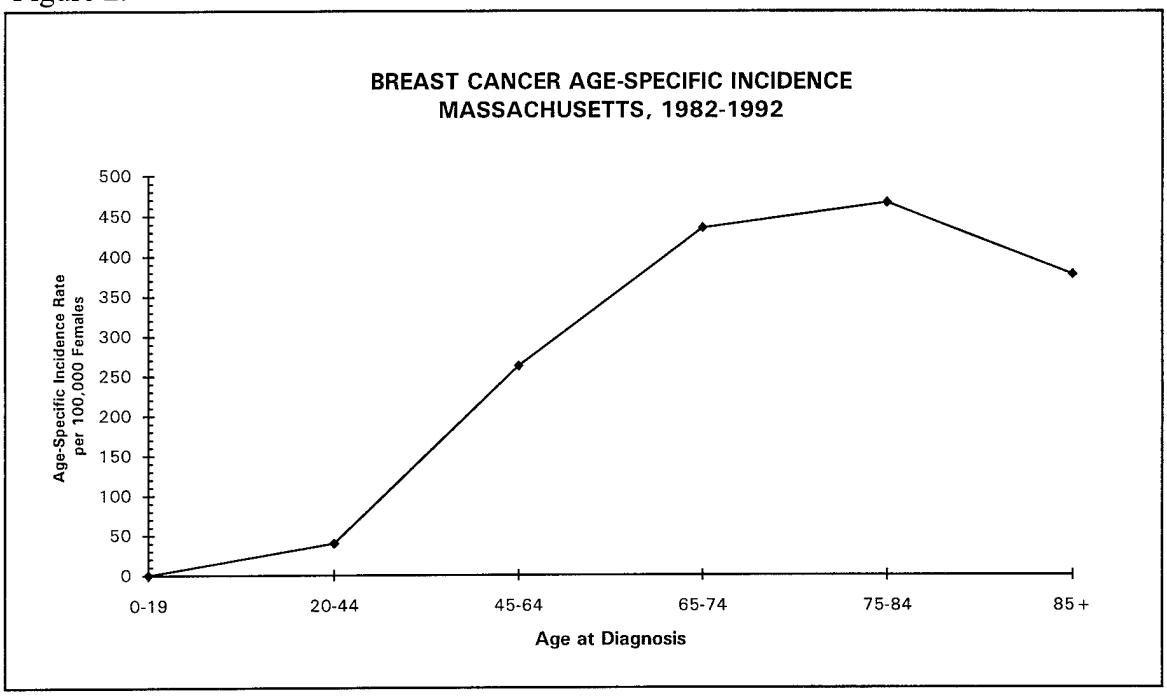
In Massachusetts, 46,070 new cases of breast cancer were reported between 1982 and 1992. Breast cancer was the leading cancer among females during this period, accounting for 30.9% of all newly diagnosed cancers. The average annual age-adjusted incidence rate for Massachusetts females for 1982-1992 was 109.4 per 100,000, and incidence increased more than 30% during this period (Figure 1).

Figure 1.



Among Massachusetts females breast cancer incidence increases steadily with age, reaching about 466 per 100,000 in ages 75-84, and then decreases in ages 85 and over (Figure 2).

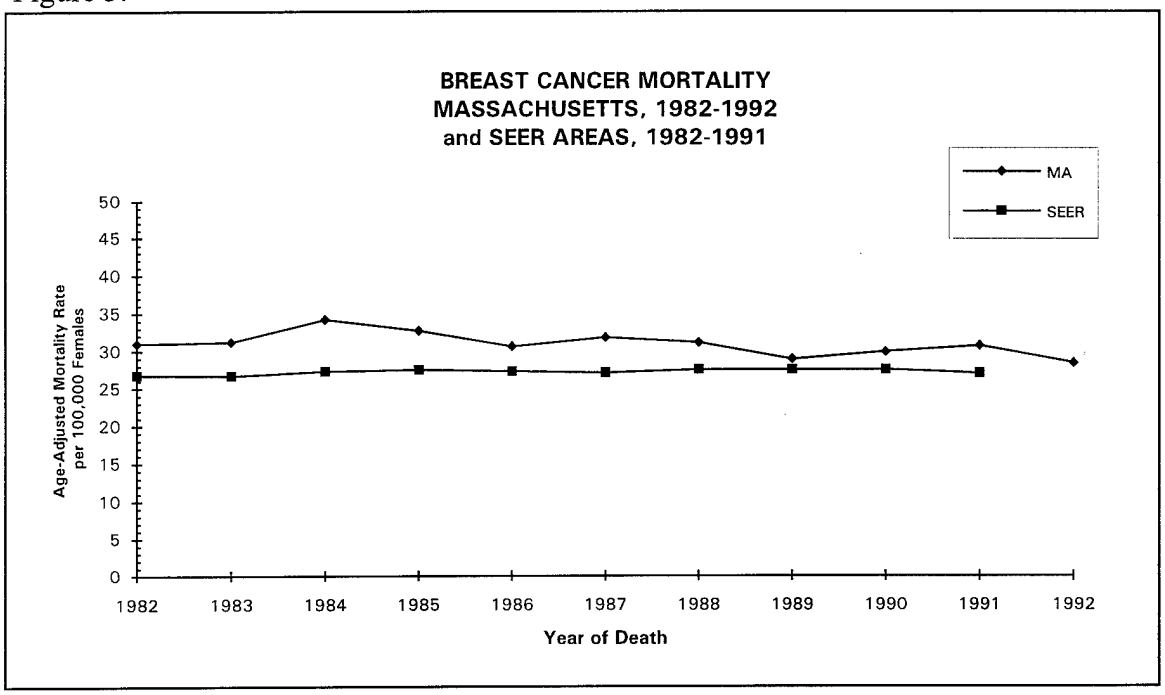
Figure 2.



For many years, breast cancer ranked as the number one killer of women both statewide and nationally. In recent years, however, lung cancer has overtaken breast cancer as the leading cause of cancer deaths in women. The ACS estimates that 44,300 women in the US will die of breast cancer in 1996, a slight decrease from the 46,000 deaths projected for 1995. The US mortality rate changed little between 1973 (the first year for which SEER data is available) and 1989, when the US mortality rate was 27.5 per 100,000 (Miller, 1993). The National Cancer Institute recently announced, however, a 4.7% decline in the breast cancer mortality rate between 1989 and 1992 (Smigel, 1995).

In 1993, the Massachusetts breast cancer mortality rate was 29.5 per 100,000. That year, 1316 Massachusetts women died of breast cancer. Mortality rates due to breast cancer among Massachusetts women are, on average, 18% higher than for women nationally (Figure 3). According to data from the Centers for Disease Control and Prevention, Massachusetts has the fourth highest breast cancer mortality rate in the United States (Morbidity and Mortality Weekly Report, 1994). Its 1993 mortality rate was 17% higher than the goal established in ***Healthy People 2000*** by the federal government to decrease the breast cancer mortality rate to not more than 25.2 per 100,000 (U.S. Department of Health and Human Services, 1991).

Figure 3.



Risk Factors:

A variety of factors have been shown to be associated with an elevated risk of breast cancer, including

- *behavioral factors* such as high dietary fat intake and daily alcohol intake;
- *hormonal and reproductive events* such as early age at menarche, menstrual cycle length, late age at menopause, menopausal status (including history of oophorectomy), late age at first childbirth or nulliparity; and
- *demographic characteristics*, including increasing age, race (being white for breast cancers diagnosed at greater than 45 years of age; being black for breast cancers diagnosed at less than 40 years of age), high socioeconomic status, having never married, being Jewish, urban residence, and residence in the northern United States (vs. the southern United States) (Kelsey, 1993).

Demographic characteristics related to socioeconomic status will be the variables of primary interest in this project.

Background

Breast Cancer Staging:

Cancers are staged by site and size of tumor and the extent of spread to lymph nodes or other organs. Neoplasms are categorized as either *in situ* or invasive. *In situ* designates an epithelial tumor that is bound by an intact basement membrane and has not invaded the organ. "Invasive" designates an epithelial tumor which has broken through

the underlying basement membrane and has assumed tumorogenic potential in the underlying tissue. Invasive tumors are further categorized as being *local* (within the organ), *regional* (beyond the organ by direct extension to surrounding organs or lymph nodes, or *distant* (metastasized to other organs or distant lymph nodes). Cancers are staged according to the tumor's size, nodal status and extent of metastasis at the time of diagnostic evaluation.

Previous Research:

It is postulated that the increased incidence in breast cancer seen during the 1980s may be due to an increase in mammography utilization, with resultant detection of earlier stage cancers than would have been detected without mammography. A study by White et al. (1990) found that mammography usage explains the increased incidence in women 45-64 years of age, while it only accounts for half of the increased incidence in women 65-74. Using population-based cancer registry records of the metropolitan Atlanta SEER program from 1979-1986, Liff et al. (1991) found that increased mammography detection accounted for some but not all of the rising incidence of breast cancer in the US. Feuer and colleagues (1992) developed an alternative model incorporating estimates of differential lead time by age group and found that the increase in incidence is concordant with increased mammography usage even in the older age groups. Kessler, Feuer, and Brown (1991) have mathematically modeled long-term incidence trends for 1990-2000, using Connecticut tumor registry data and information on availability of mammography machines. They projected that breast cancer incidence would continue to rise until approximately 1990 and then decline as screening rates stabilize.

There is no known primary prevention strategy for breast cancer, thus secondary prevention through mammography screening / early detection is the only method of breast cancer control. The first and most convincing evidence demonstrating the benefits of mammography screening was the Health Insurance Plan (HIP) of New York study, in which 62,000 women were randomized into two groups; half were offered annual mammograms and breast palpation and the other half received their usual care. The 10-year mortality rate for women 50 years and older was one-third lower among screenees than among controls (Shapiro, 1982). Also, the results of a Swedish trial confirmed the HIP study when it was found that single-view mammography decreased mortality from breast cancer by 40% in 50-74 year old women, although no significant reduction was observed in 40-49 year old women (Taber, 1985). Long-term survival of women with breast cancer depends on diagnosis at the early stages (Farley, 1989; Chu, 1991).

The ultimate goal of breast cancer screening is to decrease breast cancer mortality. Breast cancer mortality has remained constant since the 1930s (Kelsey, 1993), despite the increase in mammography screening during the 1980s. It may take many years for a decrease in mortality to be seen; thus, intermediary outcomes are necessary for evaluation of breast cancer control programs. A change in the distribution of incidence of disease by stage (an increase in the proportion of *in situ* and localized cases, and a decrease in the proportion of regional and distant invasive cases) has been postulated as an appropriate

intermediate outcome. These "staging shifts" serve as an important indicator of the success of cancer control activities. Robertson et al. (1990) and others have demonstrated that mammographic screening lowers the percentage of women presenting with stage II disease from 55% to 30% and increases the percentage presenting with stage I disease from 16% to 42%.

The work of three research groups has been the basis of a number of the analyses performed thus far:

Roffers and Austin: Assessment of Cancer Incidence Data

The North American Association of Central Cancer Registries (NAACCR) has developed a set of cancer registry data measures for use in evaluating the efficacy of breast and cervical cancer control programs (Roffers, 1992; Roffers and Austin, 1993). Specifically, the NAACCR's Technical Advisory Committee (TAC) sought to identify data items from population-based cancer registries which could be used to plan, implement, monitor, and evaluate cancer control projects.

Breast cancer measures were selected to represent three different indicators of early diagnosis. These measures are:
(1) the proportion of all breast cancers of known stage diagnosed at an *in situ* stage,
(2) the proportion of all invasive breast cancers of known stage diagnosed at a localized stage,
(3) average annual age-specific and stage-specific breast cancer incidence rates, and
(4) the proportion of localized female breast cancers diagnosed with a tumor size <= 2 cm in diameter (of all cases of known stage and known tumor size).
In project analyses, these measures will be referred to as "Roffers 1", "Roffers 2", "Roffers 3" and "Roffers 4", respectively.

According to the TAC these measures provide an indication of the effectiveness of screening mammography and early detection. For example, populations with a low degree of screening mammography and a high reliance upon manual screening would be expected to have a low percentage of breast cancers diagnosed at an *in situ* stage (less than five percent). Populations with higher degrees of screening mammography have higher proportions of *in situ* cancers, up to 15 to 20 percent. Thus, Roffers 1 serves to indicate information about the relative frequency of screening mammography.

Roffers 2 indicates the degree to which manual screening methods are utilized. High percentages of localized disease (above 75%) indicate relatively high levels of manual screening, whereas lower percentages of localized disease (40 to 50%) indicate low levels of manual screening. Evaluations of these measures have shown that Roffers 1 and Roffers 2 vary independently, reflecting different aspects of cancer control.

Roffers 4 serves as an additional indicator of early detection, although the TAC notes that a degree of confounding may occur when assessing detection of cancers of size

2 cm or less, where detection by manual palpation is most difficult. This is because this measure, which uses less than 2 cm as a point of dichotomizing non-metastatic disease, would be most confounded as to which measure of early detection (mammography vs. manual palpation) it is detecting. The TAC also recommends that age-specific and stage-specific incidence rates (Roffers 3) be utilized, as they serve as yet another intermediate indicator of cancer control efforts.

Andrews et al.: Assessment of Census Data

Andrews et al. (1994) utilized combinations of census-based demographic variables and cancer-specific mortality rates to predict the incidence of cancers diagnosed at a late stage. ("Late stage" is defined as regional or distant disease.) Specifically, they developed a small-area multiple regression model which related cancer incidence to mortality, census demographics, or both, in areas where cancer registry data were available. They then used this model to estimate late-stage incidence for areas where cancer registry data were not available. Work was done for breast, cervical and colorectal cancers. Areas of interest were "health areas", administrative units of the New York City Department of Health consisting of four to six census tracts.

Demographic predictors were selected on the basis of two criteria: a known etiologic relationship to at least one of the cancers of interest, and an absence of multicollinearity among these predictors. On this basis, fourteen variables were selected. Multiple regression was then used to isolate four variables which accounted for nearly as much variability in rates as the entire set of fourteen variables. These final four variables were (1) the percentage of the population aged 65 and older, (2) the percentage of household incomes greater than $50,000, (3) the percentage of the population aged 15 and older who were divorced or separated, and (4) the percentage of women in the labor force with one or more children aged 16 years or younger.

Andrews found that good estimates of late-stage rates of breast, cervical and colorectal cancers could be developed utilizing the above census-based variables. The inclusion of site-specific mortality data further increased the accuracy of estimation. Mortality alone was also found to be valuable in targeting areas where late-stage disease is high, but adding these selected demographic variables added 10% to 20% to the explained variability.

Farley and Flannery: Association of Socioeconomic Status and Late-Stage Diagnosis

Farley and Flannery (1989) used data from the Connecticut Tumor Registry to examine trends in stage at diagnosis of breast cancer over time in relationship to socioeconomic indicators, and to project numbers of "preventable" deaths from breast cancer in various groups. They first examined the distribution of cancer stage at time of diagnosis by year for 1975-1985, utilizing the stage categories of *carcinoma in situ, local* (invasive cancer localized to the breast), *regional* (cancer in the breast with spread to regional lymph nodes or pectoral muscles) and *remote* (presence of distant metastases).

Information on the census tract of residence was available as of 1984, and stage at diagnosis was examined by race, place of residence and socioeconomic status (as estimated from census tract information) for 1984 and 1985.

Socioeconomic status (SES) was estimated for each census tract using three markers: median household income, percentage of persons below the poverty line, and percentage of adults who have completed a high school education. As these three variables were found to be highly correlated, and led to identical conclusions, one variable was selected for use as an SES indicator. Values of this variable were used to group women into quartiles, using the percentage of high school graduates in a woman's census tract as a surrogate for her socioeconomic status.

Using data on survival rates, population estimates, and the number of breast cancer cases, Farley and Flannery calculated the projected number of deaths from breast cancer in a cohort of women with breast cancer for 1984-85. They further divided this estimate into estimates of "nonpreventable" vs. "preventable" deaths.

Results of these analyses showed that between 1975 and 1981, there was little variation from year to year in stage at diagnosis, while from 1982 to 1985 there was a statistically significant increase in the proportion of cancers diagnosed at an *in situ* or local stage. For 1984 and 1985, cancer stage was significantly associated with SES. Women in lower SES tracts were significantly more likely to present with remote disease and less likely to present with *in situ* or localized disease than women in high SES tracts. These differences persisted after adjusting for race, although black women were significantly less likely to have *in situ* or local cancer, and more likely to have remote disease. In examining projected mortality, lower SES women had a 25% higher projected death cancer rate and a greater percentage of those deaths termed "preventable".

Purpose of Present Work

Massachusetts is currently one of 35 states receiving comprehensive screening funding from the Centers for Disease Control and Prevention (CDC) under the national Breast and Cervical Cancer Prevention and Control Program. Under this program, the Massachusetts Department of Public Health funds 37 sites throughout the state to provide breast and cervical cancer screening services (including mammograms, clinical breast exams, pap smears and physical exams, and instruction in breast self-examination) to uninsured and underinsured women. Public education, professional education, quality assurance and surveillance are also integral components of this program. Because of the existence of the Massachusetts Breast and Cervical Cancer Initiative (BCCI), and the availability of multiple data sources for breast cancer, this project is focusing initially on the development of a model for the assessment of breast cancer control activities.

The purpose of this project is to integrate within a cancer registry management system a component to evaluate the effectiveness of cancer control programs. The evaluation components of this model include incidence, mortality, staging shifts, health

behavior regarding mammography usage, location of and access to mammography usage, and socioeconomic factors. Often, early detection programs are implemented without a means of evaluating the program. Through this project, an efficient, effective model for program evaluation and modification is being designed.

An additional model aimed at estimating cancer incidence in small geographical areas is also being considered. A small area estimation model will be useful in targeting areas in need of cancer screening programs. An estimation of a large proportion of late stage diagnoses would be indicative of an area in which it is necessary to target screening programs. The main goal of such a program would be to reduce the proportion of late stage incidence at diagnosis by identifying *in situ* and localized cancer in individuals who would otherwise have progressed to a late stage cancer by the time of diagnosis and would have had a lower chance of survival.

Methods of Approach

This project's data is being examined by three different geographic units of analysis. Proceeding from smallest to largest, they are:

Census Tracts:

Census tracts are geographic units designed collaboratively by the Census Bureau and communities. They were initially created in 1970 so as to contain homogenous groups of 2000 to 8000 persons (on average, 4000 persons). The intent of census tracts was to create stable geographical units which do not change boundaries over time, so that communities could monitor changes in their populations below the city/town level. Since 1970, however, some census tracts have been subdivided because of population growth, and census tracts have been created in four previously untracted counties in Massachusetts.

Cities and Towns:

Massachusetts has 351 incorporated cities and towns, which account for all land in the Commonwealth. These cities and towns are equivalent to the Census Bureau's "Minor Civil Divisions" (MCDs).

Community Health Network Areas:

The Department of Public Health has divided Massachusetts into 27 Community Health Network Areas (CHNAs). CHNAs have been created by aggregating cities and towns in order to develop health networks -- consortia of health care providers, human service agencies, schools, churches, advocacy groups, and members of the public of all ages. These networks will identify and assess health needs in their communities, and evaluate responses to these health needs. The major foci of the networks are increased access to care, increased efficiency of health services delivery, and increased

communication and collaboration among health care and human service providers in these areas.

Increasingly, the Department is analyzing data on the basis of CHNAs, so as to provide area-wide data to these coalitions. In many parts of the state, the numbers of persons and number of occurrences of health conditions are small, and it is difficult to assess trends and identify problems. By examining data on a CHNA-wide level, coalitions can more readily identify and monitor health conditions and problems in their communities.

BODY:

The project staff has already analyzed multiple data sets, including Massachusetts incidence and mortality, SEER incidence, California incidence, Connecticut incidence, Massachusetts Behavioral Risk Factor Surveillance System (BRFSS) data, and Massachusetts Census data. Massachusetts data will be utilized in the development of an evaluation model, while other registries' data sets were examined in order that staff members could familiarize themselves with breast cancer incidence data, including trends and staging shifts. A summary of cancer incidence variables collected and/or available for analysis from the Massachusetts, SEER, California and Connecticut registries is provided in Table A (in Appendix). A summary of analyses conducted on these data sets is provided in Table B (in Appendix), while selected results are given in the text.

Massachusetts Incidence

The Massachusetts Cancer Registry (MCR) collects data on the incidence of breast cancer in Massachusetts. The MCR began collecting cancer incidence data on Massachusetts residents with cases diagnosed as of January 1, 1982. Currently, case reports are obtained from two sources: 92 acute care hospitals, and seven state cancer registries through reciprocal agreements (Connecticut, Maine, New Hampshire, New York, Rhode Island, Vermont and Florida). The MCR processes data on more than 30,000 cases per year, and its data base currently consists of approximately 325,000 cases. Reporting is estimated to be 90% complete.

The information collected by the MCR on its reporting form includes demographic variables such as age, sex, race, and town of residence as well as information on the primary site, histology, and stage of tumor. The MCR began collecting data on *in situ* carcinomas with cases diagnosed as of January 1, 1992. Registry data go through extensive checks for quality assurance and completeness of reporting. Case-specific data are confidential by law and are released only after a thorough review of research requests.

The Massachusetts breast cancer incidence file contains 46,859 cases collected over the 10 year span from 1982 to 1992. *In situ* cases were only collected for 1992; for

this year, 15.3% of reported cases were diagnosed at an *in situ* stage, with the proportion *in situ* decreasing with increasing age (Figure 4, in Appendix). For the overall period 1982-1992, localized cases accounted for 57.2% of the 46,859 cases, regional 29.2%, distant 6.2% and unknown 5.7%. The proportion of localized cases increased over time (Figures 5a and 6, in Appendix), while the proportion of regional and distant cases decreased over time (Figures 5b, 5c and 6, in Appendix). This trend is what one would expect after breast cancer screening programs have been implemented. When trends in staged tumors were examined for the three age groups 0-49, 50-64 and 65+, we found that the proportion of localized tumors increased as age increased (Figure 5a), the proportion of regional tumors decreased as age increased (Figure 5b), and the proportion of distant tumors increased as age increased (Figure 5c). It is somewhat surprising that younger women had more regional tumors than their older counterparts.

When annual age-specific breast cancer rates were analyzed by 5-year age groups, we see that the 40-44, 45-49 and 50-54 year age groups show a trend of increased incidence (Figure 7, in Appendix). Thus, we believe that the 30% increase in breast cancer incidence which has been observed in Massachusetts between 1982 and 1992 is primarily attributable to increased detection of localized cancers in women aged 40-54. Given this trend, we will expect to observe a similar increase in detection of *in situ* cancers, although data to evaluate this trend is not yet available.

Massachusetts Mortality

Data on deaths from breast cancer in Massachusetts are collected by the Department of Public Health's Registry of Vital Records and Statistics. Established in 1841, the Registry is responsible for the legal registration, collection, and reporting of almost 250,000 births, deaths, marriages, and divorces annually, and provides data on cancer mortality. Each year the Registry issues its **Annual Report: Vital Statistics of Massachusetts**, the oldest continually published statewide vital statistics report in the United States. In conjunction with the Registry of Vital Records, the Division of Research and Epidemiology publishes an **Advance Data** series with separate volumes for births and deaths. This series reports community-specific information as well as statewide information on variations in age-adjusted mortality rates, ethnic variations in mortality, years of life lost, and trends. Registry of Vital Records and Statistics data constitute the basis for identifying communities excessively burdened by disease -- such as breast cancer -- and for developing programs and services to address these needs.

Age-adjusted mortality rates have been calculated by CHNA for 1993; rates ranged from a low of 21.3 per 100,000 for CHNA 25 (Fall River) to a high of 42.2 per 100,000 for CHNA 16 (Medford). Age-adjusted mortality rates were also calculated by CHNA for the time periods 1982-86 and 1987-92. Due to computational errors in standardization, however, the results of these analyses are presently under revision.

The Surveillance, Epidemiology, and End Results (SEER) Program

Through the National Cancer Institute's SEER Program, data on cancer incidence are collected from nine geographic areas throughout the United States, which represent approximately 9.5% of the U.S. population. This program reports data for cancers reported in the selected areas beginning with cases diagnosed as of January 1, 1973. Areas currently participating in the SEER Program are Connecticut, Hawaii, Iowa, New Mexico, Utah, Detroit Standardized Metropolitan Statistical Area (SMSA), Atlanta SMSA, San Francisco-Oakland SMSA, Seattle-Puget Sound, Los Angeles County and four counties in San Jose-Monterey Area, California. The latter two areas became SEER registries in 1992, and data is not yet available for analysis. Analysis was conducted both overall and on each of the first nine SEER registries listed.

As seen in Figure 8 (in Appendix), age-adjusted breast cancer incidence steadily increased from 1982 to 1991 in all of the SEER areas. During this period, 10% of breast cancers were diagnosed at an *in situ* stage, 51% localized, 30% regional, 6% distant and 4% unknown. This is consistent with the stage distribution seen in the Massachusetts incidence file. As also noted in Massachusetts, the proportion of SEER cases diagnosed at a localized stage increased over time from 50.6% to 64.2% between 1982 and 1991. The proportion *in situ* increased from 4.7% to 13% during that same time period.

California Incidence

Data on breast cancer incidence in California were provided by the California Cancer Registry (CCR). The CCR first collected cancer incidence data from selected California hospitals beginning in 1947. Reporting of newly-diagnosed cancer cases has been mandated by law since 1985, and the CCR has collected information statewide since 1988. The public use tape analyzed contains breast cancer cases diagnosed among female California residents between January 1, 1988, and December 31, 1992, and reported to the CCR as of November, 1994.

The age-adjusted breast cancer incidence rate for California increased slightly from 122 per 100,000 to 125 per 100,000 between 1988 and 1992. The proportion *in situ* increased from 11.1% in 1988 to 13.2% in 1991, while the proportion of localized cases increased from 61.3% to 64.2% during this time.

Connecticut Incidence

Data on breast cancer incidence in Connecticut was provided by the Connecticut Tumor Registry (CTR). The CTR is the oldest cancer registry in the US, with initial operation in 1935 and population-based data available since 1941. It is a participant in the SEER Program. The public use tape analyzed contains data for reporting years 1973 through 1992.

As expected, the proportion of breast cancers diagnosed at an *in situ* stage increased over time from 4% in 1982 to 13% in 1992. Comparable calculations for the proportion of cancers diagnosed at a localized stage are under revision at this time.

Behavioral Risk Factor Surveillance System

Data on cancer screening practices among Massachusetts women are available through the Behavioral Risk Factor Surveillance System (BRFSS), a telephone survey conducted in nearly every state under the auspices of the CDC. At present, almost 3,000 Massachusetts residents are surveyed annually, including approximately 2,000 women. They are asked a series of questions about preventive health practices, including cancer screening. The BRFSS includes a women's health section which asks female respondents about their use of mammography, clinical breast exams (CBEs) and pap smears. Among the questions asked are whether or not the woman has ever had the exam, how recent her last exam was, and the reason for the last exam. In 1994, Massachusetts added new state-specific questions on whether women know how to perform breast self-examination, and how often they do so.

BRFSS data files have been examined for 1990, 1991 and 1992. For these years, the number of women surveyed were 737, 800, and 816 respectively. Some data were also available for the 897 women surveyed in 1993. For each year available, questions relating to use of mammography were analyzed overall and by CHNA. This CHNA-level data is included in Table C (in Appendix). Among the findings were that the percentage of women surveyed in 1993 who had ever had a mammogram ranged from a low of 35.7% in CHNA 9 (Fitchburg) to a high of 72.4% in CHNA 10 (Lowell). The proportion of women surveyed that year who had had a mammogram within the last year (among those who had ever had a mammogram) ranged from a low of 7.4% in CHNA 9 (Fitchburg) to a high of 92.1 in CHNA 24 (Taunton).

Variable Rankings by CHNA

One of the more interesting analyses conducted thus far is shown in Table C (in Appendix). As noted previously, CHNAs (Community Health Network Areas) are one of the geographic units of analysis for this project. Here, a number of variables analyzed separately are ranked by CHNA. [As an example, values for the first variable -- %insitu92, or the proportion of 1992 breast cancer incidence cases diagnosed at an *in situ* stage -- ranged from a low of 9% in CHNA 9 (Fitchburg) to a high of 23% in CHNA 15 (Woburn).] The variables utilized include breast cancer incidence data (proportion of cases diagnosed at an *in situ* stage, crude incidence rates), BRFSS data on mammography screening, breast cancer mortality data, and census demographic variables (per capita income, percentage below poverty level, percentage with household income >$50,000, percentage of home ownership, and several education variables).

In examining these multiple data sources in this way, interesting patterns emerged. For example, CHNA 9 (Fitchburg) showed a low percentage of *in situ* diagnoses, a low percentage of recent mammography or ever mammography, and a high breast cancer mortality rate. CHNA 18 (Newton/Waltham), conversely, showed a high proportion of *in situ* diagnoses, a high proportion of recent mammography and ever

mammography, and a high level of education. CHNA 15 (Woburn) showed a high proportion of *in situ* diagnoses, a high income level, high levels of home ownership, and high educational levels. Overall, we found that high mammography usage was "associated" with CHNAs that had greater than 15% *in situ* diagnoses, and had high per capita income, high percentage of home ownership, and higher levels of education.

Census Data

Population counts by sex and age for Massachusetts cities and towns for 1980 and 1990 were brought together in order to interpolate sex/age counts per town/city for the intercensal years 1982-1989. These numbers are the denominators for subsequent rate (incidence, mortality, etc.) determinations. A similar effort was made with census tracts. However, the absence of 1980 census tracts in four Massachusetts counties rendered impossible intercensal interpolations for all of the Commonwealth's census tracts.

Demographic variables deemed to be associated with health and/or access to adequate health care were identified in the Census Bureau's Summary Tape File 3A for 1980 and 1990. Most of the responses were from the "long form" of the census questionnaire sent to an approximately 16% sample of the nation's residents. The items are available for the 351 Massachusetts cities and towns, and for the state's census tracts.

These computations have resulted in a data set of 1,177 census tracts for which socioeconomic characteristics and other relevant data were collected in the 1990 US Census. Data for these 1,177 tracts remained after data from some tracts had been omitted or combined for various reasons (such as too few women -- e.g., a tract which recorded the male prison population in a community, or too little data -- e.g., college dormitories with large numbers of women with no income, no working mothers with children, etc.). The geographic location of the center of each tract was given by its latitude and longitude in degrees.

The census variables retained were measured as rates (percents) of the appropriate populations: (1) Non-whites, (2) Blacks, (3) Asians, (4) Asian language spoken at home, (5) Hispanics, (6) Spanish spoken at home, (7) Elderly (65+ years of age), (8) Mothers in the labor force with children younger than age eighteen, (9) Women separated or divorced, (10) Foreign-born; (11) Educational attainment of less than 9 grades, (12) Educational attainment of some high school, (13) Educational attainment of high school and some college, (14) Educational attainment of four years of college, (15) Unemployed, (16) Persons below poverty level, (17) Persons living in the same house for the past five years, (18) Persons owning their own home, and (19) Per capita income.

Incidence data were collected from two periods: 1982-1986 and 1987-1992; the data were the stages reported at diagnosis, from which incidence could be aggregated and Roffers' proportions could be determined. The frequency distribution of ages suggested that those women of age 30 through 94 years would be a suitable universe. This

population in each tract was the denominator for incidence and stage rates. These data sets are those used in the following statistical analyses among census tracts.

Modeling of Socioeconomic Variables

Recent attention has focused on those census variables that may reflect socioeconomic status. Some of the selected variables are shown in Table D along with their means and standard deviations. Currently analyses are being done at the level of the census tract, with 1,177 census tracts in Massachusetts. The data are also available at the level of the towns, with 351 in Massachusetts, or at the level of CHNA, with 27 in Massachusetts.

Table D. Census variables associated with socioeconomic status.

Variable Label	Census Variable	Mean	Std Dev
PBLK	% black	6.7%	15.97%
PHSP	% Hispanic	6.0%	11.32%
PFORN	% foreign-born	10.3%	8.88%
PEDCL9	% <9 grades school	9.6%	9.52%
PCOLL4	% 4-yr college degree	26.3%	17.02%
PHS13	% some hs but not completed	12.9%	6.87%
PCVUNEM	% unemployed	7.5%	4.54%
PBLWPOV	% annual income below poverty level	10.6%	10.37%
PFMKD	% women in labor force spouseless with children <18	6.8%	5.4%
PFSPDV	% females separated or divorced	11.0%	4.87%
PERCAP	Per capita annual income	$16783.50	$6663.99
POWNR	% owning own homes	56.3%	25.22%
PTEGT65	% of population 65 or over	14.0%	5.74%
STG1RTE	Stage 1 incidence rate	121.1 per 100,000	121.61
PRPSTG1	Prop Stage 1 of all diagnoses	0.57	0.36

From Table D it is evident that for many of the measures, the standard deviations are as large or larger than the means, suggesting non-normal distributions. Figures 9, 10, and 11 (in Appendix) show the distributions for several of these measures. In fact, there is statistically significant skewness and kurtosis for each of these variables separately,

and when they are examined as a multivariate set, there is serious departure from multivariate normality.

The violation of the assumption of multivariate normality poses problems to the investigation of a possible measurement structure underlying these data. Most of the approaches to exploratory factor analysis assume multivariate normality. The approach in this project, therefore, has to use confirmatory factor analysis, where possible measurement structures can be tested proactively. Furthermore, under conditions of non-normality, our approach was to analyze the covariance matrix rather than the correlation matrix since most forms of correlation assume multivariate normality. It has been shown that under conditions of non-normality, weighted least squares ensures correct estimates of model parameters (Browne, 1984).

So far, a number of alternative measurement structures have been proposed and tested. One model hypothesizes a single theoretical variable, SES, underlying the census measures. A second model, with its path diagram sketched in Figure 12 (in Appendix), hypothesizes three theoretical variables, Race/Ethnicity, Education, and Economics. A third model hypothesizes the same three theoretical variables, but a second order factor, SES, underlying those three, as sketched in Figure 13 (in Appendix).

Confirmatory factor analysis not only offers a proactive approach to testing measurement structures underlying the data, it also provides goodness of fit indices to determine how well each model fits the data. In this way, one model can be compared to another. Furthermore, it provides an extensive amount of diagnostic information to help understand where the model fits badly, or where certain measures are redundant, unreliable, or do not contribute useful information. Its most important potential contribution is parsimony, whereby, starting with some 19 candidate census tract variables, all reflecting some aspect of socioeconomic status, it provides a way of reducing these data to a much smaller set without serious loss of their information content. Table E shows a grouping of the candidate variables according to the three hypothesized variables: Race/Ethnicity, Education, Economics.

Table E. Census variables organized by hypothetical variables.

Race/Ethnicity	Education	Economics
% non-white	% < 9 grades of school	% unemployment
% black	% with some high school	% below poverty
% Asian	% with some college	% in same house
% Asian language	% with 4-yr college degree	% owning own home
% Hispanic		Per capita annual income
% Hispanic language		% women spouseless in labor force with children <18
% Foreign born		% females separated or divorced

Using the three sets of measures shown in the above table, the most recent analysis tested each set separately, to determine whether it was justifiable to hypothesize a single underlying measurement structure for each set. The process of testing each set also served to identify problems and to eliminate measures because of poor fit arising from excessive measurement error, redundancy, or information value.

The results of this process provide a Race/Ethnicity factor consisting of a combination of three of the seven candidate variables. The linear combination and their standardized regression coefficients consisted of: 0.552*PBLK, 0.714*PHSP, and 0.606*PFORN. This process reduced the education measures from four to two, PHS13 and PCOLL4, but with only two measures it is not possible to characterize the fit statistically.

By far, the measurement structure of the economic measures had the best statistical characteristics. The number of measures was reduced from seven to four: -0.875*PCVUNEM, -0.851*PLBPOV, 0.669*PERCAP, and -0.697* PFMKD. Furthermore, the fit indices were all supportive of a model with a single underlying structure. It will be possible to create a single economics variable from this analysis to use in future studies.

The economics model and the race/ethnicity model were next combined to test a two factor structure, that is, to determine whether it is reasonable to assume that two distinct, yet correlated, variables underlie these measures. The fit indices for the two factor structure were acceptable, but the race/ethnicity and economic factors are so highly correlated that the model amounts to a distinction without a difference. So, an additional model was tested to determine whether the race/ethnicity and economics factors could be combined into a single theoretical variable. According to the fit indices, the combined single factor model fits the data better than the two construct model. This analysis provides us with a second useful economics variable which also contains a race/ethnicity

component. In future analyses this will mean that we will have more than one way to control for or account for socioeconomic influences.

While a stable education factor could not be justified statistically, combinations of the educational measures were found to complement racial/ethnic measures in such a way that a useful two factor model could be developed. The two factors are correlated in the moderate to high range, 0.748. The factor score regression weights for the two measures are shown in Table F.

Table F. Factor score regressions for two factor model.

	% <9 grades	% some hs	% black	% Hispanic	% foreign
Education	0.659	0.212	0.009	0.075	0.032
Race/ Ethnicity	0.197	0.063	0.068	0.540	0.288

Tests of the three factor model produced statistically acceptable fit indices, but revealed again the highly correlated nature of these factors. The correlations between pairs of factors are shown in Table G. Clearly the correlation between the economics factor and the education factor is high, 0.930, as is the correlation between the economics factor and the race/ethnicity factor, 0.965, while the correlation between the race/ethnicity factor and the education factor drops to 0.731.

Table G. Correlations between factors in the three factor model.

	Economics	**Education**	**Race/Ethnicity**
Economics	1.000		
Education	0.930	1.000	
Race/Ethnicity	0.965	0.731	1.000

It would seem from the above analyses that the project now has available a number of alternative measures which parsimoniously capture the economics information, the race/ethnicity information, and the education information available from the census data. The extent to which the dependent variables investigated in this project are affected by these socioeconomic factors should now be accessible.

Integration of Statistical Model into MCR-CIMS

In order to ensure that the Cancer Control Automated Evaluation Model to be developed is fully integrated into the Massachusetts Cancer Registry-Cancer Information Management System (MCR-CIMS), the project software engineers have focused on

familiarizing themselves with the production system. This has included meeting with the software developers of MCR-CIMS and discussing in detail the design and implementation strategy, as well as how the system will be used by MCR staff. To ensure that the software engineers were familiar with MCR-CIMS, it was decided to have them make all the necessary modifications and/or enhancements to specific components of the system. This step was necessary to prepare MCR-CIMS for having the Cancer Control Automated Evaluation Model become an integral component to the system.

The software engineers began by first identifying the components that needed to be modified and/or enhanced to ensure compatibility with the Evaluation Model. The components identified in MCR-CIMS provide the end user with the ability to create ad hoc queries and perform statistical and mapping analyses on cancer incidence data. Once the components had been identified, a project schedule was created and implemented. Completion is expected in December 1995. During this phase, a set of requirements for the Cancer Control Automated Evaluation Model is being prepared for the software engineers to implement beginning January 1996, providing that the necessary modifications and/or enhancements to MCR-CIMS have been completed.

CONCLUSIONS:

Year 1 activities have focused on examining the distribution of breast cancer in Massachusetts and throughout the US. Using data from Massachusetts, SEER, Connecticut and California, we have explored trends in cancer incidence, staging, mortality and mammography screening, and begun integration of these data sources. We have also analyzed census data, prepared population data for multiple geographic units of analysis and multiple time periods, and examined correlations between various socioeconomic factors. Additionally, we have compiled a master file of data sources in preparation for developmental modeling. We anticipate more complete and better-founded results of this modeling because of the improved quality and completeness of data being used in these analyses, particularly census data.

Year 2 activities will focus upon completion of the statistical model, and integration of this model into the Massachusetts Cancer Registry's database (MCR-CIMS). In Year 2, the following tasks will be done in order to integrate a Cancer Control Automated Evaluation Model into MCR-CIMS:

1. End user polling
2. Requirements analysis
3. Finalization of features and capabilities
4. Comprehensive formal system design
5. Software development
6. System integration
7. Beta-testing
8. System modifications
9. Final release.

(An overview of MCR-CIMS is provided in Figures 14, 15 and 16, in the Appendix.)

Future Analyses

In addition to the socioeconomic variables created from our measurement modeling of the census tract measures, other known covariates will be analyzed. Age, for instance, is an extremely important covariate, and is available at the level of the census tract in many forms. The percent of the female population equal to or greater than 65 years of age has already emerged in our data as highly correlated with the incidence of breast cancer. The literature has also revealed that the availability of mammography facilities is also critical. The project staff is assembling mammography site information and integrating that information into the census tract database. While such information is useful in its own right as a measure of diagnostic availability, it may also be useful as a covariate in our modeling efforts.

Latitude and longitude data have also been incorporated into the census tract data base for use with spatial scan statistical analysis (Kulldorff, 1994). Kulldorff is currently incorporating the time dimension into his program, a feature that may also be useful.

While current analyses are being conducted at the level of the census tract as the unit of analysis, it will also be possible to conduct analyses at different levels, using towns or CHNAs as the unit of analysis. One scenario envisions the CHNA as client with interest in examining breast cancer information and relevant covariates for the CHNA as a whole first, and then calling for analyses at the level of the towns within the CHNA, and finally the census tracts within the CHNA. In this way, CHNAs would have available overall information as well as detailed maps of variation within their region.

REFERENCES

American Cancer Society. Breast cancer facts and figures -- 1996. Atlanta: American Cancer Society; 1995a.

American Cancer Society. Cancer facts and figures -- 1995. Atlanta: American Cancer Society; 1995b.

Andrews HF, Kerner JF, Zauber AG, Mandelblatt J, Pittman J and Streuning E. Using census and mortality data to target small areas for breast, colorectal, and cervical cancer screening. *American Journal of Public Health.* 1994;84(1):56-61.

Browne MW. Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology.* 1984;37:62-63.

Chu KC, Kramer BS, Smart CR. Analysis of the role of cancer prevention and control measures in reducing cancer mortality. *Journal of the National Cancer Institute.* 1991;83(22):1636-1643.

Connecticut Tumor Registry Breast Cancer Incidence File, 1973-1992.

Deaths from breast cancer - United States, 1991. *Morbidity and Mortality Weekly Report.* 1994;43(15):273-281.

Farley TA and Flannery JT. Late-stage diagnosis of breast cancer in women of lower socioeconomic status: public health implications. *American Journal of Public Health.* 1989;79(11):1508-1512.

Feuer EJ and Wun L-M. How much of the recent rise in breast cancer incidence can be explained by increases in mammography utilization? A dynamic population model approach. *American Journal of Epidemiology.* 1992;136(12):1423-1436.

Kelsey JL. Breast cancer epidemiology: summary and future directions. *Epidemiologic Reviews.* 1993;15(1):256-63.

Kessler LG, Feuer EJ and Brown ML. Projections of the breast cancer burden to U.S. women: 1990-2000. *Preventive Medicine.* 1991;20:170-182.

Kulldorff M. Spatial disease clusters: detection and inference. *Statistics in Medicine.* 1994;13:1-12.

Liff JM, Sung JF, Chow WH et al. Does increased detection account for the rising incidence of breast cancer? *American Journal of Public Health.* 1991;81(4):462.

Massachusetts Department of Public Health, Bureau of Health Statistics, Research and Evaluation, Division of Research and Epidemiology, Chronic Disease Surveillance Program. Massachusetts Behavioral Risk Factor Surveillance System computer files, 1995.

Massachusetts Department of Public Health, Bureau of Health Statistics, Research and Evaluation, Massachusetts Cancer Registry. Computer files, 1995.

Massachusetts Department of Public Health, Bureau of Health Statistics, Research and Evaluation, Registry of Vital Records and Statistics. Computer files, 1995.

Miller BA, Ries LAG, Hankey BF, Kosary CL, Harras A, Devesa SS, Edwards BK (eds). SEER Cancer Statistics Review: 1973-1990, National Cancer Institute. NIH Pub. No. 93-2789. Bethesda, MD, 1993.

Ries LAG, Miller BA, Hankey BF, Kosary CL, Harras A, Edwards BK (eds). SEER Cancer Statistics Review, 1973-1991: Tables and Graphs, National Cancer Institute. NIH Pub. No. 94-2789. Bethesda, MD, 1994.

Robertson FM, Romanow J, Otchy DP, Walters MJ. The effect of mass screening mammography on staging of carcinoma of the breast in women. *Gynecology and Obstetrics.* 1990;171:55.

Roffers SD and Austin DF. Cancer registry data measures for breast and cervical cancer control: definitions, applications, and analyses. *The Abstract.* 1993 (April);13-15.

Roffers SD. Analytical Report: Analysis of AACCR Data Items and Interim Outcome Measures. American Association of Central Cancer Registries Cancer Surveillance and Control Program, April 1992.

Shapiro S, Venet W, Strax P, Venet L, Roeser R. Ten- to fourteen-year effect of screening on breast cancer mortality. *Journal of the National Cancer Institute.* 1982;69(2):349-355.

Smigel K. Breast cancer death rates decline for white women. *Journal of the National Cancer Institute.* 1995;87(3):173.

State of California, Department of Health Services, California Cancer Registry 1988-1992 Public Use Tape.

Surveillance, Epidemiology and End Results (SEER) Program special public use tape (1973-91), National Cancer Institute, DCPC, Surveillance Program, Cancer Statistics Branch, July 1994.

Taber L, Gad A, Holmberg LH, Ljungquist U et al. Reduction in breast cancer mortality by mass screening by mammography; first results of a randomized trial in two Swedish counties. *Lancet*. 1985;1:829-832.

U.S. Department of Health and Human Services, Public Health Service. Healthy People 2000: National Health Promotion and Disease Prevention Objectives (Full Report, With Commentary). DHHS Publication No. (PHS) 91-50212. 1991.

White E, Lee CY, Kristal AR. Evaluation of the increase in breast cancer incidence in relation to mammography use. *Journal of the National Cancer Institute*. 1990;82(19):1546-1552.

Table A. Variables collected* and/or available for analysis** from selected registries.

Variable	MA*	SEER**	CA**	CT**
Hospital name	X			X
Hospital code	X			X
Date of admission			X	
Date of diagnosis	X	X	X	X
Record number	X	X		
Region ID			X	
Region patient number			X	
Region tumor number			X	
SEER registry		X		
Coding procedure		X		
Name	X			
Sex	X	X	X	
Race	X	X	X	X
Spanish name or origin		X	X	
Maiden name	X			
Address	X			X (town code)
County of residence		X	X	
Census tract	X	X		
Zip code	X			X
Birthdate	X	X (year)		
Age	X	X	X	X
Place of birth	X	X	X	
Smoking status	X			X
Marital status		X	X	
Primary site	X	X	X	(breast only)
Histology	X	X (in morph.)	X	X
Morphology		X	X	
Differentiation		X (in morph.)	X	
Stage	X	X	X	X (EOD)
Laterality		X	X	X
Extent of disease (EOD)		X		X
Sequence number	X	X	X	
Confirmation method	X	X	X	
Place of diagnosis	X			
Reporting source		X	X	
Treatment		X		X
Vital status	X	X	X	X
Date of last contact				X
Cause of death		X	X	X
Occupation	X			

Table B. Summary of operations performed on data files, Year 1.

File	Year(s)	Operations
92 MASS INCIDENCE	1992	Roffers prop 1
		Roffers prop 2
		Avg. annual age-specific incidence
		Avg. annual stage-specific incidence
		Roffers prop 4
82-92 MASS INCIDENCE	1982-1992	Frequency distribution for all variables
	1992	Roffers prop 1 *in situ* by age (3 grps)
		by age (18 grps)
	1982-1992	Roffers prop 2 localized by age (3 grps)
	(by single yrs)	by age (18 grps)
		proportion regional by age (3 grps)
		by age (18 grps)
		proportion distant by age (3 grps)
		by age (18 grps)
	1982-1986	Roffers prop 2 by age (3 grps)
		by age (18 grps)
	1987-1992	Roffers prop 2 by age (3 grps)
		by age (18 grps)
	1982-1992	unknown stage/ unknown race by hosp.
	1982-1992	Merged with Allcodes and aggregated byCHNA
	1992	% *in situ* by CHNA
	1982-1992	Annual age-specific inc. rates (18 grps)
82-92 MASS MORTALITY	1982-1986	Age-adjusted mortality rates by CHNA
	1987-1992	Age-adjusted mortality rates by CHNA
93 MASS MORTALITY	1993	Merged with Allcodes and aggregated by CHNA
		Age-adjusted mortality rates by CHNA

Table B. Summary of operations performed on data files, Year 1 (continued).

File	Year(s)	Operations
73-91 (9) SEER AREAS	1973-1991	Cumulative incidence by: gender
		race
		age (18 grps)
		marital status
		stage
		age x stage
		race x stage
	1982-1991	Total SEER sample:
	(1982+ has *in situ*)	Freq by year x stage
		Roffers prop 1 by year
		Roffers prop 2 by year
		Individual SEER areas:
		Freq by year x stage
		Roffers prop 1 by year
		Roffers prop 2 by year
	1982-1986	Age-adjusted incidence by SEER area
	1987-1991	Age-adjusted incidence by SEER area
	1982-1986	Age-adjusted incidence by age (18 grps)
	1987-1991	Age-adjusted incidence by age (18 grps)
	1982-1986	Age-specific incidence by SEER area
	1987-1991	Age-specific incidence by SEER area
	1982-1986	Stage-specific incidence by SEER area
	1987-1991	Stage-specific incidence by SEER area
	1988-1991	Roffers prop 4 (<2cm) by SEER area
	(1988 + has tumor size)	
	1982-1991	Annual age-specific inc. by SEER area
	1982-1991	Total age-specific inc. by SEER area
	1982-1991	Annual age-adjusted inc. by SEER area
	1982-1991	Total age-adjusted inc. by SEER area

Table B. Summary of operations performed on data files, Year 1 (continued).

File	Year(s)	Operations
88-92 CALIFORNIA INCIDENCE	1988-1992	Frequency and % of breast cancer by: stage
		race x stage
		age x stage
		Roffers proportion 1
		Roffers proportion 2
	1988-1992	Freq. and % of breast cancer by: age (by single yrs)
		race
		stage
		race x stage
		age x stage
		Roffers proportion 1
		Roffers proportion 2
	1988-1992	Cum. age-specific inc. by age (18 grps)
	1988-1992 (by single yrs)	Age-specific incidence by age (18 grps)
	1988-1992	Cumulative age-specific incidence by age (3 grps: 0-49, 50-64, 65+)
	1988-1992 (by single yrs)	Age-specific incidence by age (3 grps)
	1988-1992	Age x stage incidence (18 grps)
	1988-1992	Age x invasive stage incidence (3 grps)
	1988-1992 (by single yrs)	Age x invasive stage incidence (3 grps)
	1988-1992	Cum. age-adjusted incidence by year (18 grps) and (3 grps)
	1988-1992 (by single yrs)	Age-adjusted incidence (18 grps) and (3 grps)
	1988-1992	Cum. age-adjusted incidence by stage (18 grps) and (3 grps)
73-92 CONNECTICUT INCIDENCE	1973-1992	Roffers 1 & 2 for each year
90-92 MASS BRFSS	1990-1992	Freq. and % of mammography ques. 2, 3, 4, and 6 by CHNA

Table B. Summary of operations performed on data files, Year 1 (continued).

File	Year(s)	Operations
MASS SES	1981-1989	Population interpolations
	1991-1994	Population interpolations
	1990	Demographic/SES variables by census tract, town, and CHNA Proportion of localized, regional, distant by census tract, town and CHNA
SUMMARY	*graphs:*	*SEER time trend:* *% in situ for each SEER area* *% localized for each SEER area*
	graphs:	*California time trends:* *% in situ by age (3 grps)* *% in situ by race* *% in situ by age & race* *% localized by age* *% localized by race* *% localized by age & race*
	graphs:	*Mass incidence 1982-1992:* *% in situ by age (3 grps) 1992* *% in situ by age (18 grp)1992* *% localized by age (3 grps) & time (2 grps)* *% regional by age (3 grps) & time (2grps)* *% distant by age (3 grps) & time (2 grps)* *trends of invasive cases over time from 1982-1992* *trends: age specific incidence over time from 1982-1992*
		Variables by CHNA: Mass92 % *in situ*, Mass90 per capita $, income < poverty, income >$50,000, education, Mass90-92 BRFSS mammography questions
		Frequency distribution of variables by CHNA (histograms)

Table C. Selected Variable Rankings by CHNA

CHNA	CHNA	%insitu92	CHNA	incidence (90)	CHNA	%mam<1yr (90-91)	CHNA	%mam<1yr (92)
1 PITTSFIELD	9	0.09	19	170.5	12	21	4	16
2 GREENFIELD	1	0.09	8	175.6	5	22	1	18
3 HOLYOKE	2	0.09	9	175.9	**7**	**25**	21	21
4 SPRINGFIELD	11	0.09	1	178.9	6	27	5	24
5 SOUTHBRIDGE	13	0.11	11	179.4	**27**	**27**	16	26
6 MILFORD	24	0.11	21	182.1	25	28	9	28
7 FRAMINGHAM	25	0.11	12	187.0	26	29	10	29
8 WORCESTER	12	0.11	10	188.3	16	30	**7**	**30**
9 FITCHBURG	16	0.11	**22**	**189.1**	19	30	12	30
10 LOWELL	10	0.12	**17**	**190.6**	1	31	6	33
11 LAWRENCE	**22**	**0.12**	24	198.0	3	32	**17**	**33**
12 HAVERHILL	5	0.12	3	200.5	2	33	25	33
13 BEV/GLOUC	8	0.12	2	204.8	9	33	**27**	**33**
14 LYNN/SALEM	21	0.13	**7**	**206.5**	10	33	8	35
15 WOBURN	26	0.13	6	211.4	**20**	**33**	2	36
16 MEDFORD	6	0.14	26	211.9	13	34	**20**	**36**
17 CAMB/SOM	3	0.14	25	212.2	4	36	3	38
18 NEWTON/WALT	**14**	**0.14**	4	217.0	**17**	**36**	11	39
19 BOSTON	**27**	**0.16**	**23**	**220.9**	21	37	**15**	**40**
20 QUINCY	**7**	**0.16**	16	225.2	8	38	**14**	**41**
21 ATTLEBORO	**23**	**0.17**	**14**	**231.8**	**15**	**38**	19	41
22 BROCKTON	19	0.18	**15**	**236.8**	**23**	**40**	24	43
23 PLYMOUTH	**20**	**0.18**	**20**	**240.3**	24	41	26	46
24 TAUNTON	4	0.18	**18**	**242.8**	**14**	**42**	**23**	**47**
25 FALL RIVER	**18**	**0.20**	5	243.0	**18**	**44**	**18**	**49**
26 NEW BEDFORD	**17**	**0.22**	13	257.1	11	48	13	53
27 CAPE/ISLANDS	**15**	**0.23**	**27**	**282.2**	**22**	**53**	**22**	**54**

Table C. Selected Variable Rankings by CHNA

CHNA	%mam<1yr (93)	CHNA	%mam<2yr (92)	CHNA	%ever mam (90-91)	CHNA	%ever mam (93)	CHNA	age-adj mort (87-92)
9	7.4	19	2	5	41	9	35.7	11	34.2
2	37.2	10	3	1	44	24	39.0	**20**	**34.2**
1	38.9	23	3	6	44	**18**	**45.0**	19	32.3
12	45.2	**7**	**4**	**7**	**46**	25	45.4	**17**	**32.2**
10	50.9	11	4	10	46	19	48.0	**22**	**31.6**
26	60.3	**15**	**5**	12	46	3	50.8	9	31.4
27	61.0	20	5	16	46	8	51.7	**18**	**31.4**
7	**61.2**	1	6	9	47	**20**	**55.2**	21	31.2
8	61.4	13	6	26	47	6	55.9	**23**	**31.2**
11	61.4	2	7	21	48	**17**	**55.9**	**14**	**31.2**
18	**62.4**	6	7	19	49	**14**	**56.0**	**15**	**30.8**
21	64.3	21	7	24	49	4	57.4	**7**	**30.5**
19	66.5	24	7	**20**	**50**	12	57.8	16	30.3
4	66.7	26	8	3	51	**7**	**58.2**	1	29.8
5	66.8	**17**	**9**	**17**	**51**	26	59.2	4	29.7
6	67.3	9	10	4	52	2	59.4	**27**	**29.5**
17	**68.5**	12	10	**23**	**52**	5	62.8	26	28.9
15	**70.4**	**14**	**10**	25	53	**15**	**63.2**	3	28.9
20	**71.6**	5	12	**27**	**54**	13	63.8	5	28.6
13	76.9	4	13	8	55	11	63.9	10	28.4
3	79.4	16	13	2	56	**23**	**65.2**	13	28.1
23	**83.4**	25	13	11	56	21	65.6	8	27.8
16	84.0	**27**	**13**	**14**	**56**	**27**	**65.7**	6	27.0
25	85.1	8	14	**15**	**57**	16	67.0	2	26.5
22	**85.7**	**18**	**15**	13	59	**22**	**67.8**	24	25.9
14	**91.9**	3	16	**18**	**66**	1	69.9	12	25.4
24	92.1	**22**		**22**	**67**	10	72.4	25	23.5

Table C. Selected Variable Rankings by CHNA

CHNA	percap$	CHNA	%<povert	CHNA	%inc>50k	CHNA	% owner	CHNA	%<9th gr
25	12,388	19	18.7	25	19.0	19	30.9	25	**26.8**
26	12,956	11	14.4	26	21.1	**17**	**39.5**	26	22.4
3	13,712	4	13.0	2	21.9	16	46.1	11	12.6
2	13,760	3	12.3	19	25.5	25	48.7	24	11.7
5	14,169	26	11.6	3	26.1	11	55.1	19	10.5
4	14,491	25	10.9	**27**	**28.0**	8	55.7	5	10.4
1	14,856	8	10.7	16	28.4	3	58.8	3	9.8
24	14,999	16	10.3	5	30.2	26	59.9	10	9.2
22	**15,005**	2	9.8	8	30.2	**14**	**60.0**	16	8.9
16	15,291	10	9.0	9	31.7	4	61.4	4	8.8
8	15,350	**14**	**9.0**	4	31.8	13	63.2	8	8.8
9	15,462	1	8.7	24	32.6	**18**	**63.2**	9	8.6
19	15,581	**22**	**8.7**	11	33.3	**20**	**63.3**	1	8.5
11	16,117	**17**	**8.6**	**14**	**34.1**	9	63.5	**17**	**7.5**
10	16,183	9	7.4	**17**	**34.8**	**22**	**64.8**	**14**	**7.3**
27	**16,632**	**27**	**7.4**	**22**	**36.3**	1	65.2	6	6.7
6	17,181	5	7.2	12	37.0	10	65.5	12	6.6
12	17,375	12	6.4	1	37.6	5	65.7	**22**	**6.1**
14	**17,469**	13	5.9	13	38.1	12	66.6	2	6.0
21	17,507	24	5.7	**20**	**39.2**	2	66.9	21	5.6
20	**18,473**	**18**	**5.2**	10	39.7	21	68.3	13	4.9
23	**18,989**	**20**	**4.9**	6	44.0	**7**	**68.5**	**18**	**4.2**
13	19,413	21	4.7	21	44.5	24	71.6	**20**	**4.2**
17	**19,513**	**23**	**4.2**	**23**	**45.9**	**27**	**72.0**	**7**	**3.8**
7	**22,518**	6	3.7	**7**	**50.2**	6	74.5	**15**	**3.0**
15	**23,168**	**7**	**3.7**	**18**	**52.3**	**15**	**74.5**	**27**	**3.0**
18	**26,690**	**15**	**3.2**	**15**	**52.9**	**23**	**80.3**	**23**	**2.7**

Table C. Selected Variable Rankings by CHNA

CHNA	%hs1-3yr		CHNA	%coll 4yr
25	18.3		25	12.4
5	16.9		26	13.9
26	16.9		5	16.0
16	15.5		24	16.3
4	15.2		16	17.6
10	15.0		**22**	**17.8**
11	14.5		4	20.0
19	14.2		1	21.0
22	**14.0**		9	21.6
8	13.9		3	22.1
1	13.8		10	22.1
24	13.7		2	22.8
10	13.4		**14**	**24.0**
9	12.7		8	24.3
2	12.6		11	24.9
14	**12.4**		6	25.4
6	11.3		**20**	**25.4**
12	10.8		12	26.0
21	10.6		21	27.0
13	10.0		**27**	**28.6**
20	**9.8**		**23**	**29.5**
17	**9.4**		19	30.4
27	**8.6**		13	31.0
23	**8.5**		**15**	**38.8**
7	7.8		**7**	**39.5**
15	**7.2**		**17**	**44.6**
18	**6.0**		**18**	**51.3**

Figure 4.

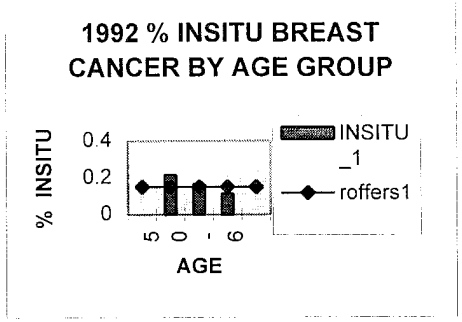


Figure 5a.

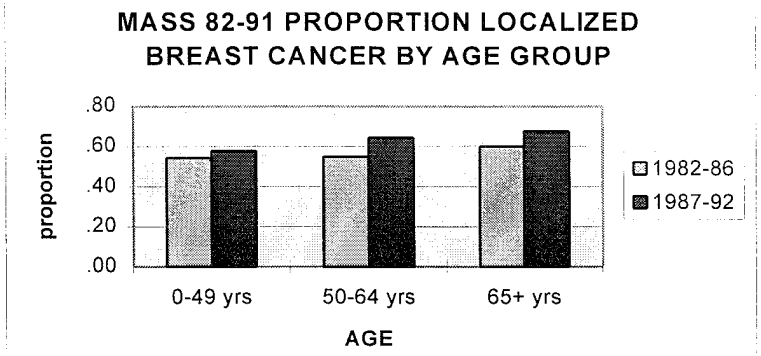


Figure 5b.

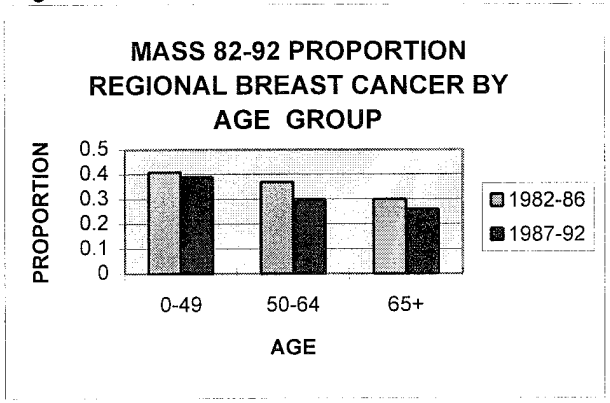


Figure 5c.

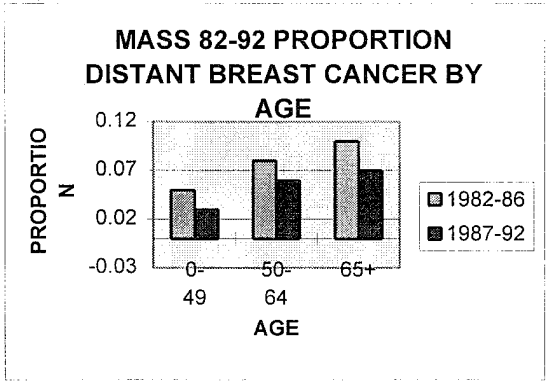


Figure 6. Invasive Breast Cancer Staging Distributions Massachusetts, 1982-92

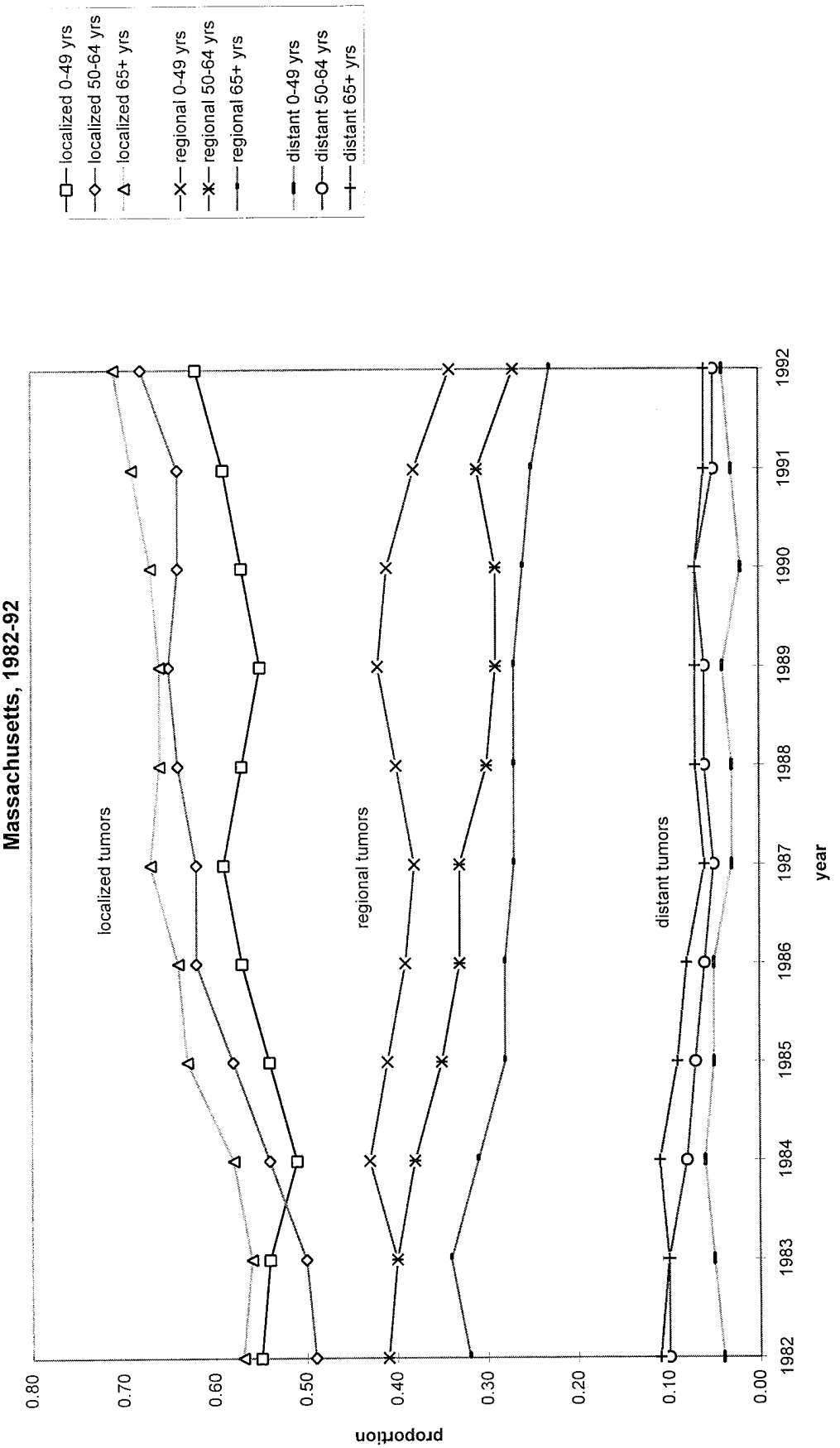


Figure 7. Age-Specific Breast Cancer Incidence Rates Massachusetts, 1982-1992

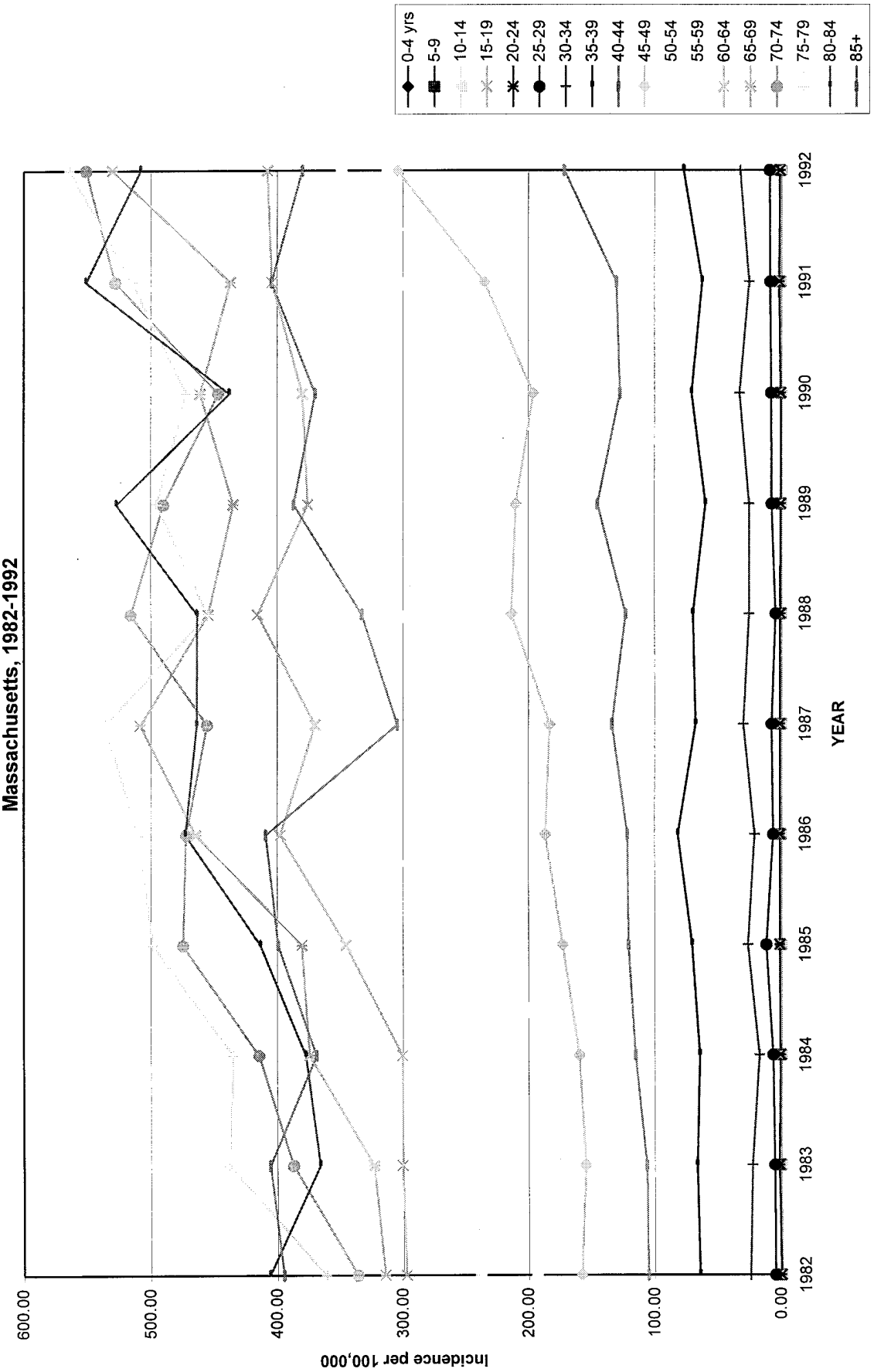


Figure 8. Age-Adjusted Breast Cancer Incidence Rates SEER Registries, 1982-91

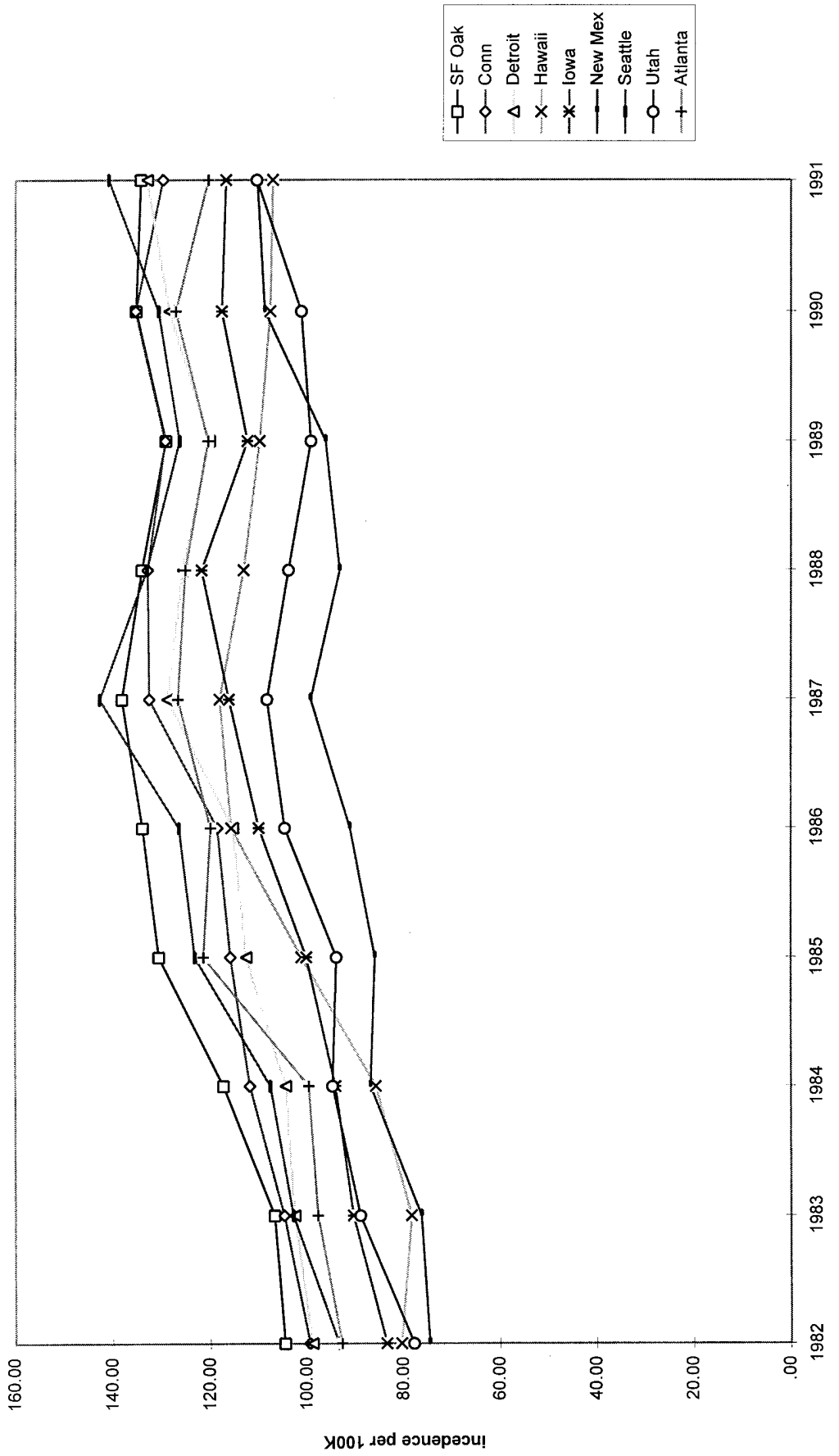


Figure 9.

Percent Black in each census tract

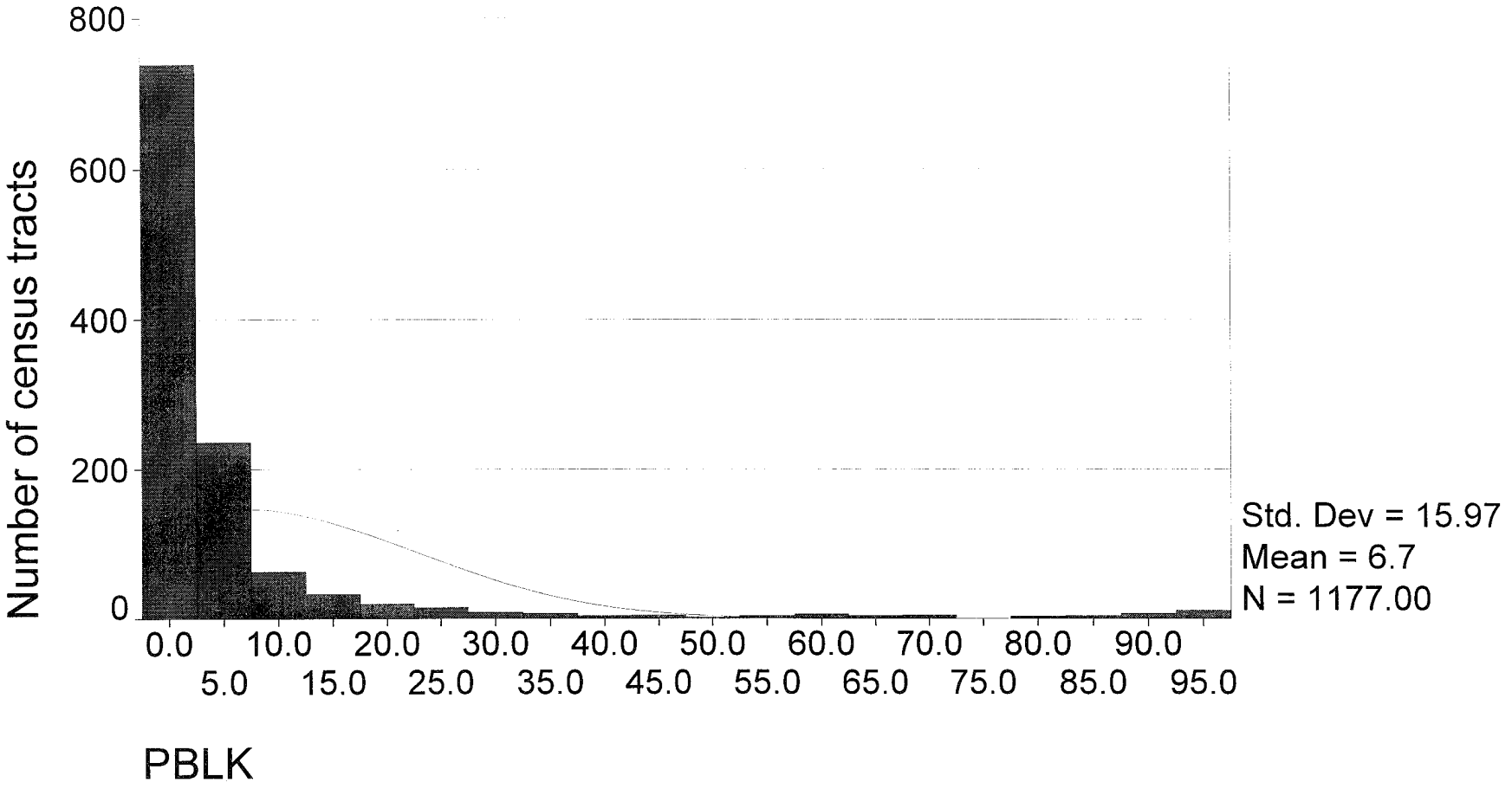


Figure 10. Percent with annual income below poverty level per census tract

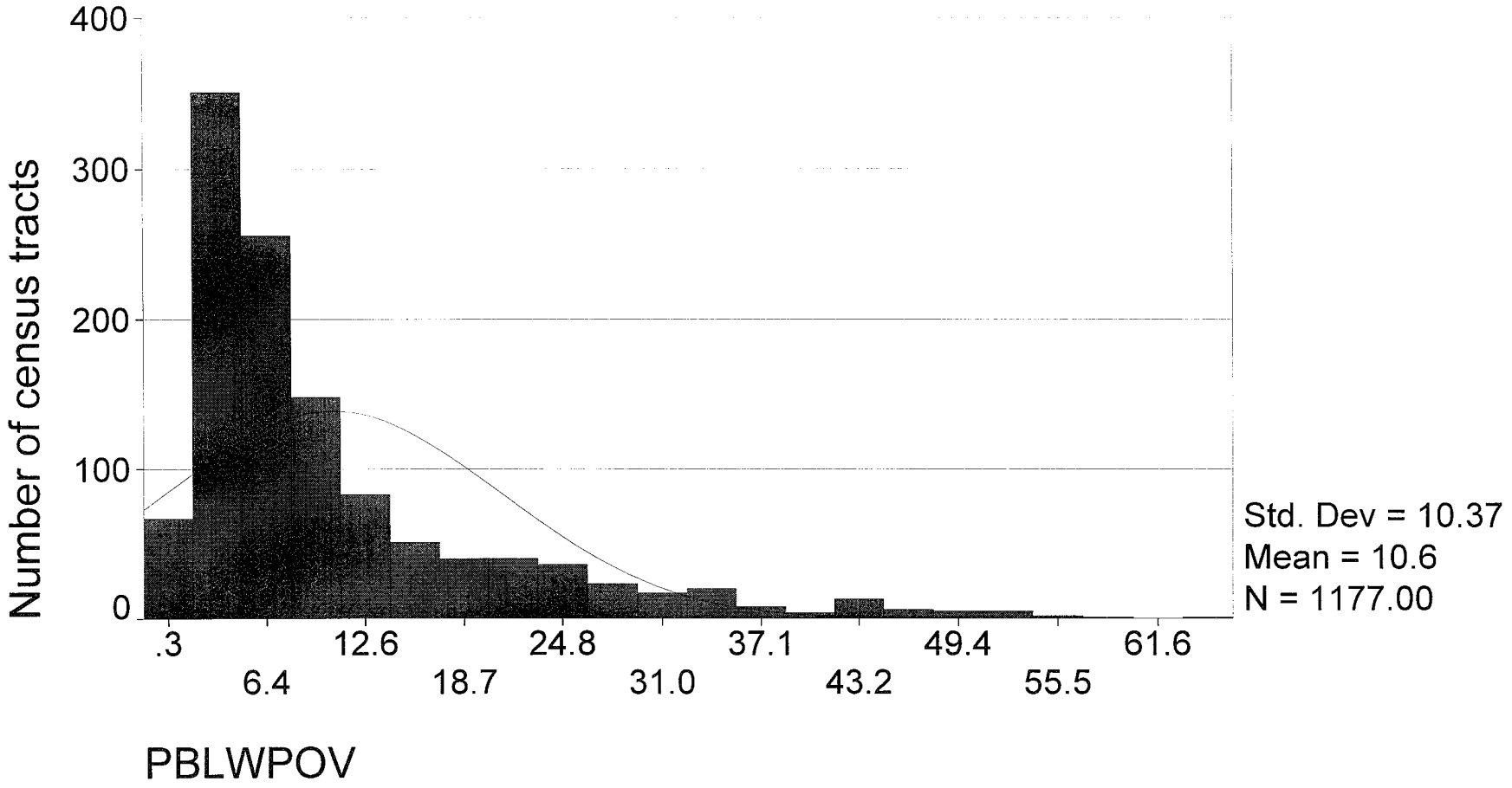


Figure 11. Percent with fewer than 9 grades of school per tract

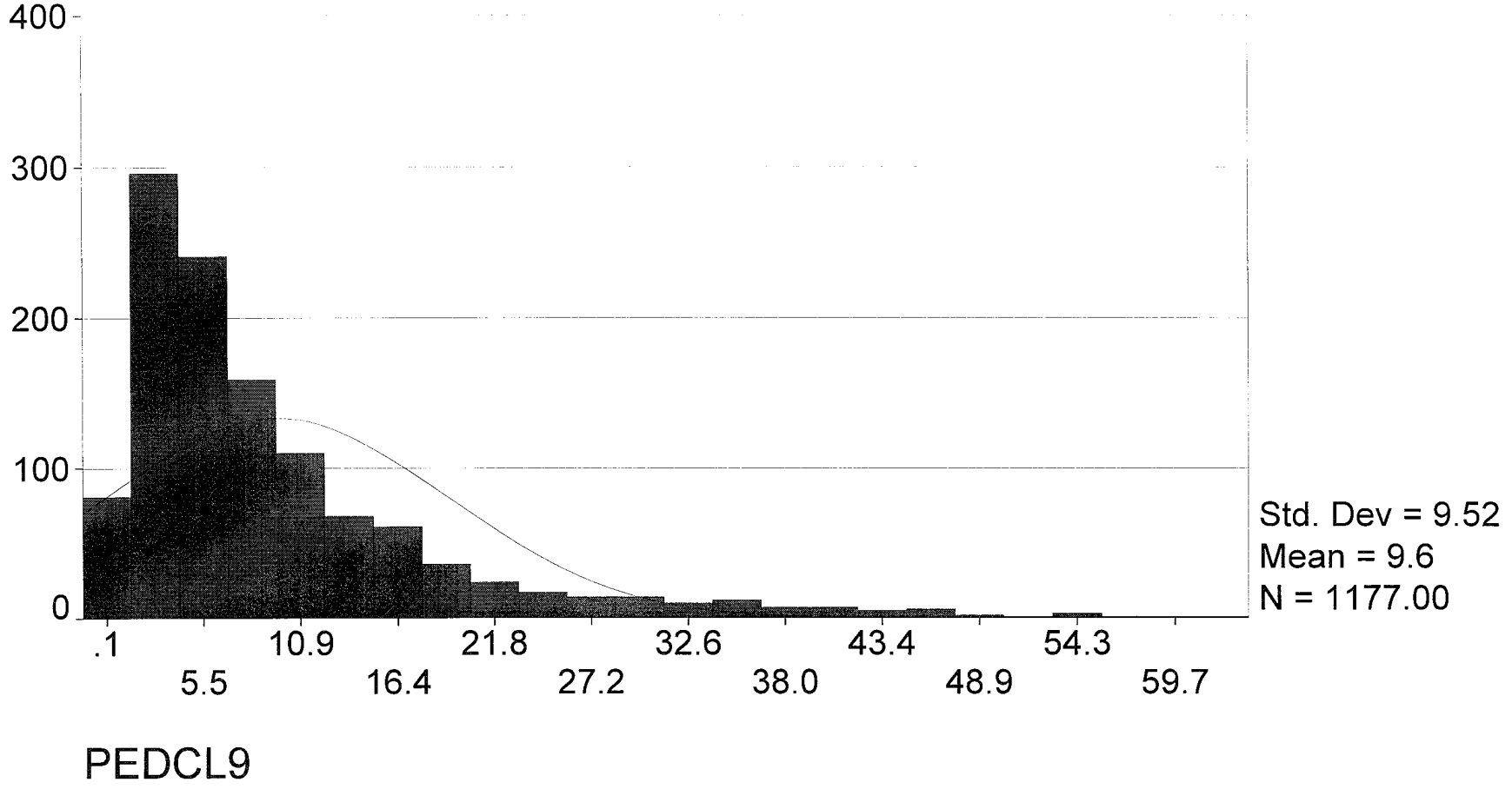


Figure 12
Three-Factor Model of SES

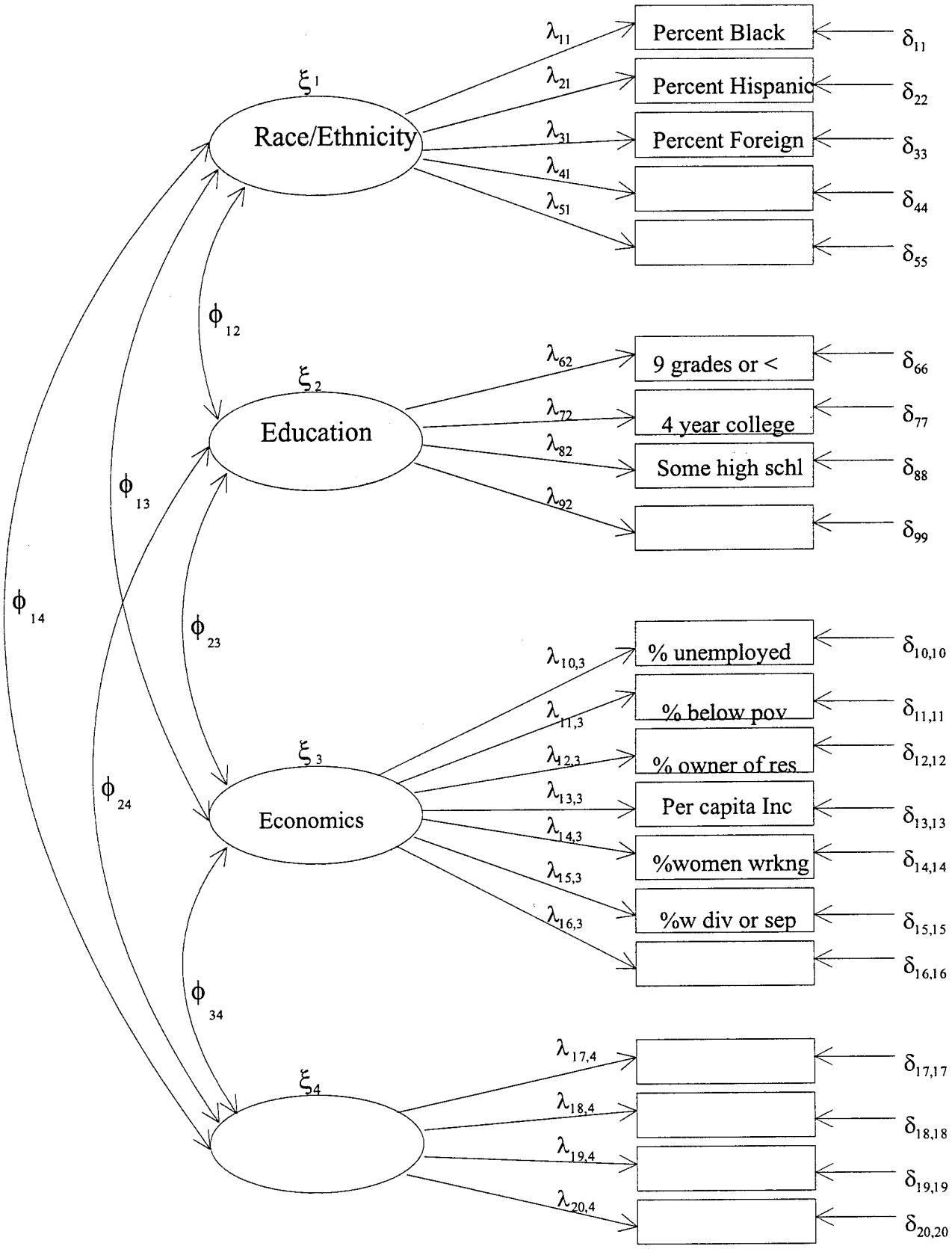


Figure 13
Second Order Factor Model

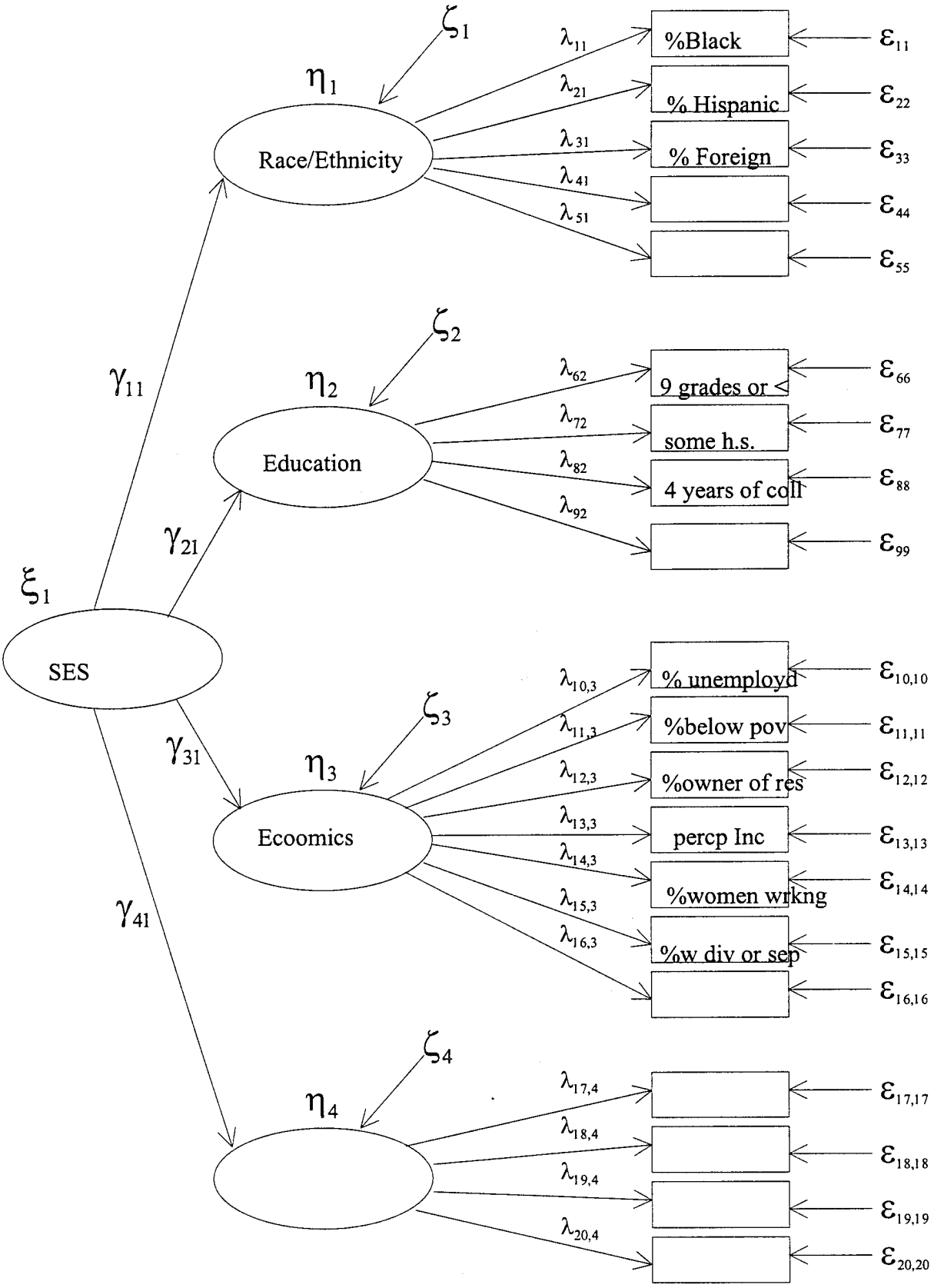
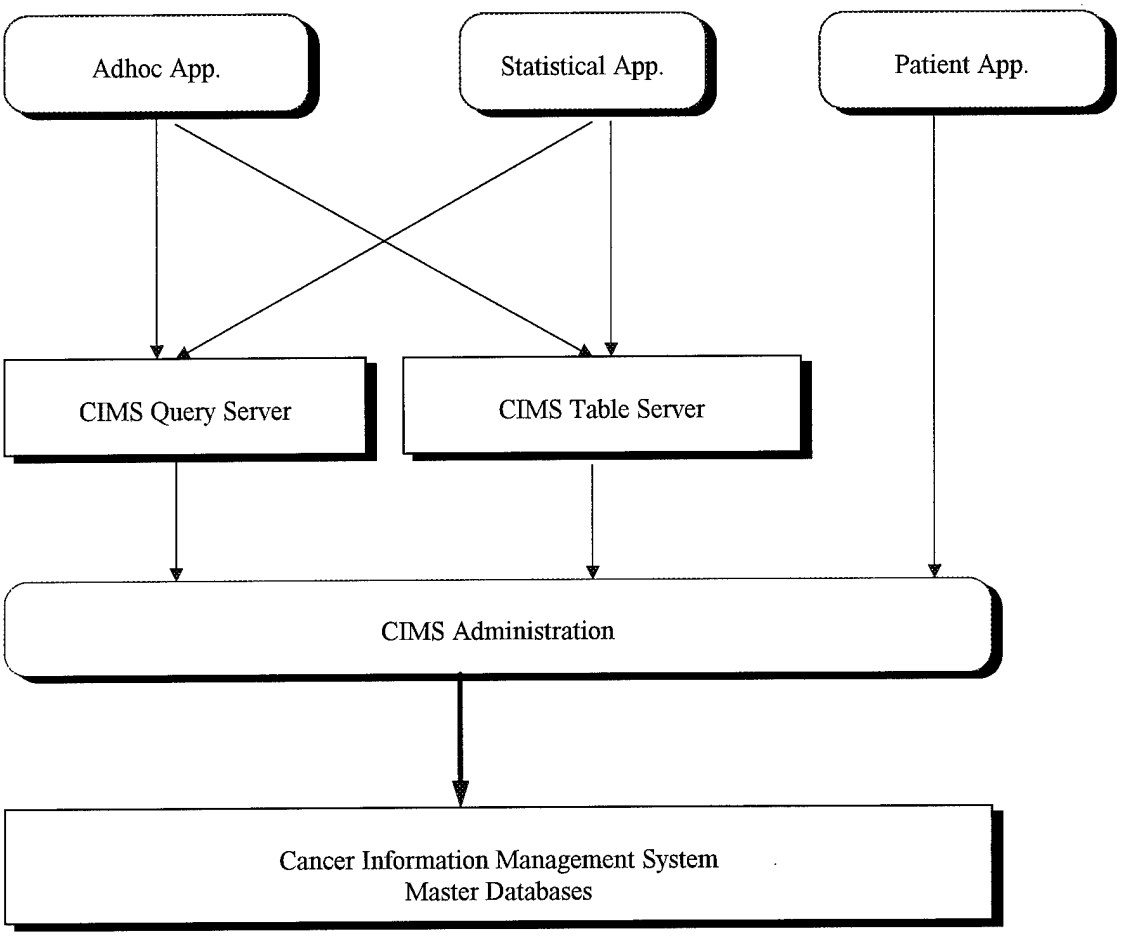


Figure 14.

Overview of Cancer Information Management System



* This is not a complete overview of the CIMS

Figure 15.

Cancer Case Data Flow

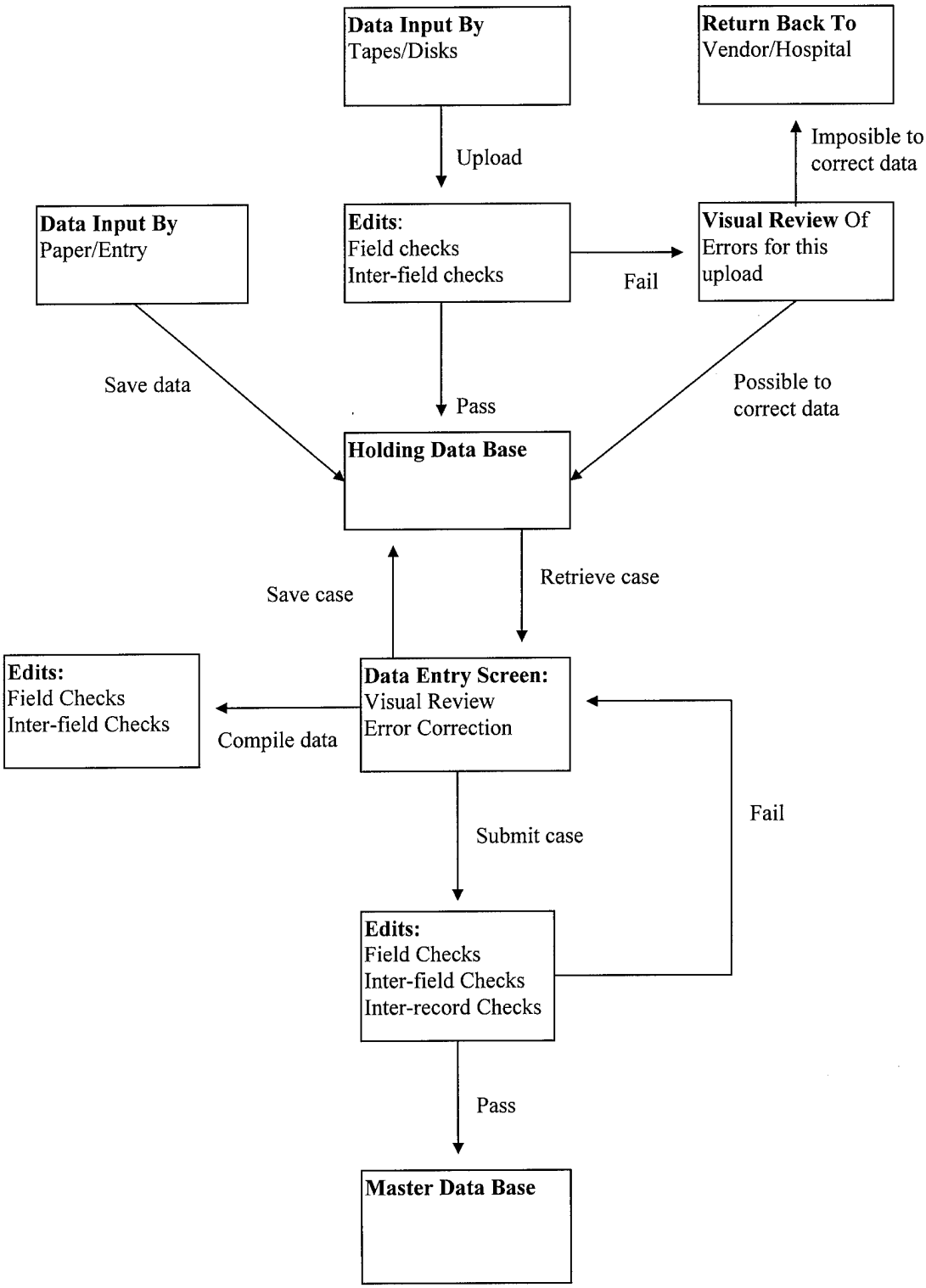


Figure 16.

Cancer Information Management System's Master Databases Structure