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

For Principal Investigators proposing to utilize the VPR-CLS in future studies

(Updated May 1,2020)

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 using a single cohort file, 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, the VPR-CLS provides a single location to facilitate minimal risk linkage studies. Use of a standard linkage methodology and streamlined application process significantly reduces the level of effort researchers must dedicate to the linkage 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 system's third party honest broker. 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 Phase I and Phase II Development and Timeline

As the VPR-CLS has been developed, 38 central registries have participated in the pilot testing. All VPR-CLS linkage requests proceed in two distinct 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 researcher application, secure data transfer protocols between researchers and IMS and registries, and a new record linkage software optimized for linkages between cancer registries and research cohorts. Phase I functionality has been successfully tested with six large, national cohort studies.

- **Phase II** supports the process of applying to registries and IRBs for release of individual-level cancer data on matched cases identified during Phase I. The system includes a web-based application, a robust and comprehensive tracking system, automated reminders, and future incorporation of a Central IRB. Phase II testing is underway with two national cohort studies and is anticipated to be completed in 2020.

Resources to Streamline the Data Release Application and Review Process

The VPR-CLS will streamline the process of applying for release of individual-level data on matched cases during Phase II of a VPR-CLS linkage by offering optional use of the following resources:

1. Templated Forms: NAACCR is leading 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 includes common questions from across 50+ individual registry and IRB applications. The TIRA can be used in lieu of state-specific applications and has been incorporated into the VPR-CLS. The TIRA has been adopted by 82% of the pilot test registries for all or part of their review process.
    b. Templated Data Use Agreement (DUA): This template is in development and, when finalized, can be used in lieu of state-specific DUAs.
2. Central IRB: For states that accept it, the Central IRB will serve as a reviewing body, in lieu state/local IRBs, for multi-site minimal risk linkage studies, such as those coming through the VPR-CLS. NCI is currently developing a mechanism to establish the Central IRB.

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.

Overview of VPR-CLS Workflow

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

Anticipated Timeline: *2-4 weeks for application review; 2-4 weeks after the researcher provides validated, edited cohort file, the VPR-CLS will return match counts by registry and diagnosis year.*

1. Researcher submits online VPR-CLS application and supporting documents. Supporting documents include the current institutional IRB approval, approved study protocol, consent form or waiver of consent, investigator's curriculum vitae, and signed data use agreement with IMS (3rd party honest broker).
2. NAACCR reviews application for completeness and resolves any issues with researcher.
3. Research Review Committee (RRC) 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 and editing software.
5. Cohort file is validated and posted for registries to securely download.

6. Registries perform linkage behind their firewalls using Match*Pro and a pre-defined, standard linkage configuration file.

7. Registries generate aggregate match counts (no patient records) of the number of high quality and potential matches by diagnosis year and upload to the VPR-CLS.

8. Match counts are provided to the research through the VPR-CLS

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 considerably based on the review process in each state and whether the Central IRB can be used (when implemented).*

9. Researcher reviews match counts and selects which registries to approach for release of individual-level data on matched cases.

10. The VPR-CLS assists the researcher in completing a streamlined application process, which includes filling out a Templated IRB/Registry Application, uploading supporting documents, and providing links to state-specific applications, if needed.

11. Central IRB (when implemented), registries, and local/state IRB, as appropriate, review application and enters the review determination into the VPR-CLS tracking system. Registries may also require a data use agreement, confidentiality form, and payment to cover costs.

12. Upon approval and fully executed DUA (if applicable), registry creates a file of individual-level data on the matched cases from Phase I.

13. Registry provides data directly to the researcher through a secure site, independent of the VPR-CLS, as specified by either the researcher or the registry.

Data Security and Protections

The VPR-CLS provides a web-based portal through which researchers submit an application to use the system to link with registries. A Data Use Agreement is signed between the researcher and IMS prior to uploading study data files containing patient identifiers to the secure data transmission service. 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 full-time 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 reviewed and approved by IMS, it will be accessible for secure download by a single 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 all registries have downloaded the study file, access to the file on the VPR-CLS will be removed. Each registry will perform the linkage behind their firewall. Both the registry and IMS will destroy the study file at the conclusion of the match.

Results of VPR-CLS Pilot Tests

As part of the VPR-CLS development, the U.S. Radiologic Technologist study volunteered to test the feasibility of using a standard methodology to link across U.S. cancer registries, assess the value of ascertaining cancer incidence through these means, and provide 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. In addition, the process of applying to individual registries informed VPR-CLS functionality.

Now that the online VPR-CLS has been developed, six national studies, ranging in size from 26K to 1.6M cohort members, have successfully completed Phase I pilot test linkages with the 38 participating registries and have received registry-specific match counts. Two of these studies have proceeded to Phase II pilot testing and are using the VPR-CLS to streamline and track their application for release of individual-level data on the matched cases. One of these studies is a cancer survivor cohort that currently captures subsequent cancers through self-report and medical record validation. The Phase I match counts suggest that VPR-CLS linkage with registries will nearly double the number of subsequent cancers known to the study and receipt of the individual-level data will allow a completed evaluation.

Registry Participation and Available Data for Linkage

Registry participation in the VPR-CLS is voluntary. Currently, 38 registries, representing 78% of the U.S. population, are participating in the pilot testing initiatives. It is anticipated that the number of participating registries will increase when the VPR-CLS moves into production.

U.S. cancer registries have varying years of complete data available for linkage as noted on the NAACCR website: https://www.naaccr.org/cina-deluxe-for-researchers/ . The majority of registries will include the most recent data in their VPR-CLS linkage file (e.g. In 2019, cancer cases diagnosed through the end of 2017 would be included). Cancer registries provide standardized, curated, population-based data on patient demographics, incidence, stage, treatment, and follow up.