



Virtual Pooled Registry (VPR) Project Instructions For Creating the Annual VPR-CLS Registry Linkage File for December 2023 Submission (as of 8/8/23)

The Virtual Pooled Registry Cancer Linkage System (VPR-CLS) facilitates streamlined linkages between research studies and participating U.S. registries. Each year as part of the NAACCR Call for Data, VPR registries create a single, standardized VPR-CLS registry linkage file. This file contains PII, is used for every VPR linkage that occurs during a calendar year and serves as a static file from which researcher-requested data items will be selected. **IMPORTANT NOTE: We request that all U.S. registries, regardless of VPR participation, create the annual VPR-CLS linkage file** to keep on file in case there are separate initiatives that utilize the VPR methodology in the upcoming year.

Timeframe for File Creation: Registries will prepare the VPR-CLS registry linkage file as part of the annual NAACCR Call for Data process. The VPR-CLS files should be created after 12-month data (cases diagnosed in 2022) have been submitted to NAACCR, NPCR and/or SEER. We request that the VPR-CLS file be created by December 31st, 2023, but it must be created **no later than January 31, 2024**. Producing the VPR-CLS file immediately following the Call for Data ensures that all registry files contain comparable timeframes for follow-up and include the most complete, high quality, edited data for linkage and data release. **NOTE: Unlike the Call for Data submission, the VPR-CLS registry linkage file contains patient identifiers and will remain at the cancer registry. This file is NOT to be uploaded to the CFD Portal.**

Case Selection Criteria: Cancer registries will extract NAACCR Version 23 Type C records for all state resident cases. It is essential that all Type C data items are populated, including patient identifiers. At a minimum, include cases from inception year to 2021 (24-month data); however, registries are encouraged to include cases from as far back as possible up through 2022 (12-month data). The extract should include all reportable cases routinely submitted in the NAACCR data submission (e.g., in situ, invasive, and benign brain/CNS-prior to 2004 if collected) and include VA cases, DCOs, and cases received through inter-state data exchange. **NOTE: Do not mask or blank out the day portion of date fields, as certain dates are essential for a high-quality linkage to occur. If registry policies do not allow release of full dates, this information can be excluded from the data that is transmitted to VPR studies.**

NOTE: Registries participating in the NAACCR Inter-Registry Data Exchange Agreement (Revised July 2021) allow the receiving registry to release inter-state-exchange-only cases for VPR studies unless otherwise stated in a registry addendum.

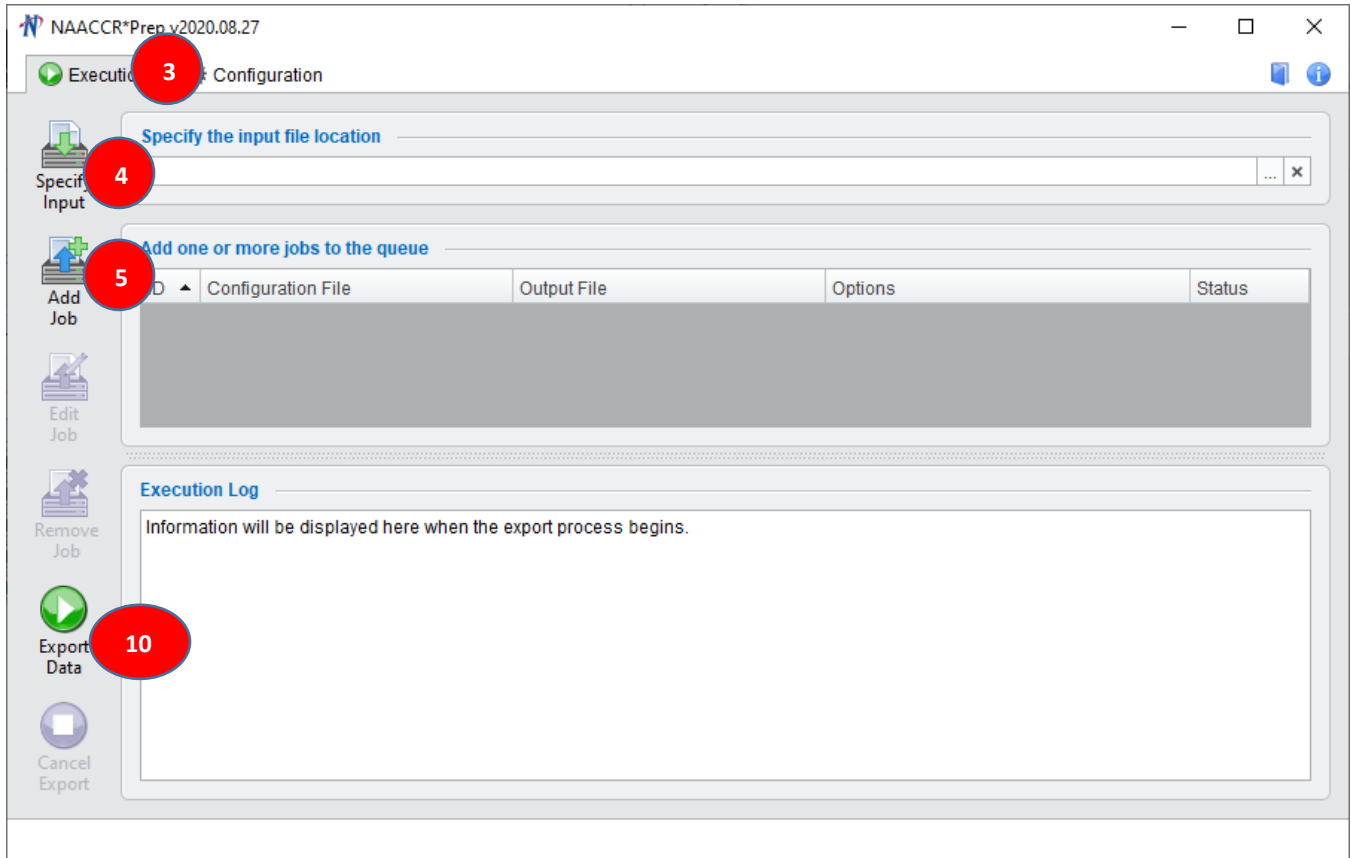
NAACCR Prep: There is a specific configuration file for NAACCR*Prep to create the VPR-CLS file. This configuration generates additional fields as needed but does not delete the confidential information, first/last/middle name/SSN/etc, required to conduct a linkage.

Steps for Creating the VPR-CLS Linkage File

1. Create a fully populated Type C NAACCR layout file using the Case Selection Criteria listed above. This file must be created using the .XML format.
 - a. We strongly encourage registries to convert their non-releasable information to the two new NAACCR variables: [#1854 Patient No Contact Flag](#) and [#1856 Reporting Facility Restriction Flag](#). However, if a registry has not implemented the new NAACCR data

- items that flag potentially non-releasable records [#1854 and #1856], their registry-defined field for tracking non-releasable records should be placed in the Restricted Release field [user defined, #9700] when the VPR-CLS linkage file is created. **Note that this is a non-standard item.** For more information on how to include non-standard items, please see the document “How to Include Non-Standard Data Items” on the NAACCR Call for Data web page.
- b. For NPCR registries participating in the NCCR and who are not on SEER*DMS (FL, OH, PA, and TN): Tumor Record Number [Item #60] must be populated in your VPR Linkage File and included in the NCCR submission. For CRS Plus users, Med Ref ID has been used to populate Tumor Record Number field. For participating NCCR registries not on SEER*DMS or CRS Plus, the registry-defined field for tracking static tumor identifier should be placed in the Registry Defined Tumor ID field [user defined, #9701] when the VPR Linkage File is created. **Note that this is a non-standard item.** For more information on how to include non-standard items, please see the document “How to Include Non-Standard Data Items” on the NAACCR Call for Data web page.
2. Download NAACCR Prep and the VPR-CLS NAACCR Prep configuration file from NAACCR Call for Data page.
 3. Open NAACCR Prep.
 4. Press the button labeled “Specify Input”. Browse to the Type C XML file created in Step 1, select the file, then press the “Open” button.
 5. Press the button labeled “Add Job”. The Add Job dialog box will appear.
 6. Provide the location of the NAACCR*Prep configuration file for the VPR-CLS.
 7. Specify where the output file should be written. Use the following naming convention to identify your state and diagnosis years (dates may vary by registry) when entering the name for the output file: ‘**XX(state)YYYY(inception year)2022**’. For example, if Idaho includes data from 1975 to 2022, the file will be named ‘ID19752022’.
 8. By now, several drop-down boxes should be visible at the bottom of the dialog but only three are relevant (the remaining will be set to “N/A” and disabled)
 - a. Inception Year (VPR): This field defaults to 1973. Change this to your registry’s inception year or the first year of high quality/complete data that you want to include in VPR linkages.
 - b. Overwrite NHAPIIA: Choose “Yes” to allow NAACCR*Prep to recalculate any existing values for NHIA or NAPIIA or “No” to keep any existing values.
 - c. NHIA Option: Choose “Option 1”, “Option 2”, or “All records”.
 - d. Output File Type: Make sure “XML” is selected.
 9. Press the OK button to close the dialog. The job will be added to the execution queue.
 10. Press the “Export Data” button.
 11. The zip file that is output by NAACCR*Prep contains several files. We are interested in the data file, the log file, and the XML user dictionary. The data file has a “.xml.gz” extension. The log file has a “.log” extension. The XML user dictionary is named “vpr-dictionary-2023.xml”. Extract the .xml.gz file, the .log file, and the XML user dictionary from the zip file. The other files, and the zip file/container, can be discarded.
 12. Review the output data file (the file with the .xml.gz extension) to verify that all Type C data items, including patient identifiers, are populated in the file. File*Pro (available on <https://seer.cancer.gov/tools/filepro/>) or other text editing tools can be used to view the file content. You will need to provide whichever tool you use with the XML user dictionary (extracted in step 11) in order to verify the existence of the user-defined fields (including the non-standard items referenced in Steps 1A and 1B) in the data file.

13. Keep a copy of the .xml.gz file created in Step 11. This file includes PII and will be retained at your registry for future use in VPR-CLS linkages and data release activities.
14. Upload the “.log” file created in Step 11 to the NAACCR Call for Data submission site under the Submission Forms section.



Add Job [Close]

Provide the location of the NAACCR*Prep configuration file [Browse] [X] **6**

Specify where the output file should be written [Browse] **7**

Job Options

Suppress Canadian Fields	N/A	Inception Year (VPR)	N/A	8A
Suppress Census Tracts	N/A	Overwrite NHAPIIA	N/A	8B
Suppress Counties	N/A	NHIA Option	N/A	8C
Suppress Days on Dates	N/A			
Suppress EPHT 5K	N/A	Output File Type	N/A	8D
Suppress EPHT 20K	N/A			
Suppress EPHT 50K	N/A			
Suppress Postal Codes	N/A			
Suppress U.S. Fields	N/A			

9 [OK] [Cancel]