Description of document:	National Institutes of Health (NIH) Contract# HHSN261201600412P Final Report: U.S. Radiologic Technologists Cohort, Cancer and Other Disease Risks in U.S. Radiologic Technologists: Virtual Pooled Cancer Registry Follow-up, 2016-2018
Requested date:	01-February-2019
Release date:	14-March-2019
Posted date:	17-June-2019
Source of document:	NIH FOIA Office Building 31, Room 5B35 31 Center Drive, MSC 2107 Bethesda, MD 20892-2107 Fax: (301) 402-4541 NIH FOIA Request Portal

DEPARTMENT OF HEALTH & HUMAN SERVICES

National Institutes of Health
National Cancer Institute
Freedom of Information / Privacy Act Office
Bldg. 31, Room 10A48
9000 Rockville Pike, MSC 2580
Bethesda, Maryland 20892-2580
Ph: (240) 781-3340 Fax: (240) 541-4519

March 14, 2019

Re: NCI 19-040; FOIA Case No. 50338

This is a final response to your February 1, 2019 Freedom of Information Act (FOIA) request addressed to the NIH FOIA office, which was received in this office on February 6, 2018. You requested a copy of the final report from Contract# HHSN261201600412P; that was awarded by the National Institutes of Health (NIH)

We searched the files of the Divisions of Cancer Epidemiology and Genetic (DCEG) for records responsive to your request. That search produced 1 CD (equivalent to 18 pages) responsive to your request. Enclosed is a copy of Contract# HHSN261201600412P Final Report. You have agreed to accept this record in its redacted form.

If you are not satisfied with the processing and handling of this request, you may contact the NCI FOIA Public Liaison:

NCI FOIA Public Liaison
M.K. Holohan
Building 31, Room 10A48
9000 Rockville Pike
Bethesda, MD 20892
240-781-3410 (phone)
240-541-4519 (fax)
ncifoia@mail.nih.gov (email)

In certain circumstances provisions of the FOIA and Department of Health and Human Services FOIA Regulations allow us to recover part of the cost of responding to your request. Because the cost is below the $25 minimum, there is no charge for the enclosed materials.

Sincerely,

Suzanne Milliard
Freedom of Information Coordinator
NCI/OGCR/FOIA

Enclosures: 1 CD = Contract# HHSN261201600412P Final Report (18 pages).

U.S. Radiologic Technologists Cohort

Cancer and Other Disease Risks in U.S. Radiologic Technologists: Virtual Pooled Cancer Registry Follow-up

Contract Number: HHSN261201600412P

Bruce H. Alexander, PhD
Principal Investigator

Final Report

July 1, 2016 through June 30, 2018

VPR Scope of Work:

Base Period (non-severable discrete work segment): Link the USRT cohort with the pooled virtual cancer registry, review total cancer summaries by state and year of diagnosis, prepare a summary of findings, and recommend to NCI from which states detailed cancer information for individual cohort members should be pursued.

1. Assist NCI in contacting NAACCR representatives to obtain information about needed approvals, transfer agreements, submission file formats, and other requirements.
2. Assist NCI in obtaining IRB approvals to conduct the initial linkage with NAACCR to obtain total cancer information by state.
3. Develop data transfer agreement packages and assist NCI in obtaining data transfer approvals.
4. Develop file formats and documentation, prepare and submit to NAACCR a file of USRT study participants for linkage with the pooled virtual cancer registry.
5. Assist in reviewing returned summaries from NAACCR on numbers of cancers identified by state, and determining from which states to pursue IRB approvals.
6. Submit interim progress reports as tasks are completed and a final summary of findings that includes a recommended list of states from which detailed cancer information for individual cohort members should be pursued upon completion of all Base Period tasks.

Option I (non-severable discrete work segment): Link the USRT cohort with selected state cancer registries through the pooled virtual cancer registry to obtain individual-specific detailed cancer information, evaluate returned matches, determine if a registry-identified cancer is new or was previously reported, and incorporate new cancers plus detailed medical information for new and previously reported cancers into the cohort medical validation database.

7. Assist NCI/NAACCR in obtaining IRB approvals from individual states to conduct the second linkage with NAACCR to obtain individual-specific cancer information by state.
8. Pending receipt of IRB and other required clearances, prepare and submit a second cohort file to NAACCR to obtain individual-specific cancer information from selected states.
9. Receive NCI-recoded cancer validation database, and prepare a new cancer validation database "Master USRT cancer validation database" that will incorporate original and recoded cancer outcome information from self-report and medical record review, cancer information from the individual registries, and new cancer outcome variables that will be developed based on the most valid information available from each of the sources.
10. From the return file provided by NAACCR, evaluate results to ensure correct matches for individual cohort members.

11. Enter NAACCR-reported breast, thyroid, brain, and hematopoietic cancers into the Master USRT cancer validation database, compare against existing reports to identify new and previously-reported cancers, identify cancers that were previously reported but misclassified, rule out duplicates, identify cancers that were previously reported and/or medically validated but not identified from the virtual cancer registry, and, after determining decision criteria with NCI investigators, create new cancer outcome variables for the entire cohort of questionnaire responders and non-responders based on the best available information from self-report, medical record review, and cancer registries
12. Submit preliminary cancer incidence files to NCI for review.
13. Submit a final cancer incidence file with detailed documentation.
14. Submit interim progress reports as tasks are completed and a final Option I progress report upon completion of all tasks.

Option II (non-severable discrete work segment): Assist in evaluating: (a) the accuracy of cancers that were self-reported by individual technologists (b) the completeness of the virtual cancer registry by determining which of the self-reported and/or medically validated cancers were not identified through the virtual registry linkage. Prepare a publication-quality summary of the completeness and accuracy of self-reported cancer information before and after medical record review, and the completeness of the virtual cancer registry.

15. Assist NCI in evaluating the accuracy of cancers that were self-reported by individual technologists.
16. Assist NCI in evaluating the completeness of the virtual cancer registry by determining which of the self-reported and/or medically validated cancers were not identified through the virtual registry linkage.
17. Work with NCI to prepare a publication-quality summary of the completeness and accuracy of self-reported cancer information before and after medical record review, and the completeness of the virtual cancer registry.
18. Submit a final progress report for Option II upon completion of all tasks.

1. Creating the linkage file

In June 2016, the UMN prepared a file to submit to NAACCR for the Virtual Pooled Registry linkage. To ensure appropriate use of data with personal identifiers and comply with the individual state cancer registry requirements a protocol was developed to use IMS as an "Honest Broker" to allow the pooling of cancer data from state registries. IMS set up user accounts on their secure portal for the Phase I linkage data exchange.

The UMN created a file for linkage to meet IMS specifications that would allow the linkages to occur This entailed re-coding of UMN codes for race, gender, and vital status to align with NAACCR coding scheme.

As part of creating this file, the data were run through an Edits Program provided by IMS to standardize the data for linkage. This multi-step process identified potential data quality issue. The program fixed some formatting errors and identified other errors manual review. The flagged records were compared with available raw source data and modified as needed. The process was repeated until the program did not identify any errors. The errors identified and corrected included all records with extra spaces in the name or street address fields, zip codes identified as invalid for the state, invalid state abbreviations, etc. The UMN ultimately worked through the review sequence 6 times to ensure that all data issues were resolved. Once all issues were resolved, the a flat file with 146,022 records for the Virtual Pooled Registry (VPR) linkage was created and uploaded to the IMS secure portal in July 2016.

2. Phase I IRB Approvals

Phase I VPR linkage was to obtain de-identified data in the form of match counts and most participating state cancer registries did not require IRB approval to complete this linkage.

PROPRIETARY INFORMATION,REQUESTER IN AGREEMENT

For the initial Phase I linkage, four states required approvals or agreements prior to participating in the Phase I linkage:

PROPRIETARY INFORMATION,REQUESTER IN AGREEMENT

Once these IRB approvals were obtained the study would be approved for the Phase II linkage as well as the Phase I linkage.

PROPRIETARY INFORMATION,REQUESTER IN AGREEMENT

3. Phase II IRB Approvals

Phase II of the VPR linkage was to request individual level VPR cancer registry data for linkage to the USRT cohort. Obtaining these data required approvals from the various cancer registries and/or IRBs overseeing the use and disclosure of VPR cancer registry

data. The requirements were different for each registry/IRB. Although the general categories, as noted above, were similar, the actual questions or wording of the sections were unique to each IRB application. As the University of Minnesota is the holder of all personal identifying information for the USRT cohort much of the negotiation with cancer registry IRBs required the University of Minnesota to be identified as the lead institution.

PROPRIETARY INFORMATION,REQUESTER IN AGREEMENT

The UMN uploaded all the required state documents to a web-based drive shared with NCI staff. The UMN submitted applications to the multiple state IRBs and coordinated the activities with NCI to ensure appropriate documentation was obtained. For example, some registries needed signatures from both the UMN, which holds the personally identifying information, and NCI, which was the contracting agency. As needed, the UMN uploaded updated and signed documents to the shared drive. This included the current UMN Staff CITI training documentation, signatures, and draft and final responses to various sections in specific IRB applications, the final versions submitted to the state IRB, responses to stipulations, and the final approved version.

4. UMN work with VPR Cancer Registry Data

Once the approvals for a state was received, the UMN worked with each registry on the secure file exchange. This involved the UMN working with a specific contact at each registry to provide them with the secure drop site, username, and password. Additionally, the cancer registry contact would provide UMN with the password to unencrypt the data once it arrived at UMN.

After receiving the cancer registry data files (VPR), the UMN downloaded the files, unencrypted the files, and extract data using standardized SAS code. Part of the SAS code automatically runs frequencies of key variables, which are posted to the shared Drive. (see Figure A for a snapshot of the type of frequency data)

To review the linked data and process the file for IMS, the UMN re-linked, VPR ID, a unique identifier assigned specifically for the VPR effort with the original USRT study ID for each record in each registry file. The USRT study ID is the identifier required to link the VPR cancer registry data, the USRT cancer validation file, and all survey and mortality data residing at IMS. Additionally, the UMN added 7 fields for tracking the outcome of the review and data comparison to create a cancer registry data file (UMN) for each state/registry (see Figure B below.)

To determine how many cancers identified were reported by participants in the USRT study the cancer registry data file was linked to the USRT cancer validation file. Seven new variables were added to the USRT cancer validation file to assist tracking the matches (Table 1) Cancers diagnoses were considered matches when the location and diagnosis year (+/- 2 years) are the same for an individual. This outcome of "match" was recorded under [umn_match_review]. The UMN also populated a variable [umn_cancer_linkage] with the corresponding record number in the USRT cancer validation file for quick reference.

Table 1. UMN Variables Added for Cancer Registry Data (VPR) Comparison to USRT Cancer Validation File

UMN Variable Added	Variable Definition	Outcome
umn_match_review	Indicates outcome of comparison of Cancer Registry Data (VPR) to USRT Cancer Validation File	New: Cancer Registry Data (VPR) record is new; no existing records in the Validation File Additional: Cancer Registry Data (VPR) record does not match existing record in the Validation File Match: Cancer Registry Data (VPR) matches existing record in the Validation File Match (manual review): Cancer Registry Data (VPR) was manually reviewed and determined a match to an existing record in the Validation File
umn_cancer_linkage	If outcome of review = 'match,' this field contains corresponding record number (linkage) in the Validation File	*[numeric value of record number (linkage) in the Validation File to which Cancer Registry Data matched]*
mismatch_site	Indicates records were not a match due to site of cancer	0 = not flagged 1 = flagged as mismatch on site
mismatch_date	Indicates records were not a match due to diagnosis date of cancer (not within +/- 2 year range)	0 = not flagged 1 = flagged as mismatch on diagnosis date
previous_nmsc_only	Indicator that the only previous record in the Validation File was for a non-melanoma skin cancer (nmsc)	0 = not flagged 1 = flagged indicating only previous record was a non-melanoma skin cancer
possible_match	For VPR records not meeting match criteria, but close either by site or diagnosis date, this field contains the corresponding possible match record number in the Validation File	*[numeric value of record number (linkage) in the Validation File to which Cancer Registry Data is possibly a match]*
umn_survey_review	UMN indicator of last survey completed	Q1: LQ1 Q2: LQ2 Q3: LQ3/SQ3 Q4: LQ4/SQ4 None: no record of completed survey

Figure A: Page 1 of Standard SAS Frequency Report Generatedfor Alaska VPR Cancer Registry Data

REGISTRY FILE -- USRT COHORT LINKAGE
CHECK TO SEE THAT THE DISTRIBUTION OF FIELD VALUES IS WHAT YOU EXPECT
DATE FIELDS FORMATTED AS YYYYMM TO SHORTEN OUTPUT. FULL DATES INCLUDED IN FILE.
PATIENT ID AND USRTID ARE FORMATTED AS 4 DIGITS. FULL VALUES INCLUDED IN FILE.

The FREQ Procedure

500_Type of Reporting Source

N16_500	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	39	90.70	39	90.70
3	1	2.33	40	93.02
4	1	2.33	41	95.35
8	2	4.65	43	100.00

90_County at DX

N16_90	Frequency	Percent	Cumulative Frequency	Cumulative Percent
020	18	41.86	18	41.86
090	3	6.98	21	48.84
110	3	6.98	24	55.81
122	6	13.95	30	69.77
130	1	2.33	31	72.09
150	2	4.65	33	76.74
170	4	9.30	37	86.05
185	2	4.65	39	90.70
220	1	2.33	40	93.02
261	3	6.98	43	100.00

94_County at DX Geocode1990

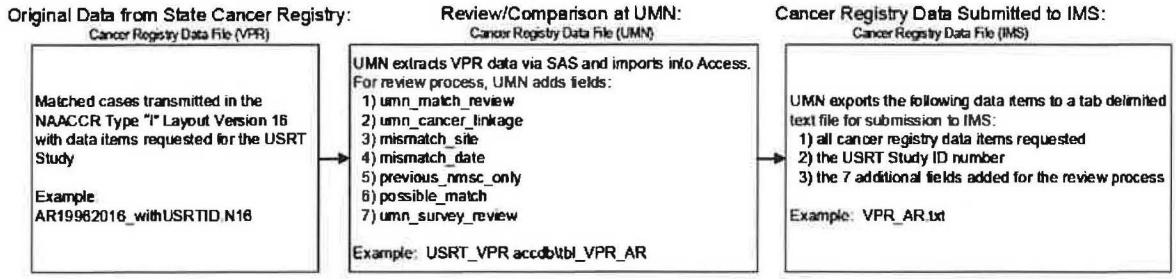
N16_94	Frequency	Percent	Cumulative Frequency	Cumulative Percent

Frequency Missing = 43

95_County at DX Geocode2000

N16_95	Frequency	Percent	Cumulative Frequency	Cumulative Percent

Figure B: Data Processing Flow of VPR Cancer Registry Data



Records for cancer diagnoses in the cancer registry data file (VPR) for which no cohort ID existed in the USRT cancer validation file were coded as new cancers (coded "new" ([umn_match_review]). The remainder of the records from each registry, which had a corresponding cohort ID in the USRT cancer validation file but did not match by year of diagnosis or diagnosis, required manual review.

The review process consistently identified about 1/3 of each cancer registry file (as "new," 1/3 as "matches," and 1/3 requiring manual review. Some of the mismatches were anticipated due to variations in the self-reporting of cancer diagnosis and date of diagnosis across the surveys, specifically for female reproductive cancers (uterine, cervical, and ovarian), hematopoetic/lymphopoetic, colorectal, and soft-tissue cancers. This variability in reporting resulted in a number of respondents having multiple cancer records in the USRT cancer1 validation database. . Similar challenges occurred with respondents reporting.

To resolve questionable matches, the UMN reviewed cancer information by person rather than cancer. In this way, the UMN was able to compare the full listing of cancers in the USRT cancer validation file for each individual to the cancer registry data and evaluate the cancer in question, within the context of all cancer data reporting and coding variations across time, to determine whether the 'match' criteria was met. The outcomes entered in [umn_match_review] were recorded as either "match-manual review" or "additional" in the cancer registry data file.

When the UMN determined that a cancer was not a match, indicators were 'checked' in variables [mismatch_site] and [mismatch_date] to indicate why the "additional" determination fell outside the match criteria. Additionally, if the match criteria wasn't met, but the UMN made the determination that the cancers were likely to be a match, the corresponding record number from the USRT cancer validation file was entered into a variable [possible_match] in the cancer registry data file.

Prior to creating the cancer registry data file for IMS, the UMN performed a quality review of the data to ensure the variable [umn_cancer_linkage] in the cancer registry data file matched the USRT cancer validation file as expected and to review for any

cancers potentially linked as duplicates. The UMN performed data cleanup as necessary if the values did not match as expected.

After completing the final quality review on each VPR cancer registry file, the UMN exported the cancer registry data, cohortid, and associated review variables to a tab delimited text file (as shown in Figure B above). The text files and pdf of frequencies were uploaded to IMS in batches. The dates UMN uploaded data to IMS are shown above in Table 2.

The UMN did not merge any VPR data into the USRT cancer validation file due to individual State Cancer Registry requirements for data sharing and destruction. The UMN tracked and stored individual cancer registry data (VPR & UMN) from each registry separately. To address the aims of Tasks 9 & 11, UMN will work closely with NCI to determine the structure and variables to be included in future analysis files.

PROPRIETARY INFORMATION,REQUESTER IN AGREEMENT

For individuals with a record in the cancer file and the registry cancer record had a different ICD or diagnosis year, the records were manually reviewed to determine whether the existing cancer file record could be a match to the registry cancer. The UMN evaluated these questionable matches, as described earlier, by viewing the entire record of an individual to take into account reporting and coding variations over time.

PROPRIETARY INFORMATION,REQUESTER IN AGREEMENT

Table 2. Summary of Records Received from State Cancer Registries for the VPR

Cancer Registry	Order Received	Year/Month Received	Years Covered	Records Received	Date to IMS
PROPRIETARY INFORMATION,REQUESTER IN AGREEMENT					
TOTAL				**19,145**	

PROPRIETARY INFORMATION,REQUESTER IN AGREEMENT

Up through June 2018, the UMN has received data from 34 cancer registries across the United States. A summary report of the cancer review for all states providing VPR data to UMN is shown above in Table 2.

Table 3 below illustrates the outcomes at Phases 1 and 2.

PROPRIETARY INFORMATION,REQUESTER IN AGREEMENT

Table 2 also illustrates the outcomes of the cancer review. The column headers shaded light green represent the records received and reviewed at the UMN. Phase II received cases is the sum of Electronic Matches, Manual Matches, and New Cases. The column New Cases represents the sum of [umn_match_review] outcomes of 'new' and 'additional.'

Table 3. Summary of State Cancer Registry Data Review

STATE	States not participating (NP)	Overall Status (A=Approved; SU= Submitted)	Do we have the data?	Est cases in USRT Cohort	Phase I USRT matched cases	Phase II received cases	Electronic Matches	Manual Matches	New Cases

PROPRIETARY INFORMATION,REQUESTER IN AGREEMENT,WITHHELD THROUGH PAGE 16

STATE	States not participating (NP)	Overall Status (A=Approved; SU= Submitted)	Do we have the data?	Est cases in USRT Cohort	Phase I USRT matched cases	Phase II received cases	Electronic Matches	Manual Matches	New Cases
PROPRIETARY INFORMATION,REQUESTER IN AGREEMENT									
TOTAL				**144893**	**25108**	**19145**	**6844**	**1248**	**10275**

5. On-going Work

VPR Cancer Registry Data

As of June 30, 2018, the UMN had received files from 34 cancer registries and reviewed 18,367 records. The UMN has processed and transmitted all cancer registry files received prior to June 17, 2018, to IMS. The UMN continues to work closely with NCI to obtain IRB approvals and data from remaining registries (Florida, Missouri, New Mexico, Vermont, and Washington) and data from IRB approved registries (District of Columbia, Iowa, New Jersey, and Oklahoma).

USRT Cancer Validation Database Files – Recoded for NCI Analyses

The UMN continues to work closely with NCI and IMS in the creation of files for analysis. Special considerations and accommodations are made in data handling to comply with agreements for data use and destruction with the various IRBs across the United States. One analysis file will be a recoded cancer validation database. Another will be a new cancer validation database that will incorporate original and recoded cancer outcome information from self-report and medical record review, cancer information from individual registries, and new cancer outcome variables that will be developed based on the most valid information available from each of the sources. These files will be created with options for both internal analysis and potential external analysis.

In preparation for a USRT cancer validation database, recoded for NCI analyses, the UMN reviewed data from state cancer registries to identify matching records in the USRT cancer validation file and new reports of cancer. Records matching current data include a linkage record number to connect with associated records in the USRT cancer validation file. Potentially misclassified or mismatched cancers are identified as possible matches. The UMN submitted each of the reviewed VPR cancer registry files to IMS with that information.

Evaluate Accuracy of VPR Cancer Registry Data

The UMN continues to work closely with NCI and IMS in evaluating the accuracy of cancers that were self-reported by individual technologists. The UMN regularly discussed the process of linkage and data quality with NCI and NAACCR. The UMN continues to assist NCI in evaluating the completeness of the VPR cancer registry data by determining which of the self-reported and/or medically validated cancers were not identified through the virtual registry linkage. The UMN tracks the range of diagnosis years received from the registries to be able to identify potential limitations in the assessment of completeness.

Communication and Study Planning

The UMN had regular calls including participants from NCI/DCEG, NAACCR and NCI/DCCPS to discuss the quality and completeness of the data and approaches to summarizing the data. The UMN worked with NCI on PROPRIETARY INFORMATION,REQUESTER IN AGREEMENT PROPRIETARY INFORMATION PowerPoint presentations for the annual NAACCR meeting and for update reviews with directors of DCEG and DCCPS. Much of the work focused on

compiling cost estimates for past medical validation to provide a comparison of the cost savings involved in linking to a Virtual Pooled Cancer Registry.

IRB Continuing Review

A large part of the on-going UMN work will be complying with annual IRB renewals and reporting requirements for the various IRBs in a timely manner to avoid non-compliance, data destruction notices, and/or study closure notices. The UMN will work closely with NCI to monitor for adherence to varying restrictions on publications, data retention, data destruction, progress report requirements, and continuing reviews.

Assist NCI in Preparing Publications

The UMN will continue to work closely with NCI in the analysis of the data from the VPR linkage and production of manuscripts on the VPR data. As manuscripts are developed the UMN will provide specific information on the methods and procedures for obtaining and linking. The UMN will also work with NCI and IMS to identify other variables in the USRT cohort that may be helpful in evaluating the comprehensive nature of the VPR linkage. This may include additional review of the individually identifiable information maintained at the University of Minnesota.

6. Lessons Learned

Obtaining Approvals

Applying for access to the registry data was challenging in that most states required their own cancer registry and/or IRB applications to be completed and submitted. Each state had different document formats, methods for submitting (i.e., email, postal mail, online), and supporting document requirements. PROPRIETARY INFORMATION,REQUESTER IN AGREEMENT

PROPRIETARY INFORMATION,REQUESTER IN AGREEMENT

It took some effort to identify key contact personnel at the various cancer registries and IRBs. The key contacts were helpful in providing information about how and what was

needed for the submissions and either provided the required templates or direct us to their online form libraries.

IRB approvals could take several months to receive, particularly if modifications or clarifications were requested.

PROPRIETARY INFORMATION,REQUESTER IN AGREEMENT

To obtain data from a virtual pooled registry where participating cancer registries accepted a standardized IRB application template and process would be more efficient and would help to accelerate the process.

Data exchange

The data exchange worked very well. For the Phase II linkage, states were able to access their initial Phase I linkage data from the secure portal, populate it with the USRT VPR ID, and submit the data to the UMN in a standardized NAACCR file format. The UMN used standardized SAS code provided by NAACCR to extract the data and run frequencies. SAS frequencies for each file were uploaded to the shared Drive.

The UMN used a secure ftp site and worked with a contact person at each registry to ensure a smooth and successful file upload.

Cancer Registry Variables

Due to differences in state cancer registries, their reporting requirements, and year of cancer diagnoses, the variables requested for this study were not always fully populated. UMN tracked these differences in a spreadsheet to capture knowledge to inform the VPR development. For example the variables pertaining to 'cause of death' were restricted by some state cancer registries and not released in the linkage. Newer variables added by NAACCR may contain information for newer cancers only.

Continuing Review

IRB approvals are subject to Annual Review or Closure as specified in the initial notice of approval. Those requirements are specific by state as well. In order to continue using the VPR cancer registry data, continuing review approval must be obtained prior to the expiration date. This usually involves providing current training records and sometimes involves providing updated Data Use or Confidentiality Agreements. Failure to timely submit or obtain approvals from IRBs may trigger data destruction procedures.

Benefits of VPR

If the process of obtaining cancer registry data can be streamlined through a mechanism like the Virtual Pooled Registry. The systematic identification and validation of diagnosed cancers for large cohorts has great potential to reduce costs in materials and staffing. A potentially substantial benefit of the VPR is to identify cancer diagnoses in all individuals in an enumerated cohort and not just those who respond to a questionnaire. The potential to at least partially characterize the potential for selection bias in cohort studies.