National Program of Cancer Registries Cancer Surveillance System (NPCR-CSS)

2015 Data Release Policy Diagnosis Years 1995–2013

Policy Revised June 2015

Cancer Surveillance Branch
Division of Cancer Prevention and Control
NCCDPHP, CDC
4770 Buford Hwy, N.E., Mailstop F-76
Atlanta, GA 30341-3717
E-mail: cancerinfo@cdc.gov (specify "NPCR-CSS" in subject line)

TABLE OF CONTENTS

National Program of Cancer Registries
Cancer Surveillance System
2015 Data Release Policy
June 2015

I. INTRODUCTION

This document describes the format and content of data that the Centers for Disease Control and Prevention's National Program of Cancer Registries (NPCR) Cancer Surveillance System (CSS) releases or shares. This multi-year policy updates the May 2014 NPCR-CSS Data Release Policy. This policy applies to data submitted to the Centers for Disease Control and Prevention (CDC) for the 2015 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 data, ensuring that the data are released according to the terms of the NPCR-CSS Data Release 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 as a result 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 what is 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 is described in this policy, NPCR central 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.

A. Summary of Changes

- Clarification of unsmoothed and smoothed rates for the Environmental Public Health Tracking Network dataset, page 5
- Information regarding the International Association of Cancer Registries' Call for Data returned to the Policy on a permanent basis, page 6
- USCS SEER*Stat Restricted Dataset (restricted to NPCR Cancer Registry staff) replaced with USCS SEER*Stat Research Dataset (open to everyone), page 8
- Race/ethnicity combination data presentation added as White Hispanic, White non-Hispanic, Black Hispanic, Black non-Hispanic (with appropriate cell suppressions), Table page 18
- CDC Nondisclosure Agreement added, Appendix E, page 27
- Data items for USCS Internal Analytic Dataset added, Appendix F, page 28
- Data items for NPCR Internal Survival Dataset added, Appendix G, page 30
- Data Items for USCS SEER*Stat Research Dataset added, Appendix J, page 36
- NPCR Research Data Use Agreement added, Appendix K, page 38
- Flowchart for NPCR-CSS Levels of Data Access added, Appendix N, page 46

II. OVERVIEW OF DATA

In 1992 Congress established NPCR by enacting the Cancer Registries Amendment Act, Public Law 102-515.[4] 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 for the planning and evaluation of 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 Program (SEER);[11,12] 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);[13] and established standards for data management and data completeness, timeliness, and quality similar to those recommended by NAACCR.[13,14] In 1994, the first 37 States received funding from CDC.[15] Currently, 45 States, the District of Columbia, Puerto Rico, and the U.S. Pacific Island Jurisdictions are funded by NPCR (appendix A).[16] NPCR-funded central registries collect data on patient demographics, primary tumor site, morphology, stage of disease at diagnosis, and first course of treatment. In addition, NPCR central registries conduct follow-up for vital status by linking with State and national death files.

Invasive and in situ cancer case reports are submitted to CDC by population-based statewide central cancer registries in all 45 participating States, the District of Columbia, Puerto Rico, and the U.S. Pacific Island Jurisdictions. In each State or territory, State laws and regulations mandate the reporting of cancer cases by facilities and practitioners who diagnose or treat cancer to the State health department or its designee.[4] The central cancer registry receives case reports from facilities and practitioners throughout the State and processes them according to standard data management procedures.[14] 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). For more information on NPCR-CSS data, see the Technical Notes as posted on the *United States Cancer Statistics* (*USCS*) Web site (http://www.cdc.gov/uscs), which is updated annually. NPCR-CSS received formal approval (protocol #2594) from CDC's Institutional Review Board (IRB) in October 1999. The approval is updated annually. 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) (http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm).

Central cancer registries and Federal agencies routinely publish cancer incidence data 22 months after the close of each diagnosis year based on data that meet standards for completeness and quality.[16,17] 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 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 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 (i.e., 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 22 months after the close of a diagnosis year are likely to be small; however, delays in reporting are more likely to impact certain cancer sites and may be important for some research studies.[18]

III. DATA RELEASE ACTIVITIES

A. 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 (i.e., 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 and have a database behind a CDC firewall that is case-specific microdata.[2, 5–10] Users are able to 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). These systems do not contain information that is identifiable or potentially identifiable according to currently accepted procedures for reducing disclosure risk.[2, 5–10] Before each system is finalized, the aggregate values are analyzed to determine whether there is a need for complementary cell suppression.[2, 5–10] If appropriate, the analysis includes consultation with a statistician with specific 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 page 11) 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 vehicles (see table). As a convenience to NPCR central registries, States may request from CDC a copy of their complete State-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:

- United States Cancer Statistics (USCS)
- USCS expanded dataset
- USCS county cancer incidence data
- Environmental Public Health Tracking Network (EPHTN) data

All NPCR-CSS public web-based query systems consist of cancer incidence data selected from the NPCR-CSS analytic database. This is the same database that provides cancer incidence data for the annual publication of USCS.[16] Data sources, case definitions, basic registry eligibility criteria in terms of required data completeness and quality, population denominator sources, methods for calculating incidence rates, and the rationale for specific cell suppression thresholds are as described in the Technical Notes for USCS, unless noted in separate documentation that accompanies the data.

Separate documentation may accompany each system that describes its unique features (e.g., the data submission date, percentage of the U.S. population covered, diagnosis years and cancer sites included, variables included, any special data quality criteria required for inclusion, and any unique statistical methods employed).

1. United States Cancer Statistics (USCS)

The USCS dataset is a hypertext markup language (HTML) file containing the aggregate counts and rates for incidence and mortality published annually (see web based report at www.cdc.gov/uscs), with text documentation. The USCS web-based report is a public web-based query system which displays the single year and 5-year aggregate counts, age-adjusted rates, and 95-percent confidence intervals at the State, regional, and national levels. The Web-based query system has the database behind a CDC firewall with pre-tabulated data using SEER*Stat queries, which allows for the display of counts and rates. Users are able to access only aggregate counts and rates with all confidentiality protections built in. Downloadable ASCII or MS Excel files, are available. Male and female combined counts and rates are also available in the HTML edition, in addition to the published sex-specific tables.

2. USCS Expanded Dataset (WONDER)

The USCS expanded dataset displays the aggregate counts, rates, and 95-percent confidence intervals published yearly in USCS, as well as additional aggregate values created from the same analytic file containing more detailed breakdowns of counts and rates based on selected variables (see table). These data are available on the Wide-ranging Online Data for Epidemiologic Research (WONDER) (**http://wonder.cdc.gov**), a Web-based query system that has a database behind a CDC firewall with case-specific microdata. However, users are able to access only aggregate counts and rates with all confidentiality protections built in. Because this system presents data in more detail than is presented in USCS, States have the option to notify NPCR if they prefer not to have their State data included.

3. USCS County Cancer Incidence Dataset (SCP & OWH)

The USCS county cancer incidence dataset consists of aggregate cancer incidence counts, crude rates, and age-adjusted rates for selected counties in the United States (see table). These data are available as an ASCII file. Because this dataset presents data at a sub-State geographic level, States have the option to notify NPCR if they prefer not to have their county data included. A limited version of this dataset has been released to a small number of users and states have the option to release age-adjusted rates only:

a) Age-adjusted rates only

Data released to NCI for the State Cancer Profiles (SCP) project (***www.statecancerprofiles.cancer.gov***) includes age-adjusted rates only.

b) Age-adjusted and crude rates

Data released to the U.S. Department of Health and Human Services, Office of Women's Health (OWH) includes crude and age-adjusted rates.

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.

4. Environmental Public Health Tracking Network (EPHTN) Dataset

The EPHTN dataset for the national EPHTN portal displays the 5-year aggregate counts, age-adjusted rates, and 95-percent confidence intervals for selected cancer sites at the State and county level (see table). This dataset is available on the public EPHTN network (http://www.cdc.gov/nceh/tracking), a Web-based query system that has the database behind a CDC firewall with case-specific microdata, which allows for the calculation of locally-weighted smoothed rates or unsmoothed rates, or both:

a) EPHTN Unsmoothed Rates.
Data published are similar to that on State Cancer Profiles.

b) EPHTN Smoothed Rates.
Smoothing is the process of averaging a measure for an area based on information about that area and areas around it. Please note that the main purpose of smoothing is to clarify spatial patterns and to improve the stability of rates, not to prevent disclosure of private information. Back-calculation of case counts from smoothed rates is sometimes possible when the method of smoothing is made known and (non-sensitive) denominator data are available from other sources.

Through EPHTN, users are able to access only aggregate counts and rates with all confidentiality protections built in. Because this system presents data using a locally weighted smoothing procedure, States have the option to notify NPCR if they prefer not to have their State and county data included.

c) EPHTN National Portal to State Portal
CDC's Environmental Health Tracking Branch (EHTB) has grantees in several NPCR-funded states that are responsible for the state-level public portals. In collaboration with EHTB, upon request, CDC-NPCR provides the state-level EPHTN dataset to the EHTB state counterpart. States have the option to notify NPCR if they prefer not to have their data provided to the EHTB state counterpart.

B. Data Release to Collaborating Partners

1. 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: http://www.cbtrus.org/reports/reports.html. The report includes age-adjusted rates and corresponding 95-percent confidence intervals on brain and other central nervous system tumors and is presented by State, histology, major histology grouping, primary site, behavior, gender, race, ethnicity, and age at diagnosis. 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 that agree to participate and meet 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 is released; only aggregated data with the corresponding confidence intervals (if applicable) and appropriate suppression criteria are provided to data inquirers. Attribution to the 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.

2. Indian Health Services (IHS)

CDC continues to use the IHS linkage results for analyses related to cancer incidence in the AI/AN population (e.g., USCS). In addition to improving cancer incidence rates presented in USCS, an analytic database is maintained by an Albuquerque CSB assignee for DCPC AI/AN activities, with access limited to approved CDC staff. These activities include responding to data requests for AI/AN cancer incidence rates from tribal epidemiology centers and tribal organizations contingent upon permission from the state registries that comprise the IHS areas of interest. Inclusion in this dataset also includes an option for IHS to ***provide the state with the date of death obtained through NDI-IHS linkage and/or the date the linkage occurred by diagnosis year***.

3. International Association of Cancer Registries (IACR)

The International Association of Cancer Registries (IACR) produce the *Cancer Incidence in Five Continents* (CI5) and the *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 CI5 databases provide access to detailed information on the incidence of cancer recorded by cancer registries (regional or national) worldwide in two formats and the IICC provides access to detailed information on the incidence of pediatric cancers:

a) CI5

Presents the basic data published in the CI5 volumes.

b) CI5plus

Contains annual incidence for selected cancer registries published in CI5 for the longest possible period.

c) IICC

Presents basic pediatric data.

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. CDC-NPCR will provide additional information regarding the CI5 Call for Data as it becomes available. There are two components of the CI5 Call for Data: 1) the questionnaire and 2) data submission. .

Data submitted for CI5 may also be used for the IICC publication making a separate data submission unnecessary. This IACR product does require a separate questionnaire to be completed by the states. An introductory text is required for this publication.

States are responsible for completing the on-line questionnaires and providing an introductory text, indicating if the CI5 data and introductory text are also used for the IICC product. States may opt to submit data on their own or have CDC-NPCR to submit data on their behalf. CDC-NPCR will submit aggregated NPCR data for central cancer registries that provide permission and meet USCS publication criteria.

C. Analytic datasets

1. CDC Internal Analytic Data

CDC staff members or contractors perform analyses of NPCR program data as needed utilizing an internal analytic database based on the USCS dataset. This dataset is available to federal employees in the Division of Cancer Prevention and Control after signing a *NPCR Analytic Data Use Agreement* (appendix D) and *CDC Nondisclosure Agreement* (appendix E). Analyses include assessment of the completeness, timeliness, and quality of cancer incidence data and analyses of the cancer burden and survival as needed for meeting national cancer control objectives. Such analyses of State and national data are conducted routinely by Federal agencies, including CDC, for programmatic or statistical purposes, as needed, to achieve the agency's mandate.

There are two internal analytic datasets routinely analyzed by CDC staff members:

a) USCS Internal Analytic Dataset

Publications or presentations describing the quality of the data or the burden of cancer are one outcome of such analyses. Examples of topics for such analyses are descriptive analyses by racial and ethnic populations for specific cancers and descriptions of cancer incidence trends. Appendix F lists the variables available in this dataset.

b) NPCR Internal Survival Dataset

Cancer survival data is critical for evaluating the progress and impact of early detection/screening programs and/or comprehensive cancer control plans as well as interventions from other sources. CDC's NPCR-CSS is now poised to calculate and publish survival rates on this population at the national, state, and regional levels. Focusing on the entire NPCR-CSS dataset supports analyses of survival estimates for rare cancers that cannot be addressed otherwise and provides data for publication on the USCS website as official statistics for the U.S. Appendix G lists the variables available in this dataset.

In compliance with the 308(d) Assurance of Confidentiality, CDC employees and contractors 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/Security Statement (appendices B, H, and I).

In addition to adhering to strict requirements for protecting confidentiality, CDC staff members notify the State cancer registry(ies) in advance whenever they plan to present, publish, or release State-specific information on cancer incidence or survival that have not been previously presented, published, or released. This notification includes, when possible, sending a pre-publication copy of the entire publication or other information to the specific States. When that is not possible (for example, if the information is embargoed), the specific State cancer registries receive a summary of the information before it is published or released. In addition, CDC staff members are required to acknowledge State cancer registries whenever NPCR-CSS data are presented, released, or published by CDC by making available the following (or similar) statement:

These data were provided by cancer registries participating in NPCR and submitted to CDC in the (insert submission date) NPCR-Cancer Surveillance System data submission.

2. USCS SEER*Stat Research Data

For purposes of this policy, the USCS SEER*Stat Research Dataset is defined as the version of the full NPCR-CSS microdata (i.e., individual case-specific data) that have been modified as needed to minimize the potential for disclosure of confidential information. It consists of a subset of data items published in USCS. This 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. Certain data items such as date of birth have also been removed from this research dataset to minimize the potential identification of individuals with the occurrence of 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 this dataset is in appendix J.

The dataset was previously only available to NPCR Registry Staff and is now available to anyone after signing an *NPCR Research Data Use Agreement* (Appendix K). A Public Release Disclosure Statement cautions users against inappropriate use of the data or inappropriate disclosure of information. Cell suppression of <16 cases is automatic and the SEER*Stat case listing function is disabled as additional data protection measures. This dataset allows the authorized counts, crude rates, age-adjusted rates, and 95-percent confidence intervals to be generated by the authorized user to meet their specific needs.

Registries have the option to notify NPCR if they prefer to not have their State and county data included in any research use dataset, or they may choose to limit release of their data to certain groups.

3. Restricted-Access Research Dataset (RDC)

For purposes of this policy, the restricted-access dataset is defined as the version of the full NPCR-CSS analytic dataset, either aggregated data or microdata (i.e., 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 is available through the NCHS RDC only after the standard data quality reviews that occur as part of the preparation for USCS and State Cancer Profiles. 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 after the researcher completes the NCHS RDC confidentiality and security requirements. The requestor must also comply with the confidentiality procedures at and data sharing agreements with the NCHS RDC.

The NCHS RDC has developed and maintains detailed data sharing agreements and procedures for user authentication and for logging and monitoring of data releases. Proposed project proposals are reviewed by staff at central cancer registries, through the NPCR Central Cancer Registry Council, and by CDC, which includes NPCR and NCHS RDC staff. User documentation including a data dictionary for every diagnosis year available at the NCHS RDC is provided.

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 data analysis. Using the NCHS RDC 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 at the RDC 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 list of the restricted-access dataset variables is in appendix M.

Because this restricted-access dataset may potentially contain identifiable information, States have the option to notify NPCR if they prefer not to have their data included in this restricted access dataset. Options for inclusion include:

a) All Variables, Including County at Diagnosis

County data will only be used in approved analyses and in the following ways: a) used as a linkage variable (linkage to census data, for example) only by RDC staff. The county variable will not be available to the researcher, but the RDC analyst would use it to create a linked dataset and then remove the county variable; b) included as a confounder or other control variable, but no data are presented by county. Again, it will be possible for the RDC data analyst to mask the actual county name and create dummy variables for this purpose; c) used in geographically aggregated form such as large metropolitan statistical areas (e.g., those with a population of 1 million or larger), multi-county regions, or geographical areas (e.g., Appalachia or IHS Contract Health Services Delivery Areas (CHSDA) counties). It will be possible for RDC analysts to create these areas for the researcher.

b) All Variables Except County at Diagnosis

The same file with the county at time of diagnosis removed.

D. Data Release Under Controlled Conditions

CDC-wide policy stipulates that a CDC program may consider release of data that cannot be released as either a public Web-based system, a research dataset, or restricted-access dataset under certain controlled conditions.[1] These controlled conditions may include a CDC-controlled data center such as the data center established at National Center for Health Statistics (NCHS) (http://www.cdc.gov/nchs/r&d/rdc.htm) or through special licensing. NPCR-CSS data are available at the NCHS RDC but will not otherwise be released under these controlled conditions while the current policy is in place. Release of data under controlled conditions will be considered as part of discussions with partners, and a determination will be made as to whether such releases of data will be considered for NPCR-CSS data.

E. 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 are shared with CDC employees and contractors, NPCR central registries, and other partners in order to facilitate quality reviews of the data. When appropriate, individuals who participate in such reviews sign a *NPCR Analytic Data Use Agreement* and a *CDC Nondisclosure Agreement* (when applicable) before accessing the data or tables.

IV. PROTECTION OF DATA

A. 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[1–3] and specific

restrictions imposed on NPCR-CSS data (appendices B, H, and I).[4] 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.[5–10]

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.

B. 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 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 and 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 less 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.

Asian/Pacific Islander and American Indian/Alaskan Native data are presented only for the nation, and states with at least 50,000 population, because of concerns regarding possible misclassification of race data and the relatively small sizes of these populations in the United States.

C. 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 or metropolitan area cancer registry, of any such discovery.*

D. Freedom of Information Act (FOIA) Data Requests

The Freedom of Information Act (FOIA) (http://www.cdc.gov/od/foia/) generally provides that, upon written request from any person, a Federal agency (i.e., 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-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 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), and
- 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. There are three reasons for this: (1) the State health departments can release cancer incidence data in accordance with locally established policies and procedures and consistent with provisions of the Cancer Registries Amendment Act (Public Health Service Act, (42 USC 280e-280e-4), as amended);[4] (2) the relative infrequency of data submission to Federal agencies assures that the State health department or its designated central cancer registry will have the most complete,

accurate, and up-to-date information; and (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, 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.

E. CDC 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, CDC staff members or contractors tabulate the data for the inquirer. For requests that identify a State, CDC staff members seek States' permission regarding use. See Appendix N for additional details.

V. REFERENCES

1. Centers for Disease Control and Prevention. *CDC/ATSDR Policy on Releasing and Sharing Data*. Atlanta: Centers for Disease Control and Prevention; 2003. Available at http://www.cdc.gov/maso/Policy/ReleasingData.pdf.
2. Centers for Disease Control and Prevention. *CDC/ATSDR/CSTE Data Release Guidelines for Re-Release of State Data*. Atlanta: Centers for Disease Control and Prevention; 2003 (available upon request).
3. 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; 1999.
4. Cancer Registries Amendment Act, Public Law 102-515, Stat. 3312 (October 22, 1992). Available at http://www.cdc.gov/cancer/npcr/npcrpdfs/publaw.pdf.
5. American Statistical Association. Data Access and Personal Privacy: Appropriate Methods of Disclosure Control. Alexandria, VA: American Statistical Association; 2008. Available at http://www.amstat.org/news/statementondataaccess.cfm.
6. 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.
7. Federal Committee on Statistical Methodology. *Checklist on Disclosure Potential of Proposed Data Releases*. Available at http://fcsm.sites.usa.gov/committees/cdac/cdac-checklist/.
8. Federal Committee on Statistical Methodology. *Report on Statistical Disclosure Limitation Methodology*. (*Statistical Working Paper* 22). Washington, DC: Office of Management and Budget; 1994. Available at http://www.fcsm.gov/reports/#fcsm.
9. McLaughlin C. Confidentiality protection in publicly released central cancer registry data. *Journal of Registry Management* 2002; 29(3):84–88.
10. 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.
11. Surveillance, Epidemiology, and End Results Program. *The SEER Program Code Manual*. 3rd ed. Bethesda, MD: National Cancer Institute,1998.
12. Percy C, Van Holten V, Muir C (eds). *International Classification of Diseases for Oncology, 2nd edition.* Geneva, Switzerland: World Health Organization; 1990.
13. Hultstrom D, editor. *Standards for Cancer Registries, vol. II: Data Standards and Data Dictionary*, version 9.1, 6th ed. Springfield, IL: North American Association of Central Cancer Registries; 2001.
14. North American Association of Central Cancer Registries. *Standards for Cancer Registries, vol. III: Standards for Completeness, Quality, Analysis, and Management of Data*. Springfield, IL: North American Association of Central Cancer Registries; 2000.
15. 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.
16. U.S. Cancer Statistics Working Group. *U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999–2011 Incidence and Mortality Web-based Report, Technical Notes.*

Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute; 2014. Available at: www.cdc.gov/uscs.
17. NAACCR Method to Estimate Completeness. A Data Analysis Tool for Calculations. Available at: http://www.naaccr.org/Research/DataAnalysisTools.aspx.
18. 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):1537–45.

TABLE
Comparison of the NPCR-CSS Datasets

Overview						
	Public Web-Based Query Systems				Analytic datasets	
	USCS	**USCS Expanded**	**USCS County**	**EPHTN**	**USCS SEER*Stat Research Dataset**	**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 and rates, 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 the National Center for Health Statistics Research Data Center (NCHS RDC)
Mode of Access	Web-based query system with downloadable ASCII files, MS Excel files, and SAS datasets	Web-based query system	Flat ASCII file and Web-based query system and separate brief text documentation	Web-based query system	SEER*Stat client-server mode only after receipt of signed Data Use Agreement	On-site at NCHS RDC, on-site at a Census Research Data Center (RDC), remotely or staff-assisted
Web Address or Contact Information	USCS Web site www.cdc.gov/uscs	CDC WONDER http://wonder.cdc.gov	Request from cancerinfo@cdc.gov (specify "USCS County" in subject line)	National Environmental Public Health Tracking http://www.cdc.gov/nceh/tracking	Online request form	Application process available at www.cdc.gov/rdc
Contains Potentially Identifiable Information	No		No	No	No	Yes
Registry Eligibility Criteria for Data Completeness and Quality	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 2015		Updated 2015	Updated 2015	Updated 2015	Updated 2015

TABLE

Comparison of the National Program of Cancer Registries-Cancer Surveillance System Datasets

Cases Included						
	Public Web-Based Query Systems				Analytic datasets	
	USCS	**USCS Expanded**	**USCS County**	**EPHTN**	**USCS SEER*Stat Research Dataset**	**USCS Restricted-Access Dataset**
States/ Territories	NPCR and SEER States that meet eligibility criteria		NPCR States that meet eligibility criteria*	NPCR States that meet eligibility criteria	Registries that meet USCS publication criteria	NPCR States that meet eligibility criteria
Diagnosis Years	1999; 2000; 2001; 2002; 2003; 2004; 2005; 2006; 2007; 2008; 2009; 2010; 2011; 2012; 2013; 2009–2013	1999; 2000; 2001; 2002; 2003; 2004; 2005; 2006; 2007; 2008; 2009; 2010; 2011; 2012; 2013	2009–2013	Individual years for 2001 through 2013for State level; 2001-2005, 2002-2006, 2003-2007, 2004–2008; 2005-2009; 2006-2010; 2007-2011; 2008-2012; 2009-2013 for county level	1999; 2000; 2001; 2002; 2003; 2004; 2005; 2006; 2007; 2008; 2009; 2010; 2011; 2012; 2013 2009-2013	1999; 2000; 2001; 2002; 2003; 2004; 2005; 2006; 2007; 2008; 2009; 2010; 2011; 2012; 2013
Cancer Sites	All reportable invasive cancers; in situ 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; stomach; ovary; corpus and uterus, NOS; leukemias; non-Hodgkin lymphoma; liver and intrahepatic bile duct; pancreas, esophagus; and childhood cancers	Female breast; lung and bronchus; bladder; brain and other nervous system; thyroid; leukemias; non-Hodgkin lymphoma; all childhood cancers (state level only); childhood leukemias (state level only); childhood CNS and miscellaneous intercranial and intraspinal neoplasms (state level only); mesothelioma (state level only); kidney and renal pelvis; prostate; melanoma of skin; liver and intrahepatic bile duct; pancreas; oral/pharynx; esophagus, larynx	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)

* Future plans may include the addition of SEER data similar to the USCS dataset.

TABLE

Comparison of the National Program of Cancer Registries-Cancer Surveillance System Datasets

Variables Included						
	Public Web-Based Query Systems				Analytic datasets	
	USCS	**USCS Expanded**	**USCS County**	**EPHTN**	**USCS SEER*Stat Research Dataset**	**USCS Restricted-Access Dataset**
Geographic Levels	All areas combined; U.S. census region and division; NPCR and SEER State or territory; SEER metropolitan area	All areas combined; NPCR and SEER State or territory; MSA for cities of ≥500,000 (additional levels may be added)*	NPCR State; county*	NPCR State; county	All areas combined; U.S. census region and division; NPCR and SEER State or territory	NPCR State or territory; county for approved data sets only†
Race	All races combined; white; black; Asian/Pacific Islander (API) ; American Indian/Alaska Native (AI/AN); white non-Hispanic; black non-Hispanic		All races combined; white; black; AI/AN; API (with appropriate 50,000 population suppression and State permission for AI/AN and A/PI); white/black non-Hispanic	All races combined; white; black; AI/AN; API (with appropriate 50,000 population suppression and State permission for AI/AN and A/PI)	All races combined; white; black; Asian/Pacific Islander (API) ; American Indian/Alaska Native (AI/AN)	All races reported
Ethnicity (Hispanic)	Yes		Yes for State Profiles only (with State permission)	Yes (with State permission)	Yes	Yes
Age Groups	All ages combined and standard 5-year age groups for adults and <15 and <20 for childhood cancers	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+	All ages combined, standard 5-year age groups	Standard 5-year age groups and individual ages (Month and day of birth not provided for confidentiality reasons. If the age at diagnosis >99, then grouped into one category. Year of birth is also grouped.)
Summary Stage	No		No	No	Yes	Yes
Histology	International Classification of Childhood Cancers, Third Revision (all geographic areas combined), Mesothelioma (national and State level), Kaposi Sarcoma (national and State level), Consensus Conf on Cancer Registration of Brain, and CNS Tumors (all geographic areas combined)		No	No	Same as USCS	Yes

* Pending further data quality investigation and discussion with the NPCR Central Cancer Registry Council.

† County data used as a: a) linkage variable only by the NCHS RDC analyst; b) confounder or other control variable, no data presented by county; c) geographically aggregated form; e.g., large MSA (population ≥1 million), multi-county regions, or geographical areas (Appalachia, IHS CHSDA counties).

TABLE

Comparison of the National Program of Cancer Registries-Cancer Surveillance System Datasets

Confidentiality Protection/Disclosure Limitation Measures Employed						
	Public Web-Based Query Systems				Analytic datasets	
	USCS	USCS Expanded	USCS County	EPHTN	USCS SEER*Stat Research Dataset	USCS Restricted-Access Dataset
Direct or Record-Level Identifiers?	No		No	No	No	Yes, but not in output which will be reviewed by NCHS RDC staff for confidentiality
Aggregation	Yes		Yes	Yes	No	No
Limited Number of Variables	Yes		Yes	Yes	Yes	Yes
Grouping/Collapsing of Variables or Response Codes	Yes		No	Yes	No	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 <16		Yes Counts and rates: 5 year total count of <16	Yes Counts and unsmoothed rates: count of <16 Smoothed rates: RSE >25%	Yes Counts and rates: count of <16 enforced Case listing disabled	Yes (output will be reviewed by NCHS RDC analyst to ensure that small cell sizes are suppressed)
Complementary Cell Suppression	As needed		As needed	As needed	As needed	No
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	Yes	Yes
Logging and Monitoring	Limited		Limited	Limited	Yes, monitoring databases used, session type and date only	Yes

APPENDIX A

State, Metropolitan Area, and Territory 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, Metropolitan Area, 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/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/SEER
Atlanta	1975	SEER
Hawaii	1973	SEER
Idaho	1995	NPCR
Illinois	1995	NPCR
Indiana	1995	NPCR
Iowa	1973	SEER
Kansas	1995	NPCR
Kentucky	1995/2000	NPCR/SEER
Louisiana	1995/2000	NPCR/SEER
Maine	1995	NPCR
Maryland	1996	NPCR
Massachusetts	1995	NPCR
Michigan	1995	NPCR
Detroit	1973	SEER
Minnesota	1995	NPCR
Mississippi	1996	NPCR
Missouri	1996	NPCR
Montana	1995	NPCR
Nebraska	1995	NPCR
Nevada	1995	NPCR
New Hampshire	1995	NPCR
New Jersey	1995/2000	NPCR/SEER

APPENDIX A

State, Metropolitan Area, and Territory 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, Metropolitan Area, or Territory	First Diagnosis Year for Which Cancer Cases Were Reportable to NPCR or SEER*	Federal Funding Source
New Mexico	1973	SEER
New York	1996	NPCR
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	NPCR
United States Pacific Island Jurisdictions	2007	NPCR
Utah	1973	SEER
Vermont	1996	NPCR
Virginia	1996	NPCR
Washington	1995	NPCR
Seattle-Puget Sound	1974	SEER
West Virginia	1995	NPCR
Wisconsin	1995	NPCR
Wyoming	1996	NPCR

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

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 ICF International. To ensure the security and confidentiality of project data, the following provisions have been incorporated into the ICF International NPCR-CSS Security Plan in accordance with the requirements of the Assurance of Confidentiality.

The NPCR-CSS server is housed in a secure facility at Terremark, Culpeper, VA with a guard on duty 24 hours a day. Only authorized staff is 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 ICF International's firewall. ICF International project staff access the server via VPN from ICF International office located at Rockville, MD. Elevator and stairwell access to Rockville office 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 ICF International project staff (see below). It is password-protected on its own security domain. No one, including non-project staff at ICF International, is allowed access to the NPCR-CSS data.
- All ICF International 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 ICF International's firewall. Each State has its own directory location so that no State has access to another State'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 once they are no longer needed.
- A comprehensive security plan has been developed by ICF International's security team. The security team consists of Kevin Zhang, Project Director; Shailendra Bhavsar, Systems Lead and Security Officer; and Gretchen Stanton, Database Administrator and LAN/WAN Security Steward. All project staff receive annual security awareness training covering security procedures. The ICF International project security team oversees operations to prevent unauthorized disclosure of the NPCR-CSS data.
- Periodic (currently quarterly, but no less than once per year) reviews and updates of ICF International security processes will be conducted to adjust for rapid changes in computer technology and to incorporate advances in security approaches. The Security Plan will be amended as needed to maintain the continued security and confidentiality of NPCR-CSS data.

APPENDIX B

NPCR-CSS Overview of Data Security

ICF International

Authorized Project Staff

Staff Member	Position
NPCR-CSS Team	
Kevin Zhang, Ph.D.	Project Director
Yuan Ren, Ph.D.	Project Manager/Sr. Statistical Programmer
Gretchen Stanton	System Administrator and LAN/WAN Security Steward
Olga Galin, M.S	Sr. SAS Programmer/QA Specialist
Shailendra Bhavsar, B.S	Security Officer/Development Manager
Xing Dong, M.S.	Sr. Statistical Programmer
David Radune, B.A.	Sr. Developer
Jonathan Stanger, M.P.A.	Database Administrator
Jing Guo, B.S.	Programmer Analyst
Haoping Jiang, M.S.	Data Analyst
Ping Huang	Developer/Database Administrator
Alternate Staff	
Nicholas Kelsch Metin Sisman Carlos Birdsong Scott Gimsley Tom Sauer Gregory Hicks	These staff will assist the ICF System Administrator and LAN/WAN Security Steward on an as-needed basis to provide support for the NPCR-CSS network, servers, virtual private network (VPN), data and system backup, and NPCR-CSS Websites. Access to the NPCR-CSS record-level datasets is not included in the planned support activities.

Appendix C

Data Items for Central Brain Tumor Registry of the United States Dataset

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

**Diagnosis Years 1995-2003 invasive cases only, 2004-2012 invasive, benign, and borderline cases*

Item Name	NAACCR Data Item Number	Comments
Patient ID (unique)	20	
NAACCR Record Version	50	
State of Residence at Diagnosis	80	
Rural/Urban Continuum/Beale Code 1993	3300	
Rural/Urban Continuum/Beale Code 2003	3310	
NPCR Race Recode	Derived based on [160], [161], and [192]	*Same as race for USCS*
NHIAv2 Derived Hispanic Origin	191	
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 (*YEAR portion only*)	390	*Day and month of diagnosis not requested*
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	
Derived Summary Stage 2000	3020	
EDITS overrides	1990–2074	
CS Site-Specific Factor 1	2880	WHO Grade

APPENDIX D

NPCR Analytic Data Use Agreement

NPCR+SEER 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+SEER Analytic Data for research purposes, it is absolutely necessary to ensure, to the extent possible, that use of the data will be limited to research or public health purposes. Any person who attempts 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, will be prosecuted to the full extent of the law.

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. Nevertheless, in rare instances, through complex analysis and with outside information, it may be possible to ascertain from the dataset the identity of particular persons. Considerable harm can ensue if this occurs.

Data users must agree to the following provisions prior to receiving access to NPCR+SEER Analytic Data:

- I understand that all NPCR and SEER data are owned by the states and territories.

As the recipient of the NPCR+SEER Analytic Data, I will:

- Not use or permit others to use the datasets in any way other than for statistical reporting and analysis.
- Not release or permit others to release the datasets or any part of them to any person except with DCPC'S written approval.
- Not attempt to link or permit others to link the datasets with individually identifiable records from any other dataset without DCPC's approval.
- 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.
- Take the following actions if the identity of any person or establishment is discovered inadvertently:
  - Make no use of this knowledge.
  - Notify DCPC's Cancer Surveillance Branch Chief.
  - As requested by DCPC, safeguard or destroy the information that identifies an individual or establishment.
  - Inform no one else of the discovered identity.

In addition, I will make every effort to release all statistical information in such a way as to avoid inadvertent disclosure. I will:

- Ensure that no data on an identifiable case can be derived through subtraction or other calculation from the combination of tables in the given publication.
- Ensure that no data permit disclosure when used in combination with other known data.
- 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 [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 1998-xxxx (excluding SEER-Metro Registry data).

My signature below indicates that:

- 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.
- I understand that the NPCR+SEER Analytic Data belong to the states and territories. The states' and territories' agreement to use of the data is obtained through the activities outlined in the general NPCR-CSS Data Release Policy and by specific requests to the states and territories through the CSB management team.
- I am familiar with the use of **SEER*Stat** in analyzing data or will complete the needed training.
- I have submitted an analysis proposal describing the intended data use to the NPCR Internal Data users Group, who will review the proposal and forward it to the Applications, Statistics, and Informatics Support Team (ASIST) when the proposal is approved.
- I will provide a courtesy copy of draft papers or abstracts to the NPCR Internal Data Users Group at npcridug@cdc.gov as they are entered into Documentum for clearance.

_____

Signature Date

Name:_____

Title/ Organization_____

CDC campus/building/room/mailstop_____

Telephone_____ E-mail:_____

Please return completed form to the NPCR Internal Data Users Group at npcridug@cdc.gov.

Appendix E

CDC Non-Disclosure Agreement

The success of CDC's operations depends upon the voluntary cooperation of States, of establishments, and of individuals who provide the information required 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 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.

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

'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 an 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.'

''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

Appendix F

Data Items for USCS Internal Analytic Dataset

Record ID and Demographic Data Items [NAACCR Data Item Number]

- Patient ID Number [20]
- Address at DX – State [80]
- NHIA [191]*
- Sex [220]
- Age at Diagnosis [230]
- Birth Date [240] - year only

Cancer Identification Data Items

- Sequence Number--Central [380]
- Date of Diagnosis [390] - month and year only
- Primary Site [400]
- Laterality [410]
- Grade [440]
- Diagnostic confirmation [490]
- Type of Reporting Source [500]
- Histologic Type ICD-O-3 [522]
- Behavior Code ICD-O-3 [523]

Stage/Prognostic Factors Data Items

- SEER Summary Stage 2000 [759]
- SEER Summary Stage 1977 [760]
- Derived Summary Stage 2000 [3020]

Geographic locations

- Rural-urban continuum 2003 [3310]

Over-ride Flags

- Over-ride Age/Site/Morph [1990]
- Over-ride SeqNo/DxConf [2000]
- Over-ride Site/Lat/Sequence Number [2010]
- Over-ride Site/Type [2030]
- Over-ride Histology [2040]
- Over-ride Report Source [2050]
- Over-ride Ill-define Site [2060]
- Over-ride Leuk,Lymphoma [2070]
- Over-ride Site/Behavior [2071]
- Over-ride Site/Lat/Morph [2074]
- Over-ride CS 20 [3769]

Derived Fields

- Site recode ICD-O-3/WHO 2008
- ICCC site recode ICD-0-3/WHO 2008
- Behavior Recode for Analysis
- Behavior Recode for Analysis Derived
- SEER-modified ICCC Recode
- Standard 5-year age recode
- Race Recode for USCS (W, B, AI/AN, API, O,Unk)

User-specified fields

- USCS standard

uscs9812
uscs9912
uscs0312
uscs0812
US Census Region
US Census Division
Econ status

* Since the NHIA algorithm continues to be refined, states may choose not to provide data for NHIA in this file.

Appendix G

Data Items for NPCR Internal Survival Dataset

Demographic Data Items

- Alternate Patient ID Number
- Address at Diagnosis – State
- Race 1
- Race Recode
- Sex
- Age at Diagnosis
- Age Recode with <1 Year Olds
- Age Recode
- Year of Birth

Cancer Identification Data Items

- Sequence Number – Central
- Day of Diagnosis
- Month of Diagnosis
- Year of Diagnosis
- Primary Site
- Laterality
- Grade
- Diagnostic Confirmation
- Type of Reporting Source
- Histologic Type ICD-O-3
- Behavior Code ICD-O-3
- Behavior Recode for Analysis Derived/WHO 2008
- Site Recode ICD-O-3/WHO 2008
- ICCC Recode ICD-O-3/WHO 2008

Stage/Prognostic Factors Data Items

- SEER Summary Stage 2000
- Derived SS2000
- NPCR Cancer Stage

Follow-up/Death Data Items

- Original Day of Last Contact
- Original Month of Last Contact
- Original Year of Last Contact
- Survival Months-Presumed Alive
- Survival Flag-Presumed Alive
- Presumed Alive Day of Last Contact Recode
- Presumed Alive Month of Last Contact Recode
- Presumed Alive Year of Last Contact Recode
- Original Day of Diagnosis Recode
- Original Month of Diagnosis Recode
- Original Year of Diagnosis Recode
- Vital Status
- Vital Status Recode
- Follow-up Source Central

Appendix H

NPCR-CSS 308(d) Assurance of Confidentiality Statement

A 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 International, 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, and private physicians on all cancer patients diagnosed in the State. This information includes patient demographics and cancer diagnosis and treatment data. Each August, CDC requests cumulative data from central cancer registries. The variables reported to CDC may vary from year to year. The cancer registries maintain these data permanently in longitudinal databases that are used for public health surveillance, program planning and evaluation, and research. CDC will update its longitudinal database each year with data received from the States. These data are used by CDC scientists for routine cancer surveillance, program planning and evaluation, and to provide data for research. 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.

Information received by CDC or its contractors as part of this 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 will be used only for purposes stated in this assurance and will not otherwise be disclosed or released, even following the death of cancer patients in this surveillance system.

Information collected by CDC will be 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 will not be made available to any group or individual. In particular, such information will not be 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 will be kept confidential and—with the exception of CDC employees, their contractors, and qualified researchers—no one will be allowed to see or have access to the information. CDC employees and contractors will be 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 will be required to sign the NCHS RDC data sharing agreements and abide by the NCHS RDC confidentiality procedures. Organizations (e.g., the North American Association of Central Cancer Registries, American Cancer Society, and National Cancer Institute) will be required to sign a detailed data release agreement to have access to restricted release data.

Appendix I

NPCR-CSS 308(d) Assurance of Confidentiality FAQ

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) (formally a separate agency) 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 within 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 50 projects have received 308(d) protection since CDC received this authority, and currently there are approximately 25 active projects with 308(d) confidentiality assurances. As a testament to the importance of this project to the mission of CDC, the 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 the CDC Office of General Counsel and the CDC 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. In 2006 and 2010, NPCR filed and received approval for continuation.

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 (e.g., Social Security number) or any combination of variables that could be used to indirectly identify an individual. 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?

Appendix I

NPCR-CSS 308(d) Assurance of Confidentiality FAQ

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 that 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 surveillance system 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, ICF International. The 308(d) confidentiality protection does not go with the data whether released publically or through restricted means, and any data released to qualified researchers by CDC 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 (attachment I). 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 contractors at ICF International 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 at ICF International with access to NPCR-CSS data are required to sign a confidentiality agreement.

Appendix J

Data Items for USCS SEER*Stat Research Dataset

The research use USCS dataset contains individual case-specific data from the USCS dataset with enforced <16 cell suppression and case listing disabled.

Name	Status [NAACCR Data Item number listed in brackets]	Notes
Demographic Data Items		
Address at Diagnosis—State	[80]	
Race Recode	Derived based on [160], [161] and [192]	
NHIA Derived Hispanic Origin	[191]	
Sex	[220]	
Age Recode	Derived based upon [230]	Recoded to 5-year age groups. Age over 99 recoded to 95-99 age group.
Cancer Identification Data Items		
Sequence Number—Central	[380]	
Date of Diagnosis	Derived based upon [390]	Day of diagnosis will not be provided.
Primary Site	[400]	
Laterality	[410]	
Grade	[440]	
Diagnostic Confirmation	[490]	
Type of Reporting Source	[500]	
Histologic Type ICD-O-3	[522]	
Behavior Code ICD-O-3	[523]	
Behavior Recode for Analysis	Derived based upon [400], [522], and [523]	
Primary Site Recode	Derived based upon [400] and [522]	
Primary Site Recode with Mesothelioma and Kaposi Sarcoma	Derived based upon [400] and [522]	
SEER-Modified International Classification of Childhood Cancer (ICCC) Recode	Derived based upon [400], [522], and [523]	
Stage/Prognostic Factors Data Items		
SEER Summary Stage 2000	[759]	
SEER Summary Stage 1977	[760]	
Derived SS2000	[3020]	
User-Defined Data Items		
USCS Standard	USCS standard for single years of analysis at the national	

	level.
uscs9913	USCS standard for diagnosis years 1999-2013 total.
uscs0913	USCS standard for diagnosis years 2009-2013 total.
U.S. Census Region	US Census Regions are Northeast, Midwest, South, and West.
U.S. Census Division	US Census Regions are further divided into Divisions: Northeast region consists of New England and Middle Atlantic divisions; Midwest region consists of East North Central and West North Central divisions; South region consists of South Atlantic, East South Central, and West South Central divisions; and West region consists of Mountain and Pacific divisions

Appendix K
NPCR Research Data Use Agreement

National Program of Cancer Registries (NPCR)
Research Data Use Agreement
For data submitted November, 2015

The Centers for Disease Control and Prevention (CDC) makes NPCR data available to the public and researchers through various data release activities. The *NPCR Research Data* is an unrestricted subset of data submitted to CDC made available only through the SEER*Stat statistical software.

The CDC has obtained an assurance of confidentiality for NPCR pursuant to Section 308(d) of the Public Health Service Act, 42 U.S.C. 242m(d). 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. All direct identifiers, as well as characteristics that might easily lead to identification of individuals, are omitted from *NPCR Research Data*. Certain demographic information has been included for research purposes; thus, all SEER*Stat 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 prior to receiving access to the *NPCR Research Data*.

As the recipient of access to NPCR Research Data through SEER*Stat, I will:

- Not use or permit others to use the analytic results in any way other than for statistical reporting and analysis.
- Use appropriate safeguards to prevent use or disclosure of the information other than as provided for by this data-use agreement.
- Ensure all members of the research team who have access to the *NPCR Research Data* through SEER*Stat have signed this data-use agreement.
- Not attempt to link or permit others to link *NPCR Research Data* with individually identifiable records from any other dataset without CDC approval.
- Not attempt to use the analytic results or permit others to use them to learn the identity of any person or establishment included in any dataset.
- Take the following actions if the identity of any person or establishment is discovered inadvertently:
    - Make no use of this knowledge.
    - Notify CDC.
    - As requested by CDC, safeguard or destroy the information that identifies an individual or establishment.
    - Inform no one else of the discovered identity.

In addition, I will make every effort to release all statistical information in such a way as to avoid inadvertent disclosure. I will:

- Ensure that no data on an identifiable case can be derived through subtraction or other calculation from the combination of tables in the given publication.
- Ensure that no data permit disclosure when used in combination with other known data.
- I will not disclose or otherwise make public data on any unit smaller than 16. If the total number of cases in a cell is <16, the cell data will be suppressed in oral and written presentations.

My signature below indicates that:

Appendix K

NPCR Research Data Use Agreement

- I will adhere to the requirements of this Data Use Agreement and understand that my access to the data will be revoked if these requirements are violated.
- I am familiar with the use of **SEER*Stat** in analyzing data or will complete the needed training.
- I understand that all NPCR data are owned by the states and territories. The states' and territories' have established agreements with CDC on the use and dissemination of the data.
- I understand that I am responsible for the results of my own analysis. The findings and conclusions resulting from the analysis of these data are those of the authors and do not necessarily represent the official position of the CDC.
- **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 2015, and/or the Surveillance, Epidemiology and End Results (SEER) program and submitted to NCI in November 2015.

_____ Click here to enter a date.

Signature **Date**

Name: Click here to enter text.

Title/organization: Click here to enter text.

CDC campus/building/room/mailstop: Click here to enter text.

Telephone: Click here to enter text. **E-Mail:** Click here to enter text.

Please print, sign, and date the agreement. Send the form to CDC (or authorized contractor):

- *By fax to xxx.xxx.xxxx*
- *Or, e-mail a scanned form to xxxxxx@xxxxx.xxx*

Appendix L

NPCR Data at the NCHS RDC Q&A

Can you summarize what CDC is planning to do?

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 data for research purposes. The data will be available through the NCHS RDC only after the standard data quality reviews that occur as part of the preparation for USCS and State Cancer Profiles.

The use of the NCHS RDC to manage data access will provide the highest level of data security and protection of confidentiality that is available for analysis of data. Any researcher must submit a proposal which will be reviewed and approved by CDC and 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 will leave the NCHS RDC facilities.

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, staff at NCHS compile statistical information to guide actions and policies to improve the health of our people. More information about NCHS is available at: http://www.cdc.gov/nchs/about.htm.

What is the Research Data Center (RDC)?

The NCHS RDC began in 1998 and has a long-standing history of managing access to health and vital statistics data through a rigorous proposal review process as well as review of the statistical output. The NCHS RDC mission is to give public access to the 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. Once the analysis is completed, the data extract is archived for 2 years and then destroyed.

There are currently three modes of access through the NCHS RDC, each with specific restrictions. Access is available on-site at two locations (Hyattsville, MD and Atlanta, GA), nine Census RDCs, or through remote electronic access. More information about the NCHS RDC is available at: http://www.cdc.gov/rdc/

Appendix L
NPCR Data at the NCHS RDC Q&A

Why does CDC use the NCHS RDC?
Maintaining confidentiality is the primary objective of the NCHS RDC. Staff at NCHS RDC have statistical expertise to address confidentiality and disclosure risk. Using the NCHS RDC will allow CDC to comply with the Assurance of Confidentiality [308(d)] that was obtained for the NPCR-CSS data. All researchers must take 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: http://www.cdc.gov/rdc/B4ConfiDisc/CfD400.htm.

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: http://www.cdc.gov/rdc/B3Prosal/PP300.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.

The use of the NCHS RDC to host the NPCR data is a win-win opportunity because of the confidence in knowing that the data are being used correctly and safely, while at the same time making the data available for external researchers in an appropriate way. In addition, this approach will not overtax resources here in the Branch 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.

If a CCR agrees to participate, who has access to the data and at what level?
The NCHS RDC analysts will have access to the individual record level data since it is easier to create an analytic dataset using these data. The NCHS RDC analysts will be bound by the same data use agreements that CDC staff sign on an annual basis. Researchers with approved proposals will be able to conduct analyses through the NCHS RDC on the created dataset or have the NCHS RDC analyst do the analysis for them. However, they will not be able to 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 would have direct access to the dataset itself including the ability to download the data and create a listing of individual record level data and all variables in the RADS.

Researchers have several possible modes of access to the data set created for their specific research proposal. More information is available at: http://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.

Will the CCRs be able to decide whether their data will be available through the NCHS RDC?
Yes. However, given the protection provided to the data and the review process the CCRs are expected to participate. 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 [NAACCR#90]) is to be available for use in the NCHS RDC?
Yes. CCRs will be able to decide if county data are used by the NCHS RDC. County data will be used only in approved analyses and in the following ways:

- Used as a linkage variable (linkage to census data, for example) only by the NCHS RDC analyst. The county variable will not be available to the researcher but the NCHS RDC analyst would use it to create a linked dataset and then remove the county variable.

- Included as a confounder or other control variable, but no data are presented by county. The NCHS RDC analyst will create dummy variables to mask the actual county name.

- Used in geographically aggregated form such as large metropolitan statistical areas (e.g., those with a population of 1 million or larger), multi-county regions, or geographical areas (e.g., Appalachia or IHS Contract Health Services Delivery Areas (CHSDA) counties). It will be 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 also includes the NCHS RDC analyst and the confidentiality officer, who will be 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: http://www.cdc.gov/rdc/B3Prosal/PP340.htm.

NPCR will obtain comments on each proposal from multiple CCRs whose data are included in the NCHS RDC analysis.

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?
Yes. NPCR staff are working with NCHS RDC staff to provide appropriate training for these data preparation and analysis tools.

Will researchers have access to SEER*Stat?
Yes. It is expected that researchers will know the basics of the analyses that they wish to carry out. NCHS RDC staff will be available for limited consultation. Since cell phones or access to the Internet are not available inside the NCHS RDC, all SEER*Stat tutorials (http://seer.cancer.gov/seerstat/tutorials/) would need to be completed beforehand.

What suppression rules will be used for the RADS?
The same suppression rules that are used for *United States Cancer Statistics.* More detailed information is available at:
http://www.cdc.gov/cancer/npcr/uscs/2006/technical_notes/stat_methods/suppression.htm.

In addition, the suppression rules for Asians/Pacific Islanders (A/PI) and American Indians/Alaska Natives (AI/AN) will also apply. The data for A/PI and AI/AN will be presented only for states or counties with at least 50,000 population because of concerns regarding possible misclassification of race data and the relatively small sizes of these populations in the United States.

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

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 the researcher 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 (http://www.cdc.gov/rdc/Data/B4/DesignatedAgent.pdf), 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, there is one office that 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 from their own institution, but it will not be a requirement in the application process given the ERB extension that the NCHS RDC provides.

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 are 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 datasets are individual case-specific data from the NPCR-CSS dataset. The data items to be included are listed below.

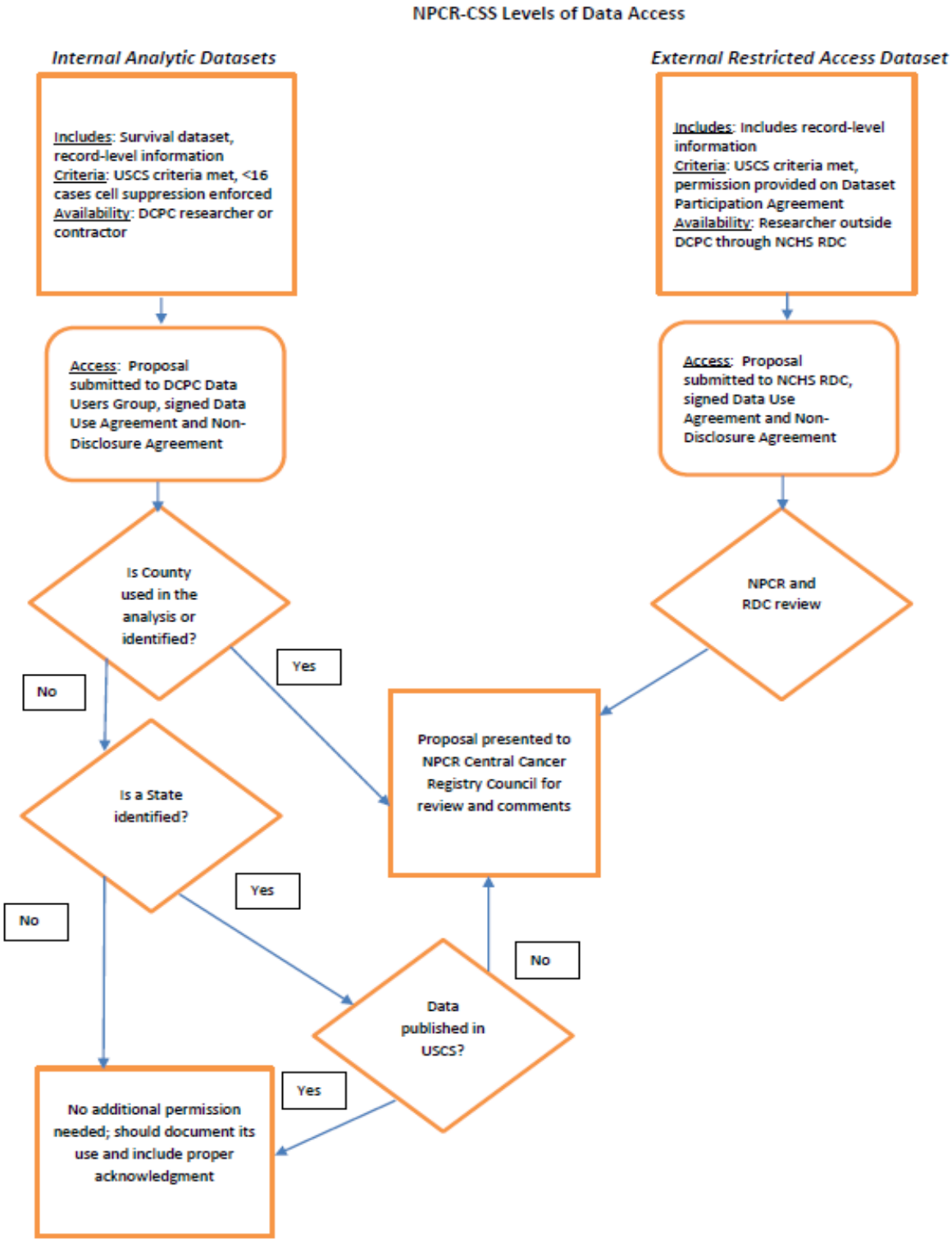
Name	Status [NAACCR Data Item number listed in brackets]	Notes
Demographic Data Items		
Patient ID Number	[20]	
Address at Diagnosis—State	[80]	
Address at Diagnosis—County	[90]	**Only used in approved analyses.***
Address at Diagnosis—Census Region	Derived based upon [80]	
NPCR Race Recode	Derived based on [160], [161] and [192]	
Spanish/Hispanic Origin	[190]	
NHIA Derived Hispanic Origin	[191]	
Sex	[220]	
NPCR Age at Diagnosis	Derived based upon [230]	Age over 99 will be recoded.
NPCR Age Recode	Derived based upon [230]	
NPCR Birth Date	Derived based upon [240]	Only year will be provided; if age is over 99, then year of birth will be recoded.
Cancer Identification Data Items		
Sequence Number—Central	[380]	
NPCR Date of Diagnosis	Derived based upon [390]	Day of diagnosis will not be provided.
Primary Site	[400]	
Laterality	[410]	
Grade	[440]	
Diagnostic Confirmation	[490]	
Type of Reporting Source	[500]	
Histologic Type ICD-O-3	[522]	
Behavior Code ICD-O-3	[523]	
NPCR Behavior Recode for Analysis	Derived based upon [400], [522], and [523]	
SEER Incidence Site Recode	Derived based upon [400] and [522]	
SEER Incidence Site Recode with Mesothelioma and Kaposi Sarcoma	Derived based upon [400] and [522]	
SEER International Classification of	Derived based upon [400],	

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

Name	Status [NAACCR Data Item number listed in brackets]	Notes
Childhood Cancer (ICCC) Recode	[522], and [523]	
Stage/Prognostic Factors Data Items		
SEER Summary Stage 2000	[759]	
SEER Summary Stage 1977	[760]	
CS Extension	[2810]	
CS Lymph Nodes	[2830]	
CS Mets at DX	[2850]	
CS Site-Specific Factor 1	[2880]	
CS Site-Specific Factor 2	[2890]	
CS Site-Specific Factor 3	[2900]	
CS Site-Specific Factor 15	[2879]	
CS Site-Specific Factor 25	[2879]	
CS Version Input Original	[2935]	
CS Version Derived	[2936]	
CS Version Input Current	[2937]	
Derived SS2000	[3020]	
Over-ride Flags		
Over-ride Age/Site/Morph	[1990]	
Over-ride SeqNo/DxConf	[2000]	
Over-ride Site/Lat/Sequence Number	[2010]	
Over-ride Site/Type	[2030]	
Over-ride Histology	[2040]	
Over-ride Report Source	[2050]	
Over-ride Ill-define Site	[2060]	
Over-ride Leuk, Lymphoma	[2070]	
Over-ride Site/Behavior	[2071]	
Over-ride Site/Lat/Morph	[2074]	

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

Appendix N
NPCR-CSS Levels of Data Access



Appendix N
NPCR-CSS Levels of Data Access