

SEER February 2013 Data Submission Requirements and Guidelines

Please direct all submission questions to:
SEERSUBM-L@list.nih.gov

1. The data submission is due **Thursday, February 28, 2013**. The data submission must include all resident cases through 2011 diagnoses.
2. **All submission files must be uploaded to the SEER Submission Reports Portal.** A February 2013 folder has been created under each registry's folder. Please add the submission files to this folder.
3. The file must be submitted in NAACCR 12.2 or 13.0 formats and should be named **rr.feb13.txd.gz** where rr represents the two character abbreviation for your registry. A list of the variables to be included in the submission can be downloaded from <http://seer.cancer.gov/tools/seer.feb13.dataitems.pdf>.
4. The file must be sorted by the following variables: Registry ID (NAACCR Item #40), Patient ID Number (NAACCR Item #20), and Sequence Number Central (NAACCR Item #380). For the Los Angeles, Greater Bay Area and Louisiana registries, the legacy ID should be put in columns 50-57.
5. The file must be edited prior to submission using the SEER*Edits program, version 6.5. The SEER*Edits software is available on the SEER Submission Reports Portal.
6. **NHAPIIA:** Use the NHAPIIA algorithms to update the API indicator (NAPIIA, NAACCR Item #193) and the Hispanic indicator (NHIA, NAACCR Item #191). These fields may be calculated by the registry's data management system or in a post-processing step. The algorithms and the NHAPIIA SAS program are available on the NAACCR Call for Data website (<http://www.naacr.org/DataandPublications/CallforData.aspx>). Every registry must run the algorithm. Please note to set the parameters needed for your registry before running NHAPIIA.
7. **Collaborative Staging:** Please run your 2004-2011 data through the CS algorithm before submission. CS Version Derived [NAACCR item # 2936] must be CS version 02.04.
8. **VA case counts:** To assess underreporting of Veterans Affairs Hospital (VA) records, specific data are being requested. The VA tumor case counts are needed for all malignant cancers combined (males only), by race, age, and cancer site. Please restrict case counts to SEER coverage areas. These data are needed for diagnosis years **2000-2011** for the current submission only.

For the February submission file, we request the following steps:

- a. Create a temporary NAACCR-format data file containing a VA indicator flag in NAACCR item 2220, Column 2468. The VA indicator may be created using facility codes to distinguish three categories of tumor records: 0) records from non-VA facilities; 1) records from a VA facility only; and 2) records from VA facility and other facilities.
- b. Use the temporary NAACCR-format data file as input to the SEER Edits program and generate a VA report. Check the option to create a modified version of the input file for the SEER Submission. Select the option to "Set the VA flag to blank." This copy is the

data file that you will submit to SEER. Use the file naming conventions defined in step 3 above.

- c. On the SEER*Edits Results tab:
 1. Export the VA Report to CSV. SEER*Edits will create a gzipped CSV file. Post this file to the SEER Submissions Portal.
 2. On the SEER*Edits Results tab, go to the tab labeled “Modified File for SEER”. Submit the modified file to SEER via the SEER Submissions Portal. Changes made to the file are documented in the Summary section of the Results Tab.
9. **In-House Record Counts:** Use the Word version of Attachment B to provide the following information to NCI SEER:
- a. Please submit count of 2011 records that are in-house but not yet processed (e.g., abstracts, path reports, death certificates, etc, for reportable cancers). If possible, please provide counts broken down by record types.
 - b. If possible, provide estimate of likely number of records that will result in 2011 cases. These are records that have not been processed but indicate a new case. The overall count is needed.
10. **Survival Calculation:** SEER will be requiring the submission of additional data in order to calculate more precise survival.
- This calculation uses the day field of the relevant dates which is not something SEER has done previously.
 - The goal is to be as consistent as possible with NAACCR and European survival calculations.
 - Specific instructions about the options are contained in Attachment C. Please direct questions to Missy Jamison at missy.jamison@nih.gov.

Attachment A
SEER Patient Follow-up Calculation
for February 2013 submission only

Note 1: The calendar years of diagnosis used for the follow-up calculation with this submission is as follows.

The last year of data being submitted is 2011. The percent of patients diagnosed during the years prior to 2011 who have current follow-up is defined as

$$P = 100(D + A)/T$$

where D is the number who died prior to January 1, 2012, A is the number with follow-up dates on or after January 1, 2012 (includes both alive and dead patients), and T is equal to A + D + the number of patients who were last known to be alive with follow-up dates prior to January 1, 2012. P can be calculated for individual years of diagnosis up through 2010 and for all years combined prior to 2011.

For all invasive cancers and calendar years specified by NCI, P shall be at least 95 percent, but must not be below 90 percent.

For patients ages 20-64, and under age 20, P shall be at least 90 percent in each case, but must not be below 80 percent.

For all in situ cancers (excluding cervix in situ) and calendar years specified by NCI, P shall be at least 90 percent, but must not be below 80 percent.

Note 2: Follow-up is expected on benign and borderline central nervous system tumor cases diagnosed 1/1/2004 and forward.

Attachment B
February 2013 Submission to NCI
Please complete for each submission and e-mail to:
SEERSUBM-L@list.nih.gov

Registry Name:
Registry ID number:
Date of file transfer to SEER:
Data file name:
Years of diagnosis submitted:
Number of cases submitted:
Number of 2011 records in-house but not submitted:
Estimate of likely 2011 cases:

NHAPIIA Options: Registries may select one of the options listed below when executing the NHIA algorithm. The option setting affects counties in which less than 5% of the population is of Hispanic/Latino ethnicity or to include all records.

Which option did you use when running NHAPIIA?

1. Option 1 – Only run the surname portion of the algorithm only on cases reported as Spanish surname only or as unknown whether Spanish (item 190 codes 7 or 9). In this choice, the surname portion will not be run on cases coded as 0, non-Hispanic.
2. Option 2 - Run the surname portion of the algorithm only on cases with a code of 7 on data element 190 (to verify that the surname is on the list of allowable Hispanic surnames) AND convert all cases with a code of 9 (unknown if Hispanic) to a code of 0 (Not Hispanic).
3. All Records – Apply all NHIA algorithms to all records. Note: this is listed as NHAPPIA option 0 in the SEER*DMS header information.

Attachment B (continued)
February 2013 Submission to NCI
Please complete for each submission and e-mail to:
SEERSUBM-L@list.nih.gov

Technical Contact

Name:

Telephone number:

e-mail address:

Registry Manager Contact

Name:

Telephone number:

e-mail address:

Known data problems including reasons:

Comments:

Attachment C – Survival Calculation
February 2013 Submission to NCI
Please complete for each submission and e-mail to:
SEERSUBM-L@list.nih.gov

There are two options for meeting the data submission requirement to enable more precise survival calculations. Additional information is available at: <http://seer.cancer.gov/survivaltime/>. Please contact Missy Jamison at Missy.Jamison@nih.gov with any questions.

Method 1:

- Use SEER*Edits or SAS to calculate the Months Survival (based on Complete Dates) fields. A SAS program is available at the link above.
- SEER*Edits and the SAS program use complete Date of Diagnosis and Date of Last Contact, including day, and create four fields. The SEER*Edits software allows you to mask the day of diagnosis and day of last contact, if your registry does not wish to include them in the submission file.
- The advantage of this method is that it enables you to retain day of diagnosis and day of last contact in your registry.

Method 2:

- Provide day of diagnosis and day of last contact in your data submission file.
- SEER had not asked for day information prior to November 2012; and this method is optional.
- The day information will be held confidentially and will only be used for survival calculations.
- This is the preferred method because it will enable SEER to evaluate multiple imputation techniques for missing dates and methods that require more continuous survival data such as flexible parametric survival methods. This approach will also harmonize SEER with other survival efforts such as CONCORD. Dr. Mariotto is planning to form a working group with interested SEER registries to assess these new methods.

Additional considerations:

- Regardless of the option chosen, please do not attempt to recode any missing date fields for Date of Diagnosis or Date of Last Contact. One purpose of the algorithm is to standardize specification of missing month and day fields.
- These data will in no way effect the Data Quality Profile.
- Choose the option that is best for your registry.

Please answer the following questions:

1. Our registry will use:

___ Method 1

___ Method 2

2. Date (month and year) of most recent National Death Index Linkage: _____