Supporting Statement A For:

Data Collection for Public Cancer Epidemiology Descriptive Cohort Database (NCI)

September 4, 2014

Seminara, Daniela, Ph.D., M.P.H.
Senior Scientist and Cohort and Consortia Coordination Team Lead,
Epidemiology and Genomics Research Program (EGRP),
Division of Cancer Control and Population Sciences (DCCPS),
National Cancer Institute (NCI),
National Institutes of Health (NIH)

9609 Medical Center Drive
Room 4E130 MSC 9763
Rockville MD 20850
Telephone: (240) 276-6748
Fax: (240) 276-7921
e-mail: seminard@mail.nih.gov

Table of Contents

A.1 Circumstances Making the Collection of Information Necessary....1

A.2 Purpose and Use of the Information....3

A.3 Use of Improved Information Technology and Burden Reduction....4

A.4 Efforts to Identify Duplication and Use of Similar Information....5

A.5 Impact on Small Businesses or Other Small Entities....6

A.6 Consequences of Collecting the Information Less Frequently....6

A.7 Special Circumstances Relating to the Guidelines of 5 CFR 1320.5....6

A.8 Comments in Response to the Federal Register Notice and Efforts to Consult Outside the Agency....7

A.9 Explanation of Any Payment or Gift to Respondents....7

A.10 Assurance of Confidentiality Provided to Respondents....7

A.11 Justification for Sensitive Questions....8

A.12 Estimates of Annualized Burden Hours and Costs....9

A.13 Estimates of Other Total Annual Cost Burden to Respondents and Record Keepers....11

A.14 Annualized Cost to the Federal Government....11

A15. Explanation for Program Changes or Adjustments....11

A16. Plans for Tabulation and Publication and Project Time Schedule....11

A17. Reason(s) Display of OMB Expiration Date Is Inappropriate....12

A18. Exceptions to Certification for Paperwork Reduction Act Submissions....13

LIST OF ATTACHMENTS

A. Approval Form

B. Data Collection Form

B1. Initial

B2. Annual Update

C. Biospecimen and Cancer Count Information Spreadsheet

C1. Initial

C2. Annual Update

D. Email Invitation for: PAR-Funded Cancer Epidemiology Descriptive Cohort Database Principal Investigators

E. Email Invitation for: Non-PAR-Funded Cancer Epidemiology Descriptive Cohort Database Principal Investigators

F. List of Consultants

G. NCI IRB Exempt

H. Privacy Act Memo

Abstract
This is a request for approval of a new information collection. The NCI Epidemiology and Genomics Research Program (EGRP) supports large-scale collaborations across numerous cancer epidemiology cohorts. This public website will allow investigators to know what data and specimens exist among other cohorts. Respondents will be cohort Principal Investigators. The data collection forms will be sent to participating cohort PIs annually to update any information that has changed so that the CEDCD website will remain current. No cohort participant-level data is being collected from any of the cohorts.

A. JUSTIFICATION

A.1 Circumstances Making the Collection of Information Necessary

Section 410 of the Public Health Service Act (42 USC § 285) authorizes the collection of the information.

The Program under which the information collection request falls is the Epidemiology and Genomics Research Program (EGRP), located in the Division of Cancer Control and Population Sciences (DCCPS), National Cancer Institute's (NCI's), National Institutes of Health (NIH). EGRP's mission includes an aim to support collaborative research among investigators and to implement research based on cancer epidemiology cohorts. Cohorts are groups of persons who are followed over a period of time. For the purposes of cancer epidemiology research, cohorts are used to evaluate whether certain aspects or characteristics of these persons, collected either at the beginning or during the middle of the follow-up period, are associated with risk of subsequent cancer or a cancer-related outcome. To this end, EGRP fosters collaborations among cohorts through which cohort investigators and researchers from the scientific community at

large can pool the large quantity of data and biospecimens necessary to conduct a wide range of cancer studies, address important scientific questions in genomics, epidemiology and translational sciences and accelerate the pace of research in cancer etiology, prevention and control.

The collection of information requested in this application is needed to populate a web-based database that will contain descriptive information about existing cohorts (e.g. study design, eligibility criteria, enrollment numbers, scope of content domains collected, numbers of biospecimens, number of cancer, and other health outcomes). The information will be housed in a searchable web-based database, the Cancer Epidemiology Descriptive Cohort Database (CEDCD), to enable the research community to locate information about the existence of data and specimens and to find potential collaborators to further important research aims. Currently, this descriptive information is partially available on individual cohort websites or in publications, or must be obtained by direct communication with individual cohort investigators. The effort required to seek out and assemble this information each time a new research project is proposed is a considerable effort and burden on the part of the researchers and the NCI. The collection of this descriptive data by EGRP and their assembly into a publicly accessible and searchable database will substantially decrease the effort needed by investigators to locate relevant information, and decrease the lead time needed to initiate collaborative research.

Cohorts funded under the PAR are mandated to provide their descriptive data for this project. In accordance to terms of agreement for PAR-11-167, "Core Infrastructure and Methodological Research for Cancer Epidemiology Cohorts (UM1)", funded cohorts (http://go.usa.gov/ZJd3)

have committed to provide information including descriptive cohort characteristics, study protocols, basic counts of participants with various characteristics such as biospecimen availability and study variable definitions, as well as any available questionnaire and tools developed by the cohort. Currently, 14 cohorts are funded under the PAR, though this number could increase. For the other cohorts, participation is voluntary.

A.2 Purpose and Use of the Information

The purpose of collecting this information is to gather descriptive information in one database that will be accessible and searchable to the public. The database will promote collaboration in the research community, increase transparency, and allow for more precise exchange of information. With approval from (Attachment A) the cohort Principal Investigators, the information would be loaded into the Cancer Epidemiology Descriptive Cohort Database (CEDCD), a searchable, web-based database. The cohort PI's will be informed when their data are loaded and will be able to review and request corrections of the information by contacting the NCI contractor. The CEDCD will be publicly accessible to increase transparency and provide the widest possible usage, and it will be of greatest interest and use by members of the research community. The CEDCD will not restrict user access; however, it will request that users provide a minimal amount of information such as name, email address, and affiliation or institution so EGRP will gain knowledge of user interest. The established website will include built-in metrics on usage and users will be asked to specify the use of the information gathered.

Though the CEDCD has a biospecimen component, the motivation of the CEDCD is not to serve as a biospecimen locator database. It is a descriptive database focusing exclusively on

descriptive data pertaining to large, prospective epidemiology cohorts. The purpose is to provide overall information to the scientific community on cancer epidemiology cohorts in order to foster multi-disciplinary collaboration within the framework of population sciences, share processes and protocols and, ultimately, map epidemiology cohorts worldwide. The information in the CEDCD will be particularly helpful for scientists interested in large scale pooled and meta-analyses, molecular and genomic epidemiology projects, and robust replication of previous results.

A.3 Use of Improved Information Technology and Burden Reduction

The Cancer Epidemiology Descriptive Cohort Database Data Collection Form (Attachment B) will be a fillable PDF document. The Biospecimen and Cancer Count Information Spreadsheet (Attachment C) will be an Excel spreadsheet. Both can be completed electronically and returned via email. The Approval Form (Attachment A) will be a PDF document that will need to be printed, signed, and returned via email, fax, or mail. The invitation will be sent by email (Attachments D and E). Respondents will complete the forms and email them back to the NCI contractor along with any other supporting documents that they agree to make public. The information will be transferred electronically from the fillable PDF and the Excel spreadsheet to the website via a maintenance database. The information from the Approval Form will be entered into an electronic format, most likely an Excel spreadsheet, retained by the NCI contractor and provided to NCI upon request.

A Privacy Impact Assessment (PIA) is currently in progress for the IT system being developed for this project.

A.4 Efforts to Identify Duplication and Use of Similar Information

While many of the cohorts have their own websites, there is no single website containing all of the descriptive data about the cohorts in one place that is current and complete. In developing this website, research was conducted to find if there were similar sites. While there are websites that focus on biospecimens availability across cohorts, there are no websites that will display as much descriptive data about cohorts as planned in the CEDCD. This data collection effort has been reviewed by many stakeholders and interested parties, both intramural and extramural, and captures the information needed by those parties. For example, this data collection effort will also provide descriptive data needed by the NCI Cohort Consortium, allowing the cohorts to provide the information once and have it be centrally maintained and accessible. The overall goal is to decrease the burden on the part of the government and the research community, and to foster greater use of data and specimens that already exist.

A.5 Impact on Small Businesses or Other Small Entities

No small businesses will be involved in this study.

A.6 Consequences of Collecting the Information Less Frequently

The respondents will initially be asked to complete the data request and update the information annually thereafter. The same Cancer Epidemiology Descriptive Cohort Database Data Collection Form and Biospecimen and Cancer Count Information Spreadsheet will be used for the annual updates, prefilled with the information from the prior year. The respondents will be asked to make any changes to the data necessary to bring the information up to date.

Maintaining the same format for the annual updates will decrease the time required by respondents to participate as they will be familiar with the format. The annual updates will keep the CEDCD accurate and up to date. Because the cohorts continue to gather data, enroll subjects, and follow up for outcomes over time, the consequences of not updating the CEDCD would result in outdated and inaccurate information in the CEDCD.

A.7 Special Circumstances Relating to the Guidelines of 5 CFR 1320.5

This project will be implemented in a manner that fully complies with the Guidelines of 5 CFR 1320.5.

A.8 Comments in Response to the Federal Register Notice and Efforts to Consult Outside the Agency

This data collection plan has been developed by EGRP staff in consultation with NCI and DCCPS leadership. Representatives of CDC and outside academic institutions participating in the NCI Cohort Consortium have also been consulted (Attachment F).

The proposed information collection was published in the Federal Register on June 18, 2014, Vol. 79, page 34766 and allowed 60-days for public comment. No public comments were received.

A.9 Explanation of Any Payment or Gift to Respondents

No payment or gifts will be given to respondents.

A.10 Assurance of Confidentiality Provided to Respondents

The cohort Principal Investigators and their Business Officials (if needed) will be requested to sign an approval form and denote what data and supporting documentation (e.g. questionnaire templates, protocols, publication guidelines) all listed on the Approval Form (Attachment A) can be publicly posted on the CEDCD website. The cohort Principal Investigators and their Business Officials can also provide more specific instructions to NCI and the NCI Contractor as to which information, if any, cannot be made publicly available. Only information approved on a signed Approval Form (Attachment A) will be loaded to the CEDCD. No cohort participant-level data is being requested.

All project staff will complete the NIH-furnished initial and refresher security and privacy education and awareness training before being granted access to project systems. The NCI contractor will furnish NCI with an IT security plan and security assessments per contract requirements. All data received will be stored on secure servers and only information approved by the cohorts will be uploaded to the CEDCD.

The NCI IRB has reviewed this project and designated the activities of this project as Exempt from further IRB review (Attachment G).

The privacy officer reviewed this project and determined the Privacy Act will apply. The data collection is covered by NIH Privacy Act Systems of Record 09-25-0200, "Clinical, Basic and Population-based Research Studies of the National Institutes of Health (NIH), HHS/NIH/OD (67 FR 60776, 9/26/2002)." (Attachment H)

A.11 Justification for Sensitive Questions

The respondents in this survey are the cohort Principal Investigators. They will be providing their name and contact information as well as the name and contact information of other staff

affiliated with their study, such as Co-Investigators and possibly their Data Managers. This information is already publicly available through publications, cohort websites, and institutional websites. This information will be posted on the CEDCD website with the signed approval of the cohort PI and their business officials (if needed) on the Approval Form (Attachment A). Sensitive information that will be collected includes the overall number and proportion of race and ethnicity of the cohort participants. The purpose of this information is to provide the opportunity for researchers to collaborate with other cohort investigators on cancer-related research questions that may pertain to specific racial or ethnic groups. No cohort data at the participant- or individual-level is being collected.

A.12 Estimates of Annualized Burden Hours and Costs

The initial survey should take each of the respondents approximately 3 hours to complete. This estimate is comprised of approximately 30 minutes to read and sign or to forward for signature to an authorized individual at their institution for signature on the Approval Form (Attachment A); approximately 60 minutes to complete the Biospecimen and Cancer Count Information Spreadsheet (Attachment C1), and approximately 90 minutes to complete the Cancer Epidemiology Descriptive Cohort Database Data Collection Form (Attachment B1). The Biospecimen and Cancer Count Information Spreadsheet (Attachment C1) will be pre-populated with data previously collected or found publicly. The annual update will take each of the respondents approximately 1 hour and 15 minutes to complete; 30 minutes to update the Biospecimen and Cancer Count Information Spreadsheet (Attachment C2) and 45 minutes to update the Cancer Epidemiology Descriptive Cohort Database Data Collection Form (Attachment B2). All annual update forms will be prepopulated with the information that was entered initially which accounts for the difference in burden in the two forms.

In Year 1, we anticipate inviting 100 respondents for initial completion (180 minutes each). In Year 2, we anticipate 100 respondents for annual updates (75 minutes each) plus 100 new respondents to be invited for initial completion (180 minutes each). In Year 3, we anticipate 200 respondents annual updates (75 minutes each) plus 100 new respondents invited for initial completion (180 minutes each). The total annualized burden hours are 550.

To ensure the burden for this project was accurately assessed, a pilot study was conducted with all of the proposed data collection materials. The forms were emailed to six PAR funded cohort PIs with the request to review, complete the forms, and provide feedback on the forms and process. Three of the PIs were unable to complete the forms within the time frame requested because of other priorities but two of three provided feedback on the forms based on reviewing the documents. The other three cohort PIs completed the forms and provided detailed feedback on the forms and the time it took to complete them.

Table A12-1. Annualized Estimates of Hour Burden

Type of Respondent	Form Name	Number of Respondents	Number of Responses Per Respondent	Average Burden Per Response (in hours)	Total Annual Burden Hour
Individual: Principal Investigator Initial Submission	Approval Form	100	1	30/60	50
	Biospecimen and Cancer Count Information Spreadsheet	100	1	1	100
	Data Collection Form	100	1	90/60	150
Individual: Principal Investigator Annual Update	Biospecimen and Cancer Count Information Spreadsheet	200	1	30/60	100
	Data Collection Form	200	1	45/60	150

Total	300		550

Table A12-2. Annual Cost to Respondent

Type of Respondent	Number of Respondents	Total Annual Burden Hours	Wage Rate*	Respondent Cost
Individual: Scientific Researcher Initial Submission	100	300	$36.14	$10,842.00
Individual: Scientific Researcher Annual Update	200	250	$36.14	$18,070.00
Totals	300	550		$28,912.00

*Hourly wage rates for 19-1029 Biologic Scientist is $36.14 (based on http://www.bls.gov/oes/current/oes191029.htm).

A.13 Estimates of Other Total Annual Cost Burden to Respondents and Record Keepers

There are no capital costs, operating costs, or maintenance costs for the respondents, other than their time.

A.14 Annualized Cost to the Federal Government

This survey is estimated to cost the Federal Government $75,500 annualized (Table A.14-1).

Table A14-1. Cost to the Federal Government

Staffing	Task	Annualized Cost
NCI	GS-13, Step 6, project manager ($104,911 x 20% time for 12 months)	$21,000
Contractor	Project Management Support, Web Management support,	$54,500

	collecting data, sending emails, and managing data	
Total		$75,500

A15. Explanation for Program Changes or Adjustments

This is a new information collection.

A16. Plans for Tabulation and Publication and Project Time Schedule

The information submitted by the cohorts will be reviewed, formatted if necessary, and uploaded to the CEDCD. We intend to write and publish in a large diffusion journal a descriptive paper illustrating the establishment of the descriptive database and its purpose/use, as well as overall statistics concerning the participating cohorts. There are multiple goals for choosing this dissemination avenue: first is to advertise the tool and its purpose as widely as possible, but with particular target to the scientific constituency that could make maximum use of it; second to achieve transparency and disseminate information on the existing cohorts infrastructures and their characteristics in a summary manner; third to provide some information in the future about time trends and research supported by cohorts infrastructures; and lastly, to implicitly extend the invitation to participate to cohorts world-wide. These are essential goals of this contract and respond to EGRP priorities. An example of such publication would be our consortia paper (Burgio et al., 2013 CEBP).

The projected timeline is as follows:

Task:	**Projected Time (after OMB approval):**
Launch initial survey	1 to 5 months

Receive initial data and upload to CEDCD	ongoing
Launch annual updates	12 to 16 months
Receive annual update data and upload to CEDCD	ongoing
Project management support and web support of CEDCD	ongoing

A17. Reason(s) Display of OMB Expiration Date Is Inappropriate

The OMB Clearance Number, Expiration Date, and Burden Disclosure Statements will be displayed on the applications.

A18. Exceptions to Certification for Paperwork Reduction Act Submissions

There are no exceptions to the Certification for Paperwork Reduction Act Submissions.