



UNIVERSITY OF SOUTHERN CALIFORNIA

Cancer Surveillance Program

VERSION 8/5/2003

Follow-up resource guide

Passive Follow-up

This list highlights major sources of routine Follow-up data available for population based cancer registries. These resources focus on large scale, record linkage, passive forms of follow-up data. Information is as accurate as possible, and will be updated over time. Groups wishing to use these resources should use this guide as a general starting point for contacting these organizations, and strict review of and adherence to the data usage guidelines for each resource goes without saying.

The Cancer Surveillance Program experience with these agencies comes, for the most part, as a result of our involvement with the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute.

Resource

Centers for Medicare & Medicaid Services (CMS)

Description

This resource, previously known as Health Care Financing Administration (HCFA), has for years provided the vast majority of routine follow-up information on patients not known to be deceased. The CMS links the registry data against the Medicare enrollment file, i.e. the list of Americans currently enrolled and getting benefits from Medicare. Since a majority of cancer patients are in the age group that Medicare is involved, this is a highly useful resource. As a prior step to processing with the enrollment file, the data is passed against the Social Security Administration 'Death Master File', which is a collection of deceased SSA members who have 'death benefit' payments made to family members after they have died. So, in addition to enrollment dates for living patients, death information is available for some patients from this resource.

Type Cost recovery based. Service provided by CMS. Registry sends data to CMS

Usage notes

CMS extracts data from their enrollment files based on the 9 digit Social Security Number (SSN) provided by the registries. The input record provided to CMS does NOT allow for registries to include their unique id number. Only unique, valid SSN's have any chance of yielding results.

Rich Pinder [✉ USC Keck School of Medicine](mailto:rpinder@usc.edu) [✉ rpinder@usc.edu](mailto:rpinder@usc.edu)

1420 San Pablo Street, Rm. B105 [✉ Los Angeles Ca 90033](#) [✉ Voice: \(323\) 442-1640](#) [✉ Fax: \(323\) 442-2485](#)

Therefore you must submit unique SSNs from your data and subsequently link the output files from CMS back to your files based on the SSN. Note – since there is a very common situation in Cancer Registries data where SSN's recorded for patients actually belong to the spouse, it is the registries responsibility to process the output data to insure follow-up is applied to the correct person. Since name and date of birth is included in output data from CMS, linkage back to registry data is possible to facilitate this.

Information obtained

For 'not known to be deceased' patients, an imputed follow-up date is used, which is based on taking the actual date of the computer run, and backing up a number of months. The logic is that there is a lag time between when a person leaves the Medicare rolls (usually due to death), and when the file is updated. For deceased SSN's, a date of death is provided.

The CSP has for years taken a conservative approach, and backs up the date for vital status of 'alive' back approximately 5 months from when the computer job is run.

Cost

SEER registries use CMS information under an Interagency Agreement between CMS and NIH, and hence do not directly pay for this process out of our own contracts. However, in 2002 an invoice for CMS to process roughly 275,000 cancer records reports an associated cost of \$1,875.

Contact information

<http://cms.hhs.gov/researchers/>
[name/address to follow]

Resource

Social Security Epidemiological Vital Status Data (SSEVSD)

Description

This resource is provided specifically for individuals doing Epidemiologic research, making it applicable for use by population based cancer registries. It includes the same administrative data files as does the CMS service described above, but adds two very important resources: the 'Master Earnings File' (MEF), which is a database of individuals paying into SSA via the payroll tax (the FICA part on your check stub!), plus the Social Security Record (SSR) for Title XVI beneficiaries, which is often referred to as SSI (disability benefits). These very valuable additions make this resource especially useful for obtaining follow-up on working age (and hence not yet Medicare aged) patients.

Type Cost recovery based. Service provided by SSAEVSD. Registry sends data to SSEVSD.

Usage notes

The SSEVSD performs a more complex linkage process on the data they extract for the user. Name and birthdate information is used to verify the SSN (a mandatory data item for this resource). Their algorithm is a 'Deterministic Record Linkage'* approach, and it accounts for some subtle disagreements in name and birth date but will still consider the record a match on SSN. Single character 'typos' in name, and an error of birth year of plus/minus one year, for example, are tolerated.

Information obtained

Just as with the CMS data described above, output from the resource for 'not known to be deceased' records require that an imputed follow-up date be used, which is based on taking the actual date of the computer run, and backing up a number of months. The logic here is that there is a lag time between when a person disappears from any of the administrative databases (usually due to death), and when the file is updated. For deceased SSN's, a date of death is provided.

The CSP has for years taken a conservative approach, and backs up the date for vital status of 'alive' back approximately 6 months from when the computer job is run.

Cost

The service is currently (2003) available at a cost of \$0.17 per record (data supplied to identify one study subject) up to 25,000 records. Additional records will be processed at a cost of \$0.013 per record.

Contact information

<http://www.ssa.gov/policy/about/epidemiology.html>

Pat McFadden

410-965-5541

ORES.Epidemiological.Requests@ssa.gov

Social Security Administration
Office of Research, Evaluation, and Statistics
4-C-15 Operations Building
6401 Security Boulevard
Baltimore, MD 21235-6401
Fax: 410-966-4071

Resource

Social Security Administration Death Master File (SSA DMF)

Description

The DMF is a huge file (45 million records!), and is available for purchase in a variety of computer formats. This data is a record of 'lump sum', (aka death benefit) payment information for deceased individuals. This is NOT a vital status resource; unlike vital status mortality data, not everyone who dies makes a claim to the SSA for the death benefit payment.

Type Data purchase – license to use

Usage notes

The key reason for using this file is the ability to perform in house record linkage processing against your registry data files. This is especially useful when using a 'Probabilistic Record Linkage'* software, which yields maximum linkage results. Keep in mind – this exact file is used by both the SSEVSD and the CMS resources mentioned above in the linkages they perform.

The DMF is obtained on a 'snapshot' type basis – there is no such thing as an 'annual' file, as reports for death benefits for any given calendar year can come in at any time. However, on any particular date, all reports that have been submitted are available (roughly two weeks after the report) – making this a timely alternative to using true vital statistic files, which usually involve inherent lag time between year of death and date the file is available.

Information obtained

From the SSA DMF website: This file includes the following information on each decedent, if the data are available to the SSA: social security number, name, date of birth, date of death, state or country of residence (2/88 and prior), ZIP code of last residence, and ZIP code of lump sum payment. The SSA does not have a death record for all persons; therefore, SSA does not guarantee the veracity of the file. Thus, the absence of a particular person is not proof this person is alive.

Cost

The DMF file is available in two ways – one is the whole file in its entirety, and is roughly \$1,750. Another option is to purchase a subscription that lets you access monthly (or even weekly) updates – even via the Internet if desired, for an annual cost of approximately \$2600.

Contact information

<http://www.ntis.gov/products/pages/ssa-death-master.asp>

US Department of Commerce
Technology Administration
National Technical Information Service
Springfield, Virginia 22161

(800) 363-2068
(703) 605-6000
(703) 605-6060

Resource

National Death Index (NDI)

Description

The NDI was started by Robert Bilgrad at the National Centers for Health Statistics (NCHS) over 20 years ago. It is a repository that aggregates vital statistic data from all 50 US States. Service is provided to epidemiologists and health researchers, and uses a 'Deterministic Record Linkage'* algorithm to link registry data against vital statistic data.

Type Cost recovery based. Service provided by NDI. Registry sends data to NDI.

Usage notes

This resource is typically used in two different ways – either to verify reported dates and causes of death for known decedents, or to submit all 'not known to be deceased' patients for searching. Due to the high cost of the resource, and the fact that a simplistic matching algorithm is used but the end user is not provided with any demographic information on matches to use in verification of potential matches, many registries prefer to use NDI for the first purpose – to augment data files with cause of death for cases where the information is missing. Accurate and complete linkage with local State vital statistic data can minimize reliance on this resource.

Information obtained

Output from the linkage process with NDI consists of a coding scheme that gives feedback on which variables matched in the linkage. Without the demographic information returned on the matches, human clerical review is not possible, and thus requires that a conservative approach to accepting matches be followed. Details on the scoring scheme is available on the NDI website.

Data items returned on matching records include date of death, place of death (state), and cause of death (which is an option included in the NDI Plus product)

Cost

Cost is based on the number of years searched – therefore for cases with known dates (or at least years) of death, a much reduced fee schedule is applicable.

From the NDI website, the following is an example of searching costs for cases NOT known to be deceased: 1,000 records searched against 10 years would cost \$350 + (\$0.30 x 1,000 x 10) or \$3,350. For sizeable registries, this kind of search becomes extremely expensive. A possible approach would be to submit only the lost to follow up group at the end of a follow-up year – this number may be manageable, cost wise.

Contact information

NATIONAL DEATH INDEX
National Center for Health Statistics
3311 Toledo Road, Room 7318
Hyattsville, Maryland 20782
301-458-4444

ndi@cdc.gov

website:

<http://www.cdc.gov/nchs/r&d/ndi/ndi.htm>

Resource

State Vital Statistics

Description

Vital statistic data is the gold standard for obtaining mortality status. Although there is usually a fairly long lag time between the year of death and the issuance date for the data files, the quality control and comprehensive nature of this resource makes it very useful for registries.

Type Governmental based data files for purchase OR Service provided by State

Usage notes

Vital statistic data originates at the county level, where it is collected and sent to a state vital statistic department. Cause of death coding is done within the state, but every record is sent to the NCHS in Hyattsville Maryland for review and correction, thus insuring consistency in cause of death information across all states. This review process is one of the reasons the final death file is issued so long after the end of the incidence year. For detailed information used in coding and classification used in all states, see <http://www.cdc.gov/nchs/about/major/dvs/im.htm>

When requesting statewide vital statistics data, be sure to request a file that contains the ‘out of state identifiers’. This file includes information on state residents who die out of state – which is an

important (albeit very small) group to include in your linkages. Often this information requires additional confidentiality and data usage agreements.

As with DMF, buying the data allows use of sophisticated record linkage software in house to obtain maximum accuracy of results.

Information obtained

Output will vary between states, but usual data elements include name, date & place & cause of death, SSN, race, place of birth and sex. Unique identifiers are provided to facilitate ordering of hard copy deaths certificate, which sometimes can be useful for special study follow-up.

Cost

Purchase: Varies by state. Typically low cost, if not provided free of charge to agencies such as population based cancer registries.

Service: Also varies by state.

Contact information

Varies by state.

Resource

Department of Motor Vehicles (DMV)

Description

This resource is referred to by different names, but driver's license information can be very helpful for follow-up. However, access to this data source may be limited in some states, and is often more accessible when the population based registry operates within the jurisdiction of a state or county governmental agency, such as a health department.

Type Service provided by State

Usage notes

Driver's record information (and optionally senior citizen ID card information in some states) can be utilized with two different approaches. Listing of a patient in a DMV computer system on any given date can be inferred to be a valid follow-up date. This liberal approach requires accepting that in the absence of any conflicting information, the valid drivers license record signifies that an individual is alive, and living in the state. The major limitation to this approach is the fact that drivers licenses are

renewed on an infrequent basis (every 5 years here in California), so it is possible for a person to move out of state, (and possibly die), without any notification to the DMV. A more conservative approach, which we use at the CSP, is to use actual dates of activity (i.e. license renewals or changes) as valid dates for follow-up.

Linkage methodologies performed by the states is (usually) strict and simple. In California, for example, an exact match on date of birth plus name (first, last, and middle initial) is required. When two records exist on the California database which are EXACT duplicates, no record is returned via the batch oriented, linkage approach. Manual DMV searches (using paper forms that also include last known addresses) can be submitted in this situation to resolve the duplicates.

Information obtained

Updated address and name information is routinely returned by DMV links.

Cost

Purchase: Varies by state. Typically low cost, if not provided free of charge to agencies such as population based cancer registries.

Service: Also varies by state.

Contact information

Varies by state.

Resource

Voter Registration Files

Description

Voter registration data is useful for registry-based follow-up. The term 'registration' doesn't tell the whole story when using this resource. Most of the information gained from using voter registration data is obtained from individual's record of actually voting in a given election. Hence, success of utilizing this resource is tied to voter turnout in a particular election – years with national presidential elections yield the best results for this linkage.

Type Governmental based data files

Usage notes

Since this resource allows for in house record linkage at the registry, manual reviewing of possible matches can increase effective yield. Another advantage from using voter registration data is the inclusion of address and phone number - data items not often available for linkage.

Information obtained

Updated address and name information is routinely returned by DMV links.

Cost

Purchase: Varies by state. Typically low cost, if not provided free of charge to agencies such as population based cancer registries

Contact information

Varies by state.