Virtual Pooled Registry Cancer Linkage System (VPR-CLS) Fact Sheet

(Updated March 8, 2022)

The Virtual Pooled Registry – Cancer Linkage System (VPR-CLS) is an online service designed to:

- efficiently connect researchers performing **minimal risk linkage studies** with multiple U.S. population-based cancer registries;
- perform linkages between a study cohort file and cancer registry data files using standard linkage software and consistent matching algorithms;
- provide initial aggregate match count results to researchers; and
- streamline the process of applying for release of individual-level data on matched cases.

Coordinated by the North American Association of Cancer Registries (NAACCR) with funding from the National Cancer Institute (NCI), the VPR-CLS provides a single location to facilitate minimal risk linkages between studies with an existing cohort and 46 U.S. registries representing approximately 90% of the U.S. population plus Puerto Rico. These registries collect high quality, complete, population-based cancer information, including patient demographics, cancer type, stage, treatment, and follow up. By providing a single point of access and streamlined processes to enable these linkages, the VPR-CLS significantly reduces the level of effort researchers must dedicate to the linkage, application, and approval process across registries.

The technology for the VPR-CLS has been developed by Information Management Services, Inc. (IMS), which also serves as the third party honest broker. IMS has more than 45 years of information technology and clinical trials experience and employs a team of 250 computer and biomedical professionals located in the Washington Metropolitan Area. Long-term clients include the National Cancer Institute, the Centers for Disease Control and Prevention, the Food and Drug Administration, pharmaceutical companies, medical device companies, and other biomedical research organizations.

VPR-CLS Availability for Linkage with U.S. Registries

In February 2022, the VPR-CLS was officially launched. Due to the number of studies that expressed interest in VPR-CLS linkages during development and testing, the system will not be placed on a public-facing website to accept new linkage requests at this time. NAACCR and NCI have prioritized those existing studies of interest and selected twelve to proceed in the coming year. Interested researchers may contact Castine Clerkin (cclerkin@naaccr.org) to get on the waiting list for future VPR-CLS linkage.

To be considered for linkage, studies must complete a linkage request and have the following in place:

1. An existing cohort with individual identifiers suitable for linkage;
2. A current IRB-approval, IRB exempt determination, or documentation of Not Human Subjects Research;
3. A protocol that includes linkage with cancer registries; and
4. A study consent form that includes linkage with cancer registries OR a specific waiver of informed consent to link with registries.

Resources to Streamline the Data Release Application and Review Process

The VPR-CLS streamlines the process of applying for release of individual-level data on matched cases by offering optional use of the following resources:

1. Templated Forms: NAACCR led efforts to create the following templated forms that can be used in lieu of the state-specific forms:

    a. Templated IRB/Registry Application (TIRA): The TIRA is a standard application that compiles common questions based on review of over 50 individual registry and IRB applications. The TIRA is used in lieu of state-specific applications and has been adopted by nearly 80% of the VPR-participating registries for all or part of their review process, thereby reducing the total number of individual applications from 58 to 16 plus the TIRA.

    b. Templated Data Use Agreement (DUA): The VPR Templated DUA is a common agreement designed to be used in lieu of the individual registry DUAs. The VPR Templated DUA has been adopted by over 50% of the VPR-participating registries to date, thereby minimizing the number of separate DUAs and ensuring consistency in the terms and conditions.

2. Central IRB: NCI recently entered into a contract with the Biomedical Research Alliance of New York (BRANY) to serve as a Central IRB (CIRB) for review of VPR linkage studies. For states that accept a CIRB for multi-site minimal risk linkage studies coming through the VPR-CLS, BRANY will perform the IRB review in lieu of state/local IRBs. BRANY is currently working with individual registries/IRBs to determine acceptance of a CIRB and develop reliance agreements.

Overview of VPR-CLS Workflow

All VPR-CLS linkage requests proceed in the following two phases:

- **Phase I** supports a secure, standardized linkage and release of aggregate match counts (by state and diagnosis year) to the researcher. Phase I includes a web-based application, secure data transfer protocols between researchers, IMS and registries, and a standard record linkage software (Match*Pro) optimized for linkages between cancer registries and research cohorts. Phase I functionality has been successfully tested with seven large, national cohort studies.

- **Phase II** supports the process of applying to registries and/or their IRBs for release of individual-level cancer data for matched cases identified during Phase I. The system includes use of the TIRA and the Templated DUA, a robust and comprehensive tracking system, automated reminders, and future incorporation of the Central IRB. Phase II has been tested by three national studies and additional enhancements are now being incorporated into the system.

The Phase I match counts allow the researcher to review the volume of matches in each registry and make an informed decision about which to proceed with into the Phase II application for release of individual-level data. A detailed description of the Phase I and Phase II workflow is provided below

Phase I: *Application to use the VPR-CLS to link with registries behind their firewall and receive aggregate match counts only (no state IRB or registry review needed).*

Anticipated Timeline: *2-4 weeks for Phase I application review plus 2-4 weeks for registries to complete the linkage and return match counts once the researcher uploads a validated, edited cohort file.*

1. Researcher submits the online VPR-CLS application and supporting documents. Supporting documents include the current IRB determination, approved study protocol, consent form or waiver of consent, investigator's curriculum vitae, and signed DUA with IMS (VPR-CLS developer and 3rd party honest broker that validates study files).
2. NAACCR reviews application for completeness and resolves any issues with researcher.
3. Research Review Committee (RRC), made up of seven representatives from cancer registries and key stakeholder organizations, reviews application and researcher is notified of decision.
4. Researcher creates, edits, and uploads a cohort linkage file to the VPR-CLS in accordance with established file specifications after performing file validation/editing with Match*Pro software.
5. IMS validates file, resolves issues with researchers, and posts for registries to download.
6. Registries perform linkage behind their firewalls using Match*Pro and a standard linkage configuration file.
7. Registries create and upload an aggregate match count report to the VPR-CLS that includes the number of high quality and uncertain matches by diagnosis year (no patient records).
8. VPR-CLS reads the reports and presents researcher with the match counts.

Phase II: *Application for release of individual-level data on matched cases identified during Phase I.*

Anticipated Timeline: *IRB/Registry approval and release of data will vary based on the review process in each state and whether the TIRA (Templated IRB/Registry Application) and Central IRB can be utilized.*

9. Researcher reviews match counts and selects which registries to approach for release of individual-level data on matched cases.
10. The VPR-CLS facilitates a streamlined application process in which the researcher fills out the TIRA, uploads supporting documents, utilizes the Templated DUA, and is provided links to state-specific applications and agreements, if required.
11. Central IRB, registries, and local/state IRBs, as appropriate, review application and enter the review determination into the VPR-CLS tracking system. Registries may also require a DUA, confidentiality form, and payment to cover costs.
12. Upon approval and fully execution of agreements (if applicable), registries create a file of individual-level data of matched cases, including the Cohort ID and requested registry variables.
13. Registries provide the data file directly to the researcher through a secure site, independent of the VPR-CLS, as specified by either the researcher or the registry.

Linkage Methodology

All VPR-CLS linkages are performed using the record linkage software, Match*Pro, developed by IMS. Match*Pro conducts probabilistic linkage based on the Fellegi and Sunter model. The following variables are used, as available, to link the study file with the registry file: First name, middle name, last name, maiden name, date of birth, social security number, telephone number, gender, and street address. After probabilistically identifying potential matches, deterministic filters classify each linked pair as a match, non-match, or uncertain.

Data Security and Protections

The VPR-CLS provides a secure, web-based portal through which researchers submit an application to use the system to link with registries. A DUA is signed between the researcher and IMS before the researcher uploads the study file containing patient identifiers. The website uses Transport Layer Security, ensuring that communication and files transferred between a client and the IMS server are securely encrypted. All files uploaded to the VPR-CLS are first scanned for viruses and then stored on a secure server behind the IMS firewall. Only authorized IMS staff can access the study files provided by researchers and all IMS staff have been trained in the handling of files that contain personal identifying information.

Once an uploaded study file has been validated by IMS, it is posted for secure download by an authorized liaison from each of the participating registries. All registry liaisons are authenticated and verified by IMS prior to receiving access to the VPR-CLS. In addition, each registry has confirmed compliance with a list of common security protections. Study files are used solely for data linkage by the participating registries. Once ach registries downloads the study file, access to the file on the VPR-CLS is removed. Each registry performs the linkage behind their firewall. Once the match count report is uploaded, registries are prompted to delete the study file and provide confirmation of destruction. IMS deletes the study file after all registry data is sent to the researcher.

Results of VPR-CLS Pilot Tests

As part of the VPR-CLS development, the U.S. Radiologic Technologist study tested the feasibility of using a standard methodology to link across U.S. cancer registries, assessed the value of ascertaining cancer incidence through these means, and provided input on the VPR-CLS functionality. The study individually applied for and received data from 43 registries and compared cancer ascertainment with their usual method of self-report, medical record validation, and death certificate review. The results indicated that 37% of the registry-identified cases had not previously been known to the study. Similar results have been reported by another study that previously relied on self-report of cancer.

Once the VPR-CLS infrastructure was developed, seven national studies, with cohort sizes ranging from 26K to 1.6M, successfully pilot tested Phase I of the VPR-CLS process and received registry-specific aggregate match counts. Three of these studies proceeded to Phase II pilot testing, using the VPR-CLS to streamline and track the status of their application and release of individual-level data on the matched cases. Data received from the VPR registries have measurably improved the completeness of cancer ascertainment and enriched existing data on self-reported cancers.