

## SEER-Medicare Data (18 selected cancer sites)

The SEER-Medicare data reflect the linkage of two large population-based sources of data that provide detailed information about Medicare beneficiaries with cancer. The data come from the Surveillance, Epidemiology and End Results (SEER) program of cancer registries that collect clinical, demographic and cause of death information for persons with cancer and the Medicare fee-for-service claims for covered health care services from the time of a person's Medicare eligibility until death. Medicare fee-for-service claims currently include Part A, B, and D.

Data currently available at UNC include cohorts of individuals diagnosed with 18 selected cancers (listed below) between 2004 and 2009. Claims from Medicare Parts A and B data are available from 2003-2010 and for Part D data from 2007-2010. A random sample of 5% of non-cancer controls is also available for the same years.

- Breast cancer, rectal cancer, non-small cell lung cancer, head/neck cancer, bladder cancer, gastrointestinal stromal tumors, small cell lung cancer, kidney cancer, lymphoma, melanoma, colon, prostate, liver, chronic myeloid leukemia, acute lymphoblastic leukemia, pancreatic cancer, multiple myeloma, thyroid cancer.

The linkage of these two data sources results in a unique population-based source of information that can be used for an array of epidemiological and health services research. For example, investigators using this combined dataset have conducted studies on patterns of care for persons with cancer before a cancer diagnosis, over the period of initial diagnosis and treatment, and during long-term follow-up. Investigators have also examined the use of cancer tests and procedures and the costs of cancer treatment. The linked SEER-Medicare data files are large and complex. Before beginning an analysis, researchers are advised to read all documentation to determine whether the data will support their proposed research question. In addition, the SEER-Medicare data have a number of particular

qualities and anomalies (see [Analytic Support for Researchers](#)). Researchers are strongly encouraged to understand the complexity of the data before undertaking any analyses or publishing findings.

The SEER-Medicare data are contained in a series of data files.

- The **Patient Entitlement and Diagnosis Summary File (PEDSF)** includes information on each person's month and year of birth, date of death (if any), sex, race, and state of residence. It also contains variables from other sources that have been added to assist investigators in their analyses. These include Medicare eligibility, reason for Medicare entitlement, and health maintenance organization (HMO) enrollment by month for the period 1991 – 2011. PEDSF also contains variables derived from Census Bureau data on basic socioeconomic status at the census tract and zip code level.
- The **Medicare Claims Files** contain fully adjudicated paid claims and dates of service for all non-capitated ambulatory, emergency department, inpatient, and outpatient encounters (including administrative claims for laboratory tests), skilled nursing facility, hospice, home health, and durable medical equipment claims for patients at the time of service. Diagnoses and procedures are identified by the International Statistical Classification of Diseases and Related Health Problems (ICD-9); Current Procedural Terminology (CPT); and Health Care Financing Administration (HCFA) Common Procedure Coding System (HCPCS) codes for both outpatient and inpatient visits/stays. Physician, specialist, and emergency room visits, as well as hospital stays, are captured in the database through CPT codes, revenue codes (e.g., room and board), used on billing forms such as the UB-92, and place of service codes. The Medicare Claims files include the following: Medicare Provider Analysis and Review claims (MEDPAR), Carrier claims (old file name Physician/Supplier Part B (NCH)), Outpatient claims (OUTPAT), Home Health Agency (HHA), Hospice, and Durable Medicare Equipment (DME).

- **Prescription Drug Claims** (Part D data) include information on dates of dispensing, National Drug Codes (NDCs), strength, amount, and days supply. Other information on prescription drug use is also available.

The content of all of these files are described in detail on the [SEER-Medicare](#) website. IMS is the SEER-Medicare information technology contractor that processes data requests for all researchers interested in using data for their research projects. IMS (on behalf of SEER-Medicare) must review and approve all projects using the SEER-Medicare data that is currently housed at UNC.

### **Gaining Access to the SEER-Medicare Data**

If you are interested in using the SEER-Medicare data, please [Request a CER Consultation](#) (note that you will need to become a member of TraCS and then log in to fill out the request form) and review the [SEER-Medicare](#) website.