Participant User File (PUF) Supplemental Documents

Version: PUF 2016 – Containing cases diagnosed in 2004-2016

Application Elements and Instructions

NCDB Participant User Files

1. Confirm that you are at a Commission on Cancer (CoC)-accredited program.

2. Before you start planning your proposal, review the information available at http://ncdbpuf.facs.org/. This resource contains an introduction and overview to the NCDB PUFs, detailed documentation of the data items included in the distributed files, a list of investigators and publications that have used the NCDB PUF.

3. Secure support from your CoC-accredited cancer program. As part of the application process principal investigators (PIs) will be asked to upload a letter of support for the proposal, written on hospital or cancer program letter head, from the Cancer Committee Chair on record with the CoC. If you are the Chair, please provide a letter of support from your Cancer Program Administrator.

4. If you do not already have a CoC Datalinks user name and password please work with your cancer registry staff or whoever manages *staff contacts* in CoC Datalinks for your cancer program for assistance in securing these credentials. PIs will need to have these credentials in order to proceed with their PUF application. The steps required to obtain a CoC Datalinks username and password may be found using the following link: https://www.facs.org/quality-programs/cancer/coc/info/datalinks/managestaff, and provide information on adding/removing roles and editing contact information. **NOTE:** Be sure your CoC Datalinks credentials identify you as a "NCDB PUF Applicant". Once this role has been designated, it may take a few business days for the designation to be approved by the CoC. If you already have CoC Datalinks credentials, confirm that in addition to your current role you are also identified as a "NCDB PUF Applicant".

5. Begin the NCDB PUF on-line application process. Once you have your CoC Datalinks credentials, log into the *NCDB PUF Application Manager* and start the on-line application. PIs can return at any time to complete, revise, or update an application until the application deadline. No facsimile or mailed copy of the application will be accepted.

6. Each NCDB PUF application will undergo a review. The review is performed by NCDB staff, who will check for the completeness of each proposal and assess the feasibility of the proposed project, given the data available in the PUFs. At the conclusion of the review process, proposals will be designated as approved; approved contingent on specified revision(s); or not approved. PIs whose proposal(s) are not approved will be asked to wait for the next Request for Application (RFA) cycle to submit a new application. PIs will be able to follow the status of their proposal on-line during the review process by logging into the *NCDB PUF Application Manager*.

7. PIs with accepted proposals will be notified when PUFs are ready to be downloaded. Files will be available through the *NCDB PUF Application Manager* only after the PI

electronically signs the Data Use Agreement (DUA). NCDB PUFs are delivered as compressed and encrypted text files, and a file-specific password is provided. *7-zip*, available for free at http://www.7-zip.org/download.html, is necessary to open these compressed files. PUFs are accompanied by SAS, SPSS, and STATA import scripts as well as informational documents describing PUF elements, changes from previous application cycles, and appropriate use of data. PUFs are available to PIs for 3 months following notification of their availability and may be downloaded up to three times within this period of time.

If you have any questions please contact NCDB technical staff at: NCDB_PUF@facs.org.

REQUIRED APPLICATION ELEMENTS

NCDB Participant User Files

The application for a NCDB PUF will require that each of the elements listed below to be completed. PUF Applications should be completed on-line, using the *NCDB PUF Application Manager*. Investigators may return at any time to complete, revise, or update an application until the application deadline. No facsimile or mailed copy of the application will be accepted.

1 - Research Project

Title 2 - Principal

Investigator:

Name; E-mail; CoC-Accredited Program;
Biosketch (2 pages or less).
Letter of Support from the Cancer Committee Chair of your cancer program (on facility letterhead). If you are the Chair, please submit a letter or support from the Cancer Program Administrator.
Terms of Agreement with electronic signature

3 - Co-Investigator(s) [members of the research team]

Name; E-mail; Institutional Affiliation; and Planned Role in the Proposed Research Project.
Biosketch (2 pages or less for each).
Please specify who will be conducting the data analysis.

4 – Disease Site and Patient Age Cohort of Interest

5 – Research Questions/Objectives (min. 50 characters, max. 1000 characters)

6 - Background (min. 50 characters, max. 1500 characters)

7 - Describe your analysis plan: NCDB needs to determine that your proposal is feasible given the data items available in the PUF file. Therefore, describe your analysis plan, naming the variables that are key to your analysis (see PUF Data Dictionary http://ncdbpuf.facs.org/). Please note that the PUF data do not include cancer specific mortality or recurrence data. (min. 50 characters, max. 2000 characters)

8 - Citations (5 or less from Background Research, min. 100 characters, max. 2000 characters)

9 - Describe prior experience in using and analyzing large data sets (min. 50 characters, max. 1000 characters)

10

11 – If you are a previous PUF participant, list, using complete citations, all publications using NCDB PUF data that you have authored.

12 - With which of the following commonly available data sets have you had experience?

NCDB Participant user File; SEER Medicare File; NAACCR CINA Deluxe; NSQIP; SEER

Public Use File; Other

13 - What statistical software package are you planning to use?
SAS; SPSS; Stata; SPLUS; Other

14 - How are you planning to fund or support your proposed PUF study?
Funded by Grant; Department Monies; Other

15 - How long do you anticipate the proposed study will take to complete?
3 Months; 6 Months; 9 Months; 1 Year, More than 1 Year

User Agreements

The Terms of Agreement Data Use Agreement are signed electronically within each application by the PI. These are provided in this document as a reference.

American College of Surgeons' Commission on Cancer
NCDB Participant Use File (PUF) Purpose and Terms of Agreement

The Participant Use File (PUF) program supports investigators at CoC-accredited institutions to conduct research derived from the National Cancer Data Base (NCDB). Specific research proposals are submitted designed around data that can be derived from the NCDB, a joint program of the American College of Surgeons Commission on Cancer (CoC) and the American Cancer Society (ACS), and is offered as an added value to clinical investigators at CoC-Accredited cancer programs who desire to conduct their own studies. Since the PUF is a perquisite of CoC accreditation, the clinical investigator of a facility that loses or withdraws from accreditation for any reason, will agree to destroy the data file(s) and contact ncdb_puf@facs.org and confirm the file(s) destruction.

The aim of the CoC and NCDB is to position investigators at CoC-accredited facilities to successfully use the PUFs to conduct relevant cancer research and should not be used to promote marketing such as using the file to compare your facility's practice to the hospitals contributing data to this PUF. The user also shall not sell, rent, loan, or otherwise grant access to the PUF files to anyone outside of their hospital without permission of the CoC NCDB.

Prior to planning to submit a PUF, you are advised to read the information provided on the Participant User Files web site. It is important that you also read the Getting Started Document, in order to understand the variables and their limitations that could impact your proposed research.

PUF applications must be focused on a specific research question that should be stated clearly in the application. PUF applications requesting large data sets that are not directed at answering a specific research question will not be considered.

PUF applications will only be entertained during the semi-annual Request for Applications (RFA) timeframe. Please note that targeted PUFs for the January 2019 RFA are only available for diagnosis years 2004-2016.

Available PUF Types:

- Pediatric: Ages 0-17
- Pediatrics / Young Adult: Ages 0-39
- Adult: Ages 18-90+
- All Ages: 0-90+

I am the Principal Investigator, and I will share these Terms of Agreement with all of the Co-Investigators, Statisticians, and Data Analysts on this proposal. If you are NOT the Principal Investigator, do NOT complete this Terms of Agreement.

I am the Principal Investigator

o Yes

I will share the Terms of Agreement

o Yes

PUF applications will be screened by NCDB staff for technical feasibility. If the data requested are sufficient to power the aims of the research question, and the research question is well defined and fits the criteria of the PUF program, the application will be approved. If there are any questions by the NCDB staff on the acceptability of an application, it will be forwarded to the Chair of the QIC Research Sub-committee for their review and decision.

NCDB PUFs are delivered electronically as compressed, encrypted text files that are accompanied by SAS, SPSS and STATA import scripts. Prior to the download of the file, a Data Use Agreement (DUA) must be signed electronically before the data can be received.

The (DUA) will specify rules for data use and must be signed prior to your download of the data. Any data breach will be subject to investigation by the CoC. If confirmation of data breaches exists, the investigator will be subject to future barring from receiving NCDB data as part of the PUF process or related processes. The institutional cancer program administrator, cancer committee chair, and cancer liaison physician will be notified. If multiple data breaches are found from an institution, investigators from that institution will be barred from receiving NCDB data as part of the PUF or related processes until a corrective action plan has been received and approved by CoC.

Do you agree to and understand this provision?

o Yes

PUF users are not allowed to share data outside of their facility.

You may share data within your facility for the same application if the investigators are included as Co-Investigators in your application. If an investigator at your institution wants to analyze PUF data for a different analysis than what you are proposing, they will need to submit a separate PUF application.

I agree not to share data outside of my CoC accredited facility.

o Yes

I am not sharing data inside my institution; OR, I am sharing data within my institution only with people that are listed on my PUF application and a bio sketch has been uploaded electronically with the application and who have agreed to the stipulations in the Data Use Agreement.

Do you agree to and understand this provision?

o Yes

If you are granted a PUF, you are prohibited from attempting to identify hospitals and/or patients in the PUF.

Do you agree to and understand this provision?

o Yes

If you receive a PUF, you must include the following disclosure in any presentation or published material using the PUF:

The National Cancer Data Base (NCDB) is a joint project of the Commission on Cancer (CoC) of the American College of Surgeons and the American Cancer Society. The CoC's NCDB and the hospitals participating in the CoC NCDB are the source of the de-identified data used herein; they have not verified and are not responsible for the statistical validity of the data analysis or the conclusions derived by the authors.

Do you agree to and understand this provision?

o Yes

The NCDB PUF data are NOT population based; they are hospital based. I will not refer to the NCDB data as population based in any presentations or publications.

Do you agree to and understand this provision?

o Yes

Upon publication or presentation of any manuscripts utilizing PUF data, researchers are requested to submit a copy of the manuscript to the NCDB.

Do you agree to and understand this provision?

o Yes

If at a future date, I want to add additional investigators, statisticians or data analyst to this proposal, I will send a request to do so to NCDB_PUF@facs.org

o Yes

If at a future date, I want to revise the proposal I have submitted I will send an email to NCDB_PUF@facs.org with a summary of the proposed changes.

o Yes

Electronic Signature By

Application ID:

Commission on Cancer's National Cancer Database Participant User File (PUF) Data Use Agreement

It is of utmost importance to ensure the confidentiality of patients who have been diagnosed with cancer, and the institutions that have reported these cases to the NCDB. This Data Use Agreement ("Agreement") by and between the American College of Surgeons ("ACS") and the Principal Investigator listed below implements the data protections of the Health Insurance Portability and Accountability Act of 1996 (HIPAA) and the American College of Surgeons, Commission on Cancer Business Associate Agreement. Any individual ("Principal Investigator (PI)") seeking to obtain or use the data in the Commission on Cancer's (CoC), National Cancer Data Base (NCDB) Participant User File ("PUF") must be representing a CoC-accredited cancer program ("Program"), must agree to the terms in this Data Use Agreement (DUA), and must submit this DUA electronically prior to the release of the designated PUF.

Consistent with the CoC Business Associate Agreement (BAA) and in accordance with HIPAA, any Protected Health Information ("PHI"), as defined by 45 CFR 164.501, submitted to the CoC NCDB by the Program for inclusion in the NCDB may be used and disclosed to create a de-identified data set as defined by 45 CFR 164.514(b), containing aggregate, patient-level data, as it is with the PUF. The NCDB PUF also retains the right to release the PUF at its discretion to CoC-accredited programs for the purpose of engaging in research activities, quality improvement analysis, and aggregate statistical reporting.

No Identification of Person(s) and Providers: The data in the PUF have been de-identified to the extent that they are no longer a limited data set as defined by Health Insurance Portability and Accountability Act (HIPAA). Any effort to determine the identity of any individual, (including but not limited to patient and other health care provider or Program), or to use the information for any purpose other than for research, quality improvement, and aggregate statistical reporting, would violate the BAA, the conditions of this Agreement, and HIPAA. PIs of the PUF are prohibited under HIPAA, the BAA, and this Agreement (DUA) from releasing, disclosing, publishing or presenting any individually identifying information. Every effort has been made to exclude patient and institutional identifying information from the NCDB PUFs. The NCDB PUF omits from the PUF all identifiers required to be excluded from de-identified data sets as defined by the HIPAA Privacy Rule. It may be possible, in limited situations, through deliberate technical analysis, and with outside information, to ascertain from the de-identified data set the identity of particular persons. Considerable harm could result if this were to occur. Therefore, any attempts to identify individuals are prohibited and information that could identify individuals directly or by inference must not be released or published. Any questions about the PUF must be referred exclusively to the NCDB staff.

Permission to use and disclose the data: Permission to use and disclose the PUF is granted from ACS to each PUF Recipient. The PI must be associated with a CoC-accredited Program. Intentional misrepresentation of association with the Program by the PI will void this Agreement, prohibit future access to the PUF by the PI, and result in legal action taken by ACS. ACS also reserves the right to deny access to the PUF at its discretion.

Expectations: In order for the ACS to provide a PUF, it is necessary that the Principal Investigator (PI) agree to the following expectations:

1. Each PUF PI must communicate with the NCDB staff if the analyses proposed in their awarded application are changed or modified.
2. Each PUF PI is expected to participate in assisting in the development and refinement of the NCDB PUF Data Dictionary for future PUFs.
3. Each PUF PI is expected to act as a resource or mentor to future PUF recipients.
4. Each PUF PI is encouraged to submit abstracts developed for review at least 5 business days prior to the abstract submission deadline. If a PUF PI chooses not to submit an abstract for review by NCDB staff, then the PI and researchers assume all responsibility for any judgment regarding the appropriateness of the statistical review and the discussion/conclusions discerned.

By signing this Agreement, the PI warrants that he/she will:

1. Use appropriate safeguards to prevent use or disclosure of the information other than as provided by this Agreement. This responsibility extends to sharing the data with other researchers involved with the proposal.
2. Assure that anyone participating with the PUF will be affiliated with a CoC-accredited program as the privilege of using a PUF is extended only to researchers in CoC-accredited programs
3. Not use the data in any way other than for research activities, quality improvement analysis, and aggregate statistical reporting for research purposes. The CoC must be notified if it is discovered that there has been any other use of the data.
4. Suppress reporting of small cell sizes (n<10).
5. Not attempt to link nor permit others to link the data with another database.
6. Not attempt to learn the identity of any person or any cancer program whose cancer data is contained in the provided file(s). If the identity of any person or any cancer program is discovered inadvertently, then the PIs must do the following:
    6.1. Will not use this knowledge.
    6.2. Will notify the Medical Director of the CoC of the incident.
    6.3. Will not disclose to anyone else the discovered identity.
7. Secure IRB permission, if a hospital or cancer program requires IRB review, in advance of using the PUF. Note: Some major clinical journals require a statement regarding this.
8. Include in all presentations/published materials the following statement: "The NCDB is a joint project of the Commission on Cancer of the American College of Surgeons and the American Cancer Society."
9. Also, include the following acknowledgement in all presentations/publications: "The data used in the study are derived from a de-identified NCDB file. The American College of Surgeons and the Commission on Cancer have not verified and are not responsible for the analytic or statistical methodology employed, or the conclusions drawn from these data by the investigator."
10. Provide, at a minimum, an abstract and reference for any published materials resulting from their analyses to the Senior Manager, CoC NCDB.
11. Not copy or distribute the data to any parties not identified in the research proposal, nor share the PUF with people outside of your hospital.
12. Not profit from the sale or use of the data as the PUF is the property of the American College of Surgeons.

13. Report to the American College of Surgeons any use or disclosure of information from the PUF not provided for in this Agreement within three (3) business days of becoming aware of such use or disclosure.
14. Ensure that any agents, including a subcontractor, to whom he/she provides any information from the PUF, agrees to the same restrictions and conditions that, apply to you under this Agreement.
15. Not use or disclose the PUF other than as permitted by this Agreement or as otherwise required by law. PI agrees that this precludes him/her from using the data for any commercial purposes.

Additional Terms:

1. The Data Recipient agrees to indemnify the ACS and its employees and agents from any liability, claims, or expenses arising from use of the PUF by the PI;

2. Except as provided below, this Agreement will remain in effect as of the date of execution and shall terminate when all copies of the PUF are destroyed and the PUF is no longer in use. PI shall notify ACS in writing when it has destroyed the PUF.

3. Any noncompliance by the PI with the terms of this Agreement or failure on the part of the PI to correct any breach or violation of this Agreement to the satisfaction of the CoC will be grounds for immediate termination of the Agreement by the American College of Surgeons, CoC.

4. The NCDB does not allow researchers to compare any data from their institution to the NCDB data, whether that be survival data, treatment or other data. In addition, you are prohibited from identifying any facility in the PUF data.

5. Since the PUF is a benefit of CoC accreditation, should the facility under which you applied for the PUF lose or withdraw from CoC accreditation for any reason, you agree to destroy the PUF data file(s) immediately.

As the undersigned, my signature confirms my affiliation at the participating CoC-accredited Program and my agreement to comply with the above stated requirements. Violators of this Agreement may also be subject to penalties under statutes that may apply to these data.

Any Inquiries about this Agreement can be sent to the American College of Surgeons:

ncdb_puf@facs.org
American College of
Surgeons Commission
on Cancer
633 N. Saint Clair Street
Chicago, IL 60611

Getting Started with the 2016 PUF

Review these pages for an overview of the data provided in the PUF, along with guidelines on how to use specific data elements. The Data Dictionary items and definitions are provided in a separate document.

Getting Started with the 2016 PUF

NCDB Participant Use File (NCDB PUF)

This document is designed to offer current and potential PUF investigators some basic guidelines and recommendations for how to approach the data provided in the NCDB PUFs. This document specifically addresses the PUF investigators as the primary reader.

Getting Started with the PUF Data Set

The Participant Use File (PUF) is pulled from the National Cancer Data Base (NCDB), a joint program of the American College of Surgeons Commission on Cancer (CoC) and the American Cancer Society (ACS), and is offered as an added value to clinical investigators at CoC-Accredited cancer programs who desire to conduct their own studies. The aim of the CoC and NCDB is to position investigators to successfully use the PUFs. There are a number of resources available to investigators. Chief among these is an on-line and publicly accessible PUF Data Dictionary (http://ncdbpuf.facs.org/), which has been developed as a resource to investigators using the NCDB PUF data files. Before proceeding with analyses, all investigators are encouraged to familiarize themselves with the on-line PUF Data Dictionary. This dictionary provides a wealth of information designed to facilitate the analytic use of the data items in the PUFs. This document is designed to expedite investigators' familiarization with the PUF data, and provides a list of the registry data items that should be considered when defining the analytic cohort of patients for proposed analyses. This introduction does not serve as a replacement for the on-line PUF Data Dictionary, which should be consulted before data analysis begins. Additionally, the NCDB provides a reference list of recent abstracts and publications that have been developed from PUF files, and categorized by cancer site.

Patients Included in the NCDB PUF Data Sets

Distributed PUFs are organ specific, based upon specific ICD-O primary site and histology combinations, and should be sufficient to address the study proposals described in reviewed applications. The site-histology combinations used to select the cases provided in the PUF data sets are documented: http://seer.cancer.gov/siterecode/icdo3_dwhoheme/index.html. Investigators are encouraged to run preliminary frequency distributions of all relevant data items, including primary site and tumor histology, in order to appropriately define the patient inclusion and exclusion criteria specific to their proposed project. The PUF data only include patient data from facilities that are currently accredited. VA and DoD facilities are excluded from PUF files. Facilities located in Puerto Rico are also excluded.

Please note that the NCDB PUF data are hospital based, NOT population based. Do not refer to the NCDB data as population based in any presentations or publications.

Reference to change in PUF file content due to active CoC Accreditation status:

If you had a previous PUF file for the same site as your new file, you may notice that the number of cases in the new file has decreased, especially for older diagnoses. The data you received this year are limited to cases reported by currently-accredited CoC hospitals. Cases reported by hospitals that are no longer accredited are excluded. Case reports for hospitals that are not currently accredited are not updated in the NCDB data, and their quality cannot be assured. The principal effect will be on cases more than 5-10 years old.

Data Items

Patient Identifiers

In compliance with the Health Insurance Portability and Accountability Act (HIPAA) regulations, PUFs have been stripped of all direct patient identifiers, de-identified according to the "Safe Harbor" rules. [1] The case identification number contained within is randomly assigned, and will change with each PUF version release. The PUF Case IDs are not the same across cancer sites, and cases cannot be linked across cancer sites. In the adult PUF file, pediatric patients have been excluded*, and patients 90 years of age or older are collapsed into the age group 90+ to maintain confidentiality. Facility Location and Facility Type are suppressed for cases aged 0-39. All dates have been removed and replaced with measures of elapsed time. The only date value appearing in the PUF is the year of diagnosis. The narrowest geographic unit available is that of the US Census Division in which the reporting hospital is located. Hospital identity has also been masked. For additional details, please see the Data De-Identification and Confidentiality document posted in the PUF on-line Data Dictionary. [2]

Year of Initial Diagnosis

The 2016 PUF Release includes data for patients diagnosed 2004 through 2016. The year of diagnosis should be used to select patients appropriate to the timeframe of the planned analysis. The availability of some data items is determined by diagnosis year, and not all data items in the PUF are available throughout the entire ten-year span of the PUF. To verify availability of data items by diagnosis year, be sure to review the description of each item in the on-line Data Dictionary.

NCDB PUFs		
File Name	**Diagnosis Years Included**	**Date of Release**
α Test Release	1998 – 2007	April 2010
βPUF Release	1998 – 2010	August 2012
2011 PUF Release	1998 – 2011	August 2013
2012 Semi-Annual PUF Releases	1998 – 2012	Fall 2014 and Spring 2015
2013 Semi-Annual PUF Releases	2004 – 2013	Fall 2015 and Spring 2016
2014 Semi-Annual PUF Releases	2004 – 2014	Fall 2016 and Spring 2017
2015 Semi-Annual PUF Releases	2004 – 2015	Fall 2017, Spring 2018, Fall 2018
2016 Semi-annual PUF Releases	2004-2016	Winter 2019

PUF Multiple Source Item

All CoC accredited programs that initially diagnose a patient or that provide all or part of first course treatment report the case to the NCDB. If more than one facility submitted a report, the "best" is provided in the PUF file (PUF_MULT_SOURCE variable, coded 1), based on the most recent patient contact with the program, completeness of coded detail and/or edit quality, where differences exist. The record used in the case of ties is arbitrary. If this item is coded 0, only one facility provided a report for this cancer. This item should be used for hospital level comparisons using surgical volume, treatment, distance, or other hospital level computations in order to take into account cases treated at more than one hospital. Researchers can choose to limit hospital level analyses to only cases that received treatment at one CoC facility, or may choose to only include variables that indicate treatment was received at the facility included in the PUF. Researchers are encouraged to consult with NCDB staff to further clarify any questions regarding duplicate records and treatment in more than one facility.

Reference Date Flag

Every facility has a reference date, from which they are accountable for the completeness of the data for cases diagnosed in that year through the present. Since a facility may request to move their reference date forward, there are some instances where a case's diagnosis year falls before the facility's reference date. This item, REFERENCE_DATE_FLAG, is coded 0 in cases where this occurs. A 1 signifies cases where the diagnosis year is on or after the reference date year. Reports for cases whose diagnosis date is prior to the reference date cannot be changed or updated by the facility. For this reason, PUF researchers may choose to omit cases where the diagnosis date precedes the reference date, depending on the nature of the study. Note that, depending on diagnosis year and cancer site, excluding cases with diagnosis year preceding the reference year may omit greater than 40% of cases.

Sequence Number

The data item Sequence Number refers to the sequence of malignant and non-malignant tumors diagnosed in a patient and is used to distinguish cases with multiple cancer diagnoses. By default, your PUF includes all sequence codes available for each reported patient. Patients with only one lifetime cancer diagnosis will have a sequence number code value of 00. Sequence number 01 indicates that the reported tumor is the first of multiple diagnoses. The NCDB has no mechanism by which to link separate case reports of the same patient. It is customary to limit analyses to patients with sequence numbers 00 and 01 to ensure that any review of treatment or outcomes of the study cohort is not confounded by treatment administered for a prior cancer diagnosis. It is not uncommon to encounter high sequence numbers, especially among melanoma patients.

Behavior

The PUF includes in situ or non-invasive (behavior code 2) and malignant or invasive (behavior code 3) primaries. Non-malignant or borderline cases (behavior codes 0 and 1) are only available for primaries of the intracranial and central nervous system tumors.

Class of Case

The PUF only includes "analytic cases" whose initial diagnosis and/or treatment were/was performed at the reporting facility. Class of Case 00 denotes cases diagnosed at the reporting facility that did not receive any treatment at that facility. Class of Case 10-14 are cases that were initially diagnosed and provided all or part of their treatment at that facility. Class of Case codes 20-22 are those patients that were diagnosed at another facility and received all or part of their treatment at the reporting facility. If the focus of a proposed project is treatment, it would be standard practice to exclude Class of Case 00 cases

from your study cohort. The Commission on Cancer does not require follow-up for Class of Case 00 cases, so they also should be excluded from survival studies.

Cancer Program Category

The PUF is limited to cancer programs currently accredited by the CoC. Facility Type provides a general classification of the reporting facility's structural characteristics, and defines a portion of the criteria required for CoC Accreditation. PUFs identify reporting facilities as one of four types: Community, Comprehensive, Academic/Research hospitals, or Integrated Network Cancer Programs. These categories follow the classification scheme used by the CoC accreditation program, and are determined by a variety of factors. If you are including facility type in your analyses, be aware that facilities in the Integrated Network Cancer Program (INCP) category are comprised of many types of facilities (such as Academic, Community, etc.), but are assigned the INCP type in the PUF data when they join a Network. . Cases reported from Veterans Affairs and Department of Defense facilities are not included in PUF data files. For additional details on data that are suppressed or cases that are omitted, please see the Data De-Identification and Confidentiality document posted in the PUF on-line Data Dictionary. This item is suppressed for cases aged 0-39.

Census and Urban/Rural Data

Area-based or environmental measures of patient income and education are provided in the PUF. These measures are derived by linking the reported ZIP code of the patient's residence at the time of diagnosis to year 2000 Census data. The data describing median household income and level of educational attainment represent the ZIP code of patient residence, not that of individual patients. Since the Census only uses the short form as of 2010, the majority of the information traditionally collected by the decennial census is now collected in the American Community Survey (ACS). The PUF will include the most recent ACS data released as of April 2016, which consists of survey years 2012-2016. The 5-year datasets are not just an average of each year in the period; the final estimate uses several weighting methods, among other adjustments. Items added are 2012-2016 median household income quartiles (MED_INC_QUAR_2016) and 2012-2016 percent without high school degree quartiles (NO_HSD_QUAR_2016.) More information about these variables are found at :

https://www.census.gov/programs-surveys/acs/guidance.html

The data are extracted from the American Fact Finder website: http://factfinder.census.gov/.

The 2013 Rural-Urban Continuum data are also included in the PUF. The 2003 Rural-Urban data are still included in the PUF, and the labels for the classification codes are the same in the 2003 and 2013 data, so a direct comparison may be made. More information can be found on the United States Department of Agriculture (USDA) website: https://www.ers.usda.gov/data-products/rural-urban-continuum-codes//

Comorbid Conditions

Comorbid disease burden is represented in the PUF as a summary value. This value is based on the Deyo adaptation (1992) of Charlson's comorbidity index and can be used as a mechanism to control for pre-existing medical conditions that may affect treatment decisions. The scores are mapped from as many as ten reported ICD-9-CM or ICD-10-CM secondary diagnosis codes and are summed to create one value for each case, categorized as a total score of 0, 1, 2, or 3 or more. The score included in the PUF is the higher of the ICD-9 or ICD-10 submitted comorbid conditions.

AJCC Stage of Disease

The AJCC clinical and pathologic stage groups included in the PUF are a TNM-based system coded or reported according to the edition corresponding to the patient's diagnosis year. The fifth edition of the AJCC Staging Manual is used to represent patients' cases diagnosed from 1998 through 2002. The sixth edition describes the anatomic extent of disease for patients diagnosed from 2003 through 2009. Patients diagnosed in 2010 are staged according to the seventh edition of the AJCC Staging Manual Data. **Exercise caution when using staging information.** Staging definitions may change between editions of the AJCC staging manuals, and rules delineating "stageable histologies" have become more specific over time, beginning with the 5th edition of the staging manual. Investigators may observe variability with the completeness of reported staging information over time. There have also been changes in CoC program standards and NCDB data reporting requirements in recent years. Rules defining which personnel within a facility were authorized to assign stage and the extent to which registry staff were directed to copy the recorded stage information have been both tightened and relaxed over the years captured in the PUF. In addition, the universal implementation of the Collaborative Stage Data Collection System (CS) in 2004 across all cancer registries in the United States contributed to a reduction in the coding of physician-reported staging in subsequent years.

Site Specific Factors from the Collaborative Stage Data Collection System

Several PUF projects examine one or more laboratory prognostic indicators. These are available as Site Specific Factors (SSF) collected as part of the Collaborative Stage Data Collection System (CS). The term "collaborative" means that the data collection tool was devised to meet the various needs of cancer registry data standard setters such as the Commission on Cancer (CoC), Surveillance Epidemiology and End Results (SEER), and the National Program of Cancer Registries (NPCR).

Up to 25 data fields are used to collect SSFs. Being site specific, they contain different information depending on the type of cancer in the report. For example, for breast cancer reports, SSF1 contains "Estrogen Receptor (ER) Assay" results, but for colon cancer reports, SSF1 contains "Carcinoembryonic Antigen (CEA)" results.

SSFs also may convey non-laboratory site-specific information that is relevant to prognosis for some cases. For example, SSF1 for gastric cancers is "Clinical Assessment of Regional Lymph Nodes", and for melanoma of skin it is "Measured Thickness (Depth), Breslow Measurement".

Some detective work is required to identify the data fields of interest, the applicable codes, and the adequacy of the data for the particular study:

I. The codes, and occasionally the fields used, for a particular prognostic factor have changed over time. In the PUF, the SSF data are retained in the form in which they were submitted. That means that it will be necessary to identify the CS Version Numbers that are used in the PUF file, and use those to identify whether the data contents for the desired SSF may have changed or moved. Links to the site-specific codes can be found at http://ncdbpuf.facs.org/?q=node/370. The CS web sites are maintained by Collaborative Stage Work Group of the American Joint Committee on Cancer. Select the applicable CS Version from the link above, and then select the schema name that applies to the cases in the project.

II. The quality of the SSF data items has undergone minimal review by NCDB, and PUF users are advised to examine the data consistency and completeness of these items carefully before proceeding with the study.

a. All SSF data items are edited for validity and internal consistency before the case report is submitted, and the submitter is required to correct any edit errors. However, some coding errors remain.

b. Case coverage of the SSFs is limited for a variety of reasons, potentially seriously affecting their applicability for some studies.

i. The availability of the measures to hospital registrars at the time of data entry is sparse for many prognostic measures. The source of information is usually the laboratory report as it appears in the hospital patient record. The information may not be available in the hospital if it was requested by a physician and the report was sent to the physician's office. Alternatively, it may be delayed and not picked up later.

ii. The individual tests are not run at all locations or for all patients, even if the test is part of an acknowledged treatment protocol.

iii. Finally, many hospital registries begin abstracting data for the years the measures were introduced prior to the hospital's upgrade of the software essential to collecting those items, and they did not necessarily return to the cases to abstract the missed data. Some of the SSFs were first introduced in 2004, and are underrepresented for cases diagnosed that year compared to later years. Most prognostic SSFs were introduced in 2010, and are certainly underrepresented for 2010 diagnoses; those are not available at all for earlier years.

The SSFs, the versions in which they were implemented, and whether the field was required for CoC registries are described in CoC SSFs for v0205.xlsx. As noted above, the fields in which these items were stored and the codes used may have changed over time. Spreadsheets for earlier versions of CS may be accessed via this link: http://seer.cancer.gov/csreqstatus/application.html (make sure to choose 'CoC' as the Standard Setter).

Treatment

The treatment information provided in the PUF is limited to "first course of treatment", which is defined as all methods of treatment recorded in the treatment plan and administered to the patient before disease progression or recurrence. "No therapy" is a treatment option that occurs if the patient refuses treatment, the family or guardian refuses treatment, the patient dies before treatment starts, or the physician recommends no treatment be given. Treatment plans describe the type(s) of therapies intended to modify, control, remove, or destroy proliferating cancer cells. Cancer registries are directed to review documentation confirming a treatment plan which may be found in several sources. Examples include medical or clinic records, consultation reports, and outpatient records. In addition:

- All therapies specified in the physician(s) treatment plan are a part of the first course of treatment if they are actually administered to the patient.
- A discharge plan must be part of the patient's record in a Joint Commission-approved program and may contain part, or all, of the treatment plan.
- An established protocol or accepted management guidelines for the disease can be considered a treatment plan in the absence of other written documentation.

- If there is no treatment plan, established protocol, or management guidelines, and consultation with a physician advisor is not possible, use the principle: "initial treatment must begin within four months of the date of initial diagnosis."
- The first course of treatment includes all therapy planned and administered by the physician(s) during the first diagnosis of cancer. Planned treatment may include multiple modes of therapy and may encompass intervals of a year or more. Any therapy administered after the discontinuation of first course treatment is subsequent treatment, and is not reported to the NCDB.
- The variable RX_SUMM_TREATMENT_STATUS indicates whether patients received any treatment or are under **active surveillance**. This variable was implemented in 2010 and is not available for prior diagnosis years.
- **Starting with the 2012 PUF (first released in 2014), additional treatment variables were included, which indicate whether treatment was received at the reporting facility included in the PUF**. Versions prior to the 2012 PUF only included summary treatment variables, which indicated whether treatment was received at any CoC accredited facility, including facilities not included in the PUF data (see PUF Multiple Source Explanation above). The new treatment variables include “Surgery at this Facility”, “Chemotherapy at this Facility”, and “Other Treatment at this Facility”.

Distance Metrics

The PUF includes a "crow-fly" or great circle distance measure between the latitude and longitude of the centroid of patient's ZIP code of residence and the latitude and longitude of the facility mailing address. The precision of this item as an indicator of the true distance between two points is dependent upon the spatial area of the ZIP code and the proximity of the facility’s administrative mailing address to the actual treatment center**. If facility level analyses of distance to treatment are conducted using this variable, researchers need to take into account whether the treatment occurred at the facility included in the PUF data, or at a facility not included in the PUF.**

Outcomes

The CoC accreditation standards require an annual 90% follow-up rate for all living, eligible, analytic patients diagnosed within the last 5 years and an 80% follow-up for all eligible analytic cases from the cancer registry's reference date. Participating registries report patient follow-up to the NCDB annually. The PUF data do not include cause of death information, so cause-specific survival cannot be calculated. It is recommended that survival analyses be restricted to patient cohorts with only one reported cancer diagnosis (Sequence Number 00) in order to avoid confounding outcomes with patients who may have been diagnosed and treated for a separate malignancy. Vital Status information is not included for patients diagnosed in 2016 due to the limited follow up for these patients.

The PUF data also include both 30 and 90 day mortality for patients undergoing surgical resection (Surgery Primary Site Codes 20-90). If analyzing these data items at the facility level, the researcher needs to limit cases to those for whom the surgery was performed at the facility, using the variable Surgery at this Facility. The 30 and 90 day mortality items also do not include data for 2016 diagnoses, due to limited follow up.

If calculating survival, note that you may not publish or present any PUF data that compares your facility’s survival to the survival in the PUF. For more information on this policy, please refer to the letter sent to Commission on Cancer facilities in February 2015: **https://www.facs.org/~/media/files/quality%20programs/cancer/ncdb/coc_survivalpublic%20reporting%20policy.ashx.**

User Defined Data Items

Hospital Volume

Beginning in 1998, all CoC-accredited cancer programs are required to submit case reports to the NCDB in response to the annual Call for Data. In the NCDB PUF, facilities are assigned a random ID, *PUF_FACILITY_ID*. This ID is assigned regardless of cancer site, so researchers may identify the same facilities across cancer sites. The number of CoC-accredited cancer programs changes from one diagnosis year to the next. Thus, not all of the hospitals available in the PUF have been accredited for every one of the diagnosis years included in the PUF. If a planned analysis includes the calculation of hospital volume, investigators should recognize that CoC reporting requirements affect the methodological approaches to computing and estimating hospital volume.

Some patients receive treatment in more than one CoC accredited program, and are noted in the **PUF Multiple Source** item; however only one of the cancer programs where diagnosis/and or treatment was received is included in the PUF, in order to exclude duplicate records for the same patient. Thus calculation of hospital volume will not include hospitals with duplicate records that were excluded from the PUF. Approximately 10% of patients in the PUF data have duplicate records, but this percentage varies by cancer site. In addition, the summary surgical procedure of the primary site variable is based on information from the hospital where the surgery was performed, but this is not necessarily the hospital report that is included in the PUF data. The *at this facility* surgery variable may be used to account for surgery performed at that particular facility included in the PUF.

Whether an analysis uses total case volume or surgical volume, the easiest way to begin assigning average volume to each hospital appearing in the PUF is to create an aggregated dataset of the number of cases by hospital and diagnosis year. Such an aggregated file can be used to assess a particular cancer program's CoC accreditation history. Hospitals that have remained accredited throughout the years covered by the PUF pose minimal challenges when attributing volume metrics. If there are observed trends (either upward or downward) or spikes in hospital case counts, investigators may deem it more appropriate to calculate an average volume from the most recent years or a select set of years. Where significant shifts in annual caseloads are observed, investigators might consider recalculating their volume metric using a minimum and maximum volume value for each hospital in the aggregated dataset.

Hospitals that have previously discontinued and subsequently re-established their CoC accreditation throughout the span of diagnosis years available in the PUF will display a seemingly inconsistent or incomplete reporting pattern across years. Investigators should be certain to check their aggregated data set to ensure that computed volume metrics appropriately account for these hospitals.

In addition, researchers need to only include surgeries performed at the facility included in the PUF, by using the variable, Surgery at this Facility.

Citing Data from the NCDB

While citation is largely at the discretion of the author(s), there are four key components of information that must be conveyed on all peer-reviewed publications that draw from NCDB data:

- The NCDB is to be cited as a joint project of the American Cancer Society and the Commission on Cancer of the American College of Surgeons.
- The data used in this study are derived from a de-identified NCDB file. The American College of Surgeons and the Commission on Cancer have not verified and are not responsible for the analytic or statistical methodology employed, or the conclusions drawn from these data by the investigator.
- The American College of Surgeons has executed a Business Associate Agreement that includes a data use agreement with each of its Commission on Cancer accredited hospitals.
- The NCDB, established in 1989, is a nationwide, facility-based, comprehensive clinical surveillance resource oncology data set that currently captures 70% of all newly diagnosed malignancies in the US annually.

Ancillary Data References

- Deyo RA, Clerkin DC, Ciol MA. Adapting a clinical comorbidity index for use with the ICD-9-CM administrative databases. J Clin Epidemiol. 1992;45:613-619.
- AJCC 5th Edition Cancer Staging Manual. Fleming ID, Cooper JS, Henson DE, Hutter RVP, Kennedy BJ, Murphy GP, O'Sullivan B, Sobin LH, Yarbro JW (eds.). Lippincott-Raven Philadelphia, 1997.
- AJCC 6th Edition Cancer Staging Manual. Greene FL, Page DL, Fleming ID, Fritz A, Balch CM, Haller DG, Morrow M (eds.). Springer, New York, 2002.
- AJCC 7th Edition Cancer Staging Manual. Edge SE, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti A (eds.). Springer, New York, 2009.
- Collaborative Staging Manual and Coding Instructions, Version 1.0. Collaborative Staging Task Force of the American Joint Committee on Cancer. NIH Publication Number 04-5496. US Department of Health and Human Services, National Institutes of Health, National Cancer Institute.
- Collaborative Stage Work Group of the American Joint Committee on Cancer. Collaborative Stage Data Collection System User Documentation and Coding Instructions, version 02.04.40. Published by American Joint Committee on Cancer (Chicago, IL).
- Facility Oncology Registry Data Standards (FORDS): Revised for 2013. Phillips JL, Stewart AK (eds.) American College of Surgeons 2002.
- International Classification of Diseases for Oncology (ICD-O), Third Edition. Percy C, Shanmugaratnam K, Whelan S, Parkin DM, Jack A, Fritz A, Sobin L (eds.) World Health Organization, Geneva, 2001.
- Registry Operations and Data Standards (ROADS), Standards of the Commission on Cancer Vol II. Johnson C (ed.). American College of Surgeons, 1998.

1. http://privacyruleandresearch.nih.gov/research_repositories.asp
2. http://ncdbpuf.facs.org/?q=node/275
3. http://www.facs.org/cancer/coc/whatis.html
4. http://www.facs.org/cancer/coc/programstandards2012.pdf
5. Bilimoria KY, Stewart AK, Winchester DP, Ko CY. The National Cancer Data Base: A Powerful Initiative to Improve Cancer Care in America. Ann Surg Oncol. 2008; on-line, Jan 9.

NCDB PUF Primary Site-Histology Groupings for PUF 2016*

Site Name	Primary Site Codes	Histologies
Head and Neck		
Lip	C000–C009	excluding 9050–9055, 9140, 9590–9992
Tongue	C019–C029	
Salivary Gland	C079–C089	
Floor of Mouth	C040–C049	
Gum and Other Mouth	C030–C039, C050–C059, C060–C069	
Nasopharynx	C110–C119	
Tonsil	C090–C099	
Oropharynx	C100–C109	
Hypopharynx	C129, C130–C139	
Other Oral Cavity and Pharynx	C140, C142, C148	
Digestive System		
Esophagus	C150–C159	excluding 9050–9055, 9140, 9590–9992
Stomach	C160–C169	
Small Intestine	C170–C179	
Colon	C180–C189, C260	
Rectosigmoid Junction	C199	
Rectum	C209	
Anus, Anal Canal and Anorectum	C210–C212, C218	
Liver	C220	
Intrahepatic Bile Duct	C221	
Gallbladder	C239	
Other Biliary	C240–C249	
Pancreas	C250–C259	
Retroperitoneum	C480	
Peritoneum, Omentum and Mesentery	C481–C482	
Other Digestive Organs	C268–C269, C488	

Respiratory System		
Nose, Nasal Cavity and Middle Ear	C300–C301, C310–C319	excluding 9050–9055, 9140, 9590–9992
Larynx	C320–C329	
Lung, Bronchus – Small Cell Carcinoma	C340–C349	8040–8045
Lung, Bronchus – Non–Small Cell Carcinoma	C340–C349	8012–8035, 8046–8576
Lung, Bronchus – Other Types	C340–C349	8000–8011, 8580–9582
Pleura	C384	excluding 9050–9055, 9140, 9590–9992
Trachea, Mediastinum and Other Respiratory Organs	C339, C381–C383, C388, C390, C398, C399	
Bones and Joints	**C400–C419**	**excluding 9050–9055, 9140, 9590–9992**
Soft Tissue including Heart	**C380, C470–C479, C490–C499**	**excluding 9050–9055, 9140, 9590–9992**
Skin NOT Basal and Squamous		
Melanoma of the Skin	C440–C449	8720–8790
Other Non–Epithelial Skin	C440–C449	All sites except 8000–8005, 8010–8046, 8050–8084, 8090–8110, 8720–8790, 9050–9055, 9140, 9590–9992
Breast	**C500–C509**	**excluding 9050–9055, 9140, 9590–9992**
Female Genital System		
Cervix Uteri	C530–C539	All sites except 9050–9055, 9140, 9590–9992
Corpus Uteri	C540–C549	
Uterus, NOS	C559	
Ovary	C569	
Vagina	C529	
Vulva	C510–C519	
Other Female Genital Organs	C570–C579, C589	

Male Genital System		
Prostate Testis	C619 C620–C629	excluding 9050–9055, 9140, 9590–9992
Penis	C600–C609	
Other Male Genital Organs	C630–C639	
Urinary System		
Urinary Bladder Kidney and Renal Pelvis	C670–C679 C649, C659	excluding 9050–9055, 9140, 9590–9992
Ureter	C669	
Other Urinary Organs	C680–C689	
Eye and Orbit	**C690–C699**	**excluding 9050–9055, 9140, 9590–9992**
Brain and Other Nervous System		
Brain	C710–C719	excluding 9050–9055, 9140, 9530–9539, 9590–9992
Cranial Nerves Other Nervous System	C710–C719 C700–C709, C720–C729	9530–9539 excluding 9050–9055, 9140, 9590–9992
Thyroid and Other Endocrine		
Thyroid Other Endocrine including Thymus	C739 C379, C740–C749, C750–C759	excluding 9050–9055, 9140, 9590–9992
Lymphoma		
Hodgkin – Nodal	C024, C098–C099, C111, C142, C379, C422, C770–C779	9650–9667
Hodgkin – Extranodal		9650–9667
Non-Hodgkin Lymphoma – Nodal	C024, C098, C099, C111, C142, C379, C422, C770–C779	9590–9597, 9670–9671, 9673, 9675, 9678–9680, 9684, 9687–9691, 9695, 9698–9702, 9705, 9708–9709, 9712, 9714–9719, 9724–9729, 9735, 9737–9738, 9811–9818, 9823, 9827, 9837

Non-Hodgkin Lymphoma - Extranodal	All sites except C024, C098-C099, C111, C142, C379, C422, C770-C779	9590-9597, 9670-9671, 9673, 9675, 9678-9680, 9684, 9687, 9688, 9689-9691, 9695, 9698-9702, 9705, 9708-9709, 9712, 9714-9719, 9724-9729, 9735, 9737, 9738
	All sites except C024, C098-C099, C111, C142, C379, C420-C422, C424, C770-C779	9811-9818, 9823, 9827, 9837
Myeloma		**9731-9732, 9734**
Leukemia		
Acute Lymphocytic Leukemia		9826, 9835-9836
	C420, C421, C424	9811-9818, 9837
Chronic Lymphocytic Leukemia	C420, C421, C424	9823
Other Lymphocytic Leukemia		9820, 9832-9834, 9940
Acute Myeloid Leukemia		9840, 9861, 9865-9867, 9869, 9871-9874, 9895-9897, 9898, 9910-9911, 9920
Chronic Myeloid Leukemia		9863, 9875-9876, 9945-9946
Other Myeloid/Monocytic Leukemia		9860, 9930
Acute Monocytic Leukemia		9891
Other Acute Leukemia		9801, 9805-9809, 9931
Aleukemic, subleukemic and NOS		9733, 9742, 9800, 9831, 9870, 9948, 9963-9964
	C420, C421, C424	9827
Mesothelioma / Kaposi Sarcoma		
Mesothelioma		9050-9055
Kaposi Sarcoma		9140
Waldenstrom macroglobulinemia	C420	9761

*Adapted from SEER Site Recode ICD-0-3/WHO 2008 Definition:
http://seer.cancer.gov/siterecode/icdo3_dwhoheme/index.html