### General Instructions for NAACCR Data File Submissions

**NAACCR Call for Data Deadline: Midnight Eastern Time Tuesday December 3, 2019**

**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)

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**

1. **Create a single data file with all reportable cancer incidence cases for residents of your coverage area, who were diagnosed from 1995 through 2017. *2018 data are also being requested to be used for evaluation purposes only******and do not impact certification or inclusion in CiNA products*.**

The list of requested data items for all submission types is available in an excel document: *Certification and CiNA Call for Data Items*.

* **Submission files**
  + 2018 data may be included in the submission file or submitted as a separate file as late as January 31, 2020. ***We encourage members to submit 12 month data as NAACCR is currently assessing the utility of including these data into the delay adjustment model and to get an assessment of current 2018 issues.***
  + For the NAACCR submission, registries should create a file including the data items listed in the document titled *Certification and CiNA Call for Data Items.*
    - *If a full, confidential record dataset is used, only the variables requested by NAACCR will be included in the submission file created by NAACCRPrep.*
    - If needed, SAS and ACCESS translation tools for NAACCR Record Layout are available here: <https://www.naaccr.org/analysis-and-data-improvement-tools/#TRANSLATION>
    - The data items highlighted in green in the document *Certification and CiNA Call for Data Items* must be included in the initial file. They are necessary to generate derived variables such the poverty indicator and NHAPIIA variables. But the data items highlighted in green will be removed by NAACCRPrep and will not be submitted to NAACCR for certification or CiNA.
  + Prior to submission, all files should be run through Edits first, followed by NAACCRprep (see steps below).
* **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 18, XML data exchange standard. Refer to the Standards for Cancer Registries Volume II, version 18 (**<https://www.naaccr.org/data-standards-data-dictionary/>) **and (**<https://www.naaccr.org/xml-data-exchange-standard/>**) for details.**
  + *If your registry is unable to submit a file in version 18, please use the conversion tools available from CDC:* <https://www.cdc.gov/cancer/npcr/tools/registryplus/up_download.htm>
  + *Contact Recinda Sherman if you have any issues at* [*rsherman@naaccr.org*](mailto:rsherman@naaccr.org)*.*

1. **Perform the Duplicate Assessment Protocol on 2017 cases for Registry Certification *and* on 1995 through 2017 cases for CiNA. A third duplicate protocol should be performed on 2018 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 CiNAand Registry Certification, you must conduct the duplicate protocol two times: once, on the 1995-2017 cases and the second time, on 2017 cases only. If you are unable to submitting all years, 1995-2017, perform the protocol on ALL the data years you are submitting. We recommend a third duplicate protocol be performed on 2018 cases (12 month data) in order for appropriate evaluation of the data to take place.

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

We recommend using the NAACCRPrep utility to generate the NHAPIIA fields.

The NAACCR Hispanic Identification Algorithm and the NAACCR Asian Pacific Islander Identification Algorithm are incorporated into NAACCRPrep. Ensure to include the NAACCR items 2230, 2240 and 2290 (first, middle and last name) in the NAACCRprep input file. These fields will be used to derive the variables and then deleted from the NAACCRprep output file.

This year registries also have the ability to use a SAS program but we will anticipate moving to a new format in subsequent years.

**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.

1. **Run the edit set appropriate for your Submission.**

Select the intra-record edit set that matches your submission. Only one error free intra-record edit run is necessary.

|  |  |
| --- | --- |
| **Type of Submission/Type Registry** | **Edit Set** |
| Certification only/All 2017 | NAACCR Certification Edits |
| Certification & CiNA/U.S. *not* Submitting to NPCR 1995-2017 | NAACCR CiNA Edits 1995-2017 |
| NAACCR 12 Month Data 2018 | NAACCR CiNA Edits 2018 |
| Certification & CiNA/Canadian | NAACCR CiNA Edits-Canada |

Please note, this year there is no combined NAACCR/NPCR edit set. All materials necessary for NPCR submission will be provided by NPCR. U.S. registries submitting data to both NPCR and NAACCR will need to use different edit sets for each agency.

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 18 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 submission type requires a variable, then the variable must be 100% error-free.

Note: The 2018 data is for evaluation purposes only. The 2018 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.

1. **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.

Note: The 2018 data is for evaluation purposes only. The 2018 cases should be as cleaned up as is practical but do not need to be error-free.

Once all edits have been corrected, use NAACCRPrep to create the NAACCR submission files. It is not required to re-run the NAACCR submission file generated by NAACCRprep through edits.

1. **Generate Derived Variables using the NAACCRPrep utility.**

The NAACCRPrep program generates the calculated survival variables (all registries) and derived Rural Urban Continuum Codes (Beale codes, RUC, and RUCA), Census Tract Poverty Indicator variable, the new poverty codes and the NHAPIIA variables (for U.S. Registries). The program automatically strips specific fields from your submission file after these derived fields are developed. The variables that are automatically stripped during processing include census tract codes, name fields, day of diagnosis and day of last follow up. If your state requires, you can also strip county of residence from the NAACCR file and cause of death.

* This year, registries also have the ability to use a SAS program to create all derived and calculated variables. Please note: these SAS codes will no longer be support after this year. We are working to provide a more efficient tool.
* 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 derive the tract-based area-based social measures like Census Tract Poverty Indicator variables (for U.S. registries), the file will need to include the data items **GeoLocationID 70/80/90, GeoLocationID 2000, GeoLocationID 2010, Census Tract 70/80/90,** **Census Tract 2000 and 2010 variables** [NAACCR items 351, 352, 353, 110, 130 and 135]. The geographic varaibles variables ***will*** ***not be included*** in the output file submitted to NAACCR.
  + Note: U.S. 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; and 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.. The YOST variables and additional poverty codes will be available, upon consent of each registry, to approved researchers. If you have any questions about area-based variables, please contact Recinda Sherman at rsherman@naaccr.org
  + In order for NAACCRPrep to generate the NAPIIA and NHIA variables, the input file must contain the following fields:
    1. Name--Last (Data Item 2230)
    2. Name--First (data Item 2240)
    3. Name—Maiden (Data Item 2390)

These fields will not be exported to the NAACCR submission files.

1. **Naming Your Data Output File**

Create ***one*** data file for your NAACCR submission. For example, if submitting 1995 through 2018 data, the output file should be labeled using the naming convention of

***ss9518v18.xml***

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 18 for the years 1995-2018 would create a file named mt9518v18.xml. If you are submitting 2018 cases later, use the same naming convention. Using the Montana example, the CiNA/Certification submission would by MT9517v18.xml and the 12 month data submission would be MT2015v18.xml. Note: NAACCRPrep will create a zip file for you with the proper extensions.

1. **If you do not use NAACCRPrep, Compress Your Data File**

NAACCRPrep outputs a zip file. If you choose not to use NAACCRPrep, you must 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>.

1. **Transmit Your Data File**

The submission site will be open on November 1, 2019. The link will be located on the NAACCR Call for Data webpage <https://www.naaccr.org/call-for-data/>.

1. **Please complete all of the electronic submission forms on the NAACCR website for your registry’s data submission by December 3, 2019.**

A link to the submission site will be available prior to November 1, 2019. The link will be located on the NAACCR Call for Data webpage <https://www.naaccr.org/call-for-data/>. Note: Registries have the option to e-sign or download, sign, and upload all submission documents.

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

The submission site will be available prior to November 1, 2019. The link will be located on the NAACCR Call for Data webpage <https://www.naaccr.org/call-for-data/>. Note: Registries have the option to e-sign or download, sign, and upload the DAA.

1. **Checklist of materials requiring signed approval**

* **NAACCR, Inc. Call for Data Assurances Agreement**
* **CiNA Public Use Dataset 1—Active Consent**
* **CiNA Public Use Dataset 2—Active Consent**
* **ACS Facts and Figures [US Only] – Active Consent for Secondary Use (US Only)**
* **Medullary Thyroid Cancer Verification- Active Consent for Secondary Use (US Only)**
* **NAACCR Synthetic Dataset—Active Consent for Secondary Use**
* **American Lung Association—Passive Consent for Secondary Use (US Only)**
* **CiNA Plus in SEER\*Stat—Passive Consent for Secondary Use**
* **Delay Adjustment—Passive Consent for Secondary Use**
* **The Affordable Care Act and Cancer Stage at Diagnosis—Passive Consent for Secondary Use (US Only)**

1. **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 by January 31, 2020. 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. If you have any questions, please contact Recinda Sherman [rsherman@naaccr.org](mailto:rsherman@naaccr.org).

Most registries input information last year. Those registries will only need to review and update if needed.

1. **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 by January 31, 2020 (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.

Note: 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).