

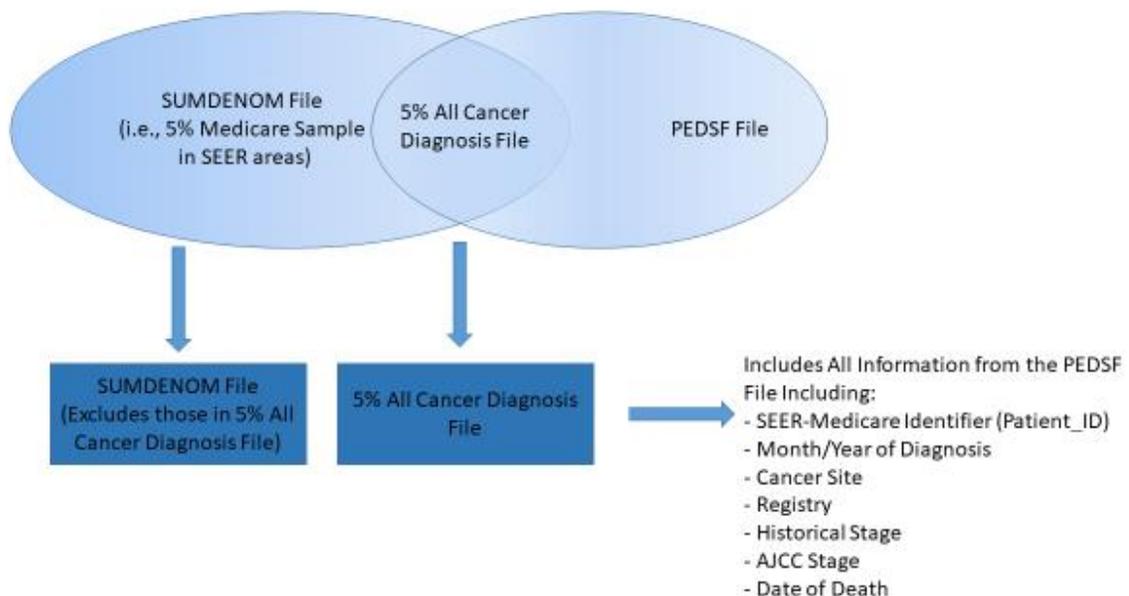
5% All Cancer Diagnosis File

What is the 5% All Cancer Diagnosis File?

The 5% All Cancer Diagnosis file includes individuals with cancer who resided in a SEER area and were in the Medicare 5% sample (i.e., included in the SUMDENOM file). In order to create a non-cancer control sample, these individuals were removed from the 5% sample when the SUMDENOM file is created. Instead, because these individuals were reported to a SEER registry as having had a cancer diagnosis, they are included in the cancer cases file (i.e., PEDSF). The information provided in the 5% All Cancer Diagnosis File allows the ability to include all individuals who were in the Medicare 5% sample. Therefore, a “true” 5% sample can be recreated at any given time by adding persons included in 5% All Cancer Diagnosis file with the SUMDENOM file. Information on cancer characteristics, including the date of cancer diagnosis and tumor characteristics (e.g., cancer site and stage), is available for the included individuals.

Who is Included in the 5% All Cancer Diagnosis File?

The 5% All Cancer Diagnosis file includes all individuals who were included in the SEER-Medicare 5% control sample (i.e., individuals who resided in a SEER area and were included in the 5% Medicare sample and thus in the SUMDENOM file) who have been diagnosed with ANY cancer, as reported to a SEER registry, from 1973-2013. The file was created by linking the SUMDENOM file with the PEDSF file to identify individuals ever diagnosed with a cancer according to SEER.



What Information is Included in the 5% All Cancer Diagnosis File?

The 5% All Cancer Diagnosis file includes all variables included in the [PEDSF](#) file from individuals diagnosed with cancer who were included in the SEER-Medicare 5% control sample (i.e., the SUMDENOM file) including:

- SEER-Medicare Identifier (Patient_ID)
- Month of diagnosis
- Year of diagnosis
- Cancer site of diagnosis
- Registry of diagnosis
- Historical Stage
- AJCC Stage
- Date of Death

How Many People are in the 5% All Cancer Diagnosis File?

The following table provides an estimate of the number of new cancer diagnoses, by cancer type and year, in the 5% All Cancer Diagnosis file. Note this table only presents a subset of diagnosis years; the file includes cancer diagnoses back to 1973, depending on when a registry joined the SEER Program.

Number of Cancer Cases in the 5% Medicare Sample, By File Year

Cancer Type	Year of diagnosis										
	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	Total
Breast	747	768	727	713	725	680	678	742	682	690	7152
Colorectal	733	707	667	632	604	555	508	527	491	421	5845
Lung	942	927	906	863	911	833	795	831	769	741	8518
Prostate	1059	954	1031	1035	939	876	854	852	693	624	8917
Other Cancers	2623	2566	2557	2589	2523	2528	2506	2468	2528	2446	25334
Total (Any Cancer)	6104	5922	5888	5832	5702	5472	5341	5420	5163	4922	55766

Table Footnote: Includes individuals 65+ at diagnosis with their first or only cancer diagnosis. Individuals must have been enrolled in Medicare Part A + B and not enrolled in Medicare Advantage in the month of diagnosis.

How Can Researchers Use the File?

Common uses for the 5% All Cancer Diagnosis file include:

- 1) **Identifying Controls for a Case-Control Study:** Researchers often use the SUMDENOM file to identify appropriate controls for a study. Individuals are not included

in the SUMDENOM file if they have been diagnosed with cancer. The 5% All Cancer Diagnosis file allows researchers to add individuals who are part of the Medicare 5% sample who have a cancer diagnosis with individuals included in the Medicare 5% sample who reside in the SEER areas who have not been diagnosed with cancer (i.e., individuals included in the SUMDENOM) to recreate a “true” 5% sample. Knowledge of the cancer type, dates of diagnosis and other cancer characteristics is available to determine control eligibility.

- 2) **Examining Cancer as an Outcome:** Researchers may be interested in identifying a cancer diagnosis (in general, not a specific cancer type) as a primary study outcome. SEER-Medicare policy does not allow researchers to request the entire data set. Using the 5% All Cancer Diagnosis File allows researchers to examine the relationship between non-cancer exposures (i.e., cancer screening, Parkinson’s, Stroke) and cancer development within the 5% Medicare sample.

Why Would I Want to Use the 5% All Cancer Diagnosis File?

- 1) **Identify all Cancers in a Representative Sample of Medicare Beneficiaries:** SEER-Medicare policy does not allow researchers to request the entire data set (i.e., all cancers included in the PEDSF file). Using the 5% All Cancer Diagnosis files allows researchers to identify all individuals who were included in the SEER-Medicare 5% sample who were diagnosed with cancer, as reported to a SEER registry, as well as identify cancer characteristics for these individuals.
- 2) **Manageable File Sizes:** SEER-Medicare data files can be extremely large and challenging to analyze without large computing capabilities (e.g., servers with large processing memory and storage space). Using a subset of cancers diagnosed in the SEER-Medicare 5% sample allows for more manageable analytic file sizes- particularly for questions evaluating cancer in general as an outcome.
- 3) **Comparable Populations for Comparing Prevalence of Cancer to Other Outcomes:** The PEDSF files contain 100% of cancers diagnosed within a SEER area. The SUMDENOM file contains 5% of individuals who reside in SEER areas who are not known to have cancer. Using the 5% All Cancer Diagnosis File in combination with the SUMDENOM file allows for evaluating comparable relative burdens of disease when comparing cancer prevalence to other disease prevalence (i.e., heart disease) in a population.