

# General Instructions for NAACCR Data File Submissions

NAACCR Call for Data Deadline: Midnight Eastern Time December 3, 2018

**For questions about the NAACCR Call for Data, contact:**

Recinda Sherman, MPH PhD CTR  
NAACCR, Inc.  
Phone: (217) 698-0800 ext. 6  
Fax: (217) 698-0188  
[rsherman@naaccr.org](mailto:rsherman@naaccr.org)

**For technical questions about sending the data file, contact:**

Rick Firth  
Information Management Services, Inc. (IMS)  
6110 Executive Blvd. Suite 310  
Rockville, MD 20852  
Phone: (301) 628-1282  
[firthr@imsweb.com](mailto:firthr@imsweb.com)

**Location of all Call for Data Materials:**

To find the forms and instructions, go to <https://www.naaccr.org/call-for-data/> .

**Submission for CiNA and Certification**

Registries submitting data only for Registry Certification, you may skip the steps 3, 6 and 11 described below.

- 1. Create a single data file with all reportable cancer incidence cases for residents of your coverage area, who were diagnosed from 1995 through 2016. In addition, 2017 data are also being requested to be used for evaluation purposes only and does not impact certification or inclusion in CiNA products.**

The list of requested data items for all uses and for a submission for Registry Certification only is available in the document *Certification and CiNA Call for Data Items*.

- **Submission files**

- U.S. registries submitting data to both NPCR and NAACCR may create a single submission file containing all data items requested for both CDC-NPCR and NAACCR for steps 2-7.
- 2017 data may be included in the submission file or submitted as a separate file as late as January 31, 2019. ***We strongly encourage members to submit 12 month data as NAACCR is currently assessing the utility of included these data into the delay adjustment model.***
- U.S. registries *not* submitting to NPCR should create a file including the data items listed in the document titled *Certification and CiNA Call for Data Items*.
- All files should be run through NAACCRprep prior to submission (see step 6).
  - The data items highlighted in green in the document *Certification and CiNA Call for Data Items* are necessary to generate poverty indicator variables and NAPIIA and NHIA variable. The data items highlighted in green will not be included in the NAACCRprep

output file and should not be submitted to NAACCR for certification or CiNA.

- **Data submitted must adhere to the NAACCR standard definitions and codes.**
  - All cases as defined in the document *Data Selection Criteria* for each diagnostic year should be included on the file, regardless of whether they are reportable in only some of the years.
- **Data must be submitted in NAACCR Record Layout Version 16. Refer to the Standards for Cancer Registries Volume II, version 16 for details.**
  - *If your registry is unable to submit a file in version 16 please contact Recinda Sherman [rsherman@naaccr.org](mailto:rsherman@naaccr.org).*
- **In preparation for the upcoming XML standard, please submit your file in an .xml format.**
  - *If you are unable to submit your file in an .xml format, please contact Recinda Sherman at [rsherman@naaccr.org](mailto:rsherman@naaccr.org) or 217 698 0800 x 6.*

**2. Perform the Duplicate Assessment Protocol on 2016 cases for Registry Certification and on 1995 through 2016 cases for CiNA. A third duplicate protocol should be performed on 2017 cases.**

The duplicate protocol is required for certification and for inclusion in any use requiring high quality data assessments. Instructions for performing this step are included in the document *Duplicate Protocol Combined Years Call for Data*.

**IMPORTANT:** If you are submitting for both CiNA and Registry Certification, you must conduct the duplicate protocol two times: once, on the 1995-2016 cases and the second time, on 2016 cases only. If you are not submitting all years, 1995-2016, for CiNA, perform the protocol on ALL the data years you are submitting. We recommend a third duplicate protocol be performed on 2017 cases (12 month data) in order for appropriate evaluation of the data to take place.

**3. U.S. registries should run the NAACCR Hispanic and Asian/Pacific Islander Identification Algorithm on all cases submitted.**

The NAACCR Hispanic Identification Algorithm and the NAACCR Asian Pacific Islander Identification Algorithm are combined into one SAS program, called NHAPIIA, for your convenience. Instructions for applying these algorithms are available on the NAACCR call for data [website](#).

The NAACCRprep utility can generate the NHAPIIA fields, avoiding the need for this separate step. Just include the NAACCR items 2230, 2240 and 2290 (first, middle and last name) in the NAACCRprep input file. These fields will be used by the algorithm then deleted from the NAACCRprep output file.

**If you elect to run NHIA/NAPIIA/NHAPIIA outside of NAACCRPrep, it is recommended that registries use Option 1, the current default for the SAS routine and in NAACCRPrep. This setting limits the Spanish surname portion of the algorithm in counties that are less than 5% Hispanic to cases coded as surname only (item 190=7) or unknown whether Hispanic (item 190=9).** Registries selecting some other option are asked to note which option they selected and to provide a rationale for using another option. Questions asking which option you choose will be included in the electronic submission pages (see instruction #10

below). If your registry software incorporates NHIA/NAPIIA, please specify the version on the Verification Form.

**4. Run the edit set appropriate for your submission and the NAACCR format being submitted.**

Select the intra-record edit set that matches your submission. Only one error free intra-record edit run is necessary. In particular, if submitting to NPCR and NAACCR, run the Combined NAACCR NPCR edits against a file with all of the data items requested by NAACCR and NPCR. Once all edits have been corrected, use NAACCRprep to create both the NPCR and NAACCR submission files. It is not required to run the NAACCR or NPCR submission files generated by NAACCRprep through edits. However, if a registry opts to run their submission file through edits prior to their final submission, they should use the NAACCR CiNA Edit set for the NAACCR submission and one of the Combined NAACCR/NPCR edits sets on their NPCR Submission file.

<b>Type of Submission/Type Registry</b>	<b>Edit Set</b>
Certification only/All 2016	NAACCR Certification Edits
Certification & CiNA/U.S. <i>not</i> Submitting to NPCR 1995-2016	NAACCR CiNA Edits 1995-2016
NAACCR 12 Month Data 2017	NAACCR CiNA Edits 2017
NPCR Submission Only 1995-2106	NPCR-CSS-1995-2016
NPCR 12 Month Data 2107	NPCR-CSS-2017
NPCR Registry submitting for certification, CiNA, and NPCR 1995-2016	Combined NAACCR-NPCR -1995-2016
Certification & CiNA/Canadian	NAACCR CiNA Edits-Canada

The EDITS metafiles may be downloaded from the NAACCR Call for Data website. All edit errors on data items reviewed for certification or included in CINA must be reconciled prior to submission. The edits are based on the version 16 EDITS metafile releases.

The edit sets use NAACCR standard definitions and are based on submission of all requested data items. **WARNING!** If you choose not to submit a data item, you may get errors because of blank fields. If you like, you can turn these edits off. Be aware that we run EDITS on your submitted files and use our results to determine Registry Certification status and fitness of the data for various uses. If the specific use of the file requires a variable to be present, then the variable must be 100% error-free. Note that the 2017 data is for evaluation purposes only. The 2017 cases should be as cleaned up as is practical but do not need to be error-free.

Submission of an edit summary report is not required.

**5. Run the Inter-record EDITS Utility.**

This application must be run after running the intra-record edits described above, but before the file is run through NAACCRprep. Please resolve all errors.

**6. Generate Derived Variables using the NAACCRprep utility.**

The NAACCRprep program can be downloaded to run against the processed data file to generate survival variables (all registries) and, for U.S. registries, derived Rural Urban Continuum Codes (Beale codes, RUC, and RUCA), Census Tract Poverty Indicator variable, and the NHAPIIA variables. The program can be instructed to strip specific fields from your submission file after these derived fields are developed. The variables

that can be stripped during processing include census tract codes, name fields, NPCR variables (BCCEDP MDE Link), day of diagnosis and day of last follow up. If your state requires, you can also strip residence county from the NAACCR file.

- *The SAS program developed for NAACCRprep is available for download. The processed file may also be run through SEER\*Edits in order to generate the data items. In addition, your central registry software vendor may include the algorithm in your call for data programs.*
- For those registries using NAACCRprep:
  - The file should be free of edit errors prior to running the NAACCRprep utility (see steps 4 &5).
    - In order for NAACCRprep to generate length of survival (all registries) and, for U.S. registries, derive Rural Urban Continuum Codes (Beale codes, RUC, and RUCA), Census Tract Poverty Indicator variables, the file will need to include the data items **Census Tract 2000 and 2010 variables** [NAACCR items 130 and 135]. Census Tract 2000 and 2010 variables **will not be included** in the output file submitted to NAACCR.
    - Note: Registries again have the option to add non-required, tract-level variables this year. The full list is included with the Call for Data materials. The purpose of the collection of these data are 1) to establish and evaluate new codes for poverty that limit the impact of residual confounding when conducting analysis by race/ethnicity; 2) to create an SES index (Yost index) from multiple census variables that will allow additional assessment by area-based SES without identifying census tract; 3) assess the feasibility and utility of additional tract-level variables for CiNA research. The variables **will NOT be released** to researchers. These data are collected for NAACCR quality assessment and evaluation only. If you have any questions about area-based variables, please contact Recinda Sherman.
    - In order for NAACCRPrep to generate the NAPIIA and NHIA variables, the input file must contain the following fields:
      - i. Name--Last (Data Item 2230)
      - ii. Name--First (data Item 2240)
      - iii. Name—Maiden (Data Item 2390)
    - These fields will not be exported to the NAACCR or NPCR submission files.
    - Canadian registries should create a file including the applicable data items listed in the document titled Certification and CiNA Call for Data Items. This file should be run through the NAACCRprep utility prior to submission (see step 6).
    - The output file generated by NAACCRprep does not need to be run against the GenEDITS file prior to submission. However, if a registry elects to run edits against the NAACCRprep output file, the NAACCR CiNA Edit set should be used on the file being submitted to NAACCR and one of the Combined NAACCR/NPCR sets should be run on the file being submitted to CDC-NPCR. Do not run the Combined NAACCR NPCR Edits set on the output file for CINA/Certification.

## 7. Create Your Data File.

Create **one** data file for your NAACCR submission. For example, if submitting 1995 through 2016 data, the file should be labeled using the naming convention of  
**ss9516v16.dat**

where ss is the two letter postal code for your registry followed by the years of data submitted, followed by the NAACCR version layout. For example, Montana submitting data in NAACCR version 16 for the years 1995-2016 would create a file named mt9516v16.dat. The file should have a record length of 3,339 characters.

## **8. Compress Your Data File**

Compress the data file with WinZip, PKZip, or Gzip. Gzip is the preferred method for compression. Gzip executables for Microsoft Windows and other operating systems are available at <http://www.gzip.org/#exe>.

## **9. Transmit Your Data File**

NAACCR has continued to make developments on the site for transmission of your CiNA and/or Certification file. A link to the submission will be available November 1, 2018. The link will be located on the NAACCR Call for Data webpage <https://www.naacr.org/call-for-data/>.

## **10. Please complete all of the electronic submission forms on the NAACCR website for your registry's data submission by December 3, 2018.**

NAACCR is developing a new site for the electronic submission forms. A link to the electronic submission forms will be available November 1, 2018 on the NAACCR Call for Data webpage.

## **11. Review, sign, and submit the NAACCR Call for Data Assurances Agreement.**

The agreement is available on the NAACCR website and should be uploaded or sent to Recinda Sherman by December 3, 2018. The link is located on the NAACCR Call for Data webpage <https://www.naacr.org/call-for-data/>.

## **12. Checklist of materials requiring hard copy submission *or* on-line approval.**

- **NAACCR, Inc. Call for Data Assurances Agreement**
- **CiNA Public Use Dataset 1—Active Consent**  
Please see *Appendix A. Description of CiNA Public Use Dataset* for additional information.
- **CiNA Public Use Dataset 2—Active Consent**  
Please see *Appendix A. Description of CiNA Public Use Dataset* for additional information.
- **ACS Facts and Figures [US Only] – Active Consent for Secondary Use**
- **Medullary Thyroid Cancer Verification- Active Consent for Secondary Use**
- **NAACCR Synthetic Dataset—Active Consent for Secondary Use**
- **American Lung Association—Passive Consent for Secondary Use**
- **CiNA Plus in SEER\*Stat—Passive Consent for Secondary Use**
- **Delay Adjustment—Passive Consent for Secondary Use**

## **13. CaRRI (Cancer Registry Researcher Information) Database**

After other Call for Data requirements are complete, please log into the Call For Data submission site and review and complete the registry information for the new CaRRI database December 1, 2018 and January 31, 2019. This database and data collection replaces the historic Registry Research Information Database (RRIDB) completed in previous Call for Data years. This system also collects additional data related to the procedures for outside researchers to get registry-level data for research.

## **14. Virtual Pooled Registry-Cancer Linkage System (VPR-CLS)**

After other Call for Data requirements are complete, please create your linkage file for the VPR-CLS between December 3, 2018 and January 31, 2019 (U.S. Registries only). VPR-CLS is a voluntary NAACCR initiative to streamline the process of research cohort linkage among multiple cancer registries as well as de-duplication among central cancer registries. Specific instructions for creating the VPR-CLS linkage file are included with

the Call for Data materials. If you have any questions, please contact Castine Clerkin  
[cclerkin@naaccr.org](mailto:cclerkin@naaccr.org).