Virtual Pooled Registry (VPR) Project
Instructions For Creating the Annual VPR-CLS Registry Linkage File for December 2025 Submission (as of 8/25/25)

The Virtual Pooled Registry Cancer Linkage System (VPR-CLS) facilitates streamlined linkages between research studies and participating U.S. registries. Starting in January 2026, the primary method for VPR linkages will be privacy-preserving record linkage (PPRL) which does not require personally identifiable information (PII). The PPRL linkages will be performed either by 1.) IMS, VPR Honest Broker, if a registry has signed the required VPR-CLS PPRL DUA and received internal approval to upload a hashed file during the CFD submission process; or 2.) registry behind their firewall if permission has not been granted to send a PPRL file to IMS during the CFD submission. In rare situations (small cohort, rare cancer, low completeness of linkage variables), studies may qualify for PII linkage, like previous VPR linkages.

As part of the NAACCR Call for Data, VPR registries will create and retain the following two standardized VPR linkage files:

1. PPRL Linkage File: This is a small file that contains only the Registry ID [Item 40], Patient ID Number [Item 20], hash tokens, and metadata fields (used to validate the hash tokens). No PII or incidence variables are included in this file. If a registry is able, they will submit the PPRL Linkage File to IMS so that IMS can perform PPRL linkages on their behalf. This means that the smaller file will only be used by IMS (if submitted); however, this file should be retained by all VPR registries.

2. PII Linkage File: This is a very large file that contains all NAACCR incidence variables and confidential data items (including PII), as well as hash tokens and metadata fields (used to validate the hash tokens). The PII Linkage File will be retained by all VPR registries and serve three purposes:

- Used internally by registries to perform PPRL linkages when they cannot submit the PPRL Linkage File to IMS.
- Used by all registries when a VPR linkage qualifies for PII linkage (which will be very few).
- Serves as a static file from which researcher-requested data items will be selected and exported for matched cases identified during the PPRL or PII linkage process.

IMPORTANT NOTE: We request that all U.S. registries, regardless of VPR participation, create both of the annual VPR linkage files and keep them on file at the registry in case they are needed during the upcoming year.

When are the Files Created? Registries will prepare the VPR-CLS registry linkage file as part of the annual NAACCR Call for Data process.

- Create the file after 24-month data (cases diagnosed in 2023) have been submitted to NAACCR, NPCR and/or SEER.
- Create the file **no later than December 31, 2025**.

NOTE: Unlike the Call for Data submission, the PII Linkage File contains patient identifiers and will remain at the cancer registry. **This file is NOT to be uploaded to the CFD Portal. Only the PPRL Linkage File should ever be submitted. This file has a .pprl.xml.gz extension.**

NOTE (for SEER registries): SEER submission files are created before IHS linkage is conducted. IMS inserts the IHS linkage results into your SEER submission file before the file is used. If you created your database extract at the same time as the SEER submission file, please insert the IHS linkage results into

the extract before running it through NAACCR*Prep to create the final VPR linkage files. IHS linkage results will be used to calculate a derived Race 1 Recode field but not retained on the file.

What is the Case Selection Criteria? Cancer registries will use the following case selection criteria:
- NAACCR Version 25 Type C records for all state resident cases with all data items populated, including patient identifiers.
- Cases from registry inception year to 2023 (24-month data). Registries should NOT include their 12-month data.
- 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. The NAACCR Inter-Registry Data Exchange Agreement (Revised July 2021) allows the receiving registry to release inter-state-exchange-only cases for VPR studies unless specified in a registry addendum.

NOTE: Do not mask or blank out the day portion of date fields, as certain dates are essential for performing a high-quality linkage with the PII Linkage File. If registry policies do not allow release of full dates, this information can be excluded from the data that is transmitted to VPR studies.

NAACCR Prep: A VPR configuration file for NAACCR*Prep will create the two VPR linkage files that will both include hash tokens. When creating the PII Linkage File, this configuration generates additional fields as needed (e.g. recodes) but does not delete the confidential information, first/last/middle name/SSN/etc, required to conduct a PII-based linkage. When creating the PII Linkage File, NAACCR*Prep will remove the IHS Link variable after calculating a derived race recode as IHS Link cannot be shared with external partners. The PPRL Linkage File will only include the Registry ID [Item 40], Patient ID Number [Item 20], hash tokens, and metadata fields (used to validate the hash tokens).
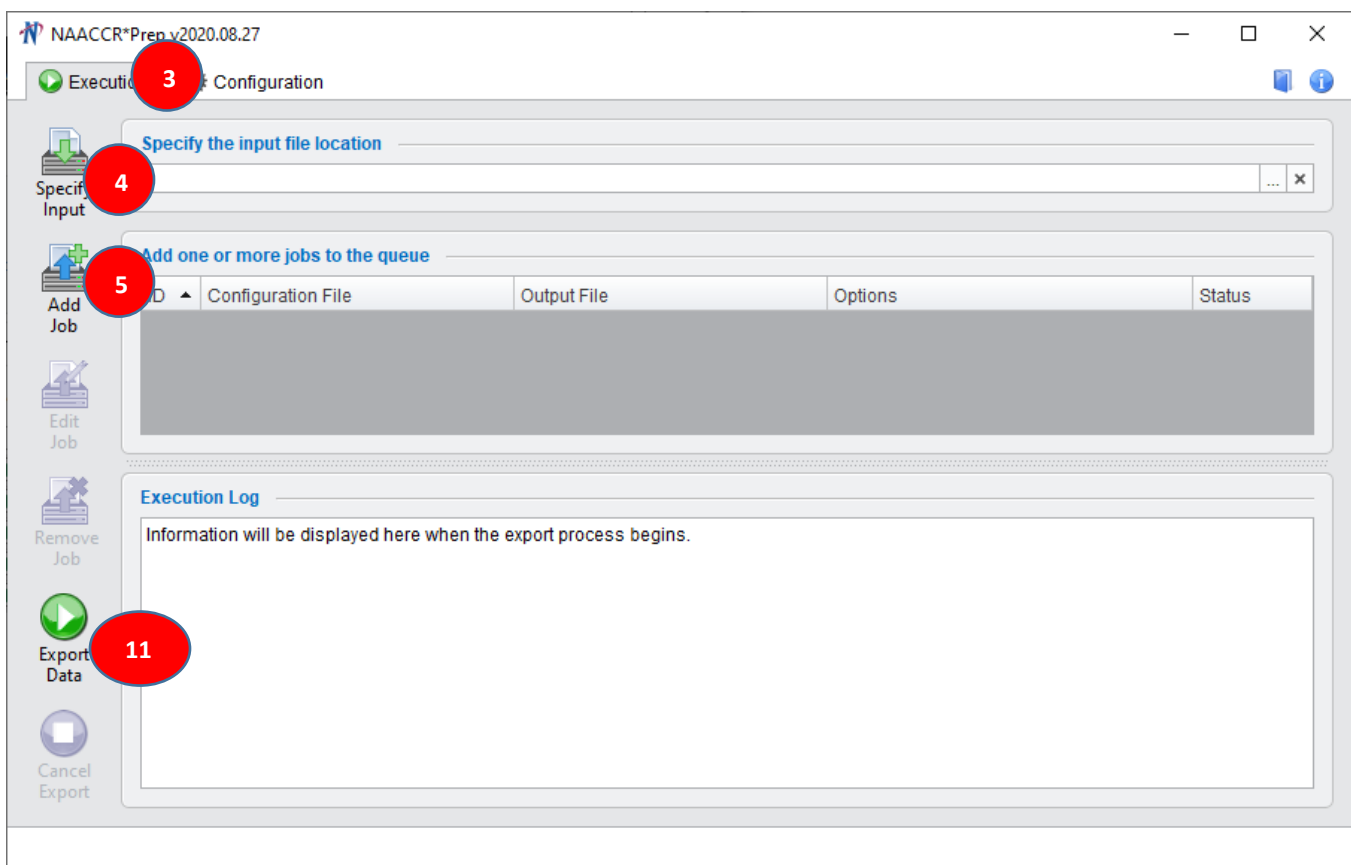
Steps for Creating the VPR-CLS Linkage File

1. In your central registry software, create a <u>fully populated</u> Type C NAACCR layout file using the Case Selection Criteria listed above. This file must be created using the .XML format.
   a. **All registries should convert their non-releasable information to the two new NAACCR variables implemented in 2023: #1854 Patient No Contact Flag and #1856 Reporting Facility Restriction Flag.**
   b. **All registries should also be populating Tumor Record Number [Item #60], which is a static identifier (meaning it should never change or be recycled) that is unique to each tumor belonging to a patient.** For NPCR registries participating in the NCCR and who are not on SEER*DMS: Tumor Record Number [Item #60] must be populated in your PII Linkage File and included in the NCCR submission. For CRS Plus users, initial population of Tumor Record Number has been based on the logic used to assign MedRefID since it is a sequential and unique identifier for each tumor in the CRS Plus database. All other registries who are not on SEER*DMS or CRS Plus will need to devise their own strategies for populating the Tumor Record Number field. For more information [click here](#).
2. Download NAACCR Prep and the NAACCR Prep VPR configuration file from NAACCR Call for Data page.
3. Open NAACCR Prep (See Screenshot #1)

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 (see Screenshot #2)
6. Provide the location of the NAACCR*Prep VPR configuration file.
7. Provide the location of the salt, which NAACCR Prep will use to generate hash tokens. This file will be sent via secure email to registry staff closer to the December timeframe.
8. Specify where the output Zip 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 Zip file: '**XX(state)YYYY(inception year)2023**'. For example, if Idaho includes data from 1975 to 2023, the file will be named 'ID19752023'. This file will contain the PPRL and PII data files, dictionaries, log, etc.
9. 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.
10. Press the OK button to close the dialog. The job will be added to the execution queue.
11. Press the "Export Data" button to create the output file.
12. The zip file that is output by NAACCR*Prep contains several files. We are interested in the data files, the log file, and the XML user dictionary. The PII Linkage File has a ".xml.gz" extension. The PPRL Linkage File has a ".pprl.xml.gz" extension. The log file has a ".log" extension. The XML user dictionary for the PII Linkage File is named "vpr-dictionary-2025.xml". The XML user dictionary for the PPRL Linkage File is named "vpr-pprl-dictionary-2025.xml". Extract the .xml.gz file, the .pprl.xml.gz file, the .log file, and the XML user dictionaries from the zip file. The other files, and the zip file/container, can be discarded.
13. Review the PII Linkage 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 corresponding XML user dictionary (extracted in step 12 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.
14. Keep a copy of the .xml.gz file created in Step 12. This file includes PII and will be retained at your registry for future use in VPR linkages and data release activities.
15. Review the PPRL Linkage File (the file with the .pprl.xml.gz extension) to verify that the Registry ID, Patient ID Number, hash tokens, and metadata 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 corresponding XML user dictionary (extracted in step 12) to verify the existence of the hash tokens and metadata.
16. On the NAACCR Call for Data submission site, upload the ".log" file created in Step 12 to the site under the Submission Forms section. For registries using SEER*DMS, the log file will be included in the data submission extract package.

17. If your registry elects to have IMS perform PPRL linkages on your behalf, upload the PPRL Linkage File, which has a .pprl.xml.gz extension and was created in step 12 to the NAACCR Call for Data submission site.
    a. Make sure to indicate if you are using this option in the US Special Files (VPR-CLS & NCCR) section of the SUBMISSION FORMS.
    b. The VPR file upload will be under the FILE SUBMISSION FORM section, along with your general submission file.
    c. The VPR-CLS PPRL DUA will be sent to your signing official via DocuSign under the DAA & CONSENT.
18. Once you have finished generating the VPR linkage files, delete the salt file that was sent to you via secure email (see Step 7).

Screenshot #1:

Screenshot #2: