

---

# Overview of the SEER—Medicare Health Outcomes Survey Linked Dataset

Anita Ambs, M.P.H., Joan L. Warren, Ph.D., Keith M. Bellizzi, Ph.D., M.P.H., Marie Topor, Samuel C. (Chris) Haffer, Ph.D, and Steven B. Clauser, Ph.D., M.P.H.

---

*The Surveillance, Epidemiology, and End Results (SEER)—Medicare Health Outcomes Survey (MHOS) links cancer registry data with survey data from Medicare managed care enrollees. The linked file includes clinical information about the cancer with self-reported data about symptoms, functional status and health-related quality of life (HRQOL) for Medicare managed care enrollees. This article provides a description of the SEER-MHOS data as a tool to study cancer among Medicare enrollees. In order to highlight the strengths of the database, we also present some descriptive statistics from the database.*

## INTRODUCTION

Although cancer strikes both young and old, it is well recognized in the medical community that the disease disproportionately affects the elderly. Almost 60 percent of incident cases (all sites combined) occur in males and females age 65 or over (Surveillance Research Program, National Cancer Institute, 2003). Of the 10.8 million prevalent cancer population, 60 percent (6.5 million) of cancer survivors are age 65 or over (Surveillance, Epidemiology, and End Results Program, 2007). Forty-three percent of these 6.5 million elderly males and females with cancer survive more than 10 years and 17 percent of them survive more than 20 years

from their initial diagnosis (Surveillance, Epidemiology, and End Results Program, 2007). Although age-adjusted cancer incidence rates in the U.S. have declined significantly from 1995-2004 (Ries et al., 2007), the number of people newly diagnosed with cancer is expected to increase because of population growth and the aging of the U.S. population. It is estimated that the number of persons with cancer will double by the year 2050 (Edwards et al., 2002).

The burden that cancer places on the U.S. population has resulted in much interest in assessing cancer treatment and outcomes, especially for older persons. A number of population-based studies have used secondary data to evaluate patterns and quality of care, outcomes, and health care costs for elderly persons with cancer. One aspect of the cancer experience of older people that cannot be addressed with studies based on available secondary data, however, is HRQOL. Data from pediatric and young adult cancer populations show that cancer and its treatment can result in years of physical and mental distress for some individuals (Robison, 2005; Zabora et al., 2001; Oeffinger et al., 2006), but little is known about how cancer treatment affects HRQOL in the elderly. Moreover, the elderly population differs from younger groups in a number of respects, including the fact that they have higher levels of comorbidities that may affect HRQOL even in the absence of a cancer diagnosis. The addition of a cancer diagnosis to other health conditions may have an interactive affect on HRQOL. The role of comorbidity

---

Anita Ambs, Joan L. Warren, Keith Bellizzi, and Steven B. Clauser are with the National Cancer Institute (NCI). Marie Topor is with IMS. Samuel C. (Chris) Haffer is with the Centers for Medicare & Medicaid Services (CMS). The statements expressed in this article are those of the authors and do not necessarily reflect the views or policies of NCI, IMS, or CMS.

on HRQOL is an emerging area of research interest only recently receiving attention among investigators (Bellizzi and Rowland, 2007; Lichtman, Balducci, and Aapro, 2007; Rao and Demark-Wahnefried, 2006; Aziz and Bellizzi, 2008; Yancik et al., 1996; Extermann, 2007; Hewitt, Rowland, and Yancik, 2003).

In 1998, CMS began to monitor the quality-of-care provided by Medicare managed care plans through the MHOS. The MHOS was designed to gather valid, reliable, and clinically meaningful data on health outcomes, including functional status, comorbid conditions, symptoms, and HRQOL for the approximately 8.3 million Medicare managed care enrollees, who account for 19 percent of Medicare beneficiaries in 2007 (U.S. Department of Health and Human Services, 2008). MHOS data provide a unique opportunity to assess HRQOL for patients with selected clinical conditions, including cancer.

To assess HRQOL for cancer patients, NCI and CMS have collaborated to link the MHOS with cancer registry data from the SEER program of population-based registries (Espey et al., 2007). This linked dataset provides a powerful and efficient way to collect data on HRQOL of older cancer patients enrolled in health maintenance organizations (HMOs).

## OVERVIEW

### Data Sources

#### SEER

The SEER program collects information about all newly diagnosed cancer (incident) cases in populations within defined geographic areas. The program, which began in 1973, now includes registries that cover about 26 percent of the U.S. population. Detailed information can be found at

<http://seer.cancer.gov/>. The SEER areas include Connecticut, Hawaii, Iowa, New Mexico, Utah, Kentucky, Louisiana, New Jersey, California, the metropolitan areas of Detroit, Atlanta, Seattle-Puget Sound, and rural Georgia. Data collection for some registries, Kentucky, Louisiana, New Jersey, and greater California, began in 2000. The information collected by the registries includes patient's age, sex, race, and marital status. Information about the cancer consists of the month and year of diagnosis, the site, behavior, and stage. Staging in the SEER data is based on classification schemata that vary by cancer site and year of diagnosis. These methods include SEER historic staging and American Joint Committee on Cancer staging system, the latter available for all cancers other than lymphoma and leukemia. The SEER historic staging variable consists of the categories of in situ, localized, regional, distant and unstaged and can be used to track trends in stage over time. SEER registries also collect information about surgical and radiation treatment recommended or provided within 12 months of diagnosis. Information about chemotherapy is not reported because of concerns about under ascertainment. Followup is limited to vital status, and cause of death, if applicable. The data collected by the registries comes primarily from medical records and reports from health professionals. The registries do not collect information on non-melanoma skin cancer, use of screening, or how the cancer was detected, cancer recurrence or progression, sequela of disease or treatment, or cancer-specific symptoms related to HRQOL (Warren et al., 2002).

#### MHOS

The National Committee for Quality Assurance began the MHOS in 1996 under contract to CMS. The original intent of

the MHOS was to measure health outcomes of Medicare beneficiaries who are enrolled in Medicare Advantage (previously Medicare+Choice) health plans for use in monitoring plan performance and improving health outcomes. The MHOS includes a 95-item core questionnaire that is administered to 1,000 randomly selected beneficiaries including institutionalized and disabled beneficiaries who were members of the participating managed care organization (MCO) in the Medicare Advantage program. In plans with fewer than 1,000 enrollees, all eligible members were surveyed. A baseline survey was administered to the first cohort in May 1998, and a followup survey was administered to the same cohort in spring 2000. A new cohort is randomly selected each year for baseline measurement and a 2-year followup assessment.

Since 1998, CMS has conducted 10 baseline surveys and 8 followup surveys. The MHOS data used for the SEER-MHOS project include people who were selected to respond to the MHOS between 1998-2001 and their 2-year followup surveys, if available. These years were selected as it allowed us to use an established link between Medicare enrollment data and persons in SEER. The core MHOS survey contains questions related to demographics, socio-economic status, health problems, functional status (activities of daily living [ADLs]), and symptoms<sup>1</sup>. It also assesses measures of HRQOL through the Medical Outcomes Study Short Form-36 (SF-36<sup>®</sup>, version 1). This instrument is widely used and has been validated in several studies, including many that have assessed cancer care (Turner-Bowker et al., 2002). As shown in Table 1, the SF-36<sup>®</sup> has eight scales, with each representing a separate construct of HRQOL (Jones, Jones, and

Miller, 2004). The eight scales provide the basis for calculating two summary measures, the physical component summary (PCS) and the mental component summary (MCS). Scores on the two summary measures include data from all eight scales, but vary by the order in weight applied to each scale. The PCS score is mostly determined by the physical functioning, role-physical, bodily pain, and the general health scales. The MCS is mostly determined by the mental health, role-emotional, social functioning, and vitality scales.

The SF-36<sup>®</sup> uses norm-based scaling, meaning that scores on the instrument are standardized using an algorithm that builds in normative values for the U.S. population. Thus, all scores above or below 50 can be interpreted as above or below the general population norm. Moreover, because the standard deviations for each scale are equalized at 10, it is relatively easy to see exactly how far above (or below) the mean any particular score is in standard deviation units (Ware et al., 2004). This aspect of the SF-36<sup>®</sup> is a major advantage over other HRQOL instruments used in this type of research, because other instruments typically do not allow direct comparisons to U.S. population norms. The extensive self-reported information on the MHOS can be used to compare HRQOL differences between cancer survivors, cancer patients, and the general elderly population enrolled in MCOs, as well as differences within these groups by demographic, socio-economic status, and type of MCO health plan. The available information on the MCO health plans include type of plan, the plan's State, name, and organization name, Medicare product name, and CMS region. It also includes information on the population of people it served by MCO. In addition, information on the health plan's start date, duration of the plan contract, and duration of the health plan categories are also available.

<sup>1</sup> The full survey is available on CMS' Web site at <http://www.cms.hhs.gov/>.

**Table 1**  
**Medical Outcomes Study Short Form SF-36®**

Scale	Description
Physical Component Summary	Summary measure which includes: physical functioning, role-physical, bodily pain, vitality, social functioning, role emotional, mental health, and general health.
Physical Functioning	Ten questions asking for ratings of the extent to which the respondent's health limits them in their performance of physical activities.
Role-Physical	Four questions asking the extent to which the physical health of the person completing the scale limits them in their work or other usual activities in terms of time and performance.
Bodily Pain	Two questions asking for the severity of pain experienced by the respondent and for the extent to which pain interferes with normal work, including work outside the home and housework.
General Health	Five questions that ask the individuals to rate their current health status overall, their susceptibility to disease, and their expectations for health in the future.
Mental Component Summary	Summary measure which includes: mental health, role-emotional, social functioning, vitality, general health, bodily pain, role-physical, and physical functioning.
Vitality	Four questions asking for subjective well-being ratings in terms of energy and fatigue
Social Functioning	Two questions on limitations in normal social functioning due specifically to health-related problems.
Mental Health	Five questions about the frequency of feelings representing the 4 major mental health dimensions.
Role-Emotional	Three questions about whether emotional problems have interfered with accomplishments at work or other usual activities in terms of time, as well as performance.

SOURCE: National Committee for Quality Assurance. Healthcare Effectiveness Data and Information Set (HEDIS®) 2000 Volume 6: Specifications for the Medicare Health Outcomes Survey. Washington, DC. 2000.

The MHOS contains limited information about cancers; there are only questions that ask if the person has been diagnosed with any of the major cancers (colorectal, lung, breast, and prostate). Other than this, the survey instrument does not identify persons with specific cancers. Additionally, no questions related to the nature of initial treatment, length of time between diagnosis and survey administration, and severity of cancer as measured by stage of disease are included in the survey.

### **Linkage of MHOS and SEER**

The linkage of MHOS and SEER data was accomplished by using an existing file that links persons in the SEER data to Medicare's Master Enrollment File. This existing file was initially constructed for the SEER-Medicare linked database,

another collaborative project between CMS and NCI (<http://healthservices.cancer.gov/seermedicare/>). The linkage of SEER cases to Medicare's Enrollment File is based on an algorithm that involved a match of a respondent's Social Security number (SSN), sex, last name, first name, and month of birth. In the absence of a match on the SSN, respondents were matched on their last name, first name, sex, and month of birth. This algorithm required an agreement between 7 or 8 digits of the SSN or a match on two or more of the following identifiers: year of birth, day of birth, middle initial, and date of death (month and year). For persons in the SEER data age 65 or over, 93 percent were matched to Medicare's enrollment data. For persons found in both data sources, a SEER-Medicare crosswalk file was created that linked each person's unique SEER case number

to their Medicare health insurance claim number (HICNUM).

To create the SEER-MHOS linked database, we took the HICNUMs from MHOS respondents and attempted to match these numbers to HICNUMs for persons in the SEER-Medicare crosswalk file. Persons found in the MHOS group who were found in the SEER-Medicare crosswalk were classified as cancer cases. Those who were not matched to the SEER-Medicare data were considered possible controls.

## SEER-MHOS Analytic File

Using this linked SEER-MHOS database, NCI and CMS have constructed a file that is designed to meet the analytical needs of investigators who wish to use the MHOS data to examine cancer-related issues. This file includes only persons who have responded to at least one MHOS survey for persons with and without cancer. Up to eight records per person may be available in this file, in the case of individuals who were included in the MHOS sample every year and responded to all four baseline and followup surveys. Figure 1 provides a flow

**Figure 1**  
**Number of Patients in MHOS (1998-2001)**  
**All Ages**

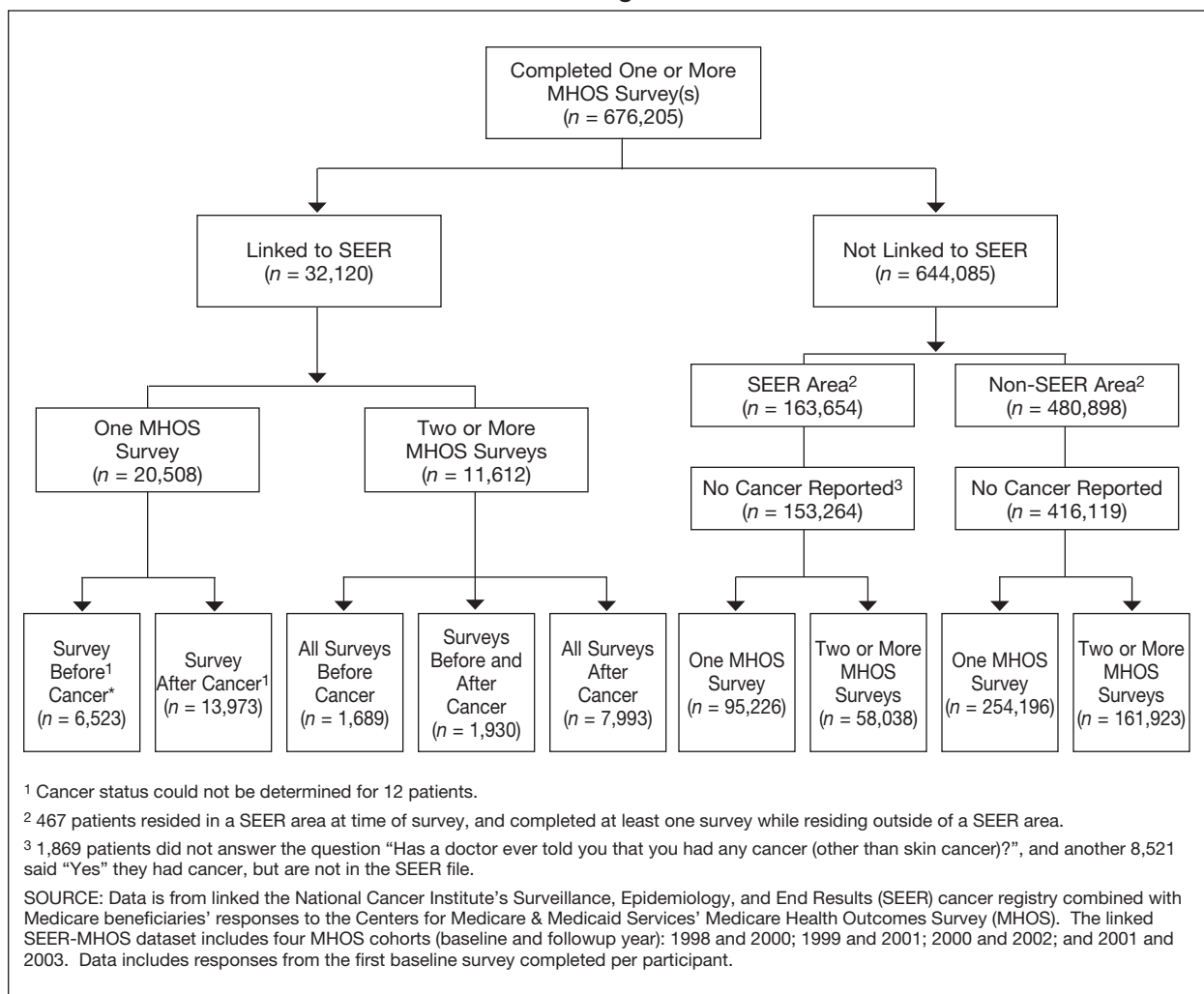


chart that describes the number of people included in the file. For persons with cancer, the chart provides information about the timing of surveys in relation to the cancer diagnosis.

Not everyone who was selected for inclusion in the MHOS cohort agreed to participate in the survey. Table 2 provides information on the total number of surveys included in the MHOS sample and the percentage of surveys completed at baseline or followup. A completed survey was defined as one in which 80 percent of all questions had been answered. Followup surveys were only provided for people who were alive and still in the plan at the 2-year followup. MHOS respondents who had discrepancies on sex or a difference of more than 6 months on date of birth or date of death were deleted from the linked cohort ( $n=88$ ). In addition, because the MHOS survey could have been completed by a proxy (such as a family member or friend), an additional verification was done to confirm that the respondent was alive at the time of survey administration. A small number of respondents were deleted because their date of death was before their survey date

( $n=82$ ). MHOS respondents not found in the crosswalk file (i.e., respondents without cancer) also were checked for consistency between date of death (if any) and MHOS survey date administration. As a result, an additional 719 respondents were deleted from the sample because their date of death was before their survey date. The number of postmortem surveys appears to be relatively large, but is still a very small fraction of the MHOS respondents (0.07 percent). Approximately one-half of those postmortem surveys were completed by the respondents themselves, indicating an erroneous survey date. Thus, most of the discrepancies could be due to an erroneous survey date or date of death in the Medicare file, and it was decided to exclude all patients with a survey date after date of death.

The SEER-MHOS Analytic File includes demographic information obtained from several sources including the Enrollment Data Base (EDB) File maintained by CMS for Medicare enrollees, SEER files, and self-reported information. For persons with cancer, information from the SEER data is provided about the date of diagnosis and

**Table 2**  
**Response Rates to the MHOS Baseline and Followup Surveys, by Cohort and Linkage to SEER Status: 1998 to 2003**

Baseline Survey Cohort <sup>1</sup>	Linked to SEER Data			Not Linked to SEER Data		
	Number of Surveys		Response Rate	Number of Surveys		Response Rate
	Distributed	Completed	Percent	Distributed	Completed	Percent
1998	14,228	9,121	64.1	264,230	166,944	63.2
1999	14,174	9,557	67.4	286,361	194,775	68.0
2000	14,119	10,103	71.6	284,169	202,206	71.2
2001	7,677	5,142	67.0	182,537	122,696	67.2
Followup Survey Cohort						
2000	4,341	3,684	84.9	84,792	72,622	85.6
2001	4,220	3,477	82.4	84,079	70,336	83.7
2002	3,465	2,644	76.3	83,499	66,520	79.7
2003	2,329	1,878	80.6	69,129	56,723	82.1

<sup>1</sup> Patients in more than one cohort are counted more than once in the table.

SOURCE: Data is from linked the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) cancer registry combined with Medicare beneficiaries' responses to the Centers for Medicare & Medicaid Services' Medicare Health Outcomes Survey (MHOS). The linked SEER-MHOS dataset includes four MHOS cohorts (baseline and followup year): 1998 and 2000; 1999 and 2001; 2000 and 2002; and 2001 and 2003. Data includes responses from the first baseline survey completed per participant.

clinical information for up to 10 cancers, vital status information, and some census tract information, such as median household income and education at time of first cancer diagnosis at age 65 or over (Warren et al., 2002). The file has been constructed in a way to facilitate the creation of cohorts. Respondents can be selected on a number of factors including cancer status, residency in SEER area, and a variety of other survey and cancer-related indicators. In addition, to support both cross-sectional and longitudinal study designs, the file allows investigators to identify individuals who participated in multiple surveys. (A file layout with a data dictionary is available on request from the NCI authors.)

### **Data Confidentiality**

The SEER-MHOS data is a valuable resource. However, it is not in the public domain at this time and is available only to Federal Government scientists and collaborators. SEER-MHOS region-specific data are available to SEER principal investigators. Before the data become available to the general public, safeguards to protect patient and provider confidentiality, such as de-identification of the data, and establishing data use agreements, must be taken. NCI intends to work with CMS to develop a public use resource for these data in the near future.

### **DESCRIPTIVE DATA FROM THE SEER-MHOS**

Table 3 presents information on selected characteristics of all MHOS respondents, comparing those persons who never had a cancer diagnosis with cancer patients and stratifying by those who had completed surveys before and following a cancer diagnosis. Respondent's age, sex, and race/ethnicity, were collected on the MHOS

survey, as well as on the EDB File and the SEER database for cancer cases. A respondent's age was derived by subtracting the date of birth from the survey administration date. If available, a respondent's race/ethnicity was constructed based on their self-reported information from the MHOS survey. Otherwise, race/ethnicity was obtained from the CMS database or the SEER File. In instances where respondents have multiple MHOS surveys (i.e., a baseline and a followup survey), race/ethnicity information was retained from the first survey. The race/ethnicity variable is a six-category variable; patients of Hispanic origin receive the Hispanic classification regardless of race. All others fall into one of the other categories: White, Black (or African-American), Asian (or Pacific Islander), American Indian (or Alaskan Native), and another race or multiracial (also referred to as other). It should be noted that the number of subjects is low from certain population groups, e.g., American Indian/Alaskan Native, that participated in the MHOS survey. Because of this, stratification of an analysis by race/ethnicity may have resulted in less reliable estimates for those groups than for better represented groups, even after combining all cancer sites in an analysis. Marital status, income, and education were self-reported by the respondent.

To measure the health condition of participants in the MHOS survey, a comorbidity score was developed from a series of questions on pre-existing chronic conditions, at the time of the survey, including hypertension, coronary artery disease, congestive heart failure, heart attack, other heart conditions, stroke, pulmonary disease, bowel disease, arthritis, diabetes, sciatica, and any cancer other than skin cancer. To be sure the answers were based on a clinical diagnosis, the questions were prefaced with "Has a doctor ever told you that you had: ... the condition." Each positive response was

**Table 3**

**Demographic Characteristics for Medicare Health Outcome Survey (MHOS) Respondents (Age 65 or Over) With Completed Baseline Survey Before or After Their First Cancer Diagnosis, and Those Without Cancer**

Characteristic	Non-Cancer <sup>1</sup> (N = 141,194)		Survey Before Cancer Diagnosis <sup>2</sup> (N = 9,827)		Survey After Cancer Diagnosis <sup>3</sup> (N = 21,307)	
	Number	Percent	Number	Percent	Number	Percent
<b>Age</b>						
65-74 Years	83,586	59.2	5,650	57.5	10,522	49.4
75-84 Years	46,605	33.0	3,501	35.6	8,743	41.0
85 Years or Over	11,003	7.8	676	6.9	2,042	9.6
<b>Sex</b>						
Male	57,701	40.9	5,298	53.9	10,771	50.6
Female	83,493	59.1	4,529	46.1	10,536	49.4
<b>Race/Ethnicity</b>						
Unknown	4	0.0	0	0.0	0	0.0
White	110,525	78.3	7,981	81.2	17,421	81.8
Asian or Pacific Islander	6,871	4.9	360	3.7	986	4.6
Black or African-American	8,604	6.1	684	7.0	1,227	5.8
Hispanic	11,833	8.4	569	5.8	1,171	5.5
American Indian or Alaskan Native	785	0.6	54	0.5	105	0.5
Another Race or Multi-Race	2,572	1.8	179	1.8	397	1.9
<b>Marital Status</b>						
Unknown	2,572	1.8	187	1.9	461	2.2
Married	79,772	56.5	5,806	59.1	12,492	58.6
Divorced/Separated	13,782	9.8	973	9.9	1,964	9.2
Widowed	41,294	29.2	2,607	26.5	5,816	27.3
Single/Never Married	3,774	2.7	254	2.6	574	2.7
<b>Income</b>						
Unknown	29,872	21.2	1,935	19.7	4,377	20.5
<\$10,000	18,098	12.8	1,134	11.5	2,288	10.7
\$10,000 to \$19,999	31,134	22.1	2,252	22.9	4,745	22.3
\$20,000 to \$49,999	47,229	33.4	3,375	34.3	7,478	35.1
\$50,000 to \$79,999	9,823	7.0	723	7.4	1,560	7.3
\$80,000>	5,038	3.6	408	4.2	859	4.0
<b>Education</b>						
Unknown	3,311	2.3	204	2.1	500	2.3
Less Than High School	40,003	28.3	2,854	29.0	5,629	26.4
High School Graduate	45,207	32.0	2,968	30.2	6,676	31.3
Some College	31,224	22.1	2,202	22.4	4,898	23.0
College Graduate	10,185	7.2	761	7.7	1,680	7.9
Post College	11,264	8.0	838	8.5	1,924	9.0
<b>Smoking Status</b>						
Non-Smoker	61,495	43.6	3,249	33.1	7,891	37.0
Former Smoker	46,759	33.1	3,770	38.4	8,461	39.7
Smoker	20,408	14.5	1,984	20.2	3,011	14.1
Unknown	12,532	8.9	824	8.4	1,944	9.1
<b>Performance Measures<sup>4</sup></b>						
Physical Component Summary Score (PCS)		42.3		Mean Score 42.1		39.9
Mental Component Summary Score (MCS)		51.1		51.3		50.0

<sup>1</sup> Individuals not linked to Surveillance, Epidemiology, and End Results (SEER), no self-report of cancer and lived in SEER area at the time of baseline survey.

<sup>2</sup> Individuals with baseline surveys before the first cancer diagnosis.

<sup>3</sup> Individuals with the first baseline survey after any cancer diagnosis.

<sup>4</sup> Overall, 96.5 percent of people in the sample have PCS and MCS scores.

NOTE: Respondents are counted once in the table.

$p < 0.0001$  from the Chi-Square statistics on demographic characteristics by cancer status.

SOURCE: Data is from linked the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) cancer registry combined with Medicare beneficiaries' responses to the Centers for Medicare & Medicaid Services' MHOS. The linked SEER-MHOS dataset includes four MHOS cohorts (baseline and followup year): 1998 and 2000; 1999 and 2001; 2000 and 2002; and 2001 and 2003. Data includes responses from the first baseline survey completed per participant.



assigned a point. The comorbidity score is the sum of all responses, and ranges from 0 (fewest comorbidities) to 12 (most comorbidities). The comorbidity score was reset to unknown if any one of the disease condition questions was not answered (data not shown) (Smith et al., 2008).

Survey respondents were categorized into four smoking groups: (1) non-smoker, (2) former smoker, (3) smoker, or (4) those with unknown smoking status. Respondents who reported smoking less than 100 cigarettes in their lifetime were categorized as a non-smoker. Former smokers were those who reported smoking 100 cigarettes or more over their lifetime but were not currently smoking and had not smoked in the past 6 months (Hays et al., 2008).

Table 3 presents data on the number of respondents with and without cancer. For persons with cancer, the data are divided into those who completed a survey before and after a cancer diagnosis. The non-cancer group includes only persons who lived in a SEER area at the time of the baseline survey. The non-cancer group in the SEER area is similar to non-cancer cases living in non-SEER areas other than those people in the SEER area are more likely to be Latino or Asian and less likely to be White. In addition to demographic information, Table 3 includes information about the mean range of scores for the PCS and MCS. The PCS mean score ranges from 39.9 to 42.3, below the general population mean of 50. The MCS mean score is slightly above the general population mean and ranges from 50.0 to 51.3. Table 3 also notes the percentage of unknown values in each variable in the table. Item non-response for most variables is less than 2 percent, except for income, which has a non-response averaging about 21 percent, and smoking status, which is missing in approximately 9 percent of cases. Work by McCall and colleagues (2004) have noted that non-response bias

in the survey is relatively modest, in spite of differences in response rates in baseline and followup surveys, and differential item non-response.

Table 4 includes information about the number of persons who have responded to at least one MHOS survey by specific type of cancer. The table presents information for all cancer patients who completed a baseline survey as well as the number of cancer patients who completed a baseline survey prior to their cancer diagnosis. For persons with multiple cancers, data are presented for the first cancer reported in the SEER data. Data on the number of cancer cases is presented differently in Table 5. This table provides information about the number of respondents that have completed the survey by stage for persons with colorectal, lung, breast, and prostate cancers. Tables 4 and 5 show that the SEER-MHOS Analytic File has a sufficiently large sample size to examine some questions related to cancer treatment and outcomes both cross-sectionally and longitudinally. This is especially true for the most common cancers, prostate, breast, colorectal, and lung.

We provide three examples of functional measures available in the SEER-MHOS data. In Figure 2, we used information from MHOS to compare the self-reported ADL status of cancer respondents who completed the survey before their cancer diagnosis to those who completed the survey after their cancer diagnosis. The data show that ADLs are more difficult to perform for persons who completed the survey after a cancer diagnosis. We performed a similar analysis of self-reported depression in relation to the timing of the cancer diagnosis and compared with Medicare beneficiaries who do not have cancer (Figure 3). The data show that those who completed the survey after their first cancer diagnosis were slightly more depressed than their

Table 4

Number of SEER-MHOS Respondents Age 65 or Over, by First Cancer Site

First Cancer	Total Number of Linked Patients	Baseline Survey (Group A)		Baseline and Followup Surveys (Subset of Group A)		Baseline Survey Before Cancer Diagnosis (Group B)		Survey Before and After Cancer Diagnosis <sup>1</sup> (Subset of Group B)		Baseline Survey After Cancer Diagnosis (Group C)	
		N	Percent	N	Percent	N	Percent	N	Percent	N	Percent
Prostate	10,475	7,261	69.3	2,591	24.7	1,840	17.6	471	4.5	5,891	56.2
Breast	8,472	5,789	68.3	2,250	26.6	1,253	14.8	346	4.1	4,881	57.6
Colorectal	6,157	4,077	66.2	1,496	24.3	1,195	19.4	263	4.3	3,144	51.1
Lung and Bronchus	3,756	2,322	61.8	608	16.2	1,442	38.4	132	3.5	1,010	26.9
Gynecological Cancers	2,941	1,915	65.1	741	25.2	343	11.7	78	2.7	1,650	56.1
Bladder	2,534	1,687	66.6	572	22.6	477	18.8	97	3.8	1,307	51.6
Melanomas—Skin	1,951	1,393	71.4	522	26.8	375	19.2	84	4.3	1,096	56.2
Head and Neck	1,945	1,228	63.1	434	22.3	371	19.1	75	3.9	932	47.9
Kidney and Renal Pelvis	938	630	67.2	215	22.9	237	25.3	51	5.4	444	47.3
Non-Hodgkin's Lymphomas—Nodal	826	534	64.6	172	20.8	207	25.1	34	4.1	361	43.7
Stomach	523	311	59.5	82	15.7	170	32.5	13	2.5	154	29.4
Pancreas	508	305	60.0	66	13.0	255	50.2	12	2.4	62	12.2

<sup>1</sup> Number of individuals with a baseline survey before their first cancer diagnosis and also a baseline or followup survey after their first cancer diagnosis. Excludes persons age 65 or under.

NOTES: The denominator for the percentages in columns 3 through 6 is the total number of linked patients by cancer site. The groups are not mutually exclusive, therefore, the percentages do not sum to 100 percent. (Group B + Group C) > Group A because some patients participated in more than one Medicare Health Outcome Survey (MHOS) cohort and completed at least one baseline survey before cancer diagnosis and at least one baseline survey after first cancer diagnosis.

SOURCE: Data is from linked the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) cancer registry combined with Medicare beneficiaries' responses to the Centers for Medicare & Medicaid Services' MHOS. The linked SEER-MHOS dataset includes four MHOS cohorts (baseline and followup year): 1998 and 2000; 1999 and 2001; 2000 and 2002; and 2001 and 2003. Data includes responses from the first baseline survey completed per participant.

Table 5

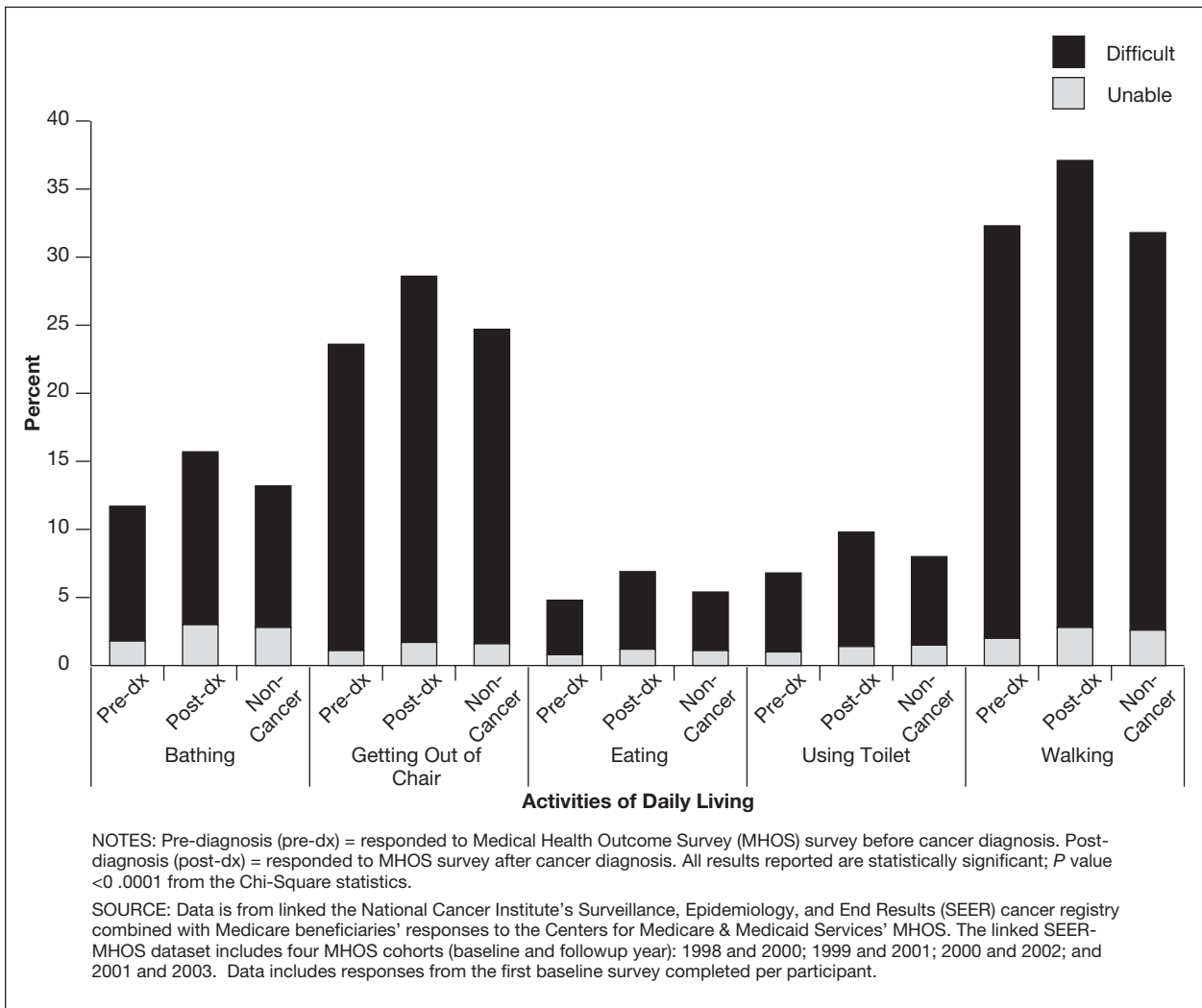
**Number of SEER-MHOS Respondents Age 65 or Over With Completed or Partially Completed Baseline Surveys Before and After Cancer Diagnosis, by Stage and Treatment**

Cancer	Survey Before or After Diagnosis	Historic Stage							
		In situ	Localized	Regional	Distant	Localized/Regional	Unstaged/Missing	Had Surgery	Had Radiation
Prostate	Survey Before Diagnosis	N	NA	NA	66	1,626	148	443	823
	Survey After Diagnosis	%	NA	NA	3.59	88.37	8.04	24.08	44.73
	Survey After Diagnosis	%	4	NA	89	3,662	1,714	2,569	1,983
Breast	Survey Before Diagnosis	N	733	243	37	NA	46	1,179	540
	Survey After Diagnosis	%	58.50	19.39	2.95	NA	3.67	94.09	43.10
	Survey After Diagnosis	%	2,206	691	29	1,125	24.58	4,319	1,503
Colorectal	Survey Before Diagnosis	N	445	429	178	NA	75	1,055	98
	Survey After Diagnosis	%	37.24	35.90	14.90	NA	6.28	88.28	8.20
	Survey After Diagnosis	%	1,114	760	69	769	26.51	2,695	240
Lung	Survey Before Diagnosis	N	241	559	477	NA	163	337	493
	Survey After Diagnosis	%	16.71	38.77	33.08	NA	11.30	23.37	34.19
	Survey After Diagnosis	%	308	268	70	248	27.74	580	232
		%	34.45	29.98	7.83	NA	27.74	64.88	25.95

NOTES: SEER is Surveillance, Epidemiology, and End Results. MHOS is Medicare Health Outcome Survey. NA is not applicable.

SOURCE: Data is from linked the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) cancer registry combined with Medicare beneficiaries' responses to the Centers for Medicare & Medicaid Services' MHOS. The linked SEER-MHOS dataset includes four MHOS cohorts (baseline and followup year): 1998 and 2000; 1999 and 2001; 2000 and 2002; and 2001 and 2003. Data includes responses from the first baseline survey completed per participant.

**Figure 2**  
**Activities of Daily Living Distribution, by Cancer Diagnosis Status**



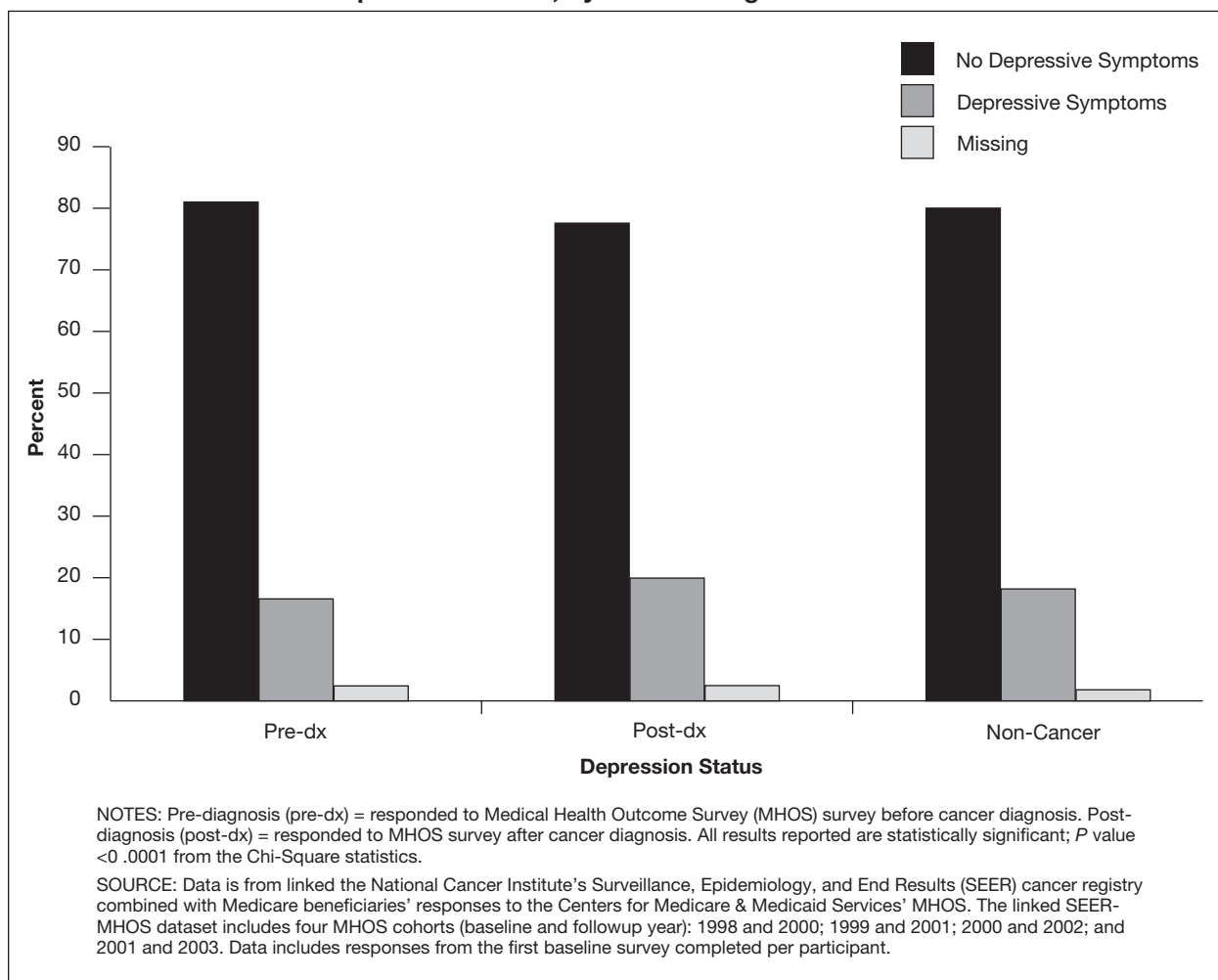
pre-diagnosis status or compared with non-cancer cases. Figure 4 presents information on comorbidity scores based on a count of selected chronic conditions. One can see that among those who completed the survey after their first cancer diagnosis, the percent of respondents who reported having four or more comorbidities was somewhat higher compared with cancer-free patients and patients who did not have cancer at the time of survey administration. Smith et al. (2008) further discuss the role of comorbidity and HRQOL for cancer patients. These examples show how MHOS data from a group of elderly Medicare ben-

eficiaries who do not have cancer can be used to evaluate the status of elderly beneficiaries with cancer.

### SEER-MHOS STRENGTHS

The SEER-MHOS data are a unique resource. The information it contains, such as HRQOL and comorbidities, can be used for analyses that cannot be performed on other secondary data. Because the dataset uses the SF-36<sup>®</sup>, it enables investigators to compare across populations of cancer survivors, as well as to compare individuals with and without cancer. Analyses of the

**Figure 3**  
**Depression Status, by Cancer Diagnosis Status**



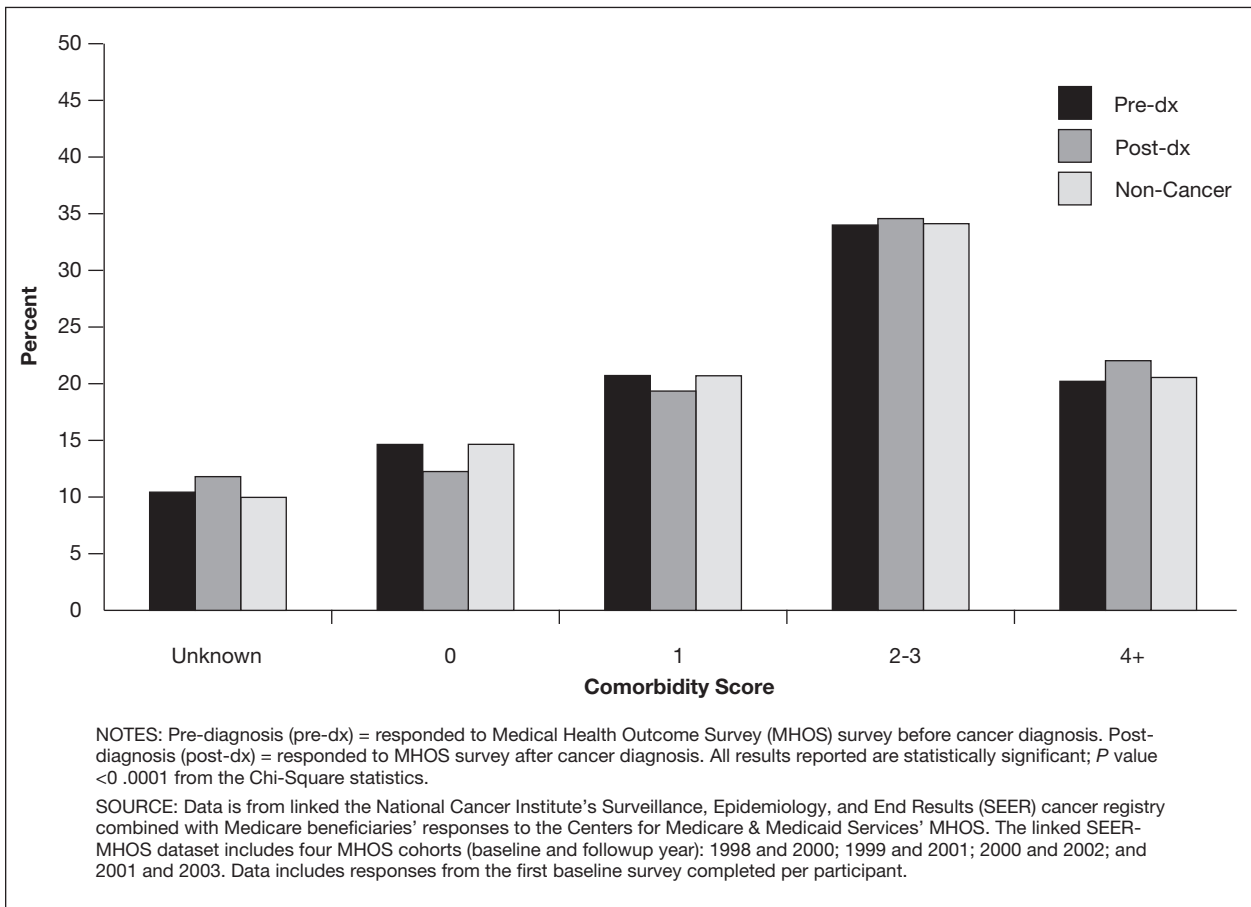
SEER-MHOS data can explore behavioral issues (e.g., smoking behavior) as well as treatment issues (e.g., surgery or radiation oncology).

Existing research on health outcomes of older adults with cancer often examines this population as one homogeneous group. Longitudinal studies of health and aging clearly demonstrate the existence of heterogeneity in respect to the health status of the elderly population (Baltes and Smith, 2003), however, suggesting that differences in health outcomes by cancer status could also exist. The SEER-MHOS data are derived from a large sample, which permits examination of data by age strata

for males and females age 65 or over. The large samples also enable investigators to stratify the elderly population by other important factors, such as race/ethnicity, and socio-economic status. Further, few HRQOL datasets exist that facilitate comparisons across multiple tumor types as well as comparisons with individuals who were never diagnosed with cancer.

Another strength of the SEER-MHOS is the inclusion of a 2-year longitudinal panel of respondents in the dataset. Many of the most important HRQOL research questions have to do with changes in health status among cancer survivors, especially in response to treatment, to the diagnosis

**Figure 4**  
**Comorbidity Score Distribution, by Cancer Diagnosis Status**



of cancer, or in response to other health events, such as a second primary cancer diagnosis. The SEER-MHOS potentially enables the exploration of these types of research questions, at least within the 2-year window of the followup sample.

Finally, as described in Table 4, the SEER-MHOS data can be used to explore HRQOL issues in tumor sites beyond breast, prostate, colorectal, and lung. The dataset has more than 1,000 cases each of gynecological cancers, bladder cancers, melanomas, and head and neck cancers. It also includes more than 2,000 individuals who have been diagnosed with multiple primary cancers (Clauser et al., 2008) and a limited number of respondents with rare cancers. Little is currently known about the HRQOL of sur-

vivors for rare or multiple cancers, and continued efforts to expand the dataset with additional cohorts of survey respondents will increase the power to examine HRQOL issues with these respondents.

### SEER-MHOS Limitations

The SEER-MHOS data also have limitations that users should consider before working with them. The SEER-MHOS is designed to allow standard comparisons across different types of cancers and other clinical characteristics of the study population (e.g., other diseases, individuals who self-report being disease free). However, cancer-specific HRQOL measures would likely be more sensitive than the SF-36®

to the impact of cancer but they would not allow comparisons to those in the sample who do not have cancer. Many researchers would recommend using these refined HRQOL instruments in studies or interventions of selected tumor types where they are available and valid. Despite this limitation, SEER-MHOS studies can still complement the use of other HRQOL instruments by providing a comparative assessment, thereby highlighting opportunities in which researchers and health plans could conduct detailed research and intervention using more sensitive instruments.

Another limitation of the SEER-MHOS is its lack of representativeness of the Medicare Program in general. The primary purpose of the MHOS is to serve as a program monitoring and accountability mechanism for the Medicare managed care program. Therefore, it is not used to collect HRQOL information on Medicare beneficiaries in the fee-for-service program where the vast majority of Medicare beneficiaries are enrolled. An earlier study by Riley (2000) found that Medicare beneficiaries with fee-for-service coverage had more risk factors and lower functional status. Also, because the data are limited to managed care, no administrative claims or utilization data are available from Medicare on this population. Utilization data are limited to what is available in SEER and to specific services associated with initial cancer treatment, such as surgery and radiation. Some types of cancer treatment, i.e., chemotherapy and hormonal therapy, are not reported by the SEER program because of concerns about under ascertainment. Nevertheless, the SEER-MHOS dataset does provide the potential to evaluate selective issues related to HRQOL and receipt of initial cancer treatment, especially with respect to surgical interventions and radiation therapy.

The SEER-MHOS dataset also is not representative of the Medicare Advan-

tage program. Medicare Advantage allows beneficiaries to obtain their Medicare services from HMOs and preferred provider organizations. Overall, the SEER regions in this dataset represent about 27 percent of all Medicare Advantage enrollees (Kaiser Family Foundation, 2007). Certain regions—Florida and Minnesota, for example—that have a large proportion of Medicare managed care enrollees are not included in SEER. Conversely, Medicare Advantage plans are not represented in all SEER regions. The SEER regions with the largest overlap with the Medicare Advantage program are California, Detroit, and Seattle. Nevertheless, these SEER regions account for approximately 1.5 million Medicare health plan enrollees annually and reflect areas that are racially and ethnically diverse. Methods for weighting these data to improve their generalizability in the Medicare Advantage program have not been developed. As an interim measure, Medicare enrollment numbers available at the plan level that might be used to adjust for different plan sizes.

## CONCLUSIONS

The SEER-MHOS data are a potentially valuable resource that combines health-related quality of life, sociodemographic, and clinical data for Medicare health plan enrollees. The data are population-based, with a large number of cancer cases and controls, and the dataset allows for both cross-sectional and longitudinal studies. The inclusion of control cases of individuals who were never diagnosed with cancer is a major advantage, because it allows investigators to evaluate HRQOL in individuals before and after a cancer diagnosis, as well as between cancer survivors and individuals never diagnosed with cancer. The strength of the dataset is its ability to reveal insights about HRQOL, both across different types

of cancers and between respondents with and without cancer (Clauser et al., 2008; Smith et al., 2008; Hays et al., 2008).

Finally, because survey respondents are representative of the health plans in which they enroll, the data may serve important policy research functions as well. The SEER-MHOS provide an attractive window into examining the HRQOL of cancer survivors enrolled in Medicare managed care. Improved cancer therapies and cancer screening technology is allowing Medicare beneficiaries diagnosed with cancer to live longer with the disease. Increased survival will place even greater emphasis on the need for Medicare managed care plans to assist cancer survivors in dealing with quality of life implications of living with cancer. The SEER-MHOS is one tool to inform health care plans on gaps in health status and intervention opportunities that can improve the HRQOL of cancer patients and survivors for whom they are accountable.

## ACKNOWLEDGMENTS

The authors wish to thank Gigi Yuan, Bryce Reeve, Ron Hays, Neeraj Arora, Ashley Smith, Arnold Potosky, and Anne Rodgers for their help with this SEER-MHOS effort.

## REFERENCES

Aziz, N.M. and Bellizzi, K.: Older Survivors and Cancer Care. *Journal of the National Cancer Institute* 100:4-6, 2008.

Baltes, P.B. and Smith, J.: New Frontiers in the Future of Aging. From Successful Aging of the Young Old to the Dilemmas of the Fourth Age. *Gerontology* 49:123-135, 2003.

Bellizzi, K.M. and Rowland, J.H.: The Role of Comorbidity, Symptoms and Age in the Health of Older Survivors Following Treatment for Cancer. *Aging Health* 3:625-635, 2007.

Clauser, S.B., Arora, N.K., Bellizzi, K.M., et al.: Disparities in HRQOL of Cancer Survivors and Non-Cancer Managed Care Enrollees. *Health Care Financing Review* 29(4):23-40, Summer 2008.

Edwards, B.K., Howe, H.L., Ries, L.A.G., et al: Annual Report to the Nation on the Status of Cancer, 1973-1999, Featuring Implications of Age and Aging on U.S. Cancer Burden. *Cancer* 94:2766-2792, 2002.

Espey, D.K., Wu, X.C., Swan, J., et al: Annual Report to the Nation on the Status of Cancer, 1975-2004, Featuring Cancer in American Indians and Alaska Natives. *Cancer* 110:2119-2152, 2007.

Extermann, M.: Interaction Between Comorbidity and Cancer. *Cancer Control* 14:13-22, 2007.

Hays, R.D., Smith, A.W., Reeve, B.B., et al.: Cigarette Smoking and Health-Related Quality of Life in Medicaid Beneficiaries. *Health Care Financing Review* 29(4):57-68, Summer 2008.

Hewitt, M., Rowland, J.H., Yancik, R.: Cancer Survivors in the United States: Age, Health, and Disability. *Journal of Gerontology: Medical Sciences* 58:82-91, 2003.

Jones, N., Jones, S.L., and Miller, N.A.: The Medicare Health Outcomes Survey Program: Overview, Context, and Near-Term Prospects. *Health and Quality of Life Outcomes*. 12(2):33, July 2004.

Kaiser Family Foundation: *State Health Facts.Org Total Medicare Advantage (MA) Enrollment*. 2007. Internet address: <http://www.statehealthfacts.org/comparable.jsp?ind=327&cat=6> (Accessed 2008.)

Lichtman, S.M., Balducci, L., and Aapro, M.: Geriatric Oncology: A Field Coming of Age. *Journal of Clinical Oncology* 25:1821-1823, 2007.

McCall, N., Bonito, A., and Trofimovich, L.: *Estimation and Analysis of Non-response Bias in Medicare Surveys*. RTI International. Durham, NC. 2004.

Oeffinger, K.C., Mertens, A.C., Sklar, C.A., et al.: Chronic Health Conditions in Survivors of Childhood Cancer. *New England Journal of Medicine* 355:1572-1582, 2006.

Rao, A.V. and Demark-Wahnefried, W.: The Older Cancer Survivor. *Critical Reviews. Oncology/Hematology* 60:131-143, 2006.

Ries, L.A.G., Melbert, D., Krapcho, M., et al.: *SEER Cancer Statistics Review, 1975-2004*. National Cancer Institute. Bethesda, MD. Based on November 2006 SEER Data Submission Posted to the SEER Web Site, 2007. Internet address: [http://seer.cancer.gov/csr/1975\\_2004/](http://seer.cancer.gov/csr/1975_2004/) (Accessed 2008.)

Riley, G.: Two-Year Changes in Health and Functional Status among Elderly Medicare Beneficiaries in HMOs and Fee-for-Service. *Health Services Research* 35(5 Pt 3):44-59, December 2000.

Robison, L.: The Childhood Cancer Survivor Study: A Resource for Research of Long-Term Outcomes among Adult Survivors of Childhood Cancer. *Minnesota Medicine* 88(4):45-49, April 2005.



Smith, A.W., Reeve, B.B., Bellizzi, K., et al.: Cancer, Comorbidities, and Health-Related Quality of Life of Older Adults. *Health Care Financing Review* 29(4):41-56, Summer 2008.

Surveillance, Epidemiology, and End Results (SEER) Program: *Prevalence Database: US Estimated Complete Prevalence Counts on 1/1/2004*. National Cancer Institute, DCCPS, Surveillance Research Program, Statistical Research and Applications Branch, Released April 2007, Based on the November 2006 SEER Data Submission. Internet address: <http://www.seer.cancer.gov> (Accessed 2008.)

Surveillance Research Program, National Cancer Institute: *SEER Stat Software Version 5.3.0*. November 2003. Internet address: <http://www.seer.cancer.gov/seerstat> (Accessed 2008.)

Turner-Bowker, D.M, Bartley, P.J., et al.: *SF-36® Health Survey & "SF" Bibliography: Third Edition (1988-2000)*. QualityMetric Incorporated. Lincoln, RI. 2002.

U.S. Department of Health and Human Services: *FY2008 Budget in Brief* 2008. Internet address: [www.dhhs.gov/budget/08budget/2008BudgetInBrief.pdf](http://www.dhhs.gov/budget/08budget/2008BudgetInBrief.pdf). (Accessed 2008.)

Ware, J.E., Gandek, B., Sinclair, S.J., et al.: *Measuring and Improving Health Outcomes: An SF-36® Primer for the Medicare Health Outcomes Survey*. Health Assessment Lab and Quality Metric Incorporated. Waltham MA. 2004.

Warren, J.L., Klabunde, C.N., Schrag, D., et al.: Overview of the SEER-Medicare Data: Content, Research Applications, and Generalizability to the United States Elderly Population. *Medical Care* 40(8 Suppl):IV-3-18, August 2002.

Yancik, R., Havlik, R.J., Wesley, M.N., et al: Cancer and Comorbidity in Older Patients: A Descriptive Profile. *Annals of Epidemiology* 6:399-412, 1996.

Zabora, J., BrintzenhofeSzoc, K., Curbow, B., et al: The Prevalence of Psychological Distress by Cancer Site. *Psycho-Oncology* 10(1):19-28, January/February 2001.

---

Reprint Requests: Anita Ambs, M.P.H., National Cancer Institute, EPN 4005, 6130 Executive Blvd., MSC 7344 Bethesda, MD 20892-7344. E-mail: [ambsa@mail.nih.gov](mailto:ambsa@mail.nih.gov)

