National Program of Cancer Registries Cancer Surveillance System (NPCR-CSS) 2025 Data Release Policy Diagnosis Years 1995–2024

Policy Revised July 2025

Cancer Surveillance Branch
Division of Cancer Prevention and Control
National Center for Chronic Disease Prevention and Health Promotion
CDC

4770 Buford Hwy, N.E., Mailstop S107-4
Atlanta, GA 30341-3717
E-mail: uscsdata@cdc.gov (specify "NPCR-CSS" in subject line)

Table of Contents

Introduction

This document describes the format and content of data that the Centers for Disease Control and Prevention's (CDC's) National Program of Cancer Registries (NPCR) Cancer Surveillance System (CSS) releases or shares. This multi-year policy updates the July 2024 NPCR-CSS Data Release Policy. This policy applies to data submitted to CDC for the 2025 NPCR-CSS data submission and for all future data submissions until a new policy is provided.

The NPCR-CSS Privacy Steward, as authorized by the Chief of the Cancer Surveillance Branch, clears all releases of state and territory data, ensuring that the data are released according to the terms of this policy.

It is possible that, in future years, data release practices or the content and format of released data may vary from those described in these guidelines. Such changes may occur because of improvements in the quality of the data, changes in information technology, and evolving data needs. However, if such variations occur, the data release practices will provide comparable protection (or more protection) for patient confidentiality to that described in this policy. If it is anticipated that any data will be released with less protection (as determined by the NPCR-CSS Privacy Steward) for patient confidentiality than described in this policy, NPCR central cancer registries will be notified and have ample time to respond before the data are released. This policy is reviewed annually by the NPCR-CSS Privacy Steward and other appropriate CDC staff members to determine whether revisions are needed.

Summary of Changes

- In the Analytic Datasets, updated the information to indicate that the survival and prevalence databases will now be combined, rather than separate files (page 9).
- Updated data item lists for Data Items for CBTRUS (Appendix C), NPCR/SEER USCS Incidence Analytic Dataset (Appendix F), NPCR Internal Survival Dataset (Appendix G), NPCR/SEER USCS Incidence Public-Use Research Dataset (Appendix J), NPCR/SEER USCS Public Use Research Database User Agreement (Appendix K), Data Items for Restricted-Access Dataset (Appendix M), and Data Items for NPCR Prevalence Database (Appendix O).

Overview of Data

In 1992 Congress established NPCR by enacting the Cancer Registries Amendment Act, Public Law 102-515.[1] The law authorized CDC to provide funds and technical assistance to states and territories to improve or enhance existing cancer registries and to plan for and implement population-based central cancer registries where they did not exist. NPCR's purpose is to assure the availability of more complete local, state, regional, and national cancer incidence data to plan and evaluate cancer control interventions and for research. NPCR adopted reporting requirements and definitions consistent with the National Cancer Institute's (NCI) Surveillance, Epidemiology, and End Results (SEER) program;[2,3] required the use of uniform data items, codes, and record layouts as defined by the consensus of members of the North American Association of Central Cancer Registries (NAACCR);[4] and established standards for data management and data completeness, timeliness, and quality similar to those NAACCR recommended.[4,5] In 1994, the first 37 states received funding from CDC.[6] Currently, 46 states, the District of Columbia, Puerto Rico, the U.S. Virgin Islands, and the U.S. Pacific Island Jurisdictions are funded by NPCR (Appendix A).[7] NPCR-funded central cancer registries collect data on patient demographics, primary tumor site, morphology, stage of disease at diagnosis, and first course of treatment. In addition, NPCR central cancer registries conduct follow-up for vital status by linking with state and national death files or active case follow-up.

Invasive and *in situ* cancer case reports are submitted to CDC by population-based statewide central cancer registries in all 46 participating states, the District of Columbia, Puerto Rico, Virgin Islands, and the U.S. Pacific Island Jurisdictions. In each state or territory, laws and regulations mandate the reporting of cancer cases by facilities and practitioners who diagnose or treat cancer to the state or territory health department or its designee. The central cancer registry receives case reports from facilities and practitioners throughout the state or territory and processes them according to standard data management procedures.[5] Personal identifiers including the patient's name, Social Security number, and street

address are removed from the NPCR-CSS submission prior to the encryption and electronic transmission of these case reports to a contractor acting on behalf of CDC. CDC and the contractor adhere to strict data security procedures when receiving, processing, and managing the data (Appendix B). CDC has an Office for Human Research Protections (OHRP)-approved, federal-wide assurance of compliance with rules for the protection of human subjects in research (45 Code of Federal Regulations 46, available at www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/). NPCR-CSS received formal approval (protocol #2594) from CDC's Institutional Review Board (IRB) in October 1999 and annual approval was approved through 2020. In 2021, under the Common Rule (45 Code of Federal Regulations Part 46, Common Rule 2018), NPCR-CSS was deemed to be a non-research public health surveillance project and annual IRB review is no longer necessary.

Central cancer registries and federal agencies routinely publish cancer incidence data 23 months after the close of each diagnosis year based on data that meet data quality standards.[7,8] However, other versions of the same data, based on the data file as it exists at different time periods, are usually available. For example, some central cancer registries have preliminary data available as soon as 12 months after the close of each diagnosis year. After the publication of official statistics, central cancer registries, as well as CDC and the National Cancer Institute (NCI), continue to update and republish data with new information incorporated. When cancer incidence data are published, it is common practice to document either the data submission date (when the data were submitted to CDC or NCI) or the date that the file was prepared. Changes in central cancer registry incidence data that occur more than 23 months after the close of a diagnosis year are likely to be small; however, delays in reporting are more likely to affect certain cancer sites and may be important for some research studies.[9]

CDC generates multiple data products using NPCR-only data and combined NPCR and SEER data. The combined NPCR and SEER data are referred to as U.S. Cancer Statistics (USCS). USCS is the official federal cancer statistics, providing the most up-to-date information on the U.S. population.

Data Release Activities

As described in DP22-2202, participation in all CDC-created and hosted analytic datasets and web-based data query systems, as outlined in this policy, is a required strategy and activity.[10] Therefore, the NPCR-CSS Dataset Participation Agreement is no longer provided.

Public Web-based Query Systems

For purposes of this policy, public web-based query systems are defined as datasets that are comprised of aggregated data (not individual case-specific data or microdata) that have been modified according to accepted procedures to block breaches of confidentiality and prevent disclosure of the patient's identity or confidential information. They are stored in a database behind a CDC firewall that is either case-specific microdata or pre-analyzed data tables.[11, 12-17] Users can access only aggregate counts and rates with all confidentiality protections built in. A combination of confidentiality protection measures is employed for each public web-based query system (see Table 1). These systems do not contain information that is identifiable or potentially identifiable according to currently accepted procedures for reducing disclosure risk.[11, 12-17] Before each system is finalized, the aggregate values are analyzed to determine whether there is a need for complementary cell suppression.[11, 12-17] If appropriate, the analysis includes consultation with a statistician with expertise in statistical disclosure limitation techniques. Following the analysis, complementary cell suppression is applied as needed.

There are no restrictions on access to public web-based query systems. A public release disclosure statement (see Public Release Disclosure Statement) cautions users against inappropriate use of the data or inappropriate disclosure of information. Data are released as delimited ASCII files, a web-based query system, or possibly through other data products (see Table 1). As a convenience to NPCR central cancer registries, states and territories may request from CDC a copy of their complete state- or territory-specific analytic database that is used to populate each public web-based query system. The following public web-based query systems are currently being released:

- USCS Data Visualizations tool
- CDC WONDER – USCS incidence and incidence/mortality rate ratios
- Federal partners' web-based query systems:
  - NCI's State Cancer Profiles
  - CDC's Environmental Public Health Program's Tracking Network
  - CDC's Chronic Disease Indicators (CDI) tool
  - CDC's Melanoma Dashboard

All NPCR-CSS public web-based query systems consist of cancer incidence data selected from the NPCR/SEER analytic database. This is the same database that provides cancer incidence data for the annual release of USCS data products, including the Data Visualizations tool, public use database and State Cancer Profiles. Data sources, case definitions, basic registry eligibility criteria in terms of required data quality, population denominator sources, methods for calculating incidence rates, and the rationale for specific cell suppression thresholds are as described in the USCS Data Visualizations Tool Technical Notes, unless noted in separate documentation that accompanies the data.

Separate documentation may accompany each data product that describes its unique features, such as the data submission date, percentage of the U.S. population covered, diagnosis years and cancer sites included, variables included, data suppression rules, any special data quality criteria required for inclusion, and any unique statistical methods employed.

USCS Data Visualizations Tool

The USCS Data Visualizations tool is a web-based application built with D3 Javascript libraries, a React framework, and web application programming interfaces (APIs) that outputs data in hypertext markup language (HTML) containing the aggregate counts and rates for incidence, mortality, prevalence, and survival estimates published annually, along with text documentation and data visualizations. The tool is available at www.cdc.gov/cancer/dataviz.

It displays single year and 5-year aggregate counts, age-adjusted rates, and 95% confidence intervals by primary site, sex, race and ethnicity (see Table 1 for race and ethnicity categories) at the county (5-year aggregate), Congressional district (5-year aggregate), state or territory, and national levels. Congressional district estimates (estimated 5-year aggregate counts and age-adjusted rates) are presented by sex, race, and ethnicity (all races/ethnicities, non-Hispanic White, Black, and Hispanic). In addition, cancers grouped by associated risk factors are presented by state, sex, race, and age-group (single year and 5-year aggregate) in Data Visualizations tool. National and state prevalence, data by stage at diagnosis, and survival by stage for select sites are presented by sex, race/ethnicity, and age at the national and state level. Stage at diagnosis is categorized as localized, regional, distant, and unknown or unstaged. Preliminary and delay-adjusted incidence rates and counts, as well as other new indicators, may be published in the tool.

The Data Visualizations tool's database is behind a CDC firewall with pre-tabulated data created using SEER*Stat queries, which allows for the display of counts and rates that meet suppression and confidentiality protections. Users can access only aggregate counts and rates with all confidentiality protections built in. Downloadable ASCII files with the pre-tabulated data are available from CDC's website. States and territories may request a state- or territory-specific web API.

CDC WONDER – USCS Incidence, Mortality, and Incidence/Mortality Ratios

The USCS dataset available on CDC WONDER (https://wonder.cdc.gov/cancer.html) displays the aggregate incidence and mortality counts, rates, and 95% confidence intervals by primary site, sex, race, and ethnicity at the state, county, regional, metropolitan statistical area (MSA), and national levels. Cancer incidence and mortality rate ratios are also available by year, state, MSA, race, ethnicity, sex, and cancer site. The WONDER database is stored behind a CDC firewall with case-specific microdata. Users can access only aggregate counts and rates with all confidentiality protections built in.

CDC WONDER allows users more flexibility in creating cross-tabulations than the Data Visualizations tool. While the same underlying USCS data are available in both tools, more detailed breakdowns of counts and rates are available

through WONDER. The additional values result from variable selections that are not currently available in the Data Visualizations tool (see Table 1) and include results for MSAs that have a population of 50,000 or higher and standard 5-year age groups that the user can combine.

Federal Partners' Web-based Systems

CDC shares aggregated data with federal partners for display in their web-based query systems. The data are generated specifically for the partners' needs and are shared via ASCII files.

Unless otherwise noted below, the data generally consist of aggregate cancer incidence counts, crude rates, and age-adjusted rates for selected primary sites, age groups, and counties in the United States (see Table 1 for more details).

Future versions may contain more detail about cancer at the county level. Beginning in 2008, CDC began routinely publishing county data averaged over 5 years.

State Cancer Profiles

State Cancer Profiles is a web-based query tool that public health professionals and others can use to prioritize cancer control efforts at the county, state, and national level. Data are released to NCI's Surveillance, Epidemiology, and End Results (SEER) program for the State Cancer Profiles data product, which presents average annual counts and age-adjusted incidence and mortality rates only.

Environmental Public Health Tracking Program

USCS data are provided to CDC's National Center for Environmental Health's Environmental Public Health Tracking Program for display on a public facing web-based query system, called the Tracking Network. The Tracking Network uses a database behind a CDC firewall with case-specific microdata to display incidence counts, age-adjusted rates, and 95% confidence intervals, as well as other indicators (e.g., standardized incidence ratio), for selected primary sites and age groups for selected geographic areas and time periods (see Table 1). Suppression will follow the NPCR-CSS standard with results suppressed if the observed case count is <16. Single-year and 5-year aggregate data can be viewed at the state level. Data, including smoothed rates, by 5-year aggregation are available at the county level.

Incidence counts and age-adjusted incidence rates for select cancers are displayed for various sub-county geographies. Incidence rates are based on incidence counts stratified by census tract, year of diagnosis, age group (standard 19 groups), sex, and Census-based population estimates. These incidence counts and rates are displayed using aggregation schemas recommended by a sub-county cancer data workgroup: spatial (census tract and geographies with a minimum of 5,000, 20,000, or 50,000 persons) and temporal (3-, 5-, 7-, or 10-year periods). Counts and rates are suppressed when there are fewer than 16 cases or fewer than 100 persons in the geographic area. Specific to this project, an additional suppression is applied when the relative standard error of the rate is greater than 30%.

For registries that opt-in, county-level standardized incidence ratios (SIR) are displayed on the Tracking Network. An SIR is displayed for cancers displayed at the county-level in the Tracking Network Data Explorer Tool. Data are aggregated for the most recent 3-year or 5-year periods (depending on a registry's opt-in response). The SIRs are not stratified by sex, age, or racial/ethnic group.

Chronic Disease Indicator Tool

This query system provides chronic disease and risk factor indicators in a uniform way. U.S. Cancer Statistics incidence and mortality rates for invasive cancers, all sites combined and by selected sites (such as female breast, prostate, and colorectal cancers).

Melanoma Dashboard

This dashboard was created in partnership with CDC's Environmental Public Health Tracking Program to help communities address their unique melanoma prevention needs. It provides state- and county-level data on melanoma incidence (using U.S. Cancer Statistics incidence data; 5- and 10-year aggregated data respectively), melanoma

mortality, and UV irradiance. It also provides information about state policies regarding minors' access to indoor tanning devices and sunscreen use at schools.

Data Release to Federal and Trusted Partners

American Cancer Society (ACS)

CDC shares NPCR and USCS data with ACS to promote collaborations on cancer surveillance and epidemiological research efforts. ACS's Surveillance and Health Services Research (SHSR) Program analyzes and disseminates cancer statistics and identifies gaps and opportunities for cancer prevention, early detection, and treatment. The SHSR annually publishes the statistical report, *Facts & Figures*, and peer-reviewed journal articles that are used by public health experts, clinicians, and scientists.

In 2018, a Memorandum of Understanding was implemented with ACS, and ACS staff members must sign a Data Use Agreement form and complete annual Assurance of Confidentiality training before they are given access to the data. Beginning in 2020, due to changes in SEER's data release policy, CDC also obtains approval from SEER before releasing USCS data. CDC provides ACS staff access to the following databases with record-level data through SEER*Stat software: USCS Delay-Adjusted database, NPCR Survival database, NPCR Prevalence database, and selected variables from the NPCR and SEER Quality Control database. The Quality Control database shared with ACS is restricted to 24-month data, excludes postal code and census tract variables, and excludes "day" fields for date of birth and date of death.

Central Brain Tumor Registry of the United States (CBTRUS)

CBTRUS annually publishes the print and web versions of the statistical report, *Primary Brain Tumors in the United States Statistical Report Supplement*; a previous version of the report is available at: https://www.cbtrus.org/reports. The report includes age-adjusted rates and corresponding 95% confidence intervals on brain and other central nervous system tumors and is presented by state, histology, major histology grouping, primary site, behavior, sex, race, ethnicity, and age at diagnosis. As a trusted partner, CBTRUS is provided access to the NPCR Survival Dataset to include survival estimates in the annual report, conduct in-depth analyses, and respond to queries. CDC provides individual, record-level data to CBTRUS for the publication of this report; Appendix C lists the variables included in this dataset. Only states meeting the USCS publication criteria are included in the dataset.

In addition, CBTRUS uses these data to respond to inquiries that are more specific than those that are provided by the report. For these inquiries, no individual record-level data are released; only aggregated data with the corresponding confidence intervals (if applicable) and appropriate suppression criteria are provided to data inquirers. Attribution to NPCR is provided. CBTRUS signs data use agreements before data are released for their report and future inquiries. For questions, contact CBTRUS staff at cbtrus@aol.com.

International Association of Cancer Registries (IACR)

The International Association of Cancer Registries (http://www.iacr.com.fr/) (IACR) produces the *Cancer Incidence in Five Continents* (CI5) and other special reports such as pediatric data reports including International Incidence of Childhood Cancer (IICC). The *CI5* series of monographs, published every five years, has become the reference source of data on the international incidence of cancer. The most recent version was published in 2017. The CI5 databases provide access to detailed information on the incidence of cancer recorded by cancer registries (regional or national) worldwide in two formats (CI5 and CI5plus) and other reports also provide access to detailed information on the incidence of pediatric cancers:

- **CI5** presents the basic data published in the CI5 volumes.
- **CI5plus** contains annual incidence for selected cancer registries published in CI5 for the longest possible period.

When IACR requests data, the formal Call for Data Submission giving information on the evaluation procedure, likely layout of how data will be presented, and questionnaire on registry operations will be available from the IACR website.

NPCR may facilitate the call for data on behalf of awardees. NPCR will provide additional information regarding the CI5 Call for Data as it becomes available. The CI5 Call for Data has two components: the questionnaire and introductory text and data submission.

Data submitted for CI5 may also be used for the IICC publication, making a separate data submission unnecessary. This IACR product requires states to complete a separate questionnaire and introductory text.

States are responsible for completing the online questionnaires and providing introductory text indicating if the CI5 data and introductory text are also used for the IICC product. NPCR will submit aggregated NPCR data for central cancer registries meeting USCS publication criteria.

CONCORD

CONCORD is the program for worldwide surveillance of cancer survival, led by the London School of Hygiene & Tropical Medicine and supported by the Union for International Cancer Control (UICC). CONCORD monitors progress toward the UICC's World Cancer Declaration, made in 2013: "major reductions in premature deaths from cancer, and improvements in quality of life and cancer survival".

A call for participation in the CONCORD studies is periodically issued and extends examination of worldwide cancer survival trends for certain cancer sites. The protocol and dataset specifications will be posted to the NPCR-CSS Document Server, CONCORD tab as they become available.

NPCR may facilitate the call for data on behalf of awardees by submitting NPCR data for central cancer registries meeting USCS publication criteria for survival analyses: they meet USCS data quality criteria and have conducted active patient follow-up or linked records with the National Death Index.

Agency for Healthcare Research and Quality (AHRQ)

The U.S. Department of Health and Human Services' Agency for Healthcare Research and Quality (AHRQ) is the lead federal agency charged with improving the safety and quality of America's health care system. It develops and disseminates knowledge, tools, and data to improve health care systems and help Americans, health care professionals, and policy makers make informed health decisions. NPCR-CSS data are shared with AHRQ for reports on national health care quality and disparities.

Analytic Datasets

USCS Analytic Data

Combined NPCR and SEER incidence data are referred to as USCS. CDC creates USCS analytic datasets each year that include data from central cancer registries meeting USCS publication criteria and diagnosis year coverage. CDC, NCI staff members, and contractors analyze USCS data as needed using these internal analytic databases.

The datasets are made available via SEER*Stat software to federal employees, fellows, and contractors in CDC's Division of Cancer Prevention and Control and NCI's SEER program after obtaining SEER Research Plus access, signing an NPCR Analytic Data Use Agreement (Appendix D) and CDC Nondisclosure Agreement (Appendix E) and completing annual Assurance of Confidentiality training. The dataset is also available to approved partnering organizations and state and territory central cancer registries after a Memorandum of Understanding and Data Use Agreements are signed (see Appendix H and Appendix I).

In specially established collaborative relationships, researchers external to CDC, NCI, and ACS may be provided access to the USCS analytic datasets. In these relationships, CDC staff must be included in the analytic project as a co-author, Data Use Agreements must be signed, and Assurance of Confidentiality training must be completed before access is provided. Additionally, access will only be allowed on-site at CDC's Cancer Surveillance Branch offices. See the section "External Data Requests".

Cancer surveillance and epidemiological analyses include assessment of the completeness, timeliness, and quality of cancer incidence data and analyses of the cancer burden and survival as needed to meet national cancer control objectives. Such analyses of state or territory and national data are conducted routinely by federal agencies including CDC and NCI for programmatic or statistical purposes, as needed, to achieve the agencies' mandates.

Five internal analytic datasets are routinely analyzed by CDC and NCI staff members:

NPCR/SEER USCS Incidence Analytic Dataset

CDC and NCI staff members and contractors conduct cancer surveillance and epidemiological research that results in publications, data briefs, and presentations. Examples of research include descriptive analyses by racial and ethnic populations for specific cancers, descriptions of cancer incidence trends, and descriptive analyses of the quality of the data. Appendix F lists the variables available in this dataset.

NPCR Internal Survival and Prevalence Dataset

A dataset will be created that allows for both survival and prevalence estimates to be calculated. Cancer survival data are critical for evaluating the progress and effect of early detection and screening programs, comprehensive cancer control plans, and interventions from other sources. CDC's NPCR-CSS calculates and publishes survival estimates on this population at the national, state, and regional levels. This NPCR-CSS dataset supports analyses of survival estimates for rare cancers that cannot be addressed otherwise and provides data for publication on the USCS Data Visualizations tool. The database also provides limited-duration prevalence estimates for NPCR registries who meet USCS publication criteria for all years included in the database and that have completed National Death Index linkages or active patient follow-up for all years included in the database. Statistics generated from this dataset are published on the USCS Data Visualizations tool. Appendix G lists the variables available in this combined survival and prevalence dataset.

NPCR/SEER Survival Dataset

This database contains data from NPCR- and SEER-funded registries that have completed National Death Index linkages or active patient follow-up for all years included in the database and meet 95% completeness estimates. This dataset will be used to assess Healthy People 2030 cancer objective C-11: Increase the proportion of cancer survivors who are living 5 years or longer after diagnosis. The variables included in the dataset are the same as the NPCR Internal Survival Data, which are listed in Appendix G.

NPCR/SEER USCS Delay-Adjusted Dataset

Case-reporting delay may result in an underestimate of true incidence. Researchers can adjust for this delay using composite delay factors, thus producing more precise cancer incidence trends. The composite delay factors (https://surveillance.cancer.gov/delay/model.html) used in this database were developed by SEER and are used by NPCR, SEER, and NAACCR. The delay-adjustment factors account for cancer site, registry, age, race, ethnicity, and diagnosis year, and are used to estimate delay-adjusted counts and rates. The variables available in this dataset are listed in Appendix O.

USCS American Indian and Alaska Native Incidence Analytic Database

CDC uses IHS linkage results for analyses of cancer incidence among American Indian and Alaska Native populations. In addition to improving cancer incidence rates presented in USCS Data Visualizations tool, an employee in CDC's Division of Cancer Prevention and Control assigned to IHS maintains an analytic database, the USCS American Indian and Alaska Native Incidence Analytic Database (AIAD). Access to this database is limited to approved CDC staff.

The data are used to respond to requests for cancer incidence rates for American Indian and Alaska Native populations from tribal organizations. Five-year aggregate incidence counts, age-adjusted rates, and 95% confidence intervals for selected primary sites are displayed in the USCS Data Visualizations tool (see Table 1). These data are limited to non-Hispanic American Indian and Alaska Native people living in IHS Purchased/Referred Care Delivery Areas (PRCDA) counties. Inclusion in this dataset also allows IHS to provide the state with the date of death obtained through National

Death Index (NDI)-IHS linkage or the date the linkage occurred by diagnosis year for registries that complete an NDI supplemental confidentiality agreement for application Y9-0033.

Requirements for Staff

In compliance with the 308(d) Assurance of Confidentiality, CDC and NCI employees and contractors and partner organizations conducting these analyses are required to handle the information in accordance with principles outlined in the CDC Staff Manual on Confidentiality and to follow the specific procedures documented in the NPCR-CSS Confidentiality and Security Statement (appendices B, H, and I).

In addition, CDC, SEER, and partner organization staff members are required to acknowledge state and territory cancer registries whenever NPCR-CSS data are presented, released, or published by making the following (or similar) statement:

> *These data were provided by central cancer registries participating in the National Program of Cancer Registries (NPCR) and submitted to CDC in [Month, Year], and/or the Surveillance, Epidemiology, and End Results (SEER) program and submitted to NCI in [Month, Year]. The dataset includes data for diagnosis years xxxx–xxxx (excluding SEER-Metro Registry data).*

NPCR/SEER USCS Incidence and Survival Public-Use Research Dataset

For purposes of this policy, the NPCR/SEER USCS Incidence Public-Use Research Dataset (Incidence PUD) and the NPCR Survival Public-Use Research Dataset (Survival PUD) are defined as the version of the full NPCR/SEER USCS microdata (individual case-specific data) that have been modified as needed to minimize the potential for disclosure of confidential information. These datasets contain a subset of data items published in the NPCR/SEER USCS Incidence Analytic dataset. Personal identifiers, such as a patient's name, street address, and Social Security number, are not included in these datasets as this information is not transmitted by central cancer registries to CDC as part of their annual data submission. Certain data items, such as date of birth and reporting-source (death certificate only and autopsy only) cases, may be removed from these research datasets to minimize the potential identification of individuals with a rare cancer in a person of certain age or racial or ethnic group or living in a specific county. The list of the variables included in the NPCR/SEER USCS Incidence Public-Use Dataset is in Appendix J. The NPCR Survival PUD is under development. The NPCR-CSS Data Release Policy will be updated before its release.

The Incidence PUD dataset is available publicly through SEER*Stat software. Upon completion, the Survival PUD will be made available through the same mechanism. Researchers are given access to the data after obtaining SEER Research Plus access and signing an NPCR and SEER – U.S. Cancer Statistics Research Data Request Form (Appendix K). Every time a user opens the public use research SEER*Stat databases, they will be required to electronically acknowledge the database's user requirements, which include that the data will be used for statistical reporting and analysis only and the user will not attempt to identify any person or entity included in the database (Appendix K). Cell suppression of fewer than 16 cases is automatic and the SEER*Stat case listing function is disabled as additional data protection measures. The Incidence PUD dataset allows authorized users to generate the authorized counts, crude rates, age-adjusted incidence rates, and 95% confidence intervals to meet their specific needs.

Restricted-Access Research Dataset (RDC)

For purposes of this policy, the restricted-access dataset is defined as the version of the full NPCR/SEER USCS analytic dataset, either aggregated data or microdata (individual case-specific data) that has been modified as needed to minimize (but may not remove entirely) the potential for disclosure of confidential information.

CDC uses the National Center for Health Statistics' Research Data Center (NCHS RDC) as a mechanism for researchers outside of the Division of Cancer Prevention and Control (DCPC) to request and gain access to NPCR data for research purposes. The data are available through the NCHS RDC only after the standard data quality reviews that occur as part of the preparation for USCS. The restricted-access dataset is released to researchers through the NCHS RDC after CDC authenticates the requestor's identity and research intent through an extensive proposal review process, and the

researcher completes the NCHS RDC's confidentiality and security requirements. The requestor must also comply with the NCHS RDC's confidentiality procedures and data-sharing agreements.

The NCHS RDC has developed and maintains detailed data-sharing agreements and procedures for user authentication and for logging and monitoring data releases. NPCR and NCHS RDC staff review project proposals. Proposals may also be shared for review with central cancer registry staff whose data are included in the proposed project. User documentation includes a data dictionary for every diagnosis year available at the NCHS RDC.

Using the NCHS RDC to manage data access provides the highest level of data security and protection of confidentiality available for data analysis and allows CDC to comply with the Assurance of Confidentiality [308(d)] that was obtained for the NPCR-CSS data. The NCHS RDC is also covered by a separate Assurance of Confidentiality [308(d)]. For further information regarding the NCHS RDC, refer to Appendix L of this policy.

The restricted-access dataset does not contain personal identifiers such as a patient's name, street address, or Social Security number, as this information is not transmitted by central cancer registries to CDC as part of their annual data submission. However, the dataset may contain information that is potentially identifiable especially when linked with other datasets, such as the occurrence of a rare cancer in a person of a certain age or racial or ethnic group or living in a specific county. The data are made available to researchers through a SAS dataset specific to each project created by NCHS RDC staff. Researchers must include a data dictionary in their proposal, and only the requested variables are included in the SAS file.

Data Release Under Controlled Conditions

CDC policy stipulates that a CDC program may consider release data that cannot be released as a public web-based system, a research dataset, or a restricted-access dataset under certain controlled conditions.[18] These controlled conditions may include a CDC-controlled data center such as the NCHS RDC, on-site at CDC's Cancer Surveillance Branch offices, or through special licensing. NPCR-CSS data will not be released except as described above while this policy is in place. Release of data under controlled conditions will be considered in discussions with partners, and a determination will be made as to whether such releases of data will be considered for NPCR-CSS data.

Emergency and Provisional Data Releases

It is not anticipated that CDC will need to release NPCR-CSS data before the files have been modified as needed to protect confidentiality as described in this policy. This is prohibited by the 308(d) Assurance of Confidentiality (appendices B, H, and I).

Provisional data and draft data tables may be shared with CDC employees and contractors, NPCR central cancer registries, and other partners to facilitate data quality reviews. When appropriate, individuals who participate in such reviews sign an NPCR Analytic Data Use Agreement and a CDC Nondisclosure Agreement (when applicable) before accessing the data or tables.

Protection of Data

Assurance of Confidentiality

All data collected and maintained by NPCR-CSS must be managed, presented, published, and released with strict attention to confidentiality and security, consistent with the general principles and guidelines established by CDC for confidential case data[11,18-19] and specific restrictions imposed on NPCR-CSS data (appendices B, H, and I).[1] Special care must be given to cancer incidence data that are not directly identifiable because geographic and small cell data may be indirectly identifying when combined with detailed information in case reports, laboratory reports, medical records, or linkage with other data files.[12-17]

NPCR-CSS has approval for protection under section 308(d) of the Public Health Services (PHS) Act (42 U.S.C. 242m(d)) (appendices B, H, and I). The 308(d) confidentiality assurance protects identifiable and potentially identifiable information from being used for any purpose other than the purpose for which it was collected (unless the person or establishment from which it was obtained has consented to such use). This assurance protects against disclosures under a court order and provides protections that the Privacy Act of 1974 (5 U.S.C. 552a) does not. For example, the Privacy Act of 1974 protects individual participants, but the 308(d) confidentiality assurance also protects institutions. Confidentiality protection granted by CDC promises participants and institutions that their data will be shared only with those individuals and institutions listed in the project's consent form or in its specified policies.

Suppression of Rates and Counts

When the numbers of cases or deaths used to compute rates are small, those rates tend to have poor reliability. Another important reason for using a threshold value for suppressing cells is to protect the confidentiality of patients whose data are included in a report by reducing or eliminating the risk of disclosing their identity.

Therefore, to discourage misinterpretation or misuse of rates or counts that are unstable because case or death counts are small, annual incidence and death rates and counts in publicly available datasets and web-based query systems are suppressed if the case or death counts are below 16. A count of fewer than about 16 results in a standard error of the rate that is approximately 25% or more as large as the rate itself. Similarly, a case count below 16 results in the width of the 95% confidence interval around the rate being at least as large as the rate itself. These relationships were derived under the assumption of a Poisson process with the standard population age distribution assumed to be similar to the observed population age distribution. For aggregated time periods, counts and rates are suppressed for fewer than 16 cases. However, average annual rates and counts may not be suppressed if the total case count for the time period exceeds 16.

The cell suppression threshold value of 16, which was selected to reduce misuse and misinterpretation of unstable rates and counts, is more than sufficient to protect patient confidentiality.

Per the Data Use Agreements, researchers using restricted-access data files are required to suppress count and statistical results that are based on cells with fewer than 6 cases in publications and presentations. Researchers are advised to use caution when presenting or interpreting results based on fewer than 16 cases.

Complementary cell suppression and suppression of certain race and ethnicity combinations are required as additional measures to assure patient confidentiality and rate stability.

Public Release Disclosure Statement

The following (or similar) public release disclosure statement is prominently displayed for users of all NPCR-CSS public web-based query systems, research datasets, and restricted-access datasets:

> *Data Use Restrictions: Read Carefully Before Using*
>
> ***By using these data, you signify your agreement to comply with the following statutorily based requirements.*** *The National Program of Cancer Registries (NPCR), Centers for Disease Control and Prevention (CDC), has obtained an assurance of confidentiality pursuant to Section 308(d) of the Public Health Service Act, 42 U.S.C. 242m(d). This assurance provides that identifiable or potentially identifiable data collected by the NPCR may be used only for the purpose for which they were obtained unless the person or establishment from which they were obtained has consented to such use. Any effort to determine the identity of any reported cases, or to use the information for any purpose other than statistical reporting and analysis, is a violation of the assurance. Therefore users will:*
>
> - *Use the data for statistical reporting and analysis only.*
> - *Make no attempt to learn the identity of any person or establishment included in these data.*

- *Make no disclosure or other use of the identity of any person or establishment discovered inadvertently, and advise the Associate Director for Science, Office of Science Policy and Technology Transfer, CDC, Mailstop D-50, 1600 Clifton Road, N.E., Atlanta, Georgia, 30333, Phone: 404-639-7240 (or NCI's SEER Program if SEER data) and the relevant state, territory, or metropolitan area cancer registry of any such discovery.*

Freedom of Information Act (FOIA) Data Requests

The Freedom of Information Act (FOIA) (www.cdc.gov/od/foia/) generally provides that, upon written request from any person, a federal agency such as CDC must release any agency record unless that record falls (in whole or part) within one of nine exemptions. FOIA applies to federal agencies only and covers only records in the possession and control of those agencies at the time of the FOIA request (except in certain instances involving grantee-held data). Because state- and territory-based data become a federal record in CDC's possession, such records are subject to disclosure in response to a FOIA request. The FOIA exemptions that may be available to protect some aspects of state and territory data from public disclosures in response to a FOIA request are:

- Exemption 3, which specifically exempts information from disclosure by statute; in this instance, pursuant to an Assurance of Confidentiality under Section 308(d) of the Public Health Service Act.

- Exemption 6, which exempts from disclosure personnel and medical files and similar files, which would constitute an unwarranted invasion of personal privacy.

In general, non-FOIA requests to CDC from the public, media, and other government agencies for local cancer incidence data are referred to the state health department for a reply for three reasons:

1. The state health departments can release cancer incidence data in accordance with local policies and procedures and consistent with provisions of the Cancer Registries Amendment Act (Public Health Service Act, (42 USC 280e-280e-4), as amended).[1]

2. The relative infrequency of data submission to federal agencies assures that the state or territory health department or its designated central cancer registry will have the most complete, accurate, and up-to-date information.

3. The central registry may be able to provide more detailed data that can better meet the needs of the requestor.

When the request is for data regarding cancer incidence involving more than one state or territory, CDC will refer the requestor to published reports or to NPCR-CSS datasets that are released in accordance with practices described in this document, if relevant.

External Data Requests

Individuals, agencies, or organizations outside CDC may request data not available from a public web-based query system or research dataset. When the requests do not identify a state or territory, CDC staff members or contractors tabulate the data for the inquirer. For requests that identify a state or territory, CDC staff members may seek states' or territories' permission regarding use. See Appendix N for additional details.

Researchers may submit data query or study proposal requests for the NPCR/SEER USCS Incidence Analytic Dataset to CDC. These requests must include:

- Names of individuals who will need access to the data.
- Purpose and public health significance of the investigation.
- Research question(s).
- Variables required beyond those in the freely available research data.

- Subset of cases needed (specific cancer type, data years, registries).
- Planned use of data, such as a manuscript, poster, or presentation.

After CDC authenticates the requestor's identity and research intent and verifies that confidentiality is maintained, a CDC analyst will process the data query and provide results to the researcher. The requestor must comply with all confidentiality and data suppression procedures outlined in the NPCR-CSS Assurance of Confidentiality [308(d)].

In circumstances where the researcher requires access to the USCS Analytic Datasets:

- CDC staff must be included in the analytic project as a co-author.
- Data Use Agreements must be signed.
- Assurance of Confidentiality training must be completed.
- Access is only allowed on-site at CDC's Cancer Surveillance Branch offices.

References

1. Cancer Registries Amendment Act, Public Law 102-515, Stat. 3312 (October 22, 1992). Available at https://www.cdc.gov/cancer/npcr/about.htm#Authorization.

2. Surveillance, Epidemiology, and End Results Program. SEER Program Coding and Staging Manual. Available at: https://seer.cancer.gov/tools/codingmanuals/.

3. North American Association of Central Cancer Registries. ICD-O-3 Coding Updates. Available at: https://www.naaccr.org/icdo3/.

4. North American Association of Central Cancer Registries. The Data Standards and Data Dictionary. Available at: https://apps.naaccr.org/data-dictionary/home.

5. Hofferkamp, J (Ed). Standards for Cancer Registries Volume III: Standards for Completeness, Quality, Analysis, Management, Security and Confidentiality of Data. Springfield (IL): North American Association of Central Cancer Registries, August 2008..

6. Hutton MD, Simpson, LD, Miller DS, Weir HK, McDavid K, Hall HI, Progress toward nationwide cancer surveillance: an evaluation of the National Program of Cancer Registries, 1994–1999. Journal of Registry Management 2001;28(3):113–120.

7. Centers for Disease Control and Prevention. U.S. Cancer Statistics Publication Criteria. Available at: https://www.cdc.gov/cancer/uscs/technical_notes/criteria/.

8. NAACCR Method to Estimate Completeness. A Data Analysis Tool for Calculations. Available at: https://www.naaccr.org/analysis-and-data-improvement-tools/#COMPLETENESS.

9. Clegg LX, Fueur EJ, Midthune DN, Fay MP, Hankey BF. Impact of reporting delay and reporting error on cancer incidence rates and trends. Journal of the National Cancer Institute 2002;94(20):1537–1545.

10. Centers for Disease Control and Prevention. Cancer Prevention and Control Program for State, Territorial and Tribal Organizations. Notice of Award, DP22-2202. Available on pages 32-33: Program 3: NPCR, Strategy 1: Enhance National Program of Cancer Registries (NPCR) data quality, completeness, use, and dissemination; Data Submission: Submit cancer data to CDC each year in accordance with CDC standards and requirements.

11. Centers for Disease Control and Prevention. CDC/ATSDR Policy on Releasing and Sharing Data. Atlanta: Centers for Disease Control and Prevention; 2005. Available at https://stacks.cdc.gov/view/cdc/7563.

12. American Statistical Association. Data Access and Personal Privacy: Appropriate Methods of Disclosure Control. Alexandria, VA: American Statistical Association; 2008. Available at https://www.amstat.org/asa/files/pdfs/POL-DataAccess-PersonalPrivacy.pdf.

13. Doyle P, Lane JI, Theeuwes JM, Zayatz LM (eds). Confidentiality, Disclosure, and Data Access: Theory and Practical Application for Statistical Agencies. Amsterdam: Elsevier Science BV; 2001.

14. Federal Committee on Statistical Methodology. Checklist on Disclosure Potential of Proposed Data Releases. Available at https://nces.ed.gov/FCSM/doc/checklist_799.doc.

15. Federal Committee on Statistical Methodology. Report on Statistical Disclosure Limitation Methodology. (Statistical Working Paper 22). Washington, DC: Office of Management and Budget; 1994 (Revised 2005). Available at https://nces.ed.gov/FCSM/pdf/spwp22.pdf.

16. McLaughlin C. Confidentiality protection in publicly released central cancer registry data. Journal of Registry Management 2002; 29(3):84–88.

17. Stoto M. Statistical Issues in Interactive Web-Based Public Health Data Dissemination Systems. Draft report prepared for the National Association of Public Health Statistics and Information Systems, Rand Corporation; September 2002.

18. Centers for Disease Control and Prevention. Data Sharing and Collection Principles and Standards; 2024. Available at https://www.cdc.gov/program-collaboration-service-integration/php/data-security/principles.html.

19. Centers for Disease Control and Prevention. CDC Staff Manual on Confidentiality. Atlanta: Centers for Disease Control and Prevention; 1984 and National Center for Health Statistics. NCHS Staff Manual on Confidentiality. Hyattsville, MD: National Center for Health Statistics; 2004.

Table 1: Comparison of the National Program of Cancer Registries-Cancer Surveillance System Datasets

Overview						
	Public Web-Based Query Systems				Analytic Datasets	
	USCS Data Visualizations Tool	**USCS WONDER**	**USCS Data for Partners [1]**	**NCEH's Tracking Network**	**USCS Public-Use Research Database**	**USCS Restricted-Access Dataset**
Format	Database of aggregate counts and rates, with text documentation	Database of aggregate counts and rates, with text documentation. The database behind the CDC firewall is case-specific microdata.	Database of aggregate counts and rates, with text documentation	Database of aggregate counts, rates, and other measures (e.g., standardized incidence ratio) with text documentation. The database behind the CDC firewall is case-specific microdata.	Customized, analytic database. The database behind the SEER*Stat firewall is case-specific microdata with enforced cell suppression and case listing disabled.	Customized, analytic database available through proposal process
Mode of Access	Web-based query system with downloadable ASCII files	Web-based query system	Web-based query system	Web-based query system	SEER*Stat client-server mode only after receipt of signed Data Use Agreement	On-site at CDC or through CDC staff assistance
Web Address or Contact Information	Website www.cdc.gov/cancer/dataviz	CDC WONDER https://wonder.cdc.gov/cancer.html	Request from uscsdata@cdc.gov (specify "USCS County" in subject line)	National Environmental Public Health Tracking Program https://ephtracking.cdc.gov/ and https://ephtracking.cdc.gov/Applications/melanomadashboard/	www.cdc.gov/cancer/public-use	Application process available at www.cdc.gov/rdc
Contains Potentially Identifiable Information?	No	No	No	No	No	Yes
Registry Eligibility Criteria for Data Completeness and Quality	USCS publication criteria; data meet criteria for unknown county	USCS publication criteria	USCS publication criteria; data meet criteria for unknown county	USCS publication criteria; data meet criteria for unknown county	USCS publication criteria	USCS publication criteria; data meet criteria for unknown county
When Available	Updated 2026	Updated 2026	Updated 2026	Updated 2026	Updated 2026	Updated 2026

[1] This column described data shared with CDC's Chronic Disease Indicator tool and NCI's State Cancer Profiles.

Table 1, continued

Cases Included						
	Public Web-Based Query Systems				Analytic Datasets	
	USCS Data Visualizations Tool	**USCS WONDER**	**USCS Data for Partners**	**NCEH's Tracking Network**	**USCS Public-Use Research Database**	**USCS Restricted-Access Dataset**
States/ Territories	NPCR/SEER states and territories meeting USCS publication criteria	NPCR/SEER states and territories meeting USCS publication criteria	NPCR/SEER states and territories meeting USCS publication criteria	NPCR states and territories meeting USCS publication criteria	NPCR/SEER states and territories meeting USCS publication criteria	NPCR/SEER states and territories meeting USCS publication criteria
Diagnosis Years	Individual years 1999 through 2023; 2018-2023	Individual years 1999 through 2023	2017–2023	Individual years 2001 through 2023	Individual years 2001 through 2023	Individual years 1999 through 2023; 2024 preliminary results
Cancer Sites	All reportable invasive cancers; *in situ* female breast, *in situ* male and female breast, and benign and borderline primary intracranial and central nervous system tumors (diagnosis year 2004)	All reportable invasive cancers; *in situ* female breast, *in situ* male and female breast, and benign and borderline primary intracranial and central nervous system tumors (diagnosis year 2004)	All reportable cancer sites combined; female breast; *in situ* female breast; cervix uteri; colon and rectum; lung and bronchus; melanoma; bladder; prostate; oral cavity and pharynx; brain and other nervous system; thyroid; kidney and renal pelvis; stomach; ovary; corpus and uterus, not otherwise specified; leukemias; non-Hodgkin lymphoma; liver and intrahepatic bile duct; pancreas, esophagus; and childhood cancers	Same as USCS Data Visualizations tool	All reportable invasive cancers; *in situ* female breast, and benign and borderline primary intracranial and central nervous system tumors (diagnosis year 2004)	All reportable invasive and *in situ* cancers and benign and borderline primary intracranial and central nervous system tumors (diagnosis year 2004)

Table 1, continued

Variables Included						
	Public Web-Based Query Systems				Analytic Datasets	
	USCS Data Visualizations Tool	USCS WONDER	USCS Data for Partners	NCEH's Tracking Network	USCS Public-Use Research Database	USCS Restricted-Access Dataset
Geographic Levels	All areas combined; NPCR/SEER state, territory, congressional districts, county; SEER metropolitan area, IHS regions (AI/AN data only) (additional levels may be added)	All areas combined; NPCR and SEER state or territory; county; region; MSA for cities of ≥500,000 population (additional levels may be added)	NPCR and SEER state or territory; county	NPCR state or territory; county/county equivalent; sub-county (to include census tract, 5,000, 20,000, and 50,000 population aggregations)	All areas combined; U.S. census region; NPCR and SEER state or territory	NPCR and SEER state or territory; county for approved requests only
Race/Ethnicity	All races combined, race groups exclusive of Hispanic ethnicity (e.g., White), race groups inclusive of Hispanic ethnicity (e.g., non-Hispanic White), ethnic groups (i.e., Hispanic or non-Hispanic)	All races combined, race groups exclusive of Hispanic ethnicity (e.g., White), race groups inclusive of Hispanic ethnicity (e.g., non-Hispanic White), ethnic groups (i.e., Hispanic or non-Hispanic)	All races combined, race groups exclusive of Hispanic ethnicity (e.g., White), race groups inclusive of Hispanic ethnicity (e.g., non-Hispanic White), ethnic groups (i.e., Hispanic or non-Hispanic)	All races combined, race groups inclusive of Hispanic ethnicity (e.g., non-Hispanic White), ethnic groups (i.e., Hispanic or non-Hispanic) Sub-county displayed for all races combined	All races combined, race groups exclusive of Hispanic ethnicity (e.g., White), race groups inclusive of Hispanic ethnicity (e.g., non-Hispanic White), ethnic groups (i.e., Hispanic or non-Hispanic)	All races reported; All races combined, race groups exclusive of Hispanic ethnicity (e.g., White), race groups inclusive of Hispanic ethnicity (e.g., non-Hispanic White), ethnic groups (i.e., Hispanic or non-Hispanic)
Age Groups	All ages combined and standard 5-year age groups and age groups of <45, 45-54, ≥75 and >65 and ≥65 years for adults. For childhood cancers age groups include <15, <20, and 5-year age groups.	All ages combined and standard 5-year age groups that can be combined by the user	Childhood cancers: <15 and <20; all other cancers: <50, 50–64, 65+	Childhood cancers: <15 and <20 Breast cancer: <50, 50+, 40-64	All ages combined, standard 5-year age groups	Standard 5-year age groups and individual ages. Month and day of birth are not provided for confidentiality reasons. If the age at diagnosis is >99, then grouped into one category. Year of birth is also grouped.
Summary Stage	Yes (localized, regional, distant, and unknown or unstaged)	Yes	Yes	Yes (late-stage screening-amenable cancers)	Yes	Yes
Histology	International Classification of Childhood Cancers, Third Revision based on ICD-O-3/IARC 2017 (all geographic areas combined), mesothelioma (national and state or territory level), Kaposi sarcoma (national and state or territory level), Consensus Conference on Cancer Registration of Brain, and central nervous system tumors (all geographic areas combined)	Same as USCS Data Visualizations tool	No	Same as USCS Data Visualizations tool	Same as USCS Data Visualizations tool	Yes

Table 1, continued

Confidentiality Protection and Disclosure Limitation Measures Employed					
	Public Web-Based Query Systems			Analytic Datasets	
	USCS Data Visualizations Tool and USCS WONDER	USCS Data for Partners	Tracking Network	USCS Public-Use Research Database	USCS Restricted-Access Dataset
Direct or Record-Level Identifiers?	No	No	No	No	Yes, but not in output which will be reviewed by CDC staff for confidentiality
Aggregation	Yes	Yes	Yes; 5-year increments for county level; 10-year increments for DCPC melanoma dashboard; 3-, 5-, 7- and 10-year increments for sub-county level	No	No
Limited Number of Variables?	Yes	Yes	Yes	Yes	Yes
Grouping or Collapsing of Variables or Response Codes such as race and age recode	Yes	No	Yes	Yes	Yes
(1) Average Annual Counts Rounded to the Nearest Whole Number **(2) Average Annual Rates** **(3) Annual Averages Are Based on At Least 5 Years of Data**	No	Yes	Yes	No	No
Cell Suppression	Yes: Counts and rates: count of fewer than16	Yes: Counts and rates: 5-year total count of <16	Yes: Counts, unsmoothed rates, and other measures (e.g., standardized incidence ratio): count less than 16 or RSE greater than the limit (25% for state or territory and county-level, 30% for sub-county level) Smoothed rates: RSE greater than the limit (25% for state or territory and county-level, 30% for sub-county level)	Yes: Counts and rates: count of fewer than 16 enforced, case listing disabled	Yes (output reviewed by CDC analyst to ensure counts of fewer than 6 are suppressed)
Complementary Cell Suppression	As needed	As needed	As needed	As needed	As needed
Public Release Disclosure Statement	Yes	Yes	Yes	Yes	Yes
Data Sharing Agreement and/or IRB Approval	No	No	No	Yes	Yes
User Authentication	No	No	No	No	Yes
Logging and Monitoring	Limited	Limited	Limited	Yes, monitoring databases used, session type and date only	Yes

Appendix A: State, Territory, and Metro Area Cancer Registries

State, Territory, and Metro Area Cancer Registries by Federal Funding Source, and First Diagnosis Year* for which cancer cases were reportable to CDC's NPCR or NCI's SEER Program

State or Territory	First Diagnosis Year for Which Cancer Cases Were Reportable to NPCR or SEER*	Federal Funding Source
Alabama	1996	NPCR
Alaska	1996	NPCR
Arizona	1995	NPCR
Arkansas**	1996	NPCR
California	1995/2000	NPCR and SEER
Los Angeles	1992	SEER
San Francisco-Oakland	1973	SEER
San Jose-Monterey	1992	SEER
Colorado**	1995	NPCR
Connecticut	1973	SEER
Delaware	1997	NPCR
District of Columbia	1996	NPCR
Florida	1995	NPCR
Georgia	1995/2010	NPCR and SEER
Atlanta	1975	SEER
Hawaii	1973	SEER
Idaho	1995/2018	NPCR and SEER
Illinois	1995/2022	NPCR and SEER
Indiana	1995	NPCR
Iowa	1973	SEER
Kansas	1995	NPCR
Kentucky	1995/2000	NPCR and SEER
Louisiana	1995/2000	NPCR and SEER
Maine	1995	NPCR
Maryland	1996	NPCR
Massachusetts	1995	NPCR
Michigan**	1995	NPCR
Minnesota	1995	NPCR
Mississippi	1996	NPCR
Missouri**	1996	NPCR
Montana	1995	NPCR
Nebraska	1995	NPCR
Nevada	1995	NPCR
New Hampshire**	1995	NPCR

State or Territory	First Diagnosis Year for Which Cancer Cases Were Reportable to NPCR or SEER*	Federal Funding Source
New Jersey	1995/2000	NPCR and SEER
New Mexico	1973	SEER
New York	1996/2018	NPCR and SEER
North Carolina	1995	NPCR
North Dakota	1997	NPCR
Ohio	1996	NPCR
Oklahoma	1997	NPCR
Oregon**	1996	NPCR
Pennsylvania	1995	NPCR
Puerto Rico	1998	NPCR
Rhode Island	1995	NPCR
South Carolina	1996	NPCR
South Dakota	2000	NPCR
Tennessee**	1999	NPCR
Texas	1995/2022	NPCR and SEER
U.S. Pacific Island Jurisdictions	2007	NPCR
Utah	1973/2016	SEER and NPCR
Vermont	1996	NPCR
Virginia	1996	NPCR
Virgin Islands	2016	NPCR
Washington	1995	NPCR
Seattle-Puget Sound	1975	SEER
West Virginia	1995	NPCR
Wisconsin**	1995	NPCR and SEER
Wyoming	1996	NPCR

* Diagnosis year is the year during which a reported cancer case was first diagnosed.

** Arkansas, Colorado, Michigan, Missouri, New Hampshire, Oregon, Tennessee, and Wisconsin receive research support from SEER but are not under contract to submit data.

CDC = Centers for Disease Control and Prevention

NCI = National Cancer Institute

NPCR = National Program of Cancer Registries

SEER = Surveillance, Epidemiology, and End Results Program

Appendix B: NPCR-CSS Overview of Data Security

The NPCR-CSS project data reside on a dedicated server maintained by the NPCR-CSS contractor. To ensure the security and confidentiality of project data, the following provisions have been incorporated into the NPCR-CSS Security Plan in accordance with the requirements of the Assurance of Confidentiality.

The NPCR-CSS server is housed in a secure facility with a guard on duty 24 hours a day. Only authorized staff are allowed to access the facility. Support people are escorted by an authorized staff member if needed. The server resides on its own local area network (LAN) behind the NPCR-CSS contractor's firewall. NPCR-CSS contractor project staff access the server via a virtual private network (VPN) from their primary office location. Elevator and stairwell access is controlled by card key 24 hours. During business hours, an attendant is always present at the reception desk to guide visitors.

- Access to the NPCR-CSS server is limited to authorized NPCR-CSS contractor project staff. It is password-protected on its own security domain. No one else is allowed access to the NPCR-CSS data.
- All NPCR-CSS contractor project staff must sign a confidentiality agreement before passwords and keys are assigned. All staff must pass background checks appropriate to their responsibilities for a public trust position.
- NPCR-CSS data that are submitted electronically are encrypted during transmission from the States. They arrive on a document server behind the NPCR-CSS contractor's firewall. Each state or territory has its own directory location so that no state or territory has access to another state or territory's data. The data are moved automatically from the document server to the NPCR-CSS server.
- Receipt and processing logs are maintained to document data receipt, file processing, and report production. All reports and electronic storage media containing NPCR-CSS data are stored under lock and key when not in use and will be destroyed when they are no longer needed.
- The NPCR-CSS contractor's security team has developed a comprehensive security plan. The security team consists of the Project Director, Project Manager, Systems Lead and Security Officer, Database Administrator, and LAN/WAN Security Steward. All project staff receive annual security awareness training covering security procedures. The security team oversees operations to prevent unauthorized disclosure of the NPCR-CSS data.
- Periodic (currently quarterly, but at least once per year) reviews and updates of the NPCR-CSS contractor's security processes are conducted to adjust for rapid changes in computer technology and to incorporate advances in security approaches. The security plan is amended as needed to maintain the continued security and confidentiality of NPCR-CSS data.

Appendix C: Data Items for CBTRUS

The dataset for CBTRUS includes individual case-specific data from the NPCR-CSS dataset. The data items included are listed below.

Diagnosis years 1995 through 2003 include invasive cases only. Starting with cases diagnosed in 2004, invasive, benign, and borderline cases are included.

Item Name	NAACCR Data Item Number	Comments
Patient ID	20	
NAACCR Record Version	50	
State of Residence at Diagnosis	80	
County at Diagnosis—Analysis	89	Results presented as 5-year average annual rates as the smallest time period with <16 cell and complementary cell suppression required
Rural/Urban Continuum/Beale Code 2013	3312	
NPCR Race Recode	Derived based on [160], [161], and [192]	Same as race for USCS
NHIA Derived Hispanic Origin	191	Results of NAACCR Hispanic/Latino Identification Algorithm
NAPIIA	193	
Sex	220	
Age at Diagnosis	230	Single year up to age 84; 85+ grouped into one category
Sequence Number—Central	380	
Date of Diagnosis	390	Day and month of diagnosis not included, only year provided
Date of Diagnosis	390	Full date
Primary Site	400	
Laterality	410	
Grade	440	
Diagnostic Confirmation	490	
Type of Reporting Source	500	
Histologic Type (ICD-O-3)	522	
Behavior (ICD-O-3)	523	
Summary Stage 2018	764	
SEER Summary Stage 2000	759	
Derived Summary Stage 2000	3020	
Merged Summary Stage	Derived based on [764], [759], [3020]	
NPCR Cancer Stage		Based on 759 and 3020
RX Summ--Surgery Primary Site	1290	≥2010 diagnosis years

Item Name	NAACCR Data Item Number	Comments
Reason for no surgery	1340	≥2010 diagnosis years
RX Summ—Radiation	1360	≥2010 diagnosis years
RX Summ--Chemo	1390	≥2010 diagnosis years
RX Summ--BRM	1410	≥2010 diagnosis years
Rad–Regional RX Modality	1570	≥2010 diagnosis years
Merged Radiation		Based on 1360 and 1570 1 = had radiation 2 = did not have radiation 3 = patient or guardian refused radiation 4 = radiation recommended but unknown if received Applied only for selection below: **8000**≤I522_HistTypeICDO3≤**9049** | **9056**≤I522_HistTypeICDO3≤**9139** | **9141**≤I522_HistTypeICDO3≤**9589**
EDITS overrides	1990–2074	
CS Site-Specific Factor 1	2880	WHO Grade
Date of Last Contact	1750	Diagnosis years 2001–2019 for states included in the NPCR RSA file. Cause of Death items (1910, 1914, 1915) are not included when review has determined that high-quality cause of death information is not available for specific states or territories.
Vital Status	1760	
Vital Status Recode	1762	
Record Number Recode	1775	
Surv-Date Active Followup	1782	
Surv-Flag Active Followup	1783	
Survival Months Active Followup	1784	
Surv-Date Presumed Alive	1785	
Surv-Flag Presumed Alive	1786	
Survival Months Presumed Alive	1787	
Surv-Date Dx Recode	1788	
Follow-Up Source	1790	
Follow-Up Source Central	1791	
Cause of Death	1910	
SEER Cause-Specific COD	1914	
SEER Other COD	1915	
ICD Revision Number	1920	
Brain Molecular Markers	3816	
Site Recode ICD-O-3/WHO 2008	9410	
ICCC site recode 3rd edition/IARC 2017	9420	
ICCC site recode extended 3rd edition/IARC 2017	9422	

Item Name	NAACCR Data Item Number	Comments
AYA Site Recode 2020	9445	
SEER Brain/CNS Site Recode	9455	

Appendix D: NPCR/SEER USCS Analytic Data Use Agreement

U.S Cancer Statistics Analytic Data

Submitted [Month,Year] (diagnosis years 1998–xxxx)

To protect the confidentiality of the individuals represented within the National Program of Cancer Registries – Cancer Surveillance System (NPCR-CSS) data, the Centers for Disease Control and Prevention (CDC) has obtained an Assurance of Confidentiality under Section 308(d) of the Public Health Service Act (42 U.S.C. 242m(d)), which provides that these data can only be used for the purpose for which they were obtained.

When using NPCR and U.S. Cancer Statistics analytic data for research purposes, it is necessary to ensure, to the extent possible, that use of the data will be limited to research or public health purposes. In accordance with applicable federal law, there must be no attempt to determine the identity of individuals represented by reported cases, or to use the information for any purpose other than for health statistical reporting and analysis.

CDC's Division of Cancer Prevention and Control (DCPC) takes every possible measure to ensure that the identity of data subjects cannot be determined. All direct identifiers, as well as characteristics that might lead to identification of individuals, are omitted from the dataset. Certain demographic and clinical information has been included for research purposes; thus, all results must be presented or published in a manner that ensures that no individual can be identified. In addition, there must be no attempt to identify individuals from any computer file or to link with a computer file containing patient identifiers.

Data users must agree to the following provisions before receiving access to U.S. Cancer Statistics Incidence, U.S. Cancer Statistics Delay-Adjusted, NPCR Prevalence, and/or NPCR Survival Analytic Data. ***Please initial after each statement to indicate agreement.***

As the recipient of the U.S. Cancer Statistics Incidence (diagnosis years {year}–{year}), U.S. Cancer Statistics Delay-Adjusted (diagnosis years {year}–{year}), NPCR Prevalence (diagnosis years {year}–{year}), and/or NPCR Survival Analytic Data (diagnosis years {year}–{year}):

- I will adhere to the requirements of the Data Use Agreement and understand that my access to the data will be revoked if these requirements are violated. **Initials:** _____
- I understand that NPCR data belong to the states and territories. The states' and territories' agreement to use of the data are obtained through the activities outlined in the general NPCR-CSS Data Release Policy and by specific requests to the states and territories through the management team of DCPC's Cancer Surveillance Branch. **Initials:** _____
- I will not use or permit others to use the datasets in any way other than for statistical reporting and analysis. **Initials:** _____
- I will not release or permit others to release the datasets or any part of them to any person except with DCPC's written approval. **Initials:** _____
- I will not attempt to link or permit others to link the datasets with individually identifiable records from any other dataset without DCPC's approval. **Initials:** _____
- I will not access nor permit others to access (directly or remotely) the data outside the United States. **Initials:** _____
- I will not attempt to use the datasets or permit others to use them to learn the identity of any person or establishment included in any dataset. **Initials:** _____

- I will protect the data file(s) I receive with a password and/or encryption. In addition, any temporary or permanent analysis files, such as those produced with analytic software, will be protected in the same manner(s). **Initials:** _____

- I will take the following actions if the identity of any person or establishment is discovered inadvertently:
  - Make no use of this knowledge.
  - Notify DCPC's Internal Data Users Group by emailing npcridug@cdc.gov.
  - As requested by DCPC, safeguard or destroy the information that identifies an individual or establishment.
  - Inform no one else of the discovered identity. **Initials:** _____

- In addition, I will make every effort to release all statistical information in such a way as to avoid inadvertent disclosure. In order to do this:
  - I agree that all oral or written reports will contain only aggregate data and I will not report counts of fewer than 6 cases or statistics generated from fewer than 6 cases. **Initials:** _____
  - I understand that calculating rates or other statistics based on small numbers can raise statistical issues concerning stability and confidentiality. I will use appropriate caution when presenting and interpreting results based on fewer than 16 cases. **Initials:** _____
  - I will use complementary cell suppression to ensure that no data on an identifiable case can be derived through subtraction or other calculation from the combination of tables in all oral and written presentations. **Initials:** _____

- I have completed the Assurance of Confidentiality Overview Course (https://intranet.cdc.gov/os/osi/pcu/aoc/training/) available through HHS Learning Portal and have e-mailed my certificate of completion to npcridug@cdc.gov. **Initials:** _____

- I have added my project to the NPCR Internal Analysis SharePoint table and, if applicable, I will notify and obtain permission from the Internal Data Users Group to analyze state- and county-level data. **Initials:** _____

- I will acknowledge central cancer registries whenever data are presented, released, or published by including the following (or similar) statement:

  These data were provided by central cancer registries participating in the National Program of Cancer Registries (NPCR) and submitted to CDC in November {year}, and/or the Surveillance, Epidemiology and End Results (SEER) program and submitted to NCI in November {year}. The U.S. Cancer Statistics Incidence Analytic dataset includes diagnosis years {year}–{year} (excluding SEER-Metro Registry data); U.S. Cancer Statistics Delay Adjusted Analytic dataset includes diagnosis years {year}–{year} (excluding SEER-Metro Registry data), NPCR Prevalence Analytic dataset includes diagnosis years {year}–{year} and the NPCR Survival Analytic dataset includes diagnosis years {year}–{year}. **Initials:** _____

- As appropriate, I will cite the data:

  *National Program of Cancer Registries SEER*Stat Database: {Database file name} – {year}–{year}. United States Department of Health and Human Services, Centers for Disease Control and Prevention. Released {date}, based on the {year} submission.* **Initials:** _____

- I understand that if I require technical assistance in analyzing or interpreting the data and when such assistance goes beyond providing non-manipulated data, IDUG members reserve the right to request to be considered as a research collaborator or co-author in any resulting publications or presentations. **Initials:** _____

- I will provide a courtesy copy of papers or abstracts to the NPCR Internal Data Users Group at npcridug@cdc.gov as they are entered into Documentum for clearance. **Initials:** _____

- I am familiar with the use of **SEER*Stat** in analyzing data or will complete the needed training. **Initials:** _____

If you are requesting access to a U.S. Cancer Statistics database, you must first set-up SEER Research Plus (https://seer.cancer.gov/data/access.html) access as the database includes SEER data.

After you have access to SEER Research Plus, complete the fields below, sign and date the agreement, and e-mail all pages to npcridug@cdc.gov.

The e-mail address you provide must be the same one used during the SEER Research Plus verification process.

My signature below indicates that I agree to comply with all the above stated provisions.

_____

Signature Date

Name:_____

Title_____

Branch_____

Telephone_____ E-mail:_____

Appendix E: CDC Non-Disclosure Agreement

Nondisclosure Agreement for Data Covered by an Assurance of Confidentiality

For use with CDC employees involved in activities with information covered by a Section 308(d) Assurance of Confidentiality

The success of CDC's operations depends upon the voluntary cooperation of establishments, including States, and of persons who provide information requested by CDC programs under an assurance that such information will be kept confidential and be used only for epidemiological or statistical purposes.

When confidentiality is authorized, CDC operates under the restrictions of Section 308(d) of the Public Health Service Act (42 U.S.C. §242m(d)), which provides in summary that no information obtained in the course of its activities may be used for any purpose other than the purpose for which it was supplied, and that such information may not be published or released in a manner in which the establishment or person supplying the information or described in it is identifiable unless such establishment or person has consented. As a CDC employee granted access to information covered by Section 308(d), I understand and acknowledge that I am bound to comply with the restrictions provided to the information under Section 308(d).

I am aware that unauthorized disclosure of information covered by Section 308(d) of the Public Health Service Act may subject me to disciplinary action.

"I am aware that unauthorized disclosure of confidential information is punishable under Title 18, Section 1905 of the U.S. Code, which reads, in relevant part:

'Whoever, being an officer or employee of the United States or of any department or agency thereof…publishes, divulges, discloses, or makes known in any manner or to any extent not authorized by law any information coming to him in the course of his employment or official duties or by reason of any examination or investigation made by, or return, report or record made to or filed with, such department or agency or officer or employee thereof, which information concerns or relates to the trade secrets, processes, operations, style of work, or apparatus, or to the identity, confidential statistical data, amount or source of any income, profits, losses, or expenditures of any person, firm, partnership, corporation, or association; or permits any income return or copy thereof or any book containing any abstract or particulars thereof to be seen or examined by any person except as provided by law; shall be fined not more than $1,000, or imprisoned not more than one year, or both; and shall be removed from office or employment.'

"I understand that unauthorized disclosure of confidential information is also punishable under the Privacy Act of 1974, Subsection 552a (i) (1), which reads:

'Any officer or employee of any agency, who by virtue of his employment or official position, has possession of, or access to, agency records which contain individually identifiable information the disclosure of which is prohibited by this section or by rules or regulations established thereunder, and who knowing that disclosure of the specific material is so prohibited, willfully discloses the material in any manner to any person or agency not entitled to receive it, shall be guilty of a misdemeanor and fined not more than $5,000.'

These provisions are consistent with and do not supersede, conflict with, or otherwise alter the employee obligations, rights, or liabilities created by existing statute or Executive order relating to (1) classified information, (2) communications to Congress, (3) the reporting to an Inspector General of a violation of any law, rule, or regulation, or mismanagement, a gross waste of funds, an abuse of authority, or a substantial and specific danger to public health or safety, or (4) any other whistleblower protection. The definitions, requirements, obligations, rights, sanctions, and liabilities created by controlling Executive orders and statutory provisions are incorporated into this agreement and are controlling.

'My signature below indicates that I have read, understood, and agreed to comply with the above statements.

_____	_____	_____
Typed/Printed Name	Signature	Date

_____
Center/Institute/Office

Non-Employee 308(d) Pledge of Confidentiality

For use when non-CDC employees are provided access to data covered by a 308(d) Assurance of Confidentiality

I, as a non-CDC Employee (e.g., Guest Researcher, Visiting Fellow, Student, Trainee, employee of a federal agency other than CDC, etc.) may be given access to information that is identifiable or potentially identifiable to a person and that is covered by Section 308(d) of the Public Health Service Act (42 U.S.C. §242m(d)), or an Assurance of Confidentiality. As a condition of this access, I am required to comply with the following safeguards for the protection of this covered data.

1. I agree to be bound by the following assurance:

   In accordance with Section 308(d) of the Public Health Service Act (42 U.S.C. §242m(d)), I agree that no information obtained in the course of the activity described in the Assurance of Confidentiality will be used for any purpose other than the purpose for which it was supplied, unless I am informed in writing that such person has consented to its use for such other purposes. Further, I agree that no information obtained in the course of the activity described in the Assurance of Confidentiality will be disclosed in a manner in which the establishment or person supplying the information or described in it is identifiable, unless I am informed in writing that the establishment or person has consented to such disclosure, to anyone other than authorized staff of CDC or staff covered under this 308(d) Assurance.

2. I agree to maintain the following safeguards to assure that confidentiality is protected and to provide for the physical security of the records:

   To preclude observation of confidential information by persons not authorized to have access to the information on the project, I shall maintain all records that I am provided access to that identify establishments or persons covered by this Assurance of Confidentiality or from which establishments or persons covered by this Assurance of Confidentiality could be identified in locked containers or protected computer files when not under immediate supervision by me or another authorized member of the project. The keys or means of access to these containers or files are not to be given to anyone other than those authorized to have access. I further agree to abide by any additional requirements imposed by CDC for safeguarding the identity of establishments or persons covered by this Assurance of Confidentiality.

My signature below indicates that I have carefully read and understand this agreement and the Assurance of Confidentiality, which pertains to the confidential nature of this project. As a(n) _____ (for example, visiting scientist, guest researcher, fellow, trainee, employee of a federal agency other than CDC), I understand that I am prohibited from disclosing any such confidential information that has been obtained under this project to anyone other than authorized staff of CDC or persons covered under this Section308(d) Assurance of Confidentiality. I understand that any disclosure in violation of this Confidentiality Pledge may lead to termination of my employment, fellowship, training experience, or scientific collaboration, as well as other penalties.

_____

Printed Name

_____

Signature

_____

Date

Agreement of CDC Contractors for Safeguards Against Invasions of Privacy for Certain Establishments or Persons Covered by an Assurance of Confidentiality

For use when contractors or subcontracts have access to information covered by a 308(d) Assurance of Confidentiality

Access to data covered by an Assurance of Confidentially, titled Assurance of Confidentiality for the National Program of Cancer Registries Cancer Surveillance System, ("Assurance") as provided by Section 308(d) of the Public Health Service Act (42 U.S.C. §242m(d)), is necessary for certain projects funded through contract task order number _____. Consistent with Section 308(d), the contractor is required to give an assurance of confidentiality and to provide for safeguards to assure that confidentiality of the data covered by the Assurance is maintained.

To provide this assurance and these safeguards in performance of the contract, the contractor shall

1. Be bound by the following assurances:

    a. No information that is identifiable or potentially identifiable to an establishment or person covered by the Assurance and obtained in the course of this activity may be used for any purpose other than the purpose for which it was supplied, unless CDC informs contractor in writing that such establishment or person has consented to its use for such other purposes.

    b. No information that is identifiable or potentially identifiable to an establishment or person covered by the Assurance and obtained in the course of this activity may be disclosed to anyone other than authorized staff of CDC or others noted in the Assurance, unless CDC informs contractor in writing that such establishment or person has consented to its disclosure to such other persons.

    c. No preliminary data from studies or projects that identifies or potentially identifies an establishment or person covered by the Assurance may be disclosed to anyone other than authorized staff of CDC or others noted in the Assurance of Confidentiality statement, unless this information is otherwise in the public domain or CDC has provided written permission for use of this information to be made public. For example, if CDC clears an abstract for a scientific presentation, this constitutes permission for public presentation.

    d. New research study ideas that are not already funded through the above-referenced contract task order may be discussed or presented during calls/meetings as part of normal communications and coordination between CDC and the contractor; should these ideas lead to further activities with information covered by this Assurance, these protections will extend to those activities only if agreed to in writing by CDC.

2. Maintain the following safeguards to assure that the confidentiality provided by Section 308(d) and the Assurance is protected by the contractor and to provide for the physical security of the records:

    a. After having read the above Assurance, each employee of the contractor participating in this project is to sign the following pledge of confidentiality:

        I have carefully read and understand the CDC assurance, which pertains to the confidential nature of identifiable or potentially identifiable data covered by the Assurance of Confidentiality to be handled in regard to these studies and reviewed as part of activities under task order _____. As an employee of the contractor, I understand that I am prohibited by law from disclosing any such confidential information that identifies or potentially identifies an establishment or person covered by the Assurance of Confidentiality, which has been obtained under the terms of this contract, to anyone other than authorized staff of CDC and that I may use this information only for the purposes for which it was obtained and consistent with the task order.

    b. To preclude observation of confidential information that identifies or potentially identifies an establishment or person covered by the Assurance by persons not employed on the project, the contractor shall maintain all

confidential records that identify establishments or persons or from which establishments or persons could be identified under lock and key.

Specifically, at each site where these items are processed or maintained, all confidential records that will permit identification of establishments or persons are to be kept in locked containers when not in use by the contractor's employees. The keys or means of access to these containers are to be held by a limited number of the contractor staff at each site. When confidential records that will permit identification of establishments or persons are being used in a room, admittance to the room is to be restricted to employees pledged to confidentiality and employed on this project. If at any time the contractor's employees are absent from the room, it is to be locked.

c. The contractor and his professional staff will take steps to insure that the intent of the pledge of confidentiality is enforced at all times through appropriate qualifications standards for all personnel working on this project and through adequate training and periodic follow-up procedures.

3. Flow down all requirements set forth in this Agreement to all subcontracts and all subcontract employees.

_____
(Typed/printed Name)

_____
(Signature)

_____
(Date)

CONFIDENTIALITY AGREEMENT for Access to Information Technology Resources at the Centers for Disease Control and Prevention and Limitation on Disclosure of Sensitive Information Under

Contract Number _____, Task Order _____

As an employee or subcontractor of _____, **THE PARTICIPANT** requires a wide range of access to confidential information and Federal information technology (IT) resources and information maintained by the Centers for Disease Control and Prevention (**CDC**), an agency of the U.S. Department of Health and Human Services.

In consideration for the following mutual covenants, the parties agree as follows:

1. Within the context of CDC Contract Number _____, Task Order _____, and in accordance with the terms of this agreement, **CDC** grants limited access to the following:
   a. The Federal information technology (IT) resources generally described in Table 1.

   b. Datasets and/or public use data tapes derived from information collected under an Assurance of Confidentiality authorized by Section 308(d) of the Public Health Service Act, also listed in Table 1.

2. **THE PARTICIPANT** acknowledges that within the CDC environment, a variety of restricted access information is held, the vast bulk of which is categorized as "Sensitive but Unclassified", and that in the performance of CDC Contract Number _____, Task Order _____, the participant may require access to such limited access information. Categories of limited access information include the following:
   - Health & health-related data on individuals, groups, entities, some of which identify individuals
   - Federal Privacy Act "systems of records"
   - Information exempted from release under Freedom of Information Act
   - Proprietary data
   - National Defense-related information
   - Information subject to contractual restrictions on access
   - Information covered by a Certificate or Assurance of Confidentiality [P.H.S. Act, Sects. 301(d) & 308(d)]
   - Data collected under other specific legislative mandates (i.e. tobacco, transfer of biological, etc.)
   - Data identified as pre-release, internal working papers, etc., of federal agency

   Therefore, **THE PARTICIPANT** further agrees to not attempt to identify any person contained in contract data and to make no use of the identity of any person or establishment discovered inadvertently and advise **CDC** of any such discovery.

3. **THE PARTICIPANT** acknowledges the sensitive and confidential nature of the information covered by this agreement and agrees to employ all reasonable efforts to maintain such information secret and confidential, such efforts to be no less than the degree of care employed by ICF Incorporated to preserve and safeguard ICF Incorporated's own information.

4. **THE PARTICIPANT** agrees to utilize any information accessed through the performance of CDC Contract Number _____, Task Order _____ solely for the purpose of performing that Contract;

5. **THE PARTICIPANT** has read and agrees to be bound by CDC policies and standards regarding confidentiality and use of Federal IT resources. Further, **THE PARTICIPANT** agrees to attend one hour of training by **CDC** on information security and the use of IT resources at **CDC**.

6. **THE PARTICIPANT** agrees to refrain from any of the following prohibited uses:
   a. Disclosing, revealing, or giving to anyone information accessed under CDC Contract Number _____, Task Order _____ except to employees of ICF Incorporated who have a need for the

information and who are bound to it by like obligation as to confidentiality, without the express written permission of **CDC**.

b. Attempting to override or avoid security and integrity procedures and devices established by **CDC**, or its components, to control access to federal IT resources.

c. Attempting to override or avoid security and integrity procedures and devices established by outside organizations to control access to their information systems and IT resources.

d. Using hardware and/or software or downloading software within the scope of the project that is not specifically authorized in writing by the Project Officer.

e. Violating copyrights or software licensing agreements.

f. Using CDC's name or logos to misrepresent, as falling under CDC auspices, personal materials, or materials one produces on behalf of an approved group.

7. Upon expiration of this Agreement or CDC Contract Number _____, Task Order _____, **THE PARTICIPANT** agrees to destroy or return to CDC any information accessed through the performance of contract that falls under one or more of the categories listed under paragraph 2 above and that was copied, printed, or otherwise duplicated.

8. **CDC** has the capability and the authority to audit its federal IT resources, and under appropriate circumstances, monitor their use.

9. **CDC** may terminate this access with or without cause at any time without advance notice.

10. **THE PARTICIPANT'S** authorized access automatically expires at the end of the contract period, or sooner if so indicated in the space at the top of Table 1. A written renewal request must be submitted *two months* prior to the termination, with appropriate justification for each access to be continued. A new Agreement for Access and Limitation on Disclosure is required for each renewal.

11. The construction, interpretation, and performance of this Agreement shall be governed by U.S. Federal law.

    Violations of this agreement or misuse of CDC's federal IT resources may subject **THE PARTICIPANT** to criminal penalties in accordance with Federal law (attached). In addition, **THE PARTICIPANT** understands that other Federal laws and regulations govern CDC's maintenance and operation of these Federal IT resources and may apply to **THE PARTICIPANT**.

12. I have read, understood, and agree to comply with the above statements.

_____	_____
Print Name: Last, First, Middle Initial (Person Requesting Access)	Print Name: Last, First, Middle Initial (Contractor's Official Witness)
_____	_____
Position	Position
_____	_____
Signature	Signature
_____	_____
Date (month, day, and year)	Date (month, day, and year)

CDC Point of Contact (Technical Monitor or Project Officer):

Print Name: Last, First, Middle Initial	Position
Signature	Date (month, day, and year)

Copies of the following CDC policy statements are to be provided to each person requesting access.

Laws, Policies and Procedures Governing Use of Electronic Mail, Intranet, Internet and Other Information Technology (IT) ADP Security Policy (Manual Guide-Information Resources Management, No. CDC-3, 3/15/89) 18 U.S.C. Sections 641 and 1030.

Table 1 Federal Information Resources Authorized

Federal IT Resource Name or Description	Location	Authorizing Official(s)
Main point of entry to CDC IT resources: Information Resources Management Office		None authorized
Other LAN account(s)		None authorized
CDC mainframe account		None authorized
CDC e-mail account		None authorized
Internet access		None authorized
CDC Intranet access		None authorized
Cancer incidence data from awardees funded by CDC's Program Announcement DP22-2202 for a cooperative agreement under for the National Program of Cancer Registries		CDC Contracting Officer's Representative CDC Project Officer
Mortality data from the National Center for Health Statistics (NCHS)[1]		NCHS
Population data from the U.S. Census Bureau		Data publicly available on Internet

[1]By signing this agreement, THE PARTICIANT agrees to abide by the conditions stipulated by NCHS in the NCHS Data Use Agreement.

Access to a specific resource does not imply access to any other resource.

Appendix 1

Access to additional resources may be granted upon written request, as described below.

A written request shall be provided to _____, who will forward the request with a statement of support of the justification provided, to _____, the CDC Contracting Officer's Representative (COR) for Contract Number _____, Task Order _____ in the _____ Branch, Centers for Disease Control and Prevention (CDC).

If the requested access involves a physically separate or limited-access device or dataset, the appropriate steward of that device or dataset shall be provided with a copy of the request for review and authorization.

Upon acceptance of the request by all appropriate parties, an amendment to the Agreement for Access and Limitation on Disclosure will be executed, and a copy of any appropriate limitations on access and use will be provided. When this has been done, access will be provided.

If effective access not contained in table 1 is recognized, or if another relationship is established with a CDC organization that may lead to additional access to federal IT resources at CDC, written notice of such shall be provided to _____, and _____, the CDC COR for Contract Number _____, Task Order _____ _____ Branch, CDC.

Appendix F: Data Items for NPCR/SEER USCS Incidence Analytic Dataset

SEER*Stat Category	SEER*Stat Variable Name	Restrictions
Age at Diagnosis	Age recode with <1 year olds	
Race, Sex, Year Dx, Registry, County	Sex	
	Year of diagnosis	
	Addr at DX – state	
	County at DX Analysis	Kansas data unavailable
	State-county	Kansas data unavailable
	USCS standard	
	USCS9823	
	USCS9923	
	USCS1423	
	USCS1923	
	Race recode (W, B, AI, API)	
	Program	
	Econ status	Derived based on Appalachian Regional Commission definition
	Region/Division	
	Region	
	Origin recode NHIA (Hispanic, Non-Hisp)	
	Race and origin recode (NHW, NHB, NHAIAN, NHAPI, Hispanic)	
Site and Morphology	Primary Site – labeled	
	Primary Site	
	Histologic Type ICD-O-3	
	Behavior Code ICD-O-3	
	Grade	
	Grade clinical	
	Grade pathological	
	Grade post therapy	
	Diagnostic confirmation	
	ICD-O-3 Hist/behavior, labeled	
	ICD-O-3 Hist/behavior, malig, labeled	
	Site recode ICD-O-3/WHO 2008	
	ICCC site recode 3rd ed ICD-O-3/IARC 2017	
	ICCC site rec extended 3rd ed ICD-O-3/IARC 2017	
	AYA site recode 2020	
	Lymphoid neoplasm recode 2021 revision	
	Schema ID	

SEER*Stat Category	SEER*Stat Variable Name	Restrictions
	Behavior recode for analysis derived/WHO2008	
Stage – LRD [Summary and Historic]	Derived SS2000	
	SEER Summary Stage 2000	
	SEER Summary Stage 1977	
	SEER Summary Stage 2018	
	Merged Summary Stage	
Therapy	RX summ – surg prim site	Diagnosis years ≥2003
	RX summ – chemo	Female and male breast only, diagnosis years ≥2010, and NPCR CCRs† only
	Phase I Radiation Treatment Modality	Female and male breast and colorectal only, diagnosis years ≥2018
	Merged radiation	Female and male breast and colorectal only, diagnosis years ≥2010, and NPCR CCRs only
Extent of Disease	CS site-specific factor 1	Brain and other CNS and diagnosis years 2011–2017
	Merged estrogen receptor	Female and male, Schema ID=breast (or equivalence site/histology) and diagnosis years ≥2004
	Merged progesterone receptor	Female and male, Schema ID=breast (or equivalence site/histology) and diagnosis years ≥2004
	Merged HER2 receptor	Female and male, Schema ID=breast (or equivalence site/histology) and diagnosis years ≥2010
	Laterality	
Multiple Primary Fields	Sequence number - central	
Race and Age (case data only)	Age at Diagnosis	
	Race 1	
	IHS Link	
Geographic Locations	Ruralurban continuum 2013	
	Census Tract Poverty Indicator	Diagnosis years ≥2014, NPCR CCRs only
Dates	Year of Birth	
	Month of diagnosis	
Other	Type of Reporting Source	
Merged System-Supplied	Alcohol-related cancers	
	HPV-related cancers	
	Obesity-related cancers	
	Physical inactivity-related cancers	
	Tobacco-related cancers	
	State race eth suppress	

Appendix G: Data Items for NPCR Internal Survival and Prevalence Dataset

SEER*Stat Category	SEER*Stat Variable Name	Restrictions
Age at Diagnosis	Age recode with single ages and 85+	
Race, Sex, Year Dx, Registry, County	Sex	
	Year of diagnosis	
	Addr at DX – state	
	County at DX Analysis	
	State-county	
	Rural-urban continuum 2013	
	NPCR project flag	
	Economic status	
	Race and origin recode (NHW, NHB, NHAIAN, NHAPI, Hispanic)	
	Race recode (White, Black, Other)	
Site and Morphology	Primary Site – labeled	
	Histologic Type ICD-O-3	
	Behavior Code ICD-O-3	
	Grade	
	Grade clinical	
	Grade pathological	
	Diagnostic confirmation	
	ICD-O-3-Hist/behavior, labeled	
	ICD-O-3-Hist/behavior, malig, labeled	
	Site recode ICD-O-3/WHO 2008	
	ICCC site recode, 3rd edition	
	Behavior recode for analysis derived/WHO2008	
Stage – LRD [Summary and Historic]	Derived SS2000	
	SEER Summary Stage 2000	
	Summary Stage 2018	
	Merged Summary Stage	
Therapy	RX summ – surg prim site	Diagnosis years ≥2010
	Merged radiation	Female breast and colorectal only, diagnosis years ≥2010, and NPCR CCRs only
Extent of Disease	CS Site-Specific Factor 1	Brain/CNS and diagnosis years 2011–2017
	Merged estrogen receptor	Female and male breast only and diagnosis years ≥2004
	Merged progesterone receptor	Female and male breast only and diagnosis years ≥2004

SEER*Stat Category	SEER*Stat Variable Name	Restrictions
	Merged HER2 receptor	Female and male breast only and diagnosis years ≥2010
	Laterality	
Cause of Death (COD) and Follow-up	Survival months – presumed alive	
	Survival months flag – presumed alive	
	Cause of death (ICD-10)	
	ICD revision number	
	Vital status	
	Follow-up source central	
	COD exclusion flag	
	COD exclusion flag – external	
	Original vital status	
	Vital status recode (study cutoff used)	
	Cause of death recode	
	COD recode with Kaposi and mesothelioma	
Multiple Primary Fields	Sequence number - central	
Race and Age (case data only)	Age at Diagnosis	
	Race 1	
	NHIA derived Hispanic origin	
	Age recode with <1 year olds	
Dates	Presumed alive year of last contact recode	
	Presumed alive month of last contact recode	
	Presumed alive day of last contact recode	
	Year of birth	
	Month of diagnosis	
	Day of diagnosis	
	Original day of last contact	
	Original month of last contact	
	Original year of last contact	
	Original year of diagnosis	
	Original day of diagnosis	
	Original month of diagnosis	
Other	Type of Reporting Source	
User-Specified	EDPMDE LinkVar	

Appendix H: NPCR-CSS 308(d) Assurance of Confidentiality Statement

A public health surveillance system of population-based cancer incidence data received from cooperative agreement holders for the National Program of Cancer Registries is being conducted by the National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP) of the Centers for Disease Control and Prevention (CDC), an agency of the U.S. Department of Health and Human Services, and ICF Incorporated, a contractor of CDC. The information to be received by CDC is a subset of a standard set of data items that the state central cancer registry routinely receives from hospitals, pathology labs, clinics, private physicians, and other mandated reporters on all cancer cases diagnosed in the state. This information includes patient demographics and cancer diagnosis, treatment, and outcome data.

Each year, CDC requests cumulative data from central cancer registries. The variables reported to CDC may vary from year to year. The data submitted to CDC do not contain any direct identifiers, such as name or Social Security Number. Though project data do not contain direct identifiers, CCRS do report indirect identifiers such as patient demographic data items (e.g., a unique identifier, birth date, sex, race, ethnicity, birth place, county of residence, census tract, zipcode) and information about the type of cancer (e.g., date of diagnosis, stage at diagnosis, treatment). The cancer registries maintain these data permanently in longitudinal databases that are used for public health surveillance, program planning and evaluation, and data analyses. CDC updates its longitudinal database each year with data received from central cancer registries. NCCDPHP, recognizing the sensitivity of the data being furnished by the states, has applied for and obtained an Assurance of Confidentiality to provide a greater level of protection for the data while at CDC and at the contractor site.

Individual record-level data received by CDC or its contractors as part of this public health surveillance system that could lead to direct or indirect identification of cancer patients is collected and maintained at CDC under Section 306 of the Public Health Service (PHS) Act (42 U.S.C. 242k) with an assurance that it will be held in strict confidence in accordance with Section 308(d) of the PHS Act (42 U.S.C. 242m). It is used only for purposes stated in this assurance and are not otherwise disclosed or released, even following the death of cancer patients in this surveillance system. These data are used by CDC scientists for routine cancer surveillance, program planning and evaluation, and to provide data for cancer-related research questions that support the purpose of this public health surveillance program, e.g., monitoring the frequency and distribution of disease, evaluating cancer prevention and control activities, program planning and evaluation.

Researchers within CDC, including contract employees and qualified organizations, will be able to access individual, record-level data (i.e., data that do not directly identify individuals but that could lead to identification when combined with other information) for legitimate cancer-related research questions and reporting purposes through the full NPCR CSS analytic dataset, a less restricted dataset with information not included in the restricted-access datasets but one that does not contain all data submitted by the CCRs. A separate complete dataset (i.e., all information submitted by the CCRs) is available for data quality assessments only. “Qualified organizations” are defined as organizations with staff qualified to undertake the proposed analyses by means of specific academic training or demonstrable, related experience in cancer epidemiologic, medical, biomedical, or statistical research and the organization is identified in the NPCR CSS Data Release Policy. These individuals and organizations will be required to adhere to a strict security and confidentiality protocol.

Restricted access will be provided to researchers outside of the CDC, its contractors, or qualified organizations through the National Center for Health Statistics Research Data Centers (NCHS RDC) or, in limited instances, an aggregated data file for federal and trusted partners. A restricted-access dataset is defined as the version of the full NPCR CSS analytic dataset, either aggregated data or individual, record-level data that have been modified as needed to minimize the potential for disclosure of confidential information. For restricted-access datasets, some variables such as county at diagnosis will only be released in a modified format. The unique identifying number assigned to each individual by the central cancer registry is replaced by a random number assigned at CDC to reduce the possibility of linkage to other state- or territory-level files with indirect identifiers. This restricted access will be controlled in such a way as to limit the researchers’ ability to publish or otherwise provide others access to data that could lead to identification of an individual

(i.e., small numbers of cases, unique cancer types in a small geographic area, or aggregated in a way that a case could be identified).

Information collected by CDC is used without personal identifiers for publication in statistical and analytic summaries and for release in restricted release datasets for research. Information that could lead to direct or indirect identification of cancer patients is not made available to any group or individual that have not met the qualifications established by CDC and are not described in the NPCR CSS Data Release Policy. In particular, such information is not disclosed to insurance companies, any party involved in civil, criminal, or administrative litigation, agencies of federal, state, or local government, or any other member of the public.

Collected information that could lead to direct or indirect identification of cancer patients is kept confidential and—with the exception of CDC employees, their contractors, and qualified researchers—no one is allowed to see or have access to the information. CDC employees and contractors are required to handle the information in accordance with principles outlined in the CDC Staff Manual on Confidentiality and to follow the specific procedures documented in the Confidentiality Security Statement for this project. Qualified researchers are required to sign the NCHS RDC data sharing agreements and abide by the NCHS RDC confidentiality procedures. Access to data released through public-use datasets requires the user to complete and return a signed data use agreement acknowledging confidentiality requirements. Qualified organizations (e.g., the North American Association of Central Cancer Registries, American Cancer Society, National Cancer Institute, and the Central Brain Tumor Registry of the United States) are required to sign a detailed data release agreement to have access to restricted release data.

Appendix I: Frequently Asked Questions About the NPCR-CSS 308(d) Assurance of Confidentiality

Background

The Centers for Disease Control and Prevention (CDC) is responsible for public health surveillance in the United States. CDC collects, compiles, and publishes a large volume of personal, medical, epidemiologic, and statistical data. The success of CDC's operations depends, in part, on the agency's ability to protect the confidentiality of these data. While it is a matter of principle for CDC to guard sensitive information, and federal statutes such as the Privacy Act of 1974 provide a degree of protection for personally identifiable data, Section 308(d) of the Public Health Service Act (42 U.S.C. 242m(d)) enables CDC to provide the highest level of confidentiality protection for sensitive and mission-significant research and surveillance data.

CDC received a formal delegation of authority from the National Center for Health Statistics (NCHS) to grant 308(d) confidentiality protection in 1983. Section 308(d) of the Public Health Service Act ensures the confidentiality of data collected under Sections 304 and 306 of the Public Health Service Act. These special legislative authorities were the provisions under which NCHS collects and safeguards most of its survey data, along with the mortality data in the National Death Index. CDC was required to establish a stringent application process and continues to use the authority sparingly. The agency has granted confidentiality assurances to projects deemed significant to CDC's mission, such as surveillance of hospital infections, AIDS and HIV infections, pregnancy-related mortality, and congenital defects. Fewer than 65 projects have received 308(d) protection since CDC received this authority, and about 25 active projects have 308(d) confidentiality assurances now. As a testament to the importance of this project to CDC's mission, its National Program of Cancer Registries (NPCR) has been afforded this special data protection.

What is stated in Public Health Service Act, Section 308(d)?

The first clause of Section 308(d) states that CDC must explain the purpose for collecting data to persons or agencies supplying information, and it guarantees that CDC will be limited to those specified uses unless an additional consent is obtained. Moreover, the information obtained may be used only by CDC staff or CDC's contractors in the pursuit of such stated purposes. The second clause states that CDC may never release identifiable information without the advance, explicit approval of the person or establishment supplying the information or by the person or establishment described in the information.

What process did NPCR undertake to obtain 308(d) confidentiality protection?

NPCR staff worked with CDC's Office of General Counsel and CDC's Confidentiality and Privacy Officer to prepare the application for the NPCR-Cancer Surveillance System (CSS) project. The application contained the following four components:

- A Justification Statement summarizing the NPCR-CSS project's programmatic purpose, the type of data to be collected, and the uses to be made of the information. This statement also included an assurance that a) the requested data would not be furnished without the guarantee of a confidentiality assurance, b) confidentiality assurance is important to protect the individuals described in the data and to reassure the institutions submitting data, c) the information cannot reliably be obtained from other sources, d) the information is essential to the project's success, e) granting the confidentiality assurance would not prohibit CDC from fulfilling its responsibilities, and f) the advantages of assuring confidentiality outweigh the disadvantages.

- An Assurance of Confidentiality Statement delineating anticipated data uses and those with whom identifiable data would be shared, along with general advisements regarding the confidentiality protection.

- A Confidentiality Security Statement detailing the stringent safeguarding measures in place to ensure that the promise of confidentiality would not be jeopardized by practices of staff handling the data.

- An Institutional Review Board (IRB) Review Status Statement verifying NPCR-CSS's exemption from CDC IRB approval. (The Human Subjects Administrator at the National Center for Chronic Disease Prevention and Health Promotion determined that NPCR-CSS activities are routine surveillance and not research on human subjects. Therefore, protocol review by CDC IRB was deemed unnecessary.)

The application was submitted to the CDC Confidentiality Officer for review and modification, prepared for presentation to the CDC Confidentiality Review Group (CRG), and in May 2000 NPCR received 308(d) confidentiality protection approval for NPCR-CSS data, including authorization for retroactive confidentiality protection beginning with diagnosis year 1995. NPCR must file for continuation every 5 years to maintain the assurance.

What makes 308(d) confidentiality assurance the best protection for NPCR-CSS data?

The 308(d) confidentiality assurance is the only confidentiality protection that covers routine surveillance activities such as those conducted by NPCR-CSS. The assurance specifies that data protected by 308(d) may be used only for statistical or epidemiological purposes and not released further in identifiable form without consent. Another exclusive advantage of 308(d) is that it also protects indirectly identifiable data. Operationally, this means that NPCR may never release a directly identifiable variable such as a Social Security number or any combination of variables that could be used to identify an individual indirectly. Finally, 308(d) provides protection for information on both living and deceased individuals.

Are there any disadvantages to individuals or institutions protected by the 308(d) confidentiality assurances?

A 308(d) confidentiality assurance does not pose a disadvantage for individuals or institutions submitting data to CDC. In fact, 308(d) provides an added benefit because it prevents CDC from freely releasing data to researchers and any other persons or entities that could request access to the data. With the confidentiality assurance protecting NPCR-CSS data, NPCR staff members are prohibited from sharing data except for the purposes stated at the time of data collection, unless consent from those who provided the assurance is obtained.

Does NPCR's 308(d) confidentiality assurance protect the data from subpoena and Freedom of Information Act (FOIA) requests?

The 308(d) assurance is the strongest protection against compulsory legal disclosure that CDC can offer. Although CDC receives FOIA requests, the FOIA (b)(6) exemption enables CDC to withhold sensitive, individually identified data that would constitute a "clearly unwarranted invasion of personal privacy." It is CDC's firm position that all projects covered by a 308(d) confidentiality assurance, including NPCR-CSS, meet this exemption.

Has a case involving 308(d) been tested in court?

Yes. CDC's ability to protect data submitted to the agency was upheld in court. The case involved a National Institute for Occupational Safety and Health project collecting death certificate information, which is widely accepted as the least sensitive data protected by 308(d). The court's ruling in favor of the non-release of these data establishes an effective precedent for restricting access to more sensitive data, such as those collected by a cancer registry.

How long are confidential data submitted to NPCR-CSS protected?

NPCR-CSS data are covered by the 308(d) confidentiality assurance forever. Individual records in the NPCR-CSS are protected even following the death of the cancer patients.

Will NPCR release CSS data to persons or agencies outside of CDC?

An assurance of confidentiality protects NPCR-CSS data held at CDC and by its contractor. The 308(d) confidentiality protection does not go with the data whether released publicly or through restricted means, and any data CDC releases to

qualified researchers are subject to the limits of any coverage afforded by the requesting agency. However, it is important to note that NPCR's confidentiality assurance prohibits the release of any data that are directly or indirectly identifiable. Therefore, CDC would not release highly sensitive NPCR-CSS data. Restricted access data that are released to external researchers are done so in accordance with the NCHS RDC proposal process and confidentiality procedures, prohibiting attempts to identify subjects within the record system. Under the 308(d), NPCR is permitted to release NPCR-CSS data to qualified researchers and organizations, such as the North American Association of Central Cancer Registries (NAACCR), American Cancer Society (ACS), and National Cancer Institute (NCI). This is so because these entities were specifically mentioned in the NPCR-CSS confidentiality assurance as anticipated recipients of identifiable data. Prior to the restricted release of NPCR-CSS data to qualified organizations, a detailed data use agreement must be signed by the requesting party. Information that could lead to the identification of cancer patients, through direct or indirect methods, cannot be made available to any other group or individual. In particular, NPCR cannot disclose information to insurance companies; any party involved in civil, criminal, or administrative litigation; agencies of federal, state, or local government; or any other member of the public.

Are there penalties for violating the confidentiality assurance?

NPCR employees and NPCR-CSS contractor staff working on the NPCR-CSS project may be subject to fine, imprisonment, and termination of employment for unauthorized disclosure of confidential information. To assure that all NPCR employees are aware of their responsibilities to maintain and protect NPCR-CSS records and the penalties for failing to comply, CDC employees must read and sign a data use agreement. Contract employees with access to NPCR-CSS data are required to sign a confidentiality agreement.

The research use NPCR/SEER USCS Incidence Public Use dataset contains individual case-specific data from the USCS dataset with enforced fewer than 16 case cell suppression and case listing disabled.

Appendix J: Data Items for NPCR/SEER USCS Incidence Public Use Research Dataset

SEER*Stat Category	SEER*Stat Variable Name	Restrictions
Age at Diagnosis	Age recode with <1 year olds	
Race, Sex, Year Dx, Registry	Sex	
	Year of diagnosis	
	Addr at DX – state	
	USCS standard	
	Program	
	Region	
	USCS0122	
	USCS1423	
	USCS1923	
	Race recode (W, B, AIAN, API)	
	Race and origin recode (NHW, NHB, NHAIAN, NHAPI, Hispanic)	
	Origin recode NHIA (Hispanic, Non-Hisp)	
Site and Morphology	Primary site – labeled	
	Histologic type ICD-O-3	
	Grade	
	Grade clinical	
	Grade pathological	
	Diagnostic confirmation	
	ICD-O-3 hist/behavior, labeled	
	Site recode ICD-O-3/WHO 2008	
	Schema ID	
	ICCC site recode 3rd ed ICD-O-3/IARC 2017	
	ICCC site rec extended 3rd ed ICD-O-3/IARC 2017	
	AYA site recode 2020	
	Lymphoid neoplasm recode 2021 revision	
	Behavior ICD-O-3	
Stage – LRD [Summary and Historic]	Merged summary stage	
Therapy	RX summ – surg prim site	Female breast only and diagnosis years ≥2010
Extent of Disease	CS site-specific factor 1	Brain and other CNS and diagnosis years 2011–2017
	Merged estrogen receptor	Female, Schema ID=breast, and diagnosis years ≥2010

SEER*Stat Category	SEER*Stat Variable Name	Restrictions
	Merged progesterone receptor	Female, Schema ID=breast, and diagnosis years ≥2010
	Merged HER2 receptor	Female, Schema ID=breast, and diagnosis years ≥2010
	Laterality	
Multiple Primary Fields	Sequence number – central	
Geographic Locations	Rural-urban continuum 2013	Grouped into 3 categories: metro (RUCC 1–3); nonmetro (RUCC 4–9); unknown
Dates	Year of birth	
	Month of diagnosis	
Merged System-Supplied	Alcohol-related cancers	
	HPV-related cancers	
	Obesity-related cancers	
	Physical inactivity-related cancers	
	Tobacco-related cancers	
	State race eth suppress	

Appendix K: NPCR/SEER – U.S. Cancer Statistics Public Use Research Database User Agreement

The following form will be completed by individuals requesting access to the NPCR/SEER – U.S. Cancer Statistics Public Use Research Database.

National Program of Cancer Registries (NPCR) and Surveillance, Epidemiology, and End Results (SEER) Incidence—U.S. Cancer Statistics Public Use Research Database Request Form 2023 Data Submission

The Centers for Disease Control and Prevention (CDC) and the National Cancer Institute (NCI) make NPCR and SEER data available to the public and researchers through various data release activities. The U.S. Cancer Statistics Public Use Research Database is an unrestricted subset of data submitted to CDC and NCI and made available only through NCI's SEER*Stat software.

NPCR has obtained an assurance of confidentiality pursuant to Section 308(d) of the Public Health Service Act, 42 U.S.C. 242m(d). This assurance provides that identifiable or potentially identifiable data collected by CDC may be used only for the purpose for which they were obtained unless the person or establishment from which they were obtained has consented to such use. Any effort to determine the identity of any reported cases, or to use the information for any purpose other than statistical reporting and analysis, is a violation of the assurance.

By using the U.S. Cancer Statistics Public Use Research Database, you agree to comply with the following requirements:

- You will use these data for statistical reporting and analysis only.
- You will not present, publish, or otherwise make public statistics based on fewer than 16 cases.
- You will not link or permit others to link the database with individually identifiable records from any other dataset without CDC's written approval.
- You will not attempt to learn the identity of any person or establishment included in these data.
- If you discover the identity of any person or establishment inadvertently, you will not disclose or use this information, and you will advise CDC at uscsdata@cdc.gov.

We suggest the following citation: National Program of Cancer Registries and Surveillance, Epidemiology, and End Results Program SEER*Stat Database: NPCR and SEER Incidence—U.S. Cancer Statistics Public Use Research Database, [*year*] submission ([*year*]–[*year*]). United States Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute. Released June [*year*]. Available at www.cdc.gov/united-states-cancer-statistics/public-use/.

Please note: You must have access to SEER Research Plus before you can be given access to the database. When you have access to SEER Research Plus data, complete the fields below and email this form to uscsdata@imsweb.com. The email address you provide must be the same one used to obtain access to SEER Research Plus.

Name _____
Title and organization _____
Email address _____

The following text will appear in a pop-up message box whenever a user opens the NPCR/SEER – U.S. Cancer Statistics Public Use Research Database.

> The National Program of Cancer Registries (NPCR), Centers for Disease Control and Prevention (CDC), has obtained an assurance of confidentiality pursuant to Section 308(d) of the Public Health Service Act, 42 U.S.C. 242m(d). This assurance provides that identifiable or potentially identifiable data collected by the NPCR may be used only for the purpose for which they were obtained unless the person or establishment from which they were obtained has consented to such use. Any effort to determine the identity of any reported cases, or to use the information for any purpose other than statistical reporting and analysis, is a violation of the assurance.
>
> By using the U.S. Cancer Statistics Public Use Research Database, you agree to comply with the following requirements:
>
> - You will use these data for statistical reporting and analysis only.
> - You will not present, publish, or otherwise make public statistics based on fewer than 16 cases.
> - You will not link or permit others to link the database with individually identifiable records from any other dataset without CDC's written approval.
> - You will not attempt to learn the identity of any person or establishment included in these data.
> - If you discover the identity of any person or establishment inadvertently, you will not disclose or use this information, and you will advise CDC at uscsdata@cdc.gov.
>
> By clicking the "OK" button I signify that I will abide by the terms of data use states above.

Appendix L: NPCR Data at the NCHS RDC Questions and Answers

Can you summarize the data access process?

CDC uses the National Center for Health Statistics' (NCHS) Research Data Center (RDC) as a mechanism for researchers outside of the Division of Cancer Prevention and Control (DCPC) to request and gain access to the Restricted-Access NPCR/SEER data for research purposes. The data are available through the NCHS RDC only after the standard data quality reviews that occur as part of the preparation for USCS.

The use of the NCHS RDC to manage data access provides the highest level of data security and protection of confidentiality that is available for analysis of data. Any researcher must submit a proposal that is reviewed and approved by CDC and may be reviewed by representatives from the participating central cancer registries (CCRs) before any data analysis begins. Trained data analysts at the NCHS RDC create a dataset that is customized to each analysis. The researcher can run his or her own statistical analysis or have the NCHS RDC analyst run the analysis. The NCHS RDC analyst reviews all output from statistical analysis to ensure that the researcher only conducts analyses relevant to the approved protocol and that small cell sizes are suppressed. Absolutely no individual-level data leave the NCHS RDC facilities. The data can only be accessed onsite; the NCHS RDC remote option is not available for the Restricted-Access NPCR/SEER data.

What is National Center for Health Statistics (NCHS)?

NCHS is one of the national centers at CDC and is located in Hyattsville, Maryland. As the nation's principal health statistics agency, it compiles statistical information to guide actions and policies to improve the health of the population. More information about NCHS is available at www.cdc.gov/nchs/about.htm.

What is the Research Data Center (RDC)?

The NCHS RDC began in 1998 and has a long history of managing access to health and vital statistics data through a rigorous proposal review process as well as review of the statistical output. Its mission is to give public access to a full range of health and vital statistics data, while protecting the confidentiality of the respondents and institutions that collected the information. There have been no breeches of confidentiality for data access through the NCHS RDC.

The NCHS RDC houses sensitive, but not classified, data. It allows access to individual data without the possibility of disclosure of identifying information. The NCHS RDC offers statistical, programming, and consulting expertise to facilitate the data analysis for research.

The NCHS RDC is a data hosting center, not a data repository. The data extracts that are hosted on the NCHS RDC are tailored specifically to the proposal and have a research life cycle. When the analysis is completed, the data extract is archived for 2 years and then destroyed. More information about the NCHS RDC is available at www.cdc.gov/rdc/.

Why does CDC use the NCHS RDC?

Maintaining confidentiality is the primary objective of the NCHS RDC. Its staff have statistical expertise to address confidentiality and disclosure risk. Using the NCHS RDC allows CDC to comply with the Assurance of Confidentiality [308(d)] that was obtained for the NPCR-CSS data. All researchers must take a confidentiality orientation, complete confidentiality forms, and review the disclosure manual, all of which outline practices that are essential to protecting the data and preventing disclosure of confidential information. Additionally, data housed at the NCHS RDC are not subject to the Freedom of Information Act (FOIA). More information about confidentiality is available at www.cdc.gov/rdc/B4ConfiDisc/CfD400.htm.

The use of the NCHS RDC to host the NPCR data provides confidence in knowing that the data are used correctly and safely by external researchers. In addition, this approach will not overtax resources in CDC's Division of Cancer

Prevention and Control or in the CCRs. The NCHS RDC provides a level of data control beyond that of any other data access system used for registry data.

What is the research proposal process?

The NCHS RDC has a rigorous review process for analyses proposed by any researchers wanting to use RADS data. All proposals will be evaluated by a Review Committee consisting of the NCHS RDC Director, the Confidentiality Officer, the assigned NCHS RDC analyst, and NPCR representatives. The iterative review and comment process may take 6 to 8 weeks.

Through this process, the NCHS RDC staff, the NPCR staff, and the CCR staff will fully understand the intended analysis and will be able to provide any needed direction or restrictions on the analysis and describe any limitations in what is proposed. It will be possible for CDC and participating registries to disapprove a proposal. However, guidance and re-direction as needed should be the norm. More Information about the review process is available at www.cdc.gov/rdc/leftbrch/userestricdt.htm.

Once a proposal has been approved, the NCHS RDC offers a secure environment for data analyses and has processes in place to review data output for small cell sizes. This will ensure that the NPCR suppression rules are properly applied. Through the NCHS RDC, the user can conduct analyses and have remote access to data but cannot download the individual record level data or obtain counts for inappropriately small cell sizes.

Who has access to the data and at what level?

The NCHS RDC analysts can access the individual record-level data, since it is easier to create an analytic dataset using these data. The NCHS RDC analysts are bound by the same data use agreements that CDC staff sign annually. Researchers with approved proposals can conduct analyses through the NCHS RDC on the created dataset or have an NCHS RDC analyst do the analysis for them. However, they cannot download any part of the data from the NCHS RDC. Any additional variables that were not included in the original analysis proposal will need a separate approval process.

Note that this is different from the process that NPCR has used in the past, where researchers with approved proposals had direct access to the dataset, including the ability to download the data and create a list of individual record-level data and all variables in the dataset.

Researchers have several possible modes of access to the dataset created for their specific research proposal. More information is available at www.cdc.gov/rdc/B2AccessMod/ACs200.htm.

When a researcher conducts an analysis, what type of output will he or she get?

If a researcher is on-site at the NCHS RDC, he or she can save the results on the hard drive of the NCHS RDC computer. The NCHS RDC analyst will review the output for disclosure then either load the output onto a flash drive supplied by the researcher or e-mail the output files to the researcher. If a researcher is accessing the NCHS RDC remotely, he or she will send program by e-mail and, after disclosure review by the NCHS RDC analyst, will receive the output files by e-mail. No individual record level data are released to the researcher.

Can CCRs decide whether their data are available through the NCHS RDC?

Participation in all CDC-created and hosted analytic datasets and web-based data query systems, as outlined in the annual NPCR-CSS Data Release Policy, is a required strategy. Therefore, data from all CCRs meeting eligibility requirements are included. Data use is important to NPCR and for continued support of the registries.

Will the CCRs be able to decide if their county-identifying variable (County at Dx-Analysis

[NAACCR#89]) is to be available for use in the NCHS RDC?

Participation in all CDC-created and hosted analytic datasets and web-based data query systems, as outlined in the annual NPCR-CSS Data Release Policy, is a required strategy. [DP22-2202, Program 3: NPCR, Strategy 1, Required Activities, Data Submission (page 33)]. Therefore, data from all CCRs meeting eligibility requirements are included. Data use is important to NPCR and for continued support of the registries. County data will be used only in approved analyses and in the following ways:

- As a linkage variable (to census data, for example) only by the NCHS RDC analyst. The county variable will not be available to the researcher. The NCHS RDC analyst uses it to create a linked dataset and then removes it.
- As a confounder or other control variable, but no data are presented by county. The NCHS RDC analyst creates dummy variables to mask the actual county name.
- In geographically aggregated form such as large metropolitan statistical areas (those with a population of 1 million or larger), multi-county regions, or geographic areas such as Appalachia or IHS Contract Health Services Delivery Areas (CHSDA) counties. The county data make it possible for the NCHS RDC analyst to create these areas for the researcher.

Previous data release policies indicate that the project proposals for RADS would be reviewed by the RADS working group, facilitated by CDC with representation by the CCRs. Does this procedure change now that the NCHS RDC is used?

The CCRs will still have input on the RADS proposals. The NCHS RDC review process includes the NCHS RDC analyst and the confidentiality officer, who are responsible mainly for disclosure review to ensure that we abide by the 308(d) assurance of confidentiality obtained for NPCR-CSS. More information about the NCHS RDC review process is available at: https://www.cdc.gov/rdc/b1datatype/rdc-Output.htm.

Will SEER data be included for analysis or will the data be limited to NPCR data?

Yes. Both NPCR and SEER data may be accessed through the NCHS RDC.

Will the NCHS RDC staff have access to SEER*Prep and SEER*Stat?

No. NPCR previously provided a SEER*Stat file to the NCHS RDC but found that researchers only used the SAS file. Therefore, the SEER*Stat file is no longer provided.

What suppression rules will be used for the RADS?

RADS use the same suppression rules that are used for *United States Cancer Statistics.* More detailed information is available at www.cdc.gov/cancer/npcr/uscs/technical_notes/stat_methods/suppression.htm.

The suppression rules for Asian and Pacific Islander (A/PI) people and American Indian and Alaska Native (AI/AN) people will also apply.

CDC doesn’t collect personal identifiers like name or Social Security number. Wouldn’t it be better for researchers to contact CCRs directly for linkage studies?

Yes, it would be best for researchers to contact CCRs directly for linkage studies that require individual identifiers. However, valuable public health research can be conducted with access to county-level data. Examples include linkage with U.S. Census data for socioeconomic analyses, or to examine regional differences in the prevalence of a specific cancer.

Will IRB review be required for each proposal? If not, will NCHS require researchers to obtain IRB approval before they submit their proposal?

The NCHS RDC has an umbrella ethics review board (ERB) protocol that covers CDC employees and can be extended to external researchers. The principal investigator and all research team members who come in contact with the data must take the confidentiality orientation and complete the confidentiality forms. One of the confidentiality forms is the designated agent form (www.cdc.gov/rdc/Data/B4/DesignatedAgent.pdf [PDF-41KB]), which extends the ERB to cover external researchers.

Note that the ERB protocol serves the same function as an institutional review board (IRB) protocol. At CDC, one office coordinates the submission and tracking of human research protocols. However, other centers such as NCHS and the National Institute of Occupational Safety and Health have different names for these review boards: Research Ethics Review Board (ERB) at NCHS and Human Subjects Review Board (HSRB) at NIOSH.

Researchers may choose to obtain an IRB protocol from their own institution, but it will not be required.

Does access to the RADS cost anything?

No. CDC covers the cost of analyzing RADS through the NCHS RDC.

As more researchers become aware of the RADS, they may want access to additional variables that CCRs submit to CDC. How will this process be handled?

The addition of new variables in RADS will be discussed with CCRs prior to their inclusion in the data release policy, which is updated annually.

How is access to the comparative effectiveness research (CER) dataset managed?

Access to the CER dataset is managed through the same NCHS RDC process. The proposal process will not differ except that staff from the specialized registries funded for CER data collection will review these proposals.

Appendix M: Data Items for Restricted-Access Dataset (RDC)

The restricted-access dataset consists of individual case-specific data derived from the NPCR-CSS dataset. The data are available to researchers at NCHS' Research Data Center (RDC) as a SAS file. SAS files are created specifically for each project's needs. The data items that researchers may request are listed below.

Variable Name

- Alternate Patient ID Number
- Address at Diagnosis – State
- Address at Diagnosis – County at Analysis*
- USCS Standard
- USCS9923
- USCS1923
- USCS9823
- USCS1423
- Address at Diagnosis – Census Region
- Race 1
- Race 2
- Race Recode (W, B, AI, AN)
- Race and origin recode (NHW, NHB, NHAIAN, NHAPI, Hispanic)
- Econ Status
- State race eth suppress
- Origin Recode NHIA (Hispanic, Non-Hispanic)
- IHS Link
- State Race Ethnicity Suppress
- Sex
- Age at Diagnosis**
- Age Recode
- Birth Date***
- Econ status
- Rural-urban continuum 2013
- Sequence Number – Central
- Date of Diagnosis****
- Primary Site
- Laterality
- Grade
- Grade Clinical
- Grade Pathological
- Grade Post Therapy
- Diagnostic Confirmation
- Type of Reporting Source
- Histologic Type ICD-O-3
- Behavior Code ICD-O-3
- Behavior Recode for Analysis Derived/WHO 2008
- Primary Site Recode
- Schema ID
- SEER International Classification of Childhood Cancer (ICCC) Recode Extended 3rd edition/IARC 2017
- AYA Site Recode 2020

- Lymphoma Neoplasm Recode 2020 Revision
- SEER Summary Stage 2000
- SEER Summary Stage 1977
- Derived SS2000
- Summary Stage 2018
- Merged Summary Stage
- RX Summ – Surg Prim Site
- Merged radiation
- CS Site-Specific Factor 1
- Merged Estrogen Receptor
- Merged Progesterone Receptor
- Merged HER2 Receptor
- Over-ride Age/Site/Morph
- Over-ride SeqNo/DxConf
- Over-ride Site/Lat/Sequence Number
- Over-ride Site/Type
- Over-ride Histology
- Over-ride Report Source
- Over-ride Ill-define Site
- Over-ride Leuk, Lymphoma
- Over-ride Site/Behavior
- Over-ride Site/Lat/Morph
- Alcohol-related cancers
- HPV-related cancers
- Obesity-related cancers
- Physical activity-related cancers
- Tobacco-related cancers

* County data will be used only in approved analyses and in the following ways: a) as a linkage variable (linkage to census data, for example) only by the NCHS RDC analyst; b) as a confounder or other control variable, but no data are presented by county; c) in geographically aggregated form such as large metropolitan statistical areas (those with a population of 1 million or larger), multi-county regions, or geographical areas such as Appalachia or IHS Contract Health Services Delivery Areas (CHSDA) counties.

**Age over 99 is recoded.

***Only year is provided; if age is over 99, year of birth is recoded.

****Day of diagnosis is not provided.

Appendix N: NPCR-CSS Levels of Data Access

USCS Data Visualizations tool

Includes: State, county, region, and Congressional-district levels, no record-level information

Criteria: USCS criteria met, permission provided on Dataset Participation Agreement, <16 cases cell suppression enforced

Availability: Public

CDC WONDER

Includes: State, county, region, and MSA levels, no record-level information

Criteria: USCS criteria met, permission provided on Dataset Participation Agreement, <16 cases cell suppression enforced

Availability: Public

State Cancer Profiles

Includes: State and county levels, no record-level information

Criteria: USCS criteria met, permission provided on Dataset Participation Agreement, <16 cases cell suppression enforced

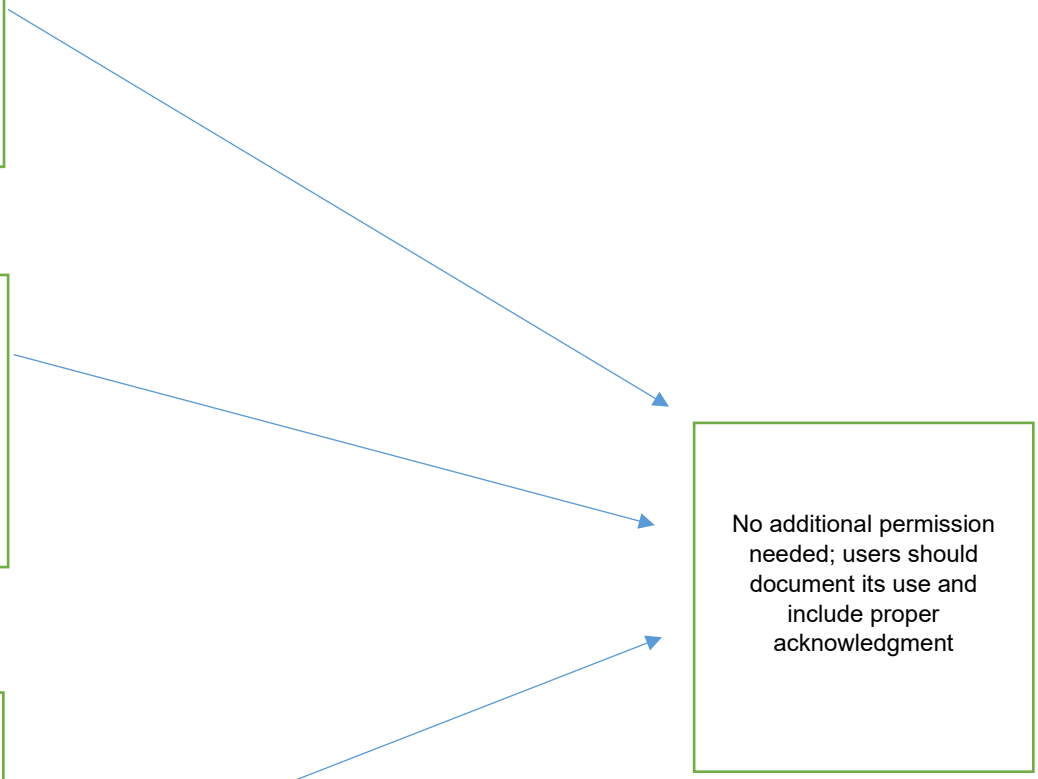
Availability: Public

NPCR/SEER USCS Public Use Dataset

Includes: State record-level information, no case listing

Criteria: USCS criteria met, permission provided on Dataset Participation Agreement, <16 cases cell suppression enforced

Availability: Public, user agreement statement must be acknowledged every time a user opens the database



Appendix O: Data Items for NPCR/SEER USCS Delay-Adjusted Database

SEER*Stat Category	SEER*Stat Variable Name
Age at Diagnosis	Delay age
	Age recode with <1 year olds
Race, Sex, Year Dx, Registry, County	Sex
	Year of diagnosis
	Addr at DX – state
	County at DX Analysis
	State-county
Required Delay Fields	Delay factor
	Delay site
	Delay race (All, Race recode (White, Black, AIAN, CHSDA, API, Hisp, Non-Hisp)
Site and Morphology	Behavior recode for analysis derived/WHO2008
Multiple Primary Fields	Sequence number - central