

Overview of the SEER-Medicare Data: Content, Research Applications, and Generalizability to the United States Elderly Population

Author(s): Joan L. Warren, Carrie N. Klabunde, Deborah Schrag, Peter B. Bach and Gerald F. Riley

Source: *Medical Care*, Aug., 2002, Vol. 40, No. 8, Supplement: Use of the SEER-Medicare Data for Cancer-Related Health Services Research (Aug., 2002), pp. IV3-IV18

Published by: Lippincott Williams & Wilkins

Stable URL: <https://www.jstor.org/stable/3767919>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



Lippincott Williams & Wilkins is collaborating with JSTOR to digitize, preserve and extend access to *Medical Care*

JSTOR

Overview of the SEER-Medicare Data

Content, Research Applications, and Generalizability to the United States Elderly Population

JOAN L. WARREN, PhD,* CARRIE N. KLABUNDE, PhD,* DEBORAH SCHRAG, PhD,†
PETER B. BACH, MD,† AND GERALD F. RILEY, MSPH‡

BACKGROUND. The Surveillance, Epidemiology and End Results (SEER)-Medicare-linked database combines clinical information from population-based cancer registries with claims information from the Medicare program. The use of this database to study cancer screening, treatment, outcomes, and costs has grown in recent years.

RESEARCH DESIGN. This paper provides an overview of the SEER-Medicare files for investigators interested in using these data for epidemiologic and health services research. The overview includes a description of the linkage of SEER and Medicare data and the files included as part of SEER-Medicare. The paper also describes the types of research projects that have been undertaken using the SEER-Medicare data. The overview concludes with a comparison of selected characteristics of elderly persons residing in the SEER areas to the US total aged.

RESULTS. The paper identifies a number of

potential uses of the SEER-Medicare data. The comparison of the elderly population in SEER areas to the US total shows that in the SEER areas there are a lower percentage of white persons and individuals living in poverty, and a higher percentage of urban-dwellers than the US total. Elderly persons in the SEER regions also have higher rates of HMO enrollment and lower rates of cancer mortality.

CONCLUSIONS. The SEER-Medicare data are a unique resource that can be used for a variety of health services research projects. Although there are some differences between the elderly residing in the SEER areas and the US total, the SEER-Medicare data offer a large population-based cohort that can be used to longitudinally track care for persons over the course of cancer diagnosis, treatment, and follow-up.

Key words: SEER Program; Medicare; cancer; health services research; epidemiology. (Med Care 2002;40[supplement]:IV-3-IV-18)

The linked Surveillance, Epidemiology and End-Results (SEER)-Medicare data are a large population-based source of information for cancer-related epidemiologic and health services research. The SEER-Medicare data have been identified by the Institute of Medicine as one of

the few population-based data resources available for analyses of the quality of cancer care.¹ The SEER-Medicare data also offer the potential to address other issues of social currency, including disparities in health care for selected subgroups and cost of cancer treatment. The use of these data

*From the Applied Research Program, National Cancer Institute, Bethesda, Maryland.

†From the Health Outcomes Research Group, Department of Epidemiology and Biostatistics Memorial Sloan-Kettering Cancer Center, New York, New York.

‡From the Centers for Medicare and Medicaid Services, Baltimore, Maryland.

Address correspondence and reprint requests to: Joan L. Warren, PhD, Applied Research Program, DCCPS, National Cancer Institute, Executive Plaza North, Room 4005, 6130 Executive Boulevard, MSC 7344, Bethesda, MD 20892-7344. E-Mail: joan_warren@nih.gov

to study cancer screening, treatment, outcomes, and costs has grown in recent years.

In this article, we provide researchers with an overview of the SEER-Medicare data. The overview includes a description of the SEER and the Medicare programs, followed by a discussion of the linkage of the SEER and Medicare data and the files included as part of these data. We then focus on the uses of the SEER-Medicare data for health services research. Finally, we address the question of how representative the SEER-Medicare data are of the US elderly population.

Materials and Methods

Data Sources, Linkage, and Files

Data Sources. SEER Program. The SEER program is an epidemiologic surveillance system sponsored by the National Cancer Institute² consisting of population-based tumor registries that routinely collect information on all newly diagnosed cancer (incident) cases that occur in persons residing in SEER areas. The information collected about each incident cancer diagnosis includes the patient's demographic characteristics, date of diagnosis, data about the cancer (eg, histology, stage, and grade), type of surgical treatment and radiation therapy recommended or provided within 4 months of diagnosis, follow-up of vital status, and cause of death, if applicable. The mortality data reported by SEER are derived primarily from data collected by the National Center for Health Statistics. The SEER data do not capture information about use of screening or about how the cancer was detected. In addition, there is no information about comorbidities, treatment provided more than 4 months after diagnosis, or about long-term disease status.

Data collection for the SEER program began in January 1, 1973, and the number of registries included in the SEER program has expanded over time. At the time of the initial linkage of the SEER-Medicare data in 1991, the SEER areas included the states of Connecticut, Hawaii, Iowa, New Mexico, and Utah, as well as the metropolitan areas of Detroit, San Francisco-Oakland, Atlanta, and Seattle-Puget Sound. These areas represented approximately 10% of the US population. In 1992, Los Angeles County and the San Jose-Monterey areas joined the SEER program. Inclusion of these areas expanded the SEER represen-

tation to approximately 14% of the US population. In 2000, the SEER program began adding four states—Kentucky, Louisiana, New Jersey, and the remainder of California. With the expansion of the SEER program, the SEER areas will capture approximately 25% of the US population.

The SEER data are considered highly valid. All of the SEER registries hold the highest level of certification of data quality as provided by the North American Association of Central Cancer Registries.³ Every year, studies are conducted in the SEER areas to evaluate the quality and completeness of the data being reported. The SEER program's standard for the completeness of case ascertainment is 98%. In addition, each year the SEER registries reabstract medical records for a sample of cases to evaluate the accuracy of each of the data elements collected from the records.

Medicare Data. Medicare is the primary health insurer for 97% of the US population 65 years and older. All Medicare beneficiaries receive Part A benefits, which cover inpatient care in short- and long-stay hospitals, skilled nursing facilities, home health, and hospice care. Ninety-five percent of beneficiaries also subscribe to Part B of Medicare to obtain benefits that cover physician services, outpatient care, durable medical equipment, and home health in some cases.⁴ Information about each beneficiary's enrollment and entitlement, demographics, and HMO membership is maintained by Medicare in a master enrollment file known as the Enrollment Database (EDB).

Data Linkage. The linkage of the SEER data with the Medicare data entails matching persons included in the SEER registries with the EDB. The linkage is based on an algorithm involving a match of social security number, name, sex, and date of birth that has been described elsewhere.⁵

The linkage of the SEER and Medicare data is the result of the collaborative effort of the National Cancer Institute (NCI), the SEER registries, and the Centers for Medicare and Medicaid Services (CMS). The linkage was initially completed in 1991 and has been updated twice, in 1995 and 1999. The first linkage attempted to match all persons in the SEER database, including cases diagnosed as far back as 1973. For each of the linkages, among persons in the SEER data who were 65 years or older at the time of the linkage, 93% were found in the Medicare enrollment file. Table 1 shows the number of cancer cases in SEER-Medicare database by type of cancer and year of diagnosis following the 1999 linkage.

TABLE 1. Number of Cancer Cases Appearing in the SEER-Medicare Data, by Type of Cancer and Year of Diagnosis

Year of Diagnosis	Total	Oral Cavity/ Pharynx	Colorectal	Pancreas	Lung	Breast	Uterus	Ovary	Prostate	Kidney/ Renal	Lymphoma (NHL)	Bladder
1996	77,152	2,190	12,908	2,536	14,605	13,383	2,668	1,524	16,591	2,122	3,795	4,830
1995	80,307	2,269	13,154	2,557	15,103	14,176	2,685	1,603	17,808	2,183	3,779	4,990
1994	82,544	2,350	13,566	2,587	14,837	14,044	2,791	1,615	19,650	2,257	3,797	5,050
1993	86,733	2,378	13,833	2,521	14,834	14,241	2,894	1,704	23,385	2,151	3,642	5,150
1992	91,065	2,441	14,064	2,484	14,852	14,924	3,002	1,649	26,724	2,204	3,575	5,146
1991	87,340	2,386	14,260	2,458	14,457	14,789	2,999	1,698	23,541	2,174	3,468	5,110
1990	81,800	2,377	14,471	2,407	14,131	14,748	3,158	1,708	18,124	2,129	3,524	5,023
1989	77,290	2,339	14,678	2,333	13,624	14,285	3,036	1,674	15,171	2,003	3,197	4,950
1988	76,864	2,379	14,583	2,378	13,423	15,210	3,065	1,658	14,009	1,991	3,187	4,981
1987	56,376	1,816	10,989	1,690	9,733	11,241	2,279	1,156	9,987	1,492	2,275	3,718
1986	52,696	1,781	10,890	1,638	9,079	10,199	2,264	998	8,702	1,418	2,162	3,565
1985	42,063	1,562	9,247	571	5,060	9,227	2,193	869	7,528	1,027	1,672	3,107
1984	32,919	1,298	7,235	134	2,596	7,968	2,064	618	6,195	821	1,314	2,676
1983	27,755	1,098	5,917	82	1,793	6,970	2,007	474	5,340	686	1,076	2,312
1982	23,682	964	5,063	47	1,415	6,048	1,951	381	4,367	526	877	2,043
1981	21,587	905	4,628	41	1,137	5,638	1,851	362	3,744	499	804	1,978
1980	18,843	831	3,978	33	914	5,060	1,819	333	3,059	475	647	1,694
1979	16,997	794	3,459	26	828	4,681	1,904	285	2,561	377	540	1,542
1978	15,471	696	3,183	16	728	4,257	1,961	278	2,096	399	435	1,422
1977	14,048	581	2,816	13	607	4,016	2,100	290	1,814	295	383	1,133
1976	13,631	594	2,607	12	525	4,017	2,361	270	1,455	344	373	1,073
1975	12,580	528	2,199	19	431	4,038	2,429	264	1,114	285	309	964
1974	10,716	393	1,903	14	365	3,618	2,120	275	751	226	233	818
1973	7,745	331	1,467	11	229	2,473	1,517	200	522	182	187	626
Total	1,108,204	35,281	201,098	26,608	165,306	219,251	57,118	21,886	234,238	28,266	45,251	73,901

There are persons younger than 65 years in the SEER-Medicare data. The under-65 group consists of people who are eligible for Medicare because they are disabled or have end-stage renal disease. Because of the unusual characteristics of the under-65 group, most investigators using the SEER-Medicare data have focused their analyses on individuals 65 years and older.

In performing the linkage, NCI uses cases from the most recent SEER data available. However, there is a lag of approximately 2 years in the reporting of cases to the SEER program. The 1999 linkage included cases diagnosed through 1996, with Medicare claims through 1999. Linkages are updated every 3 years. The next linkage will be completed in 2002, with cases through 1999 and their Medicare claims through 2001.

Data Files Included in the SEER-Medicare Data. The SEER-Medicare data are stored in a number of separate files (Table 2) because of the

large number of cases and the amount of claims for each person in the file. The SEER data included as part of the SEER-Medicare files are in a customized file known as the Patient Entitlement and Diagnosis Summary File (PEDSF). This file contains one record per person for individuals in the SEER data who have been matched with Medicare enrollment records. For persons appearing in the PEDSF file, the SEER demographic variables are available as well as clinical information for up to 10 diagnosed cancer cases. Added to the SEER data is information pertaining to each person's Medicare eligibility, Medicare demographic variables, reason for Medicare entitlement, and HMO enrollment. In addition, the PEDSF file includes information about the median household economic and educational status for the census tract or zip code where the person resides. A more detailed discussion about the demographic and socioeconomic variables included in the PEDSF file is available in this issue of *Medical Care*.⁶

The Medicare data available as part of SEER-Medicare include claims from hospital, outpatient, physician, home health, and hospice providers (Table 2). Each Medicare file varies in the data elements included and the types of procedure and diagnostic codes used, either International Classification of Diseases (ICD-9) codes⁷ for procedures and diagnoses or HCFA Common Procedure Coding System (HCPCS) codes for procedures. HCPCS are the AMA's Common Procedure Terminology codes⁸ (CPT-4) with additional codes used exclusively by CMS. In general, all Medicare files have fields for race, sex, and date of birth or age, the date(s) of service, diagnostic codes (for many files), and procedure codes in addition to the amounts for charges and reimbursement. In addition, every Medicare file contains a provider identification number for the hospital or physician. Medicare files included as part of the SEER-Medicare data contain the SEER case number on each claim, which is the unique nonidentifiable number assigned to each cancer patient by the registries. The SEER case number allows an investigator to link information for an individual across multiple SEER-Medicare files. Cases included in the SEER-Medicare data may have been diagnosed as early as 1973, although Medicare claims are only available back to 1991, with the exception of inpatient hospital data that are available back to 1986.

Data for Control Groups. The linked SEER-Medicare data include all persons in the SEER files who are Medicare eligible. For some analyses, it is helpful to have a control or comparison group of Medicare beneficiaries without cancer that can be used for comparative purposes or to provide population-based estimates of testing, treatment and costs. As a complement to the SEER-Medicare data, there are Medicare files for persons residing in the SEER areas who do not have cancer. This group is a random sample of 5% of Medicare beneficiaries residing in the SEER areas, less any beneficiaries in the 5% sample in whom cancer has been diagnosed and who appear in the PEDSF file. The Medicare files available for the control group are identical to those for the cancer cases, with the exception that files can be linked by each beneficiary's unique Medicare claim number (encrypted) in lieu of a SEER case number. By definition, there are no PEDSF data for the control group, although Medicare demographic, entitlement, and HMO enrollment data for noncancer cases similar to what is on the PEDSF file are

maintained in the Summarized Denominator (SUMDENOM) file.

Uses of the SEER-Medicare Data

Cancer control covers the spectrum from primary prevention to end-of-life care. Given the complementary nature of the SEER and Medicare data, merging these two data sources has resulted in a unique resource for numerous types of studies related to cancer control (Table 3). The SEER data provide the identification of incident cases (which is not possible from the Medicare claims) with detailed site and stage reporting and information about the cause of death. The Medicare data offer a longitudinal perspective, making it possible to look at services before, during, and after diagnosis. Claims before diagnosis can be used to measure preexisting comorbidities⁹ that might influence treatment decisions and cancer screening, to a limited extent.¹⁰ Initial treatment with surgery or radiation therapy may be found in the SEER or Medicare files.^{11,12} However, SEER only reports the most invasive surgery, while the Medicare claims can be used to capture all procedures during the peridiagnostic period. In addition, the Medicare claims can be used to measure the use of adjuvant chemotherapy, a treatment not reported by the SEER program.¹³ After initial care, the Medicare data can be used to evaluate surveillance procedures¹⁴ and other "downstream services," as well as some long-term outcomes (rehospitalizations and complications to a limited extent).¹⁵ The Medicare data also include a limited amount of information about the hospital or physician.^{16,17} Each of the claims contains the unique provider number for either the physician or hospital providing the service. This allows one to aggregate the number of procedures per provider for volume outcome studies. As the payer for services, the Medicare program also collects information about the amount charged and the amount reimbursed.¹⁸

Many published SEER-Medicare studies have focused on a specific aspect of cancer care, such as initial surgery or adjuvant therapy. More recent projects have begun to evaluate longer-term outcomes. Investigators are now using these data to measure recurrences¹⁹ and terminal care. Findings from studies using SEER-Medicare data can address issues such as the quality of care, disparities in health care, and access to care. Other studies can provide estimates of the costs of cancer care,

TABLE 2. Overview of the Files Maintained as Part of the SEER-Medicare Data*

File Name	Summary of Contents	Years Included	Diagnosis/procedure Coding	Provider Information
Patient entitlement and diagnosis summary file (PEDSF)	<ul style="list-style-type: none"> SEER data Medicare HMO and entitlement/person by month and year Census tract/zip code SES measures 	<p>Cancer cases diagnosed from 1973-1996</p> <p>Medicare status 1986-1999</p> <p>1990 Census data</p>	<p>Cancer directed surgery as collected by SEER</p> <p>Use of radiation therapy as collected by SEER</p>	None
Summarized denominator file for non-cancer cases (SUMDENOM)	<ul style="list-style-type: none"> 5% random sample of Medicare beneficiaries residing in SEER areas, less persons with cancer Medicare HMO and entitlement/person by month and year 	Medicare status 1986-1999	None	None
Medicare analysis and procedure file (MEDPAR)	<ul style="list-style-type: none"> 100% of Medicare hospitalizations occurring in a calendar year one record per hospitalization Part A service 	1986-1999	ICD-9 diagnosis and procedure codes (up to 10 each) <p>Diagnosis Related Groups codes (DRGs)</p>	Hospital identification number <p>Type of facility (short stay, long stay, SNF)</p>
Hospital outpatient file	<ul style="list-style-type: none"> 100% of Medicare outpatient claims multiple procedures for the same date of service (ex clinic visit + lab service) Part B service 	1991-1999	Procedures identified from HCPCS [†] and revenue centers	Hospital identification number
Physician/supplier data	<ul style="list-style-type: none"> 100% of bills from physicians and other providers bills are usually for a single date of service service can occur in office, hospital, etc. Part B service 	1991-1999	Procedures identified from HCPCS <p>Each procedure is accompanied by an ICD-9 diagnosis that is the reason for the service</p>	Type of facility (eg. clinic, hospital) Type of service (eg. consultation, radiology) <p>Unique physician identification number (UPIIN)</p> <p>Place of service (eg. office, hospital, etc)</p> <p>Provider specialty code.</p>
Hospice/home health files	<ul style="list-style-type: none"> 100% of claims bills often cover a span of days Part A service[‡] 	1991-1999	ICD-9 diagnosis codes (up to 10)	Provider identification number

SNF = skilled nursing facility.

*File status as of January, 2002.

[†]HCPCS = Health Care Common Procedure Classification System.

[‡]Some home health services are covered under Medicare Part B.

TABLE 3. Spectrum of SEER-Medicare Analyses Across the Course of Cancer Control Activities

Primary Prevention →	Screening/ Detection →	Diagnosis/ Treatment →	Survivorship →	Second Occurrence →	Terminal Care/Death
N/A	<ul style="list-style-type: none"> • Use of screening: mammography, PSA, sigmoidoscopy/colonoscopy • Impact of changing technology; practice patterns on cancer detection/incidence 	<ul style="list-style-type: none"> • Patterns of care • Peri-operative complications • Volume outcomes studies • Extent of staging • Factors that influence treatment-comorbidities/health disparities/health system factors 	<ul style="list-style-type: none"> • Late effects of treatment • Post-diagnostic surveillance • Treatment of prevalent cancers 	<ul style="list-style-type: none"> • Rates of recurrence/second primaries • Relationship of second events to initial treatment and ongoing surveillance 	<ul style="list-style-type: none"> • Use of hospice • End-of-life patterns of care • Survival

Studies of the costs of cancer treatment span the continuum.

within a specific phase of care or spanning the course of cancer screening and treatment. In addition to the studies related to cancer care, other analyses have focused on methodologic issues related to the use of SEER-Medicare data as a resource for cancer surveillance.

The availability of Medicare data for a control group enables studies that would not be possible if there were only data for cancer cases. For example, studies of the use of screening, on a population basis, can be done by combining screenings identified from the data for the cancer cases with the screenings identified from the data for controls. The controls can also be used to help determine costs of cancer treatment. Estimates of treatment costs are obtained by matching cancer cases with controls and then subtracting the routine costs of care, as identified from the controls.

As of January 2002, there have been more than 50 peer-reviewed publications of analyses using the SEER-Medicare data, with more under review. A summary of the types of studies conducted to date, as they relate to the continuum of cancer care, is provided in the Appendix.

Data Confidentiality

SEER-Medicare data are a valuable resource available to epidemiologists and health services researchers for research purposes. However, there is a need to provide safeguards for the confidentiality of patients and providers in these data. To ensure confidentiality, the SEER-Medicare data have been deidentified (ie, all personal identifiers have been removed from the file). In addition, variables that might allow for the reidentification, such as exact date of birth and census tract, have been removed or transposed. Despite these steps, there is a remote potential of reidentification. As a result, the SEER-Medicare data are not available as public-use files. Investigators who wish to obtain SEER-Medicare data must sign legally binding data-use agreements with CMS and SEER. In addition, CMS, NCI, and a representative of the SEER registries review all projects before data release to insure there are no concerns about confidentiality. More information about the SEER-Medicare data, including how to obtain the data, can be obtained from the SEER-Medicare WEB site at <http://healthservices.cancer.gov/seermedicare.html>.

Data Limitations

Services Not Captured in the Medicare Claims. Although the SEER-Medicare data contain information about health care for a large population-based sample of cases, there are limitations to using these data. One of the more important issues is that there may be services not captured in the Medicare claims. There will be no claim for services not covered by Medicare, so these services cannot be found in the Medicare data. Noncovered services include routine physical examinations, oral prescription drugs (including drugs such as Tamoxifen), long-term care, and until recently many types of cancer screening. Similarly, there are no Medicare claims in cases where a beneficiary receives services covered by Medicare but not billed to Medicare. This includes examples such as mammography provided by a community-funded screening program, services provided to a Medicare beneficiary by a Veterans' Administration facility, or care given to a beneficiary who is still working and is covered by an employer's health plan (in which case Medicare is the secondary payor).

Lack of Data for HMO Enrollees. The SEER-Medicare data contain claims only for fee-for-service (FFS) care. Historically, HMOs have not been required by CMS to submit claims or encounter data for specific services received by their Medicare enrollees. The lack of claims data for HMO enrollees is a significant limitation of the Medicare database. As of December 2001, HMO enrollees comprised approximately 14% of the Medicare population nationwide, with significant variation by geographic area. For example, several states have little or no Medicare HMO enrollees (including the SEER states of Iowa and Utah), whereas the highest HMO penetration for a state, 38%, was in California, which is also a SEER registry. A comparison of HMO enrollment among Medicare beneficiaries in the SEER areas compared with the US total is presented later in this paper.

Measuring HMO penetration is complicated by the fact that until recently, Medicare beneficiaries have been free to disenroll at any time from HMOs, and many plans accept new enrollees throughout the year. The freedom of movement between the HMO and FFS sectors serves to increase the number of beneficiaries for whom claims data will be missing during at least part of their period of Medicare entitlement. Of additional

concern is the fact that many HMOs have withdrawn from the Medicare program since 1998 and others have reduced their service areas. These withdrawals have caused significant fluctuations in HMO-penetration rates in some geographic areas over time.

While the lack of HMO data represents a gap in the SEER-Medicare database, the penetration of HMOs in the elderly population is much less than among younger people. Given the large number of cases in the SEER-Medicare data, there are a sufficient number of people in the SEER-Medicare data to allow for the loss of people in HMOs. However, HMO enrollees tend to be younger and healthier than beneficiaries in FFS, resulting in a biased loss of information in the claims data.

Some data about service provided by HMOs to Medicare beneficiaries may be available in the future. Beginning in July 1997, HMOs are required to submit encounter records to CMS for all inpatient hospital stays. Encounter data from HMOs for outpatient and physician services are not presently required. The quality of these hospital data is currently being evaluated by CMS, and the availability of these records for researchers is still being discussed. Data from HMOs will be included as part of the SEER-Medicare database if they become available. Finally, data from the PEDSF file for HMO enrollees permits comparison of initial treatment, as reported in the SEER data, and an evaluation of the extent to which these patients resemble their peers enrolled in the FFS plan in terms of basic tumor information, demographics, and other socioeconomic variables.

Other Limitations. Although the SEER-Medicare data permit a comprehensive picture for studying those malignancies that predominate among elderly patients (eg, lung, colorectal, breast, and prostate cancer), they yield more limited insight into cancers more common in younger populations (eg, testicular cancer, leukemia, and lymphoma). Other limitations about the data relate to the use of claims for analyses. The diagnoses and procedures found on insurance claims have agreement with the medical record, although the sensitivity of the data varies by condition.^{19,20} Medicare claims are created for payment purposes, not research. Therefore, information about the reason for a test or procedure and the outcome of a test is not known. Additionally, for hospitalizations, it is difficult to distinguish if secondary diagnoses are complications or comorbidities.²¹

Generalizability of SEER-Medicare Data to the US Elderly Population

The SEER registries have not been randomly selected; rather, they have been chosen for the quality of the registries and to facilitate adequate representation of minority populations. Therefore, to understand the generalizability of the results of SEER-Medicare analyses, it is important to evaluate how well the elderly population included in the SEER areas represents the elderly population of the entire United States. To assess how well the elderly persons in the SEER areas reflect the US elderly population, we compared sociodemographic characteristics and HMO enrollment between persons 65 years and older residing in the SEER areas and those from the United States as a whole. In addition, it is important to understand whether the cancer incidence among the elderly in the SEER areas is representative of national incidence among the elderly. However, because there is no national reporting system for cancer incidence, we examined cancer mortality rates as a proxy for incidence. We compared cancer mortality rates among the SEER elderly with cancer mortality rates for the US elderly population. Our assessment of the generalizability of the SEER elderly to the total US elderly did not include an evaluation of health care resources (physician and hospitals) between SEER and non-SEER areas, as this has been done in an earlier analysis.²²

Sociodemographic Characteristics. The comparison of sociodemographic characteristics for persons 65 years and older in the SEER population with the US population included age groups, race, sex, as well as estimates of the portion of persons residing in an urban location and the percentage living below the federal poverty level (Table 4). We used data from the EDB to compare age, race, and sex for Medicare beneficiaries residing in the SEER areas versus US totals in 1998. For the comparison of the proportion of persons living in urban areas and the proportion living in poverty, we used data from the 1990 US Census. Because the Census data provide ecologic rather than person-level estimates, we did not report the number of persons for parts of Table 4.

The age and sex distribution for individuals 65 years and older in the SEER areas is comparable with that of the US elderly population. However, the SEER-area race distribution differs from that in the US elderly population in that the SEER areas have a lower portion of white persons and a

higher concentration of persons of other races. In the SEER areas in 1998, more than 90% of persons 65 years and older (among the "other" races) were Asian/Pacific Islander.

At the time of the 1990 Census, elderly persons in the current SEER areas were much more likely to have resided in an urban setting than the average US resident 65 years and older (87% vs. 73%, respectively). The expansion of the SEER program to include the four new states will reduce but not eliminate this difference. Elderly persons in SEER areas were also more likely to be living in more affluent areas. According to the 1990 Census data, 9.5% of elderly persons in the current SEER areas resided in households with incomes below the federal poverty level, whereas the national average at that time was 12.8%. Based on Census data, the inclusion of the new SEER areas will increase the percentage of elderly persons living in poverty, but not to the national average.

The findings regarding the sociodemographic composition of the elderly SEER population are similar to those of an earlier study that compared the total population of the SEER areas with the non-SEER areas.²² The earlier study found that the total SEER population had an age and sex distribution similar to that of non-SEER areas. Moreover, for persons of all ages, the investigators found the SEER areas to be more affluent and more urban than non-SEER areas.

It should be noted that there is variation between each of the SEER registries for some demographic factors. In each of the SEER areas, the over-65 population is similar in terms of age distribution and sex (Table 5), but there is significant variation between the SEER areas in the racial composition of persons 65 years and older. Some registries, such as Iowa and Utah, are almost exclusively white, while other registries—notably San Francisco, Detroit, Hawaii, Atlanta, and Los Angeles—have greater proportions of minority populations.

HMO Enrollment. Using data from Medicare's EDB, we compared the proportion of elderly Medicare beneficiaries enrolled in HMOs in SEER areas with the proportion of elderly Medicare beneficiaries in HMOs throughout the United States. We used three points in time for which SEER-Medicare data are available: 1992, 1995, and 1998. As can be seen in Figure 1, for all three time periods, the portion of elderly SEER area residents enrolled in HMOs is considerably higher than for the US total, although the magnitude of the

TABLE 4. Demographic Characteristics for Persons Age 65 and Older* Residing in the SEER Areas Compared With United States Total

	SEER Areas		SEER with Expanded Areas		US Total	
	n =	%	n =	%	n =	%
Total population	4,030,943	—	7,889,874	—	33,694,894	—
Age group						
65–69	1,092,366	27.1	2,163,122	27.4	9,320,992	27.7
70–74	1,046,733	26.0	2,056,602	26.1	8,732,548	25.9
75–79	847,708	21.0	1,658,023	21.0	6,974,562	20.7
80–84	559,390	13.9	1,082,398	13.7	4,621,495	13.7
85+	484,746	12.0	929,729	11.8	4,045,297	12.0
Race						
White	3,216,314	79.8	6,546,016	83.0	29,212,535	86.7
Black	290,621	7.2	560,112	7.1	2,612,523	7.8
Other	504,618	12.5	747,105	9.5	1,699,017	5.0
Unknown	19,390	0.5	36,641	0.5	170,819	0.5
Gender						
Male	1,657,232	41.1	3,244,017	41.1	13,740,811	40.8
Female	2,373,711	58.9	4,645,857	58.9	19,954,083	59.2
Proportion of people in CT who live in an urban area	—	86.9	—	83.6	—	73.0
Proportion of people in CT who live in a HH with income below the poverty level	—	9.5	—	10.4	—	12.8

*The numbers for age, race and sex are from 1998 data for individuals in the Medicare master enrollment file (EDB). The numbers for urban location and poverty level are estimates from 1990 census data and do not reflect specific individuals.

CT = census tract; HH = household.

difference has diminished over time. This difference is present for both the current SEER areas and for areas to be included as part of the SEER expansion.

Cancer Mortality. Table 6 contains 1996 cancer mortality rates for all cancers as well as the four leading causes of cancer deaths, comparing the total US rates with the rates from the current SEER areas, by race and sex. In addition, we have presented mortality rates for the SEER areas that are included with SEER's expansion. Estimates were obtained from raw US mortality data from the National Center for Health Statistics maintained by the NCI (personal communication, Milton Eisner, SEER program).

As can be seen in Table 6, the US cancer mortality rate in 1996 among elderly persons for all cancers and all races was 9% higher than the mortality rate for the current SEER areas (1127.9/1038.9), with the US total for men being 11% higher than in the current SEER areas. The most notable differences in mortality between the current SEER areas and the US total were for lung cancer (especially in white males), colorectal cancer (for both black and white males), and prostate

cancer (for black males). When the expanded SEER areas were included in the comparison, the differences between the SEER areas and the US total in mortality rates for the 65-and-older population were markedly lower, although some differences persist.

An earlier study comparing 1988 cancer mortality rates for all persons in SEER with the total US cancer mortality rates found some variability between mortality rates for SEER areas compared with US totals.²³ However, the differences observed in the earlier study were less than what we found including elderly persons only. The differences we observed between the SEER and US mortality rates suggest that investigators who intend to extrapolate incidence and mortality from the current SEER-Medicare data to US totals must do so with caution and understand the limitations.

Conclusion

The SEER-Medicare data are a unique resource that combines clinical and demographic data col-

TABLE 5. Number of Persons Age 65 and Older Residing in the SEER Areas, With the Percentage in Selected Subgroups, 1998

No. of cases	Total 4,286,075 (%)	San Francisco		Detroit		Hawaii		Iowa		New Mexico		Seattle		Utah		Atlanta		San Jose- Monterrey		Los Angeles	
		(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
Age group																					
65-69 y	28.0	27.7	25.9	29.1	28.6	25.1	30.2	27.6	28.4	31.8	28.9	28.5									
70-74 y	25.3	25.1	24.6	26.3	27.1	24.0	26.0	25.0	25.5	25.0	25.5	25.5									
75-79 y	21.0	21.4	21.6	20.6	21.2	20.9	20.3	21.3	20.8	19.4	20.7	21.1									
80-84 y	13.7	13.9	14.9	13.1	12.8	15.2	12.9	14.1	13.9	12.6	13.5	13.2									
85+ y	12.0	11.9	13.1	11.0	10.3	14.8	10.6	12.0	11.3	11.1	11.4	11.7									
Race																					
White	82.9	73.8	94.5	79.6	21.3	98.8	93.2	92.6	97.5	79.3	83.8	77.9									
Black	7.6	9.4	4.6	19.3	0.8	0.9	1.9	2.2	0.5	18.7	1.6	10.1									
Asian/Pacific Islander	9.0	16.5	0.8	0.9	77.6	0.2	0.6	4.5	1.4	1.9	14.2	11.6									
American Indian/Aleutian Islander/ Eskimo	0.5	0.4	0.1	0.2	0.2	0.1	4.3	0.8	0.7	0.1	0.4	0.4									
Sex																					
Male	41.4	41.5	40.5	40.2	45.2	40.6	44.1	42.5	43.9	38.9	42.0	41.2									
Female	58.6	58.5	59.5	59.8	54.8	59.4	55.9	57.5	56.1	61.1	58.0	58.8									

Source: SEER program data.

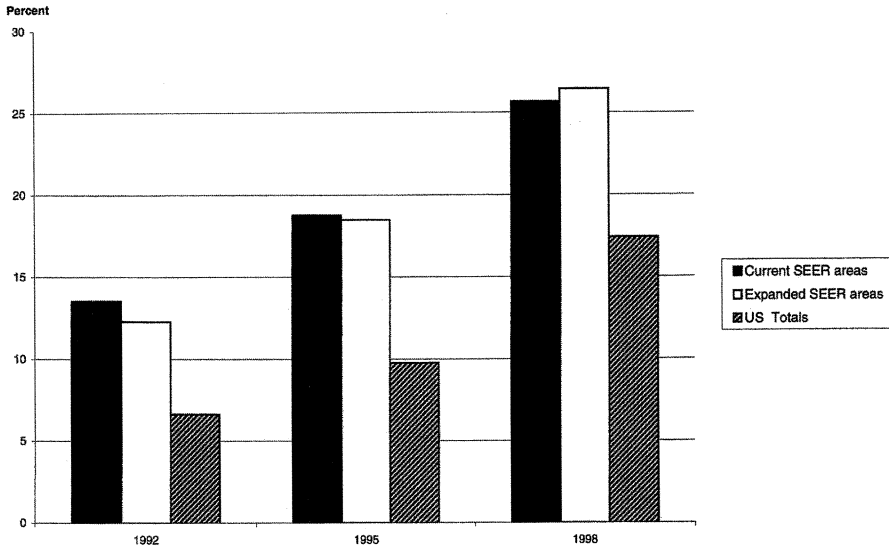


FIG. 1. Proportion of Medicare beneficiaries age 65 and older enrolled in HMOs, SEER areas compared with United States total, by selected years.

lected for public health purposes with health claims data. The data allow for a broad array of studies of cancer-related care and its outcomes in the elderly. Cancer occurs disproportionately in the elderly, with 59% of all cancer cases between 1994 to 1998

occurring in persons 65 years and older.^{24,25} Therefore, a major strength of the data is that they are population based, with a large number of cases. Extraction of all of the Medicare claims for each cancer patient makes it possible to longitudinally

TABLE 6. Comparison of Cancer Mortality Rates Per 100,000 Persons Age 65 and Older Between SEER Areas and United States Total, by Race, 1996

	All Races			White			Black		
	SEER Current Areas	SEER with Expanded Areas	US Total	SEER Current Areas	SEER with Expanded Areas	US Total	SEER Current Areas	SEER with Expanded Areas	US Total
All cancers									
Both sexes	1,038.9	1,087.5	1,127.9	1,058.3	1,099.3	1,122.1	1,216.6	1,304.4	1,334.4
Male	1,289.8	1,358.7	1,437.4	1,298.9	1,359.3	1,415.9	1,619.0	1,767.3	1,876.0
Female	863.1	895.9	912.8	890.4	915.5	916.7	946.2	995.7	988.8
Prostate									
Male	213.9	218.0	226.4	209.7	210.6	209.0	389.9	417.7	471.0
Breast									
Female	121.8	124.3	127.2	128.5	128.3	128.6	129.1	138.2	134.9
Colorectal									
Both sexes	116.8	123.3	129.4	118.7	124.7	128.8	138.2	150.0	154.2
Male	131.8	142.3	148.7	134.7	144.1	148.2	144.7	168.0	175.0
Female	106.3	109.9	116.0	107.6	111.0	115.2	133.9	137.9	141.0
Lung									
Both sexes	272.8	299.3	316.4	278.7	305.0	318.5	321.3	346.4	338.0
Male	375.8	416.2	460.9	377.5	418.1	458.2	499.9	542.3	561.2
Female	200.6	216.7	216.0	209.7	225.0	220.8	201.4	215.7	195.6

Rates are unadjusted.

track persons from the time of their Medicare eligibility until death. The longitudinal nature of these data is a major benefit, because they allow for evaluation of health care utilization both before and after a cancer diagnosis, as well as enable evaluation of long-term outcomes often missed in clinical trials. The data are also one of the few resources about the costs associated with cancer care from Medicare, the largest insurer in the United States, with over 38 million beneficiaries.

While these data have great potential, they should be used with care. It is important that investigators using these data appreciate the complexity of the data and their limitations. This appreciation includes taking time to understand Medicare's coverage policy related to treatments for the cancer of interest. In addition, researchers need to review their data to insure that all relevant diagnosis and procedure codes are included. Often coding practices do not totally reflect what is identified through the coding manuals. The SEER-Medicare data are observational, and researchers using this data should appreciate the limits of inferring findings from cases that are not randomly assigned to care. While these data are complicated and can be messy, they are a powerful tool that can provide definitive answers to some research questions or be used to help focus more in-depth studies in other areas. Used with caution, these data can serve as an important resource to advance cancer-related epidemiologic and health services research.

Acknowledgments

The authors acknowledge the programming assistance of Angela Fahey and Nicki Schussler at IMS, LeAnn Weaver at FU Associates, and Milton Eisner from the National Cancer Institute. Dr. Martin Brown provided helpful comments.

References

1. **Hewitt M, Simone JV.** *Enhancing Data Systems to Improve the Quality of Care.* Washington, DC: National Academy Press; 2000.
2. Available at: <http://seer.cancer.gov/>.
3. Available at: <http://www.naacr.org>.
4. **Gornick ME, Warren JL, Eggers PW, et al.** Thirty years of Medicare: impact on the covered population. *Health Care Financ Rev* 1996;18:179-237.
5. **Potosky AL, Riley GF, Lubitz JD, et al.** Potential for cancer related health services research using a linked Medicare-tumor registry database. *Med Care* 1993;31:732-748.
6. **Bach PB, Guadagnoli E, Schrag D, et al.** Patient demographic and socio-economic characteristics in the SEER-Medicare database: applications and limitations. *Med Care* 2002;40(Suppl 4):00-00.
7. **US Public Health Service.** *International Classification of Diseases, 9th revision.* Washington, DC: US GPO; (PHS) 92-1260, 1996.
8. **American Medical Association.** *Physicians' current procedural terminology. CPT 95, 5th ed.* Chicago, IL: American Medical Association; 1994.
9. **Klabunde CN, Warren JL, Legler JM.** Assessment of comorbidity using claims data: an overview. *Med Care* 2002;40(Suppl 4):00-00.
10. **Freeman JL, Klabunde CN, Schussler N, et al.** Measuring breast, colorectal and prostate cancer screening with Medicare claims data. *Med Care* 2002;40(Suppl 4):00-00.
11. **Cooper GS, Virnig B, Klabunde CN, et al.** Use of SEER-Medicare data for measuring cancer surgery. *Med Care* 2002;40(Suppl 4):00-00.
12. **Virnig BA, Warren JL, Cooper GS, et al.** Studying radiation therapy using SEER/Medicare-linked data. *Med Care* 2002;40(Suppl 4):00-00.
13. **Warren JL, Harlan LC, Fahey A, et al.** Utility of the SEER-Medicare Data to Identify Chemotherapy Use. *Med Care* 2002;40(Suppl 4):00-00.
14. **Nattinger AB, Schapira MM, Warren JL, et al.** Methodologic issues in the use of administrative claims data to study surveillance after cancer treatment. *Med Care* 2002;40(Suppl 4):00-00.
15. **Potosky AL, Warren JL, Riedel E, et al.** Measuring complications of cancer treatment using SEER-Medicare data. *Med Care* 2002;40(Suppl 4):00-00.
16. **Schrag D, Bach PB, Dahlman C, et al.** Identifying hospital characteristics: sources for use with SEER-Medicare and other claims data. *Med Care* 2002;40(Suppl 4):00-00.
17. **Baldwin LM, Adamache W, Klabunde CN, et al.** Linking physician characteristics and Medicare claims data: issues in data availability, quality, and measurement. *Med Care* 2002;40(Suppl 4):00-00.
18. **Brown ML, Riley GF, Schussler N, et al.** Estimating health care costs related to cancer treatment from SEER-Medicare Data. *Med Care* 2002;40(Suppl 4):00-00.
19. **Earle CC, Nattinger AB, Potosky AL, et al.** Identifying cancer relapse using SEER-Medicare data. *Med Care* 2002;40(Suppl 4):00-00.

20. **Fowles JB, Lawthers AG, Weiner JP, et al.** Agreement between physicians' office records and Medicare Part B claims data. *Health Care Financ Rev* 1995;16:189–199.

21. **Romano PS, Luft HS.** Getting the most out of messy data: problems and approaches for dealing with large administrative datasets. In: Grady ML, Schwartz, eds. *Medical Effectiveness Research Data Methods*. Rockville, MD: Agency for Health Care Policy and Research; 1992.

22. **McCarthy EP, Iezzoni LI, Davis RB, et al.** Does clinical evidence support ICD-9-CM diagnosis coding of complications? *Med Care* 2000;38:868–876.

23. **Nattinger AB, McAuliffe TL, Schapira MM.** Generalizability of the surveillance, epidemiology, and end results registry population: factors relevant to epidemiologic and health care research. *J Clin Epidemiol* 1997;50(8):939–945.

24. **Frey CM, McMillen MM, Cowan CD, et al.** Representativeness of the surveillance, epidemiology, and end results program data: recent trends in cancer mortality rates. *J Natl Cancer Inst* 1992;84(11):872–877.

25. **Ries, LAG, Eisner MP, et al (eds).** *SEER Cancer Statistics Review, 1973–1998*, National Cancer Institute, Bethesda, MD. Available at: http://seer.cancer.gov/Publications/CSR1973_1998/2001.

Appendix. Projects Using the SEER-Medicare Data by Type of Cancer-Control Activity

Cancer Screening and Detection

Studies using SEER-Medicare data to assess screening have focused on mammography, prostate-specific antigen (PSA) tests, and colonoscopy/endoscopy. The availability of claims for the control group can be used in combination with claims from the cancer cases to estimate the screening rates on a population basis, although there are some limitations on identifying screening. In addition, the Medicare claims can be used to evaluate treatment or complications subsequent to a screening procedure (eg, ultrasound or biopsy after mammography).

The SEER-Medicare data also have been used to assess the impact of screening on cancer incidence. Furthermore, several studies have used these data to evaluate the role of diagnostic tests in cancer detection.

Published Studies:

Chow WH, Devesa SS, Warren JL, et al. The rising incidence of renal cell cancer in the United States. *JAMA* 1999;281:1628–1631.

Legler JM, Ries LA, Smith MA, et al. Brain and other central nervous system cancers: recent trends in incidence and mortality. *J Natl Cancer Inst* 1999;91:1382–1390.

Legler J, Feuer E, Potosky AL, et al. The role of prostate-specific antigen (PSA) testing patterns in the recent prostate cancer incidence decline in the USA. *Cancer Causes Cont* 1998;9:519–527.

McCarthy EP, Burns RB, Coughlin SS, et al. Mammography use helps to explain differences in breast cancer stage at diagnosis between older

black and white women. *Ann Intern Med* 1998;128:9:729–736.

Merrill, RM, Feuer, EJ, Warren, JL, et al. The role of transurethral resection of the prostate in population-based prostate cancer incidence rates. *Am J Epidemiol* 1999;150:848–860.

Potosky AL, Miller BS, Kramer BS, et al. The role of increasing detection in the rising incidence of prostate cancer. *JAMA* 1995;273:548–552.

Urban N, Drescher C, Etzione R, et al. Use of a stochastic simulation model to identify an efficient strategy for ovarian cancer screening. *Cont Clin Trials* 1997;18:251–270.

Diagnosis and Treatment (Patterns of Care)

Several studies have used the SEER-Medicare data for patterns-of-care analyses during the peri-diagnostic period. The majority of analyses have included research questions addressing four topics: (1) the use of a specific treatment among persons with a selected cancer, (2) assessment of treatment with the focus on whether persons are being overtreated or undergoing test/procedures that may have no influence on the outcome of their disease, (3) disparities in the type of diagnostic evaluation or treatment that people receive, based on patient characteristics such as age, race, sex, income, or residence in an urban/rural location, and (4) influence of the provider on diagnosis and treatment.

Published Studies:

- Bach PB, Cramer LD, Schrag D, et al. The influence of hospital volume on survival after resection for lung cancer. *N Engl J Med* 2001;345:181–188.
- Bach PB, Cramer LD, Warren JL, et al. Racial differences in the treatment of early-stage lung cancer. *N Engl J Med* 1999;341:1198–1208.
- Ballard-Barbash R, Potosky AL, Harlan LC, et al. Factors associated with surgical and radiation therapy for early stage breast cancer in older women. *J Natl Cancer Inst* 1996;88:716–726.
- Deleyannis FW, Weymuller EA, Garcia I, et al. Geographic variation in the utilization of esophagoscopy and bronchoscopy in head and neck cancer. *Arch Otolaryngol Head Neck Surg* 1997;123:1203–1210.
- Du X, Goodwin JS. Patterns of use of chemotherapy for breast cancer in older women: findings from Medicare claims data. *J Clin Oncol* 2001;19:1455–1461.
- Du X, Goodwin JS. Increase of chemotherapy use in older women with breast carcinoma from 1991 to 1996. *Cancer* 2001;92:730–737.
- Earle CC, Venditti LN, Neumann PJ, et al. Who gets chemotherapy for metastatic lung cancer? *Chest* 2000;117:1239–1246.
- Earle CC, Tsai JS, Gelber RD, et al. Effectiveness of Chemotherapy for Advanced Lung Cancer in the Elderly: Instrumental Variable and Propensity Analysis. *J Clin Oncol* 2001;19:1064–1070.
- Klabunde CN, Potosky AL, Harlan LC, et al. Trends and black/white differences in treatment for nonmetastatic prostate cancer. *Med Care* 1998;36:1337–1348.
- Schrag D, Gelfand SE, Bach PB, et al. Who gets adjuvant treatment for stage II and III rectal cancer? Insight from surveillance, epidemiology, and results—Medicare. *J Clin Oncol* 2001;19:3712–3718.
- Schrag D, Cramer LD, Bach PB, et al. Age and adjuvant chemotherapy use after surgery for stage III colon cancer. *J Natl Cancer Inst* 2001;93:850–857.
- Sundararajan V, Hershman D, Grann VR, et al. Variations in the use of chemotherapy for elderly patients with advanced ovarian cancer: a population-based study. *J Clin Oncol* 2002;20:173–178.
- Sundararajan V, Grann VR, Jacobson JS, et al. Variations in the use of adjuvant chemotherapy for node-positive colon cancer in the elderly: a population-based study. *Cancer J* 2001;7:213–218.
- Warren JL, Riley GF, Potosky AL, et al. Trends and outcomes related to outpatient mastectomy in elderly women. *J Natl Cancer Inst* 1998;90:833–840.

Survivorship

The SEER-Medicare data are ideal for long-term follow-up evaluations of persons with cancer because they allow for assessment of ongoing surveillance, prevalent care, and late complications and outcomes.

Published Studies:

- Cooper GS, Yuan Z, Chak A, et al. Geographic and patient variation among Medicare beneficiaries in the use of follow-up testing after surgery for nonmetastatic colorectal carcinoma. *Cancer* 1999;85:2124–2131.
- Knopf, KB, Warren JL, Feuer EJ, et al. Bowel surveillance patterns following a diagnosis of colorectal cancer in Medicare beneficiaries. *Gastrointest Endosc* 2001;54:563–571.
- Lu-Yao GL, Potosky AL, Albertsen PC, et al. Follow up treatments after radical prostatectomy for prostate cancer: a population-based study. *J Natl Cancer Inst* 1996;88(3/4):166–172.
- Reid BC, Alberg AJ, Klassen AC, et al. Comorbidity and survival of elderly head and neck carcinoma patients. *Cancer* 2001;92:2109–2116.
- Schapira MS, McAuliffe TL, Nattinger AB. Underutilization of mammography in older breast cancer survivors. *Med Care* 2000;38:281–289.

Second Occurrence

A second cancer occurrence can be identified in the SEER data if it is a new primary, as SEER collects data on all new primaries that may develop. However, the SEER program does not collect information on recurrences. Investigators are beginning to use the Medicare data to identify second events through algorithms designed to capture recurrences. If algorithms can be validated to identify recurrences, the SEER-Medicare data can serve as a unique resource to evaluate not only the rate of recurrence, but also factors that are associated with one's risk of recurrence (eg, peri-diagnostic treatment).

Terminal Care/Death

Although there are no published studies of terminal care using the SEER-Medicare data, several projects are underway. These analyses focus primarily on the use and costs of hospice care. Other analyses are evaluating the patterns of care during the last month of life for persons not in

hospice, with emphasis on the intensity of service including the use of treatments that are likely palliative, such as chemotherapy or radiation therapy.

Health Systems/Provider Characteristics

The SEER-Medicare data can be used to assess the role of the health system and/or provider in cancer treatment and outcomes. There have been a number of studies that have compared stage, initial treatment, and survival between Medicare beneficiaries in HMOs versus FFS settings. Because there are no Medicare claims from HMOs, these studies are limited to assessing the initial treatment, as collected by the SEER program, and survival. However, some investigators have used the extent of penetration of managed care in a specific geographic area to model the influence of HMOs on regional practice patterns. Other studies have focused on regional differences in care and outcomes between SEER areas and between the US and Canada.

Investigators using the SEER-Medicare data have assessed the relationship of provider volume to postsurgical complications and mortality. These studies, as well as numerous other noncancer-related studies, have shown a consistent positive relationship between higher hospital surgical volume and favorable outcomes. More recent projects are focusing on the relationship of the surgeon's characteristics and outcomes, with special emphasis on the statistical challenges of separating the hospital's influence from that of the surgeon's.

Published Studies:

- Begg C, Cramer LD, Hoskins WJ, et al. Impact of hospital volume on operative mortality for major cancer surgery. *JAMA* 1998;280:1747-1751.
- Bach PB, Cramer LD, Schrag D, et al. The influence of hospital volume on survival after resection for lung cancer. *N Engl J Med* 2001;345:181-188.
- Groome PA, O'Sullivan B, Irish JC, et al. Glottic cancer in Ontario, Canada and the SEER areas of the United States: do different management philosophies produce different outcome profiles? *J Clin Epidemiol* 2001;54:301-315.
- Groome PA, Mackillop WJ, Rothwell DM, et al. Management and outcome of glottic cancer: a population based comparison between Ontario, Canada and the SEER Areas of the United States. *J Otolaryngol* 2000;29:67-77.
- Merrill RM, Brown ML, Potosky AL, et al. Colorectal cancer treatment and survival in Medicare

HMO and Fee-for-Service settings. *Med Care Res Rev* 1999;56:177-196.

- Potosky AL, Merrill RM, Riley GF, et al. Prostate cancer treatment and 10-year survival among group/staff HMO and FFS Medicare patients. *Health Serv Res* 1999;34:525-546.
- Potosky AL, Merrill RM, Riley GF, et al. Breast cancer survival and treatment in HMO and fee-for-service settings. *J Natl Cancer Inst* 1997;89:1683-1691.
- Riley GF, Potosky AL, Klabunde C, et al. Breast cancer stage at diagnosis and treatment patterns among older women in HMOs and fee-for-service. *JAMA* 1999;281:720-726.
- Riley GF, Potosky AL, Lubitz JD, et al. Stage of cancer at diagnosis for Medicare HMO and fee-for-service enrollees. *Am J Public Health* 1994;84:1598-1604.
- Riley GF, Feuer EJ, Lubitz JD. Disenrollment of Medicare cancer patients from HMOs. *Med Care* 1996;34:826-836.
- Schrag D, Cramer LD, Bach PB, et al. Influence of hospital procedure volume on outcomes following surgery for colon cancer. *JAMA* 2000;284:3028-3035.

Costs of Cancer Treatment

The SEER-Medicare data are one of the best data resources for studies of cancer costs. These data allow investigators to determine Medicare charges and reimbursements across the course of treatment for the universe of persons with a specific cancer. Over time, the methodology used to assess costs has evolved from measuring total costs to developing methods to assess costs related to cancer treatment. Recently, studies have focused on comparing costs by differing treatment modalities (eg, mastectomy vs. breast-conserving surgery). In addition to measuring the costs of treatment, investigators have used the SEER-Medicare data to derive estimates for costs of selected procedures as part of cost-effectiveness analyses.

Published Studies:

- Brown, ML, Riley GF, Potosky AL, et al. Obtaining long-term disease specific costs of care: application to Medicare enrollees diagnosed with colorectal cancer. *Med Care* 1999;37:1249-1259.
- Etzioni R, Ramsey SD, Berry K, et al. The impact of including future medical care costs when estimating the costs attributable to a disease: a colorectal cancer case study. *Health Econ* 2001;10:245-256.

- Etzioni R, Urban N, Baker M. Estimating the costs attributable to a disease with application to ovarian cancer. *J Clin Epidemiol* 1996;49:95–103.
- Lin DY. Linear regression analysis of censored medical costs. *Biostatistics* 2000;1:35–47.
- Riley GF, Potosky AL, Lubitz JD, et al. Medicare payments from diagnosis to death for elderly cancer patients by stage at diagnosis. *Med Care* 1995;33:828–841.
- Warren JL, Brown ML, Fay MP, et al. Costs of treatment for elderly women with early stage breast cancer. *J Clin Oncol* 2002;20:307–316.

Methods Related to the Use of SEER-Medicare Data

A number of methodological analyses have been conducted related to the use of the SEER-Medicare data for health services research. Several analyses have evaluated the completeness of each of the data sources for capturing initial surgery and adjuvant radiation by comparing what is reported in SEER with the reporting from the Medicare data.

One of the challenges of using observational data, such as SEER-Medicare, is that persons are not randomly assigned to treatment groups. Therefore, observational studies that evaluate treatment and outcomes must control for differences between groups that exist before diagnosis and treatment. Methods such as two-stage estimation and the use of instrumental variables or propensity scores can be used. A number of investigators include information from the Medicare claims to measure the preexisting health status or comorbidities of each person. Several investigators have used or are using the SEER-Medicare data

to develop improved methods to measure comorbidity.

Published Studies:

- Cooper GS, Yuan Z, Stange KC, et al. Agreement of Medicare claims and Tumor Registry data for assessment of cancer-related treatment. *Med Care* 2000;38:411–421.
- Cooper GS, Yuan Z, Strange KC, et al. The utility of Medicare claims data for measuring cancer stage. *Med Care* 1999;37:706–711.
- Cooper GS, Yuan Z, Stange KC, et al. The sensitivity of Medicare claims data for case ascertainment of six common cancers. *Med Care* 1999;37:436–444.
- Du X, Freeman JL, Goodwin, JS. Information on radiation therapy treatment in patients with breast cancer: the advantages of the linked Medicare and SEER data. *J Clin Epidemiol* 1999; 52:463–470.
- Du X, Freeman JL, Warren JL, et al. Accuracy and completeness of Medicare claims data for surgical treatment of breast cancer. *Med Care* 2000;38:719–727.
- Freeman JL, Zhang D, Freeman DH, et al. An approach to identifying incident breast cancer cases using Medicare claims data. *J Clin Epidemiol* 2000;53:605–614.
- Klabunde, CN, Potosky, AL, Legler, JM, et al. Development of a comorbidity index using physician claims data. *J Clin Epidemiol* 2000;53:1258–1267.
- Warren JL, Feuer E, Potosky A, et al. Use of Medicare hospital and physician data to assess breast cancer incidence. *Med Care* 1999;37:445–456.
- Warren, JL, Riley, GF, McBean, AM, et al. Use of Medicare data to identify incident breast cancer cases. *Health Care Financ Rev* 1996:237–246.