SEER November 2012 Data Submission Requirements and Guidelines

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

1. The data submission is due **Thursday, November 1, 2012**. 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 November 2012 folder has been created under each registry's folder. Please add the submission files to this folder.

3. The files must be submitted in NAACCR 12.2 format and should be named **rr.nov12.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.nov12.dataitems.pdf .

4. The files must be sorted by the following variables: Registry ID (NAACCR Item #40), Patient ID Number (NAACCR Item #20), 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. All cases diagnosed through 2010 must be edited prior to submission using the SEER*EDITS program, version 6.4, available on the SEER Submission Reports Portal. Use Attachment B in WORD to notify NCI SEER that the file has been submitted and whether or not SEER is to submit your data for the NAACCR Call for Data.

6. All available 2011 diagnoses must be submitted. It is not a requirement, however, that these cases be edited prior to submission. These cases will be used for in-house research only

7. **IHS indicator (NAACCR Item #192) and Record Linkage:** SEER requires the IHS indicator. Please submit the current value that you have from the November 2011 IHS linkage. We will update your file with the new IHS linkage information prior to the December submission to NAACCR. The IHS indicator should have:
   - i. 0 if sent for linkage and no match
   - ii. 1 if sent for linkage and matched
   - iii. blank if the case wasn't sent for linkage

   a. The IHS indicator will be 'required' for diagnosis years 01/01/1988 - **12/13/2010** but will be non-blank only when/if linkages were done with the records from the Indian Health Service.

   b. All SEER registries, with the exceptions of the Alaska Native Tumor Registry and the Cherokee Nation Cancer Registry, need to submit a file of SEER-reportable cases diagnosed between 01/01/88 and **12/31/2010** to IHS. The cases in this file must be identical to the 01/01/88 and **12/31/2010** data that are included in the November 1, 2012 submission file to NCI. Specific instructions for preparing and submitting this file are contained in Attachment C and the file structure is described in Table 1.

   c. Results from the IHS linkage will be returned to each registry and to the NCI (only case identification number and linkage results will be provided to NCI; patient-identifying information such as name and social security number will be sent only to the IHS for linkage and will not be sent to NCI).

    d. NCI will be responsible for updating the November submission files from the registries with the new IHS linkage results and will also supply the linkage results to NAACCR for those registries for which NCI has performed this submission in the past.

    e. Registries that submit their own data to NAACCR must update their files with the IHS linkage results since the IHS Link variable [NAACCR #192] is included as part of the NAACCR call for data.

    f. Please direct questions regarding the SEER-IHS linkage project to Ms. Melissa Jim at melissa.jim@ihs.gov (telephone: 505/248-4451) or Dr. Chuck Wiggins at cwiggins@salud.unm.edu (telephone: 505/272-3127). Please also cc SEERSUBM-L@list.nih.gov on all questions.

8. **NHAPIIA:** Use the NHAPIIA algorithms to update both 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. Every registry must run the algorithm. Please note to set the parameters needed for your registry before running NHAPIIA. Use Attachment B for further instructions.

9. **Summary stage**: If NCI will be submitting your data for the NAACCR call for data, the Summary Stage 77 and Summary Stage 2000 data must be submitted. Either field can be directly coded or recoded via the Summary Stage recode programs for Summary Stage 77 and Summary Stage 2000. Summary Stage 1977 must be on your file for all cases through 2000 and Summary Stage 2000 for all cases 2001-2003.

10. **Beale codes**: County rural-urban continuum codes (Beale codes) are not a requirement for the NAACCR Call for Data.

11. **Census Tract Poverty**: If NCI will be submitting your data for the NAACCR call for data and you wish to include the Census Tract Poverty Category (Item 2220), please include the data item in the submission file. The value may be calculated by your data management system or you may run the NAACCR SAS program to calculate the value. See http://www.naaccr.org/Research/DataAnalysisTools.aspx. Please direct questions to Missy Jamison at missy.jamison@nih.gov.

12. **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.

13. **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 **2000-2010** for the current submission only.

For the November 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; 2) records from a VA facility & 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 copy of the file without the VA flag. This copy is the data file that you will submit to SEER. Use the file naming conventions defined in steps 3.

c. On the SEER*Edits Results tab, export the VA Report to CSV. SEER*Edits will create a gzipped CSV file. Post this file to the SEER Submissions Reports Portal.

14. **Survival Calculation:** This is the first year that 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 D. Please direct questions to Missy Jamison at missy.jamison@nih.gov.

Attachment A
SEER Patient Follow-up Calculation
for November 2012 submission only

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

The percent of patients diagnosed during the years prior to 2010 who have current follow-up is defined as

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

where D is the number who died prior to January 1, 2011, A is the number with follow-up dates on or after January 1, 2011 (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, 2010. P can be calculated for individual years of diagnosis up through 2009 and for all years combined prior to 2010.

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
November 2012 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:

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 option 0 in SEER*DMS configuration settings.

Attachment B (continued)
November 2011 Submission to NCI
Please complete for each submission and e-mail to:
SEERSUBM-L@list.nih.gov

NAACCR Call for Data Dec 2012:

Do you want SEER to submit your data to NAACCR? YES or NO

IF YES:
1. What diagnosis years do you want submitted?

2. SEER Summary Stage 77(SS77):
    a. RECODED for what diagnosis years? _____
    b. DIRECTLY CODED for what diagnosis years? _____

3. SEER Summary Stage 2000:
    a. RECODED for what diagnosis years? _____
    b. DIRECTLY CODED for what diagnosis years? _____

4. NAACCR has requested certain items. Unless directed otherwise, we will send all NAACCR required data items, http://www.naaccr.org/DataandPublications/CallforData.aspx. Please indicate if there are any items you do <u>not</u> wish to include:
    _____
    _____

Attachment B (continued)
November 2012 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:

SEER November 2012 Data Submission Requirements and Guidelines

Attachment C
IHS linkage Instructions

Record selection criterion:

1. SEER-reportable cases;

2. Diagnosed between 01/01/1988 and **12/13/2010** (inclusive);

3. Cases in this file should be the _same_ as those that will be submitted in the November **2012 file** (i.e., no more, no less) to the SEER Program. In previous years, we experienced some confusion when records that were submitted to the SEER Program had not been sent for linkage with IHS.

File structure:

1. ASCII format;

2. File structure as described in Table 1 (below);

3. Encrypted/password protected file (password to be sent separately via e-mail or direct telephone contact).

File can be delivered in one of the following 2 methods:

1. Submit your files electronically via the New Mexico Tumor Registry's sFTP web-site according to the following steps:

   a. Send an e-mail message to the following address: nmtrhelp@nmtr.unm.edu

   b. Specify that you are requesting a folder on our secure sFTP web-site for the SEER-IHS linkage. Be sure to let NMTR Help Desk know which registry you are representing.

   c. NMTR Help Desk will receive an e-mail message that contains your assigned user-name.

   d. Respond to the NMTR Help Desk e-mail message; they will send you a password via a separate e-mail message. This second message will include instructions for accessing our sFTP site.

   e. Please feel free to contact Chuck Wiggins at (505) 272-3127 (direct) if you have questions.

2. Ship according to a pre-arranged schedule via express courier service to:
   IHS Division of Epidemiology and Disease Prevention
   c/o Melissa Jim
   5300 Homestead NE
   Albuquerque, NM 87110

Contacts:

Melissa Jim, M.P.H.
Epidemiologist, CDC Division of Cancer Prevention and Control
c/o Indian Health Service Division of Epidemiology and Disease Prevention
505-248-4451
melissa.jim@ihs.gov

Chuck Wiggins, Ph.D.
Director, New Mexico Tumor Registry
505-272-3127
cwiggins@salud.unm.edu

Attachment C (continued)
Table 1. Data Layout for IHS Linkage

Version 12.2		Name
Item #	Column #	
10	1-1	Record Type
20	42-49	Patient ID Number
50	17-19	NAACCR Record Version
40	30-39	Registry ID
70	95-144	Addr at DX--City
80	145-146	Addr at DX--State
100	147-155	Addr at DX--Postal Code
90	156-158	County at DX
160	177-178	Race 1
161	179-180	Race 2
162	181-182	Race 3
163	183-184	Race 4
164	185-186	Race 5
190	189-189	Spanish/Hispanic Origin
220	192-192	Sex
230	193-195	Age at Diagnosis
240	196-203	Date of Birth
241	204-205	Date of Birth Flag
192	421-421	Result of previous IHS linkage
380	528-529	Sequence Number--Cntrl
390	530-537	Date of Diagnosis
391	538-539	Date of Diagnosis Flag
1750	2116-2123	Date of Last Contact
1751	2124-2125	Date of Last Contact Flag
1760	2126-2126	Vital Status
1810	2131-2180	Addr Current--City
1820	2181-2182	Addr Current--State
1830	2183-2191	Addr Curr--Postal Code
1840	2192-2194	County--Current
2230	3340-3379	Name--Last
2240	3380-3419	Name--First
2250	3420-3459	Name--Middle
2280	3466-3505	Name--Alias
2390	3506-3545	Name--Maiden
2290	3546-3605	Name--Spouse/Parent
2320	3619-3627	Social Security Number
2330	3628-3687	Addr at DX--No & Street
2335	3688-3747	Addr at DX--Supplement
2350	3748-3807	Addr Curr--No & Street
2355	3808-3867	Addr Curr--Supplement

SEER November 2012 Data Submission Requirements and Guidelines

Attachment D – Survival Data
November 2012 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:
- Run the SAS program available at the link above.
- The SAS program uses complete Date of Diagnosis and Date of Last Contact, including day, and creates four fields.
- 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 has never asked for day information previously 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: _____