

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

Data Release Policy for 2013 Data Submission (Diagnosis Years 1995–2010)

Policy Revised August 2012

Cancer Surveillance Branch
Division of Cancer Prevention and Control
NCCDPHP, CDC
4770 Buford Hwy, N.E., Mailstop F-69
Atlanta, GA 30341-3717
Phone: (770) 488-4783
E-mail: cancerinfo@cdc.gov (specify “NPCR-CSS” in subject line)

Table of Contents

	Page
I. Introduction	1
II. Assurance of Confidentiality	1
III. Overview of the Data	2
IV. Data Requests and Notification of States	3
V. Public-Use Datasets	4
A. <i>United States Cancer Statistics (USCS)</i> Dataset	5
B. <i>USCS</i> Expanded Dataset	5
C. <i>USCS</i> County Cancer Incidence Dataset	5
D. Environmental Public Health Tracking (EPHT) Dataset	6
E. Public Release Disclosure Statement	6
VI. Restricted-Access Dataset	7
VII. Data Release to Collaborating Partners	7
A. Central Brain Tumor Registry of the United States (CBTRUS)	7
B. Veterans Health Administration (VHA) State-Level Correction Factors for the North American Association for Central Cancer Registries (NAACCR)	8
C. Indian Health Services (IHS)	8
VIII. Emergency and Provisional Data Releases	8
IX. Data Release Under Controlled Conditions	9
X. References	10
Table Comparison of NPCR-CSS Data	11
Appendixes	
Appendix A NPCR-CSS 308(d) Assurance of Confidentiality	15
Appendix B NPCR-CSS 308(d) Assurance of Confidentiality Frequently Asked Questions	16
Appendix C NPCR-CSS Overview of Data Security	20
Appendix D State and Metropolitan Area Cancer Registries by Federal Funding Source and First Diagnosis Year for Which Cancer Cases Were Reportable to the Centers for Disease Control and Prevention's (CDC) NPCR or National Cancer Institute's (NCI) Surveillance, Evaluation, and End Results Program (SEER)	22
Appendix E NPCR-CSS Levels of Data Access	24
Appendix F Data Items for Restricted Access Dataset	25
Appendix G Data Items for CBTRUS Dataset	27
Appendix H NPCR-CSS Data Sharing Agreement	28
Appendix I NPCR Data at the National Center for Health Statistics Research Data Center Questions & Answers	30

**National Program of Cancer Registries
Cancer Surveillance System
Data Release Policy
August 2012**

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) will release or share. This multi-year policy updates the August 2011 NPCR-CSS Data Release Policy. This policy applies to data submitted to the Centers for Disease Control and Prevention (CDC) for the 2013 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, will clear 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, they 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 will be reviewed annually by the NPCR-CSS Privacy Steward and other appropriate CDC staff members to determine whether revisions are needed. If revisions are needed, NPCR central registries will be notified and allowed to review and comment on the revisions before they become final.

II. 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 data¹⁻³ and specific restrictions imposed on NPCR-CSS data (appendixes A, B, and C).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)) (appendixes A and B). 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

NPCR-CSS Data Release Policy

August 2012

1995-2010 Diagnosis Years

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.

III. Overview of the 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 D).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 C). 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)) (<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

IV. Data Requests and Notification of States

A. 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. At this time, it is anticipated that two kinds of datasets will be released: public-use datasets and restricted-access datasets (see definitions below).

B. CDC Internal Analytic Data Requests

NPCR-CSS Data Release Policy

August 2012

1995–2010 Diagnosis Years

CDC staff members or contractors perform analyses of NPCR program data as needed utilizing an internal analytic database based on the *USCS* dataset. These 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.

Publications or presentations describing the quality of the data or the burden of cancer may be 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. Additionally, in order to improve cancer control in the U.S., the survival for this population must be determined and changes mapped over time. 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.

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 (appendixes A, B, and C).

In addition to adhering to strict requirements for protecting confidentiality, CDC staff members will notify the State cancer registry 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 will include, 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 will 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.

C. CDC External Data Requests

When individuals, agencies, or organizations outside CDC request data not available in a format that can be obtained from a public-use dataset, CDC staff members or contractors will tabulate the data for the inquirer for requests that do not identify a

NPCR-CSS Data Release Policy

August 2012

1995–2010 Diagnosis Years

State. For requests that identify a State, CDC staff members will seek State's permission regarding use. See appendix E for additional details.

V. Public-Use Datasets

For purposes of this policy, public use datasets (PUDS) 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.^{2, 5-10} PUDS are also defined as a Web-based query system and have a database behind a CDC firewall that is case-specific microdata; however, for this database users will be able to access only aggregate counts and rates with all confidentiality protections built in. A combination of confidentiality protection measures is employed for each PUDS (see table). PUDS will not contain information that is identifiable or potentially identifiable according to currently accepted procedures for reducing disclosure risk.^{2, 5-10} Before each PUDS is finalized, the aggregate values will be analyzed to determine whether there is a need for complementary cell suppression.^{2, 5-10} If appropriate, the analysis will include consultation with a statistician with specific expertise in statistical disclosure limitation techniques. Following the analysis, complementary cell suppression will be applied as needed.

There will be no restrictions on access to PUDS. A public release disclosure statement (see page 6) will caution users against inappropriate use of the data or inappropriate disclosure of information. PUDS will be released as delimited ASCII files, a Web-based query system, or possibly through other vehicles (see table). States will have an opportunity to review their data and have adequate time to notify CDC if they identify a problem with the data. 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 create each PUDS. The following four PUDS are currently being released:

- *USCS* dataset
- *USCS* expanded dataset
- *USCS* county cancer incidence dataset
- Environmental Public Health Tracking dataset

All NPCR-CSS PUDS will 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 PUDS.

Separate documentation may accompany each PUDS that will describe 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).

A. *USCS* Dataset

The *USCS* dataset contains the same aggregate counts and rates for incidence and mortality published annually (see table). The PUDS are a hypertext markup language (HTML) edition of *USCS*. Tables of male and female combined counts and rates are in the HTML edition in addition to the published sex-specific tables. Users can download the data in an ASCII format for use in other applications.

B. *USCS* Expanded Dataset

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). This PUDS is 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 will be able to access only aggregate counts and rates with all confidentiality protections built in. Because this PUDS will present 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.

C. *USCS* County Cancer Incidence Dataset

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). This PUDS is available as an ASCII file. Because this PUDS 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 PUDS has been released to a small number of users, including NCI for the State Cancer Profiles project (www.statecancerprofiles.cancer.gov), which used age-adjusted rates only, and the U.S. Department of Health and Human Services Women's Health Initiative, which used 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.

D. Environmental Public Health Tracking (EPHT) Dataset

The EPHT dataset for the national EPHT 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 PUDS is available on the EPHT 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 smoothed and unsmoothed rates. Users are able to access only aggregate counts and rates with all confidentiality protections built in. Because this PUDS 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.

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, CDC-NPCR will provide the state-level EPHT 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.

E. Public Release Disclosure Statement

The following (or similar) public release disclosure statement will be prominently displayed for users of all NPCR-CSS PUDS:

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:

1. Use the data for statistical reporting and analysis only.
2. Make no attempt to learn the identity of any person or establishment included in these data.
3. 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.

VI. Restricted-Access Dataset

For purposes of this policy, the restricted-access dataset (RADs) is defined as the version of the full NPCR-CSS analytic dataset, either aggregated data or microdata (i.e., individual case-specific data) that have been modified as needed to minimize (but may not remove entirely) the potential for disclosure of confidential information. RADs will 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 variables is in appendix F.

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.

CDC will use 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 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. RADS will be released to researchers through the National Center for Health Statistics Research Data Center (NCHS RDC) after CDC authenticates the requestor's identity and research intent through an extensive proposal review process, the requestor complies 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 will be reviewed by staff at central cancer registries that agree to participate in the RADS and by CDC, which will include NPCR and NCHS RDC staff. User documentation including a data dictionary for every diagnosis year available at the NCHS RDC will be provided after a project is approved.

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. Using the NCHS RDC will allow CDC to comply with 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 H of this policy.

VII. Data Release to Collaborating Partners

A. Central Brain Tumor Registry of the United States (CBTRUS)

CBTRUS will publish the print and Web versions of the statistical report, *Primary Brain Tumors in the United States Statistical Report Supplement, 2004–2010*; a previous version of the report is available at: <http://www.cbtrus.org/reports/reports.html>. The report will include age-adjusted rates and corresponding 95-percent confidence intervals on brain and other central nervous system tumors and will be presented by State, histology, major histology grouping, primary site, behavior, gender, race, ethnicity, and age at diagnosis. CDC will provide individual, record-level data to CBTRUS for the publication of this report; appendix G lists the variables to be included. Only States that agree to participate and meet the *USCS* publication criteria will be included in the dataset.

In addition, CBTRUS will use 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 will be released; only aggregated data with the corresponding confidence intervals (if applicable) and appropriate suppression criteria will be provided to data inquirers. Attribution to the NPCR will be provided. CBTRUS will sign data use agreements before data are released for

their report and future inquiries. For questions, contact CBTRUS staff at cbtrus@aol.com.

B. Veterans Hospital Administration (VHA) State-Level Correction Factors for NAACCR

CDC will continue to evaluate the ongoing effect of under-reporting by VHA. These data will be used to estimate the percent and number of cancer cases that were not reported to the central cancer registry due to delays or other issues created by the new VHA requirements. In addition to estimates for NPCR overall, corrections factors at the State level will be estimated and provided to central cancer registries that submit data for this request. NAACCR is also interested in having State-level correction factors in order to assess any effect on data completeness as part of the NAACCR certification process. If permitted, CDC will share the State-level correction factors with NAACCR *after* they have shared with the central cancer registries. Please note that NAACCR will not have access to the data that are provided to CDC.

C. Indian Health Services

CDC will continue 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 will be maintained by the IHS Division of Epidemiology and Disease Prevention 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. Additionally, incidence and mortality rates will be included in an AI/AN-focused supplement to be published in the *American Journal of Public Health*.

VIII. 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 (appendixes B and C).

Provisional data and draft data tables will be 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 will sign a data use agreement before accessing the data or tables.

IX. 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 PUDS or RADS 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 will not 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.

X. 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/od/foia/policies/sharing.htm>.
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. Privacy, Confidentiality, and Data Security Web Site. Alexandria, VA: American Statistical Association; 2005. Available at <http://www.amstat.org/committees/pc/index.html>.
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://www.fcsm.gov/committees/cdac/cdac.html>.
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/working-papers/spwp22.html>.
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: 1999–2006 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; 2010. Available at: www.cdc.gov/uscs.
17. NAACCR Method to Estimate Completeness. A Data Analysis Tool for Calculations. Available at: <http://www.naacr.org/Research/DataAnalysisTools.aspx>.
18. Clegg LX, Feuer 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 National Program of Cancer Registries-Cancer Surveillance System Datasets

Overview					
	Public Use Datasets (PUDS)				Restricted-Access Datasets (RADS)
	USCS	USCS Expanded	USCS County	Environmental Public Health Tracking	
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 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	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)	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; data meet criteria for unknown county
When Available	Updated 2011	Updated 2011	Updated 2011	Updated 2011	Updated 2011

Table. Comparison of the National Program of Cancer Registries-Cancer Surveillance System Datasets, *continued*

Cases Included					
	Public-Use Datasets (PUDS)			Environmental Public Health Tracking	Restricted-Access Datasets (RADS)
	USCS	USCS Expanded	USCS County		
States/Territories	NPCR and SEER States that meet eligibility criteria		NPCR States that meet eligibility criteria*	NPCR States that meet eligibility criteria	NPCR States that meet eligibility criteria
Diagnosis Years	1999; 2000; 2001; 2002; 2003; 2004; 2005; 2006; 2007; 2008; 2009; 2004–2009	1999; 2000; 2001; 2002; 2003; 2004; 2005; 2006; 2007; 2008; 2009	2004–2009	2001; 2002; 2003; 2004; 2005; 2006; 2007; 2008; 2009 for State level; 2001–2005, 2002–2006, 2003–2007, 2004–2008; 2005–2009 for county level	1999; 2000; 2001; 2002; 2003; 2004; 2005; 2006; 2007; 2008; 2009
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 intracranial 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 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, *continued*

Variables Included					
Public-Use Datasets (PUDS)				Restricted-Access Datasets (RADS)	
	USCS	USCS Expanded	USCS County	Environmental Public Health Tracking	
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 or more* (additional levels may be added)**	NPCR State; county*	NPCR State; county	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)		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; AI/AN; API (with appropriate 50,000 population suppression and State permission for AI/AN and A/PI)	All races reported
Ethnicity (Hispanic)	Yes		Yes for State Profiles only (with State permission)	Yes (with State permission)	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+	Standard 5-year age groups and individual ages (Month and day of birth are not provided for confidentiality reasons. If the age at diagnosis over 99, then grouped into one category. Year of birth is also grouped.)
Summary Stage	No	No	No	No	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	Yes

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

** Pending further data quality investigation and discussion with the NPCR-CSS Scientific Working Group.

† County data will be used only in approved analyses and in the following ways: a) used as a linkage variable 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).

Table. Comparison of the National Program of Cancer Registries-Cancer Surveillance System Datasets, *continued*

Confidentiality Protection/Disclosure Limitation Measures Employed					
	Public-Use Datasets (PUDS)			Restricted-Access Datasets (RADS)	
	USCS	USCS Expanded	USCS County		Environmental Public Health Tracking
Direct or Record-Level Identifiers?	No		No	No	Yes, but not in output which will be reviewed by NCHS RDC staff for confidentiality
Aggregation	Yes		Yes	Yes	No
Limited Number of Variables	Yes		Yes	Yes	Yes
Grouping/Collapsing of Variables or Response Codes	Yes		No	Yes	Yes
(1) Average Annual Counts Rounded to the Nearest Whole Number (2) Average Annual Rates (3) Annual Averages Are Based on At Least 5 Years of Data	No		Yes	Yes	No
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 (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	No
Public Release Disclosure Statement	Yes		Yes	Yes	Yes
Data Sharing Agreement and/or IRB Approval	No		No	No	Yes
User Authentication	No		No	No	Yes
Logging and Monitoring	Limited		Limited	Limited	Yes

National Program of Cancer Registries Cancer Surveillance System 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 Macro, 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 (Appendix H) to have access to restricted release data.

National Program of Cancer Registries Cancer Surveillance System 308(d) Assurance of Confidentiality Frequently Asked Questions

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.

1. 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.

2. 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

NPCR-CSS Data Release Policy

August 2012

1995–2010 Diagnosis Years

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, NPCR filed and received approval for continuation.

3. 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.

4. 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.

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

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

6. 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 nonrelease of these data establishes an effective precedent for restricting access to more sensitive data, such as that collected by a cancer registry.

7. 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.

8. 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 Macro. 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. The 308(d) confidentiality protection does not go with the data, 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. 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 H). 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.

9. Are there penalties for violating the confidentiality assurance?

NPCR employees and contractors at ICF Macro 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 Macro with access to NPCR-CSS data are required to sign a confidentiality agreement.

National Program of Cancer Registries Cancer Surveillance System Overview of Data Security

The NPCR-CSS project data reside on a dedicated server at ICF Macro. To ensure the security and confidentiality of project data, the following provisions have been incorporated into the ICF Macro 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 ICF Macro's Bethesda, MD, office with a guard on duty in the lobby 24 hours a day. Elevator and stairwell access is controlled by card key after normal business hours. The server resides on its own local area network (LAN) behind ICF Macro's firewall.

- Access to the NPCR-CSS server is limited to authorized ICF Macro project staff (see below). It is password-protected on its own security domain. No one, including non-project staff at ICF Macro, is allowed access to the NPCR-CSS data.
- All ICF Macro 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 Macro'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 Macro's security team. The security team consists of Donald McMaster, Business Steward; Kevin Zhang, Project Director; Jagruti Rana, Systems Lead and Security Officer; Gretchen Stanton, Database Administrator; and Kristopher Hall, LAN and WAN Security Steward. All project staff receive annual security awareness training covering security procedures. The ICF Macro 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 Macro 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.

**ICF Macro
Authorized Project Staff**

Staff Member	Position
Donald McMaster, M.B.A., M.S.	Business Steward
Kevin Zhang, Ph.D.	Project Director
Jagruiti Rana, M.S.	Systems Lead/Security Officer
Gretchen Stanton, M.A.	/Database Administrator
Kristopher Hall, MCSA	LAN and WAN Security Steward
Qiming He, Ph.D.	Quality Assurance (QA) Manager/Sr. Programmer Analyst
Yuan Ren, Ph.D.	Data and Statistical Manager/Senior Statistical Programmer
Olga Galin, M.S.	Sr. SAS Programmer/QA Specialist
Shailendra Bhavsar, B.S.	Programmer Analyst
Xing Dong, M.S.	Sr. Statistical Programmer
Phillip Schaeffer, M.S.	Sr. SAS Programmer
David Radune, B.A.	Sr. Developer
Jonathan Stanger, M.P.A.	SQL Programmer
Shaobin Xu, M.S.	Programmer Analyst
Jing Guo, B.S.	Programmer Analyst

**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	NPCR
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

NPCR-CSS Data Release Policy

August 2012

1995–2010 Diagnosis Years

State, Metropolitan Area, or Territory	First Diagnosis Year for Which Cancer Cases Were Reportable to NPCR or SEER*	Federal Funding Source
New Jersey	1995/2000	NPCR/SEER
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.

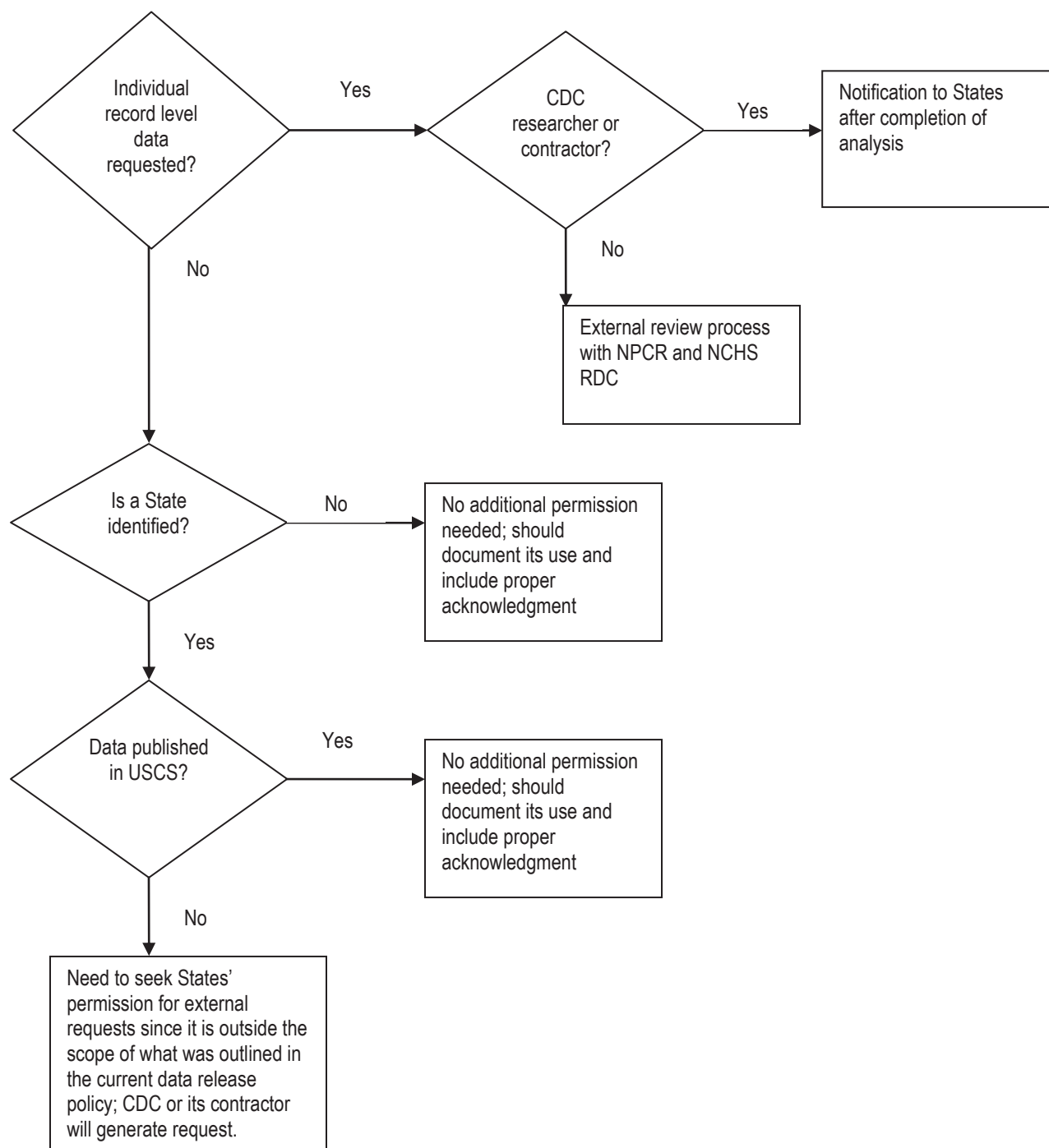
CDC = Centers for Disease Control and Prevention

NCI = National Cancer Institute

NPCR = National Program of Cancer Registries

SEER = Surveillance, Epidemiology, and End Results Program

NPCR-CSS Levels of Data Access



Data Items for Restricted-Access Dataset (RADs)

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]	

Name	Status [NAACCR Data Item number listed in brackets]	Notes
SEER 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]	
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).

Data Items for Central Brain Tumor Registry of the United States (CBTRUS) 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, 2004-2009 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]	<i>Same as race for USCS</i>
NHIAv2 Derived Hispanic Origin	191	
NAPIIA	193	
Sex	220	
Age at Diagnosis	230	<i>Single year up to age 84; 85+ grouped into one category</i>
Sequence Number—Central	380	
Date of Diagnosis (<i>YEAR portion only</i>)	390	<i>Day and month of diagnosis not requested</i>
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

**National Program of Cancer Registries
Cancer Surveillance System
Data Sharing Agreement**

It is of the utmost importance to ensure the confidentiality of individuals diagnosed with cancer when information about their cancer is entered into a database for the purpose of establishing a research resource. In order to protect these data, 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. In using data on such individuals for research purposes, it is absolutely necessary to ensure, to the extent possible, that uses of such data will be limited to research; any effort to determine the identity of any 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.

The Division of Cancer Prevention and Control (DCPC) does all it can to ensure that the identity of data subjects cannot be disclosed. All direct identifiers, as well as characteristics that might lead to identifications, are omitted from the dataset. Nevertheless, in rare instances it may be possible to ascertain from the dataset the identity of particular persons through complex analysis and with outside information. Considerable harm can ensue if this occurs.

In order for the DCPC to provide a restricted dataset to you, it is necessary that you agree to the following provisions:

1. I will not use nor permit others to use the data in any way other than for statistical reporting and analysis.
2. I will not release nor permit others to release the datasets or any part of them to any person except with the written approval of DCPC.
3. I will not attempt to link or permit others to link the dataset with individually identifiable records from any other CDC or non-CDC dataset.
4. I will not attempt to use the datasets or permit others to use them to learn the identity of any person or establishment included in any set.
5. If the identity of any person or establishment should be discovered inadvertently, then
 - a) no use will be made of this knowledge,
 - b) the Director of the DCPC will be notified of the incident,
 - c) the information that would identify an individual or establishment will be safeguarded or destroyed as requested by DCPC, and
 - d) no one else will be informed of the discovered identity.
6. In addition, I will make every effort to release all statistical information in such a way as to avoid inadvertent disclosure. For example,
 - a) No figure, including totals, should be less than 6 in tabulations for sub-State geographic areas, unless it is a tabulation routinely published by DCPC.

- b) No data on an identifiable case should be derivable through subtraction or other calculation from the combination of tables in a given publication.
- c) No data should permit disclosure when used in combination with other known data.

My signature indicates my agreement to comply with the above stated provisions with the knowledge that deliberately making a false statement regarding any matter within the jurisdiction of any department or agency of the Federal Government violates 18 U.S.C. 1001 and is punishable by a fine up to \$10,000 or up to 5 years in prison.

Signature

Date

Print or type name

Title

Organization

Mailing Address

Telephone

Fax

E-mail

Proposed use:

NPCR Data at the National Center for Health Statistics Research Data Center Questions & Answers

1. Can you summarize what CDC is planning to do?

CDC will use 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 NPCR data for research purposes. This mechanism will be included in the Data Release Policy for the 2012 submission (see page 7). A subset of the data submitted at that time (see page 11) will be the first data provided through the NCHS RDC. 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.

The NCHS RDC mechanism also provides an opportunity for CCR staff to propose and conduct analyses of the overall dataset.

2. 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>.

3. 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/>

4. What is CDC proposing for the restricted-access data set (RADS)?

CDC is proposing to release the RADS through the NCHS RDC beginning with the data submitted to CDC in November 2011. Additional details will be available in the data release policy for the 2012 data submission, which will be available in August 2011. The NPCR data will be available at the NCHS RDC in the Spring/Summer of 2012. The RADS will not be released until after CCRs have reviewed the data for State Cancer Profiles.

These services are important as we plan for ARRA/Comparative Effectiveness Research (CER) funded activities. In addition, there is continued interest in access to national data for research and cancer statistics. Wider access to the NPCR data is critical if we want continued support for these registries.

5. Why is CDC proposing to 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/B4ConfDisc/CfD400.htm>.

6. 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. A CDC-NPCR staff person will work with three CCR staff to provide comments and approval/disapproval of proposals. The CCR representatives will be selected from the participating CCRs on a rotating basis. Three CCRs will be asked to review each proposal within a given time frame. Because different CCRs will be selected each time a proposal needs review, every participating registry will have the opportunity to review proposals over time. This 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/B3Prosals/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.

7. 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>.

8. 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.

9. What data items will be included in the RADS?

The data items that will be included are listed on page 11. This is the draft version of Appendix F of the NPCR-CSS Data Release Policy for the 2012 Data Submission. The

variables are the same as previous data release policy except for the addition of County at Dx [NAACCR #90] as an option for use as an analysis variable.

10. Will the CCRs be able to decide whether their data will be available through the NCHS RDC?

Yes (see page 13). 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.

11. 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:

- a) 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.
- b) 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.
- 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 the NCHS RDC analyst to create these areas for the researcher.

12. 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 three (3) CCRs whose data are included in the NCHS RDC analysis. Staff from CCRs that participate in the RADS will be asked to review proposals on a rotating basis.

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

Only NPCR data will be available through the NCHS RDC. For researchers interested in including SEER data, they may obtain access through the SEER Website (<http://seer.cancer.gov/data/options.html>). The NCHS RDC analyst will compile the SEER data and NPCR data using SEER*Prep so that the researcher may conduct his or her analyses.

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

NPCR-CSS Data Release Policy

August 2012

1995–2010 Diagnosis Years

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

15. 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.

16. 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.

17. 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

18. 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.

19. My CCR would like to use the RADS at the NCHS RDC, but we are limited in funds to pay for the access fees as outlined at <http://www.cdc.gov/rdc/B5AprovProj/AP540.htm>.

The data access fees will not apply to CDC staff or CCRs whose data are included in the NCHS RDC dataset. A special agreement has been signed between NPCR and NCHS so that these fees do not apply.

20. 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.

21. How will access to the comparative effectiveness research (CER) dataset be managed?

Access to the CER dataset will be through the 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. A rotation system for obtaining comments will be utilized for this proposal review as well.