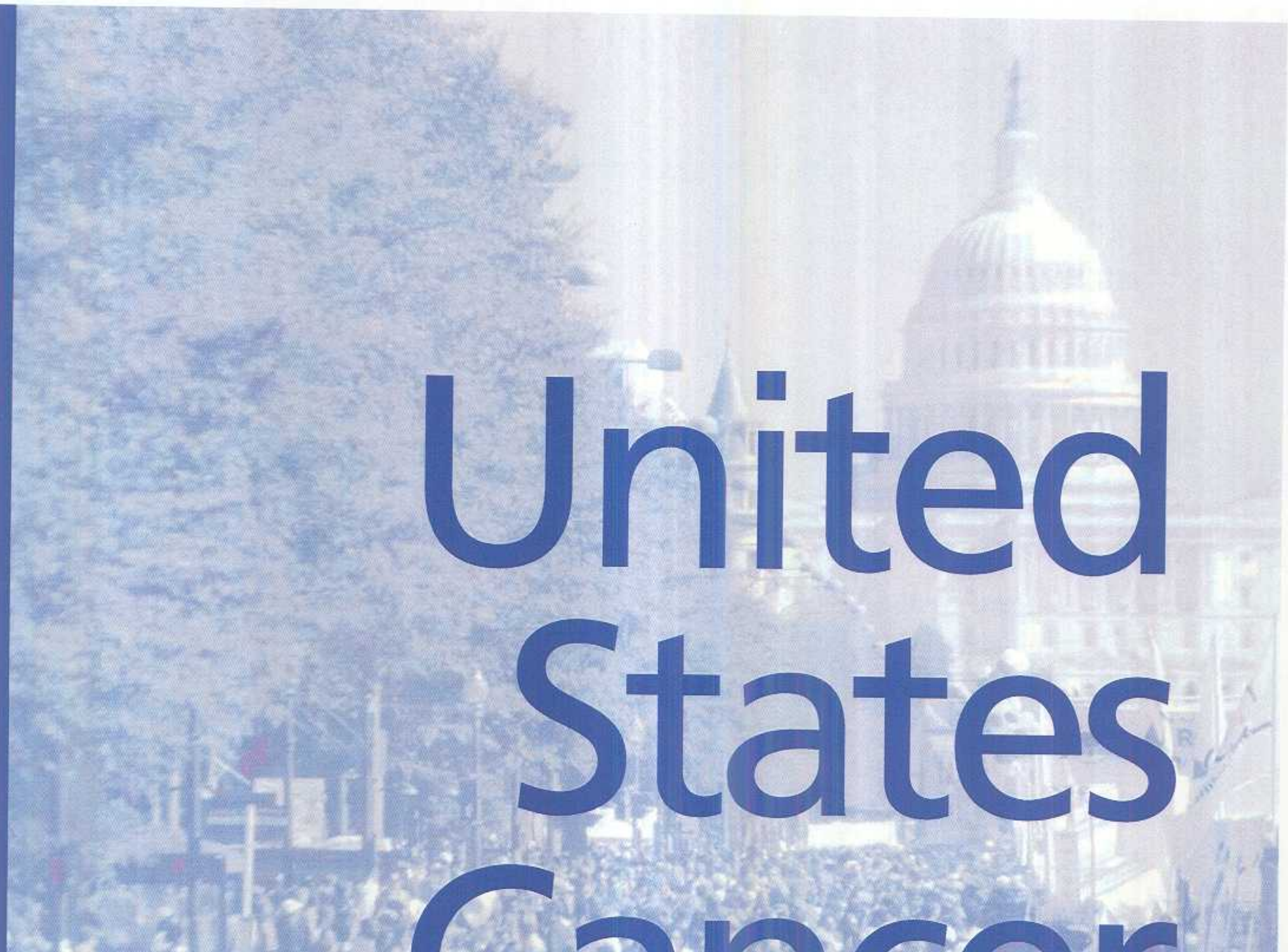
Attachment 5

United States Cancer Statistics: 2002 Incidence and Mortality



United States Cancer Statistics

2002 INCIDENCE AND MORTALITY

PUBLICATION YEAR 2005

CDC
CENTERS FOR DISEASE CONTROL AND PREVENTION

NATIONAL CANCER INSTITUTE

NORTH AMERICAN ASSOCIATION OF CENTRAL CANCER REGISTRIES

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

Suggested citation: U.S. Cancer Statistics Working Group. *United States Cancer Statistics: 2002 Incidence and Mortality*. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention *and* National Cancer Institute; 2005.

For more information, contact

National Program of Cancer Registries (NPCR)
National Center for Chronic Disease
Prevention and Health Promotion
Centers for Disease Control and Prevention
4770 Buford Highway, MS K-53
Atlanta, GA 30341-3717
Phone: (770) 488-4783
Fax: (770) 488-4759
http://www.cdc.gov/cancer/npcr

National Center for Health Statistics
Centers for Disease Control and Prevention
3311 Toledo Road
Hyattsville, MD 20782-2064
Phone: (301) 458-4000
E-mail: nchsquery@cdc.gov
http://www.cdc.gov/nchs

Surveillance, Epidemiology, and End Results (SEER) Program
Surveillance Research Program
Division of Cancer Control and Population Sciences
National Cancer Institute
Suite 504, MSC 8316
6116 Executive Boulevard
Bethesda, MD 20892-8316
Phone: (301) 496-8510
Fax: (301) 496-9949
http://seer.cancer.gov

United States Cancer Statistics will be published annually. All sections of this report are available on the Web sites listed below:

- http://www.cdc.gov/cancer/npcr/uscs
- http://seer.cancer.gov/statistics

A CD-ROM of this report may be ordered at http://www.cdc.gov/cancer/npcr/uscs.

United States Cancer Statistics

2002 Incidence and Mortality

U.S. Cancer Statistics Working Group

U.S. Cancer Statistics Working Group

Editors

Centers for Disease Control and Prevention

National Center for Chronic Disease Prevention and Health Promotion

Lyn Almon, MSPH
Cheryll Cardinez, MSPH
Temeika Fairley, PhD
Robert R. German, DrPH
Donald Haverkamp, MPH
Patricia M. Jamison, MPH
Jessica B. King, MPH
Joseph D. Rogers, BS
Sherri L. Stewart, PhD
Trevor D. Thompson, BS
Hannah K. Weir, PhD
Phyllis A. Wingo, PhD, MS

National Center for Health Statistics

Robert N. Anderson, PhD

National Cancer Institute

Brenda K. Edwards, PhD
Marsha E. Reichman, PhD

North American Association of Central Cancer Registries

Holly L. Howe, PhD
Betsy A. Kohler, MPH, CTR

Acknowledgments

The Centers for Disease Control and Prevention and the National Cancer Institute thank the staff in state and metropolitan area cancer registries for their commitment to cancer surveillance and control programs. The consistent efforts of registry staff to collect complete, timely, and high-quality cancer incidence data are essential for monitoring progress toward reducing the burden of cancer in the United States. Appendix A provides a list of state and metropolitan area cancer registries in the National Program of Cancer Registries (NPCR) and the Surveillance, Epidemiology, and End Results (SEER) program.

We thank the NPCR's Scientific Working Group members (Appendix B) for their assistance in designing and reviewing this report.

We also wish to thank vital registration and statistics staff in the 50 states, New York City, and the District of Columbia for their efforts to provide high-quality, timely, and comparable mortality data.

National partner organizations are crucial to the success of cancer registration and cancer surveillance in the United States. This report, and many advances in cancer surveillance in the United States, would not have been possible without the tireless efforts and many achievements of these organizations. The Centers for Disease Control and Prevention and the National Cancer Institute gratefully acknowledge and thank the American Cancer Society, the American College of Surgeons Commission on Cancer, the American Joint Committee on Cancer, the National Cancer Registrars Association, and our collaborating partner, the North American Association of Central Cancer Registries, for their support for cancer registration and cancer surveillance provided over many years. Appendix C has contact information for these national partner organizations.

We wish to thank several people from the Centers for Disease Control and Prevention: Kimbelian Barnes, Ryan Intlekofer, Karen Ledford, Mary Lewis, Fran Michaud, Phyllis (Janie) Nichols, and Reda Wilson for verifying the data and assisting with the appendices; Gayle Clutter and Dr. Roger E. McLendon for their assistance in preparing and reviewing the data on brain cancer by histologic group; and Dr. Harland Austin for his expert contributions to the Technical Notes and Appendix L.

Additionally, we would like to acknowledge several people who were involved in the production and distribution of this year's report: Kathleen Carey, Mary Kaeser, Diana Toomer, and the Division of Cancer Prevention and Control Web Development Team: Phuong Huynh, Aimee Murphy, Shannon Orr, Marcy Pendley, Blakeslee Potter, Wies Rafi, Corey Schuman, and Ramesh Thalluru.

Publication support was provided by Palladian Partners, Inc., under Contract No. 200-980-0415 for the National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services.

Preface

United States Cancer Statistics: 2002 Incidence and Mortality is a joint publication of the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia; the National Cancer Institute (NCI) in Bethesda, Maryland; and the North American Association of Central Cancer Registries (NAACCR) in Springfield, Illinois. This is the fourth annual joint report produced, in collaboration with NAACCR, by the two federal programs that support population-based cancer registries in the United States: CDC's National Program of Cancer Registries (NPCR) and NCI's Surveillance, Epidemiology, and End Results (SEER) Program. This report contains official federal government cancer statistics for more than 1 million invasive cancer cases diagnosed during 2002 among residents of 44 states, 6 metropolitan areas, and the District of Columbia, and these geographic areas are inhabited by about 93% of the U.S. population. Data from population-based central cancer registries in these states and metropolitan areas meet the selected criteria for inclusion in this report (see "Registry Eligibility Criteria" in the Technical Notes).

In addition to cancer incidence data, we present cancer mortality data collected and processed by CDC's National Center for Health Statistics (NCHS). Mortality statistics, based on records of deaths that occurred during 2002, are available for all 50 states and the District of Columbia.

The age-adjusted incidence and death rates are adjusted to the 2000 U.S. standard population, consistent with rates published in previous *United States Cancer Statistics* reports. Readers should not compare these data with published cancer rates that were not adjusted to the 2000 U.S. standard.

The race and ethnicity categories have been expanded this year to include national cancer incidence and mortality data for American Indians/Alaska Natives. In addition, malignant brain and central nervous system cancer incidence data are provided for the United States, categorized by histologic subtype, age, and sex. In future reports, more comprehensive cancer statistics will be provided while maintaining high-quality data standards.

A Web-based version of the report with current (2002) and updated (1999–2001) cancer incidence and mortality statistics is available at http://www.cdc.gov/cancer/npcr/uscs. Data at this Web site are provided in an HTML format that can be viewed in a variety of ways. A downloadable ASCII-formatted version of the data that can be manipulated to meet the needs of the user is also available.

Table of Contents

Technical Notes

Appendices

List of Tables, Figures, and Appendices

Cancer Incidence and Mortality for Adults and Children: All Areas Combined

Cancer Incidence and Mortality by U.S. Census Region and Division, State, and Metropolitan Area

State and Metropolitan Area Incidence Rates and State Death Rates for the Most Common Cancers

Appendices

United States Cancer Statistics: 2002 Incidence and Mortality

Executive Summary

The Centers for Disease Control and Prevention (CDC) and the National Cancer Institute (NCI), in collaboration with the North American Association of Central Cancer Registries (NAACCR), are pleased to release the fourth annual *United States Cancer Statistics* report. This report provides official federal government cancer statistics for cases diagnosed in 2002 and for cancer deaths that occurred in 2002. Cancer incidence statistics included in this report come from CDC's National Program of Cancer Registries (NPCR) and NCI's Surveillance, Epidemiology, and End Results (SEER) Program. Cancer mortality statistics are from CDC's National Vital Statistics System (NVSS).

Incidence data from 44 states, 6 metropolitan areas, and the District of Columbia are included in the report. The data obtained from NPCR and SEER registries in these areas cover 93% of the U.S. population. Mortality data from NVSS are presented for all 50 states and the District of Columbia and therefore cover 100% of the U.S. population.

Cancer incidence and mortality statistics are reported for 68 selected primary cancer sites and subsites for men of all ages and 72 selected primary cancer sites and subsites for women of all ages. These data are presented in tables and graphs in the following categories: (1) by geography: all areas combined, U.S. Census regions and divisions, states, and selected metropolitan areas and (2) by race and ethnicity: all races combined, whites, blacks, Asians/Pacific Islanders, American Indians/Alaska Natives, and Hispanics/Latinos. Age-specific analyses are also presented.

The childhood cancer section includes incidence data for more than 12,000 cancer cases and 2,000 cancer deaths among children and adolescents aged 19 years or younger. These data are presented by race and ethnicity, sex, age, and primary site, as well as by specific cancer types classified according to the *International Classification of Childhood Cancer* (ICCC).

The inclusion of cancer incidence and mortality data for American Indians/Alaska Natives is a new feature of this year's report. American Indian/Alaska Native incidence data are classified according to race information obtained from medical records and recorded in the registry. Additionally, NPCR registries linked their registry records with those of the Indian Health Service (IHS), which provides medical services to approximately 55% of the American Indian/Alaska Native population. This IHS linkage helps address racial misclassification in the registries. For mortality data, American Indian/Alaska Native race is reported as recorded on the death certificate.

Also new to this year's report is the inclusion of detailed malignant brain and central nervous system cancer incidence data. These data are grouped by age and sex, as well as by specific cancer type and subtype as defined by the Consensus Conference on Cancer Registration of Brain and Central Nervous System Tumors. In future publications, we plan to expand this list to include incidence data on benign brain and central nervous system tumors.

Age-adjusted cancer incidence and death rates, with 95% confidence intervals, are provided in each section. All rates in this report are age-adjusted to the 2000 U.S. standard population. Age-adjustment allows researchers to compare data across populations by controlling for the effect of age on populations with different age distributions. In this report, population estimates for 2002 were obtained from the 2000 U.S. Census and slightly modified by SEER for the Native Hawaiian population. These modified population estimates improve the accuracy of the rates for some racial and geographic populations. Readers should be cautious when interpreting comparisons between age-adjusted incidence rates and death rates because of differences in population coverage (93% for incidence and 100% for mortality). In addition, readers should not compare the rates published in this report with cancer rates that were calculated using different methods or standardized to different populations.

Major Findings

Rates are shown in parentheses where appropriate and are per 100,000 persons.

*Cancer Among Men**

Three most common cancers among men of all racial and Hispanic origin populations:

- **Prostate cancer** (161.2)
- **Lung cancer** (86.4)
- **Colorectal cancer** (61.3)

Three leading causes of cancer death among men:

- **Lung cancer** (73.5)
  - ➢ First among men of all racial and Hispanic origin populations.
- **Prostate cancer**
  - ➢ Second among white (25.8), black (63.0), and Hispanic (22.0) men.
  - ➢ Third among American Indian/Alaska Native men (15.7).

---

* All races combined rate is presented when ranking of cancer sites did not differ across race and ethnicity; race- or ethnic-specific rates are presented when ranking differed depending on race and ethnicity.

- **Colorectal cancer**
    - ➢ Second among American Indian/Alaska Native men (16.1).
    - ➢ Third among white (23.2), black (33.4), Asian/Pacific Islander (15.9),[†] and Hispanic (17.1) men.
- **Liver cancer**
    - ➢ Second among Asian/Pacific Islander men (15.9).[†]

*Cancer Among Women**

Three most common cancers among women:

- **Breast cancer** (124.9)
    - ➢ First among women of all racial and Hispanic origin populations.
- **Lung cancer**
    - ➢ Second among white (54.9) and American Indian/Alaska Native (32.9) women.
    - ➢ Third among black (50.3), Asian/Pacific Islander (26.7), and Hispanic (25.2) women.
- **Colorectal cancer**
    - ➢ Second among black (51.8), Asian/Pacific Islander (34.3), and Hispanic (34.9) women.
    - ➢ Third among white (43.9) and American Indian/Alaska Native women (26.8).

Three leading causes of cancer death among women:

- **Lung cancer**
    - ➢ First among white (42.6), black (40.2), Asian/Pacific Islander (17.6), and American Indian/Alaska Native (27.5) women.
    - ➢ Second among Hispanic women (14.8).
- **Breast cancer**
    - ➢ First among Hispanic women (15.7).
    - ➢ Second among white (24.9), black (34.1), Asian/Pacific Islander (12.9), and American Indian/Alaska Native (13.9) women.
- **Colorectal cancer**
    - ➢ Third among women of all racial and Hispanic origin populations (16.5).

---

* All races combined rate is presented when ranking of cancer sites did not differ across race and ethnicity; race- or ethnic-specific rates are presented when ranking differed depending on race and ethnicity.

[†] Colorectal cancer death rate for Asian/Pacific Islander men is 15.85; liver cancer death rate for Asian/Pacific Islander men is 15.86.

Cancer Among Children‡

The most common cancers in children aged 0–19 years:

- **Leukemia** (4.1)
  - ➤ Highest incidence rate found among children aged 1–4 years.
- **Brain and central nervous system cancer** (2.9)
  - ➤ Highest incidence rate of brain cancer found among children aged 1–4 years.

Leading causes of cancer death in children:

- **Leukemia** (0.8)
  - ➤ First among children aged less than 1 and aged 10–19 years.
- **Brain and central nervous system cancer** (0.7)
  - ➤ First among children aged 5–9 years.

Racial or Ethnic Variations§

All cancers combined, men:

- Incidence rates are highest among blacks (615.1), followed by whites (536.8), Hispanics (422.8), Asians/Pacific Islanders (324.3), and American Indians/Alaska Natives (267.2).
- Death rates are highest among blacks (322.9), followed by whites (236.0), Hispanics (163.9), American Indians/Alaska Natives (145.3), and Asians/Pacific Islanders (138.8).

All cancers combined, women:

- Incidence rates are highest among whites (408.9), followed by blacks (377.5), Hispanics (310.4), Asians/Pacific Islanders (264.5), and American Indians/Alaska Natives (215.4).
- Death rates are highest among blacks (190.9), followed by whites (161.9), American Indians/Alaska Natives (114.5), Hispanics (107.4), and Asians/Pacific Islanders (96.6).

Among the five races and ethnicities presented:

- American Indian/Alaska Native men have the lowest cancer incidence rates; however, Asian/Pacific Islander men have the lowest cancer death rates.
- White women have the highest cancer incidence rates; however, black women have the highest cancer death rates.
- American Indian/Alaska Native women have the lowest cancer incidence rates and the third highest cancer death rates.

‡ Rates presented are for males and females combined, all races combined, and children aged 0–19 years.

§ Race- or ethnic-specific rates are presented for all cancer sites combined.

Geographic Variations¶

Breast cancer:

- The incidence rate for the United States is 124.9; state incidence rates range from 109.0 to 147.8; approximately 53% of states have incidence rates at or above the national rate.
- The death rate for the United States is 25.5; state death rates range from 16.2 to 34.3; approximately 47% of states have death rates at or above the national rate.

Prostate cancer:

- The incidence rate for the United States is 161.2; state incidence rates range from 106.1 to 217.1; approximately 58% of states have incidence rates at or above the national rate.
- The death rate for the United States is 28.1; state death rates range from 17.6 to 51.8; approximately 59% of states have death rates at or above the national rate.

Lung cancer:

- **Men**
  - The incidence rate for the United States is 86.4; state incidence rates range from 38.1 to 133.8; approximately 47% of states have incidence rates at or above the national rate.
  - The death rate for the United States is 73.5; state death rates range from 32.2 to 112.6; approximately 47% of states have death rates at or above the national rate.
- **Women**
  - The incidence rate for the United States is 53.7; state incidence rates range from 20.9 to 73.0; approximately 49% of states have incidence rates at or above the national rate.
  - The death rate for the United States is 41.5; state death rates range from 18.7 to 57.5; approximately 47% of states have death rates at or above the national rate.

Colorectal cancer:

- **Men**
  - The incidence rate for the United States is 61.3; state incidence rates range from 43.7 to 75.1; approximately 56% of states have incidence rates at or above the national rate.
  - The death rate for the United States is 23.8; state death rates range from 16.6 to 30.5; approximately 53% of states have death rates at or above the national rate.
- **Women**
  - The incidence rate for the United States is 44.9; state incidence rates range from 32.2 to 55.0; approximately 47% of states have incidence rates at or above the national rate.

¶Geographic variations are presented for the four most common cancers.

- The death rate for the United States is 16.5; state death rates range from 12.9 to 21.5; approximately 53% of states have death rates at or above the national rate.

Two points should be kept in mind when interpreting the data in this report. First, differences in cancer incidence and death rates among racial and ethnic populations should be interpreted with caution. Recent studies involving cancer mortality data show that death rates for whites and blacks are generally reliable, whereas death rates for Asians/Pacific Islanders, American Indians/Alaska Natives, and Hispanics are underestimated. Studies involving cancer incidence data suggest similar results. Therefore, incidence and mortality data published in this report may be underestimated for some populations. Also, specific subpopulations, which are not presented in this report because of small numbers and possible misclassification, may have higher cancer incidence or death rates than the U.S. population. For example, overall cancer mortality among American Indians/Alaska Natives residing in the Alaska and the Northern Plains regions is higher compared with the U.S. population; the American Indian/Alaska Native population as a whole has lower cancer mortality than the U.S. general population. Indian Health Service coverage of American Indian/Alaska Native populations varies by region and under-represents American Indian/Alaska Natives who live in certain urban areas or who are members of non-federally recognized tribes.

Second, geographic variations in cancer incidence and death rates may be influenced by a number of factors: (1) areas in which a high percentage of the population is screened for cancer will have more cancer cases diagnosed than areas in which a low percentage of the population is screened; (2) rates for certain cancers differ among different racial and ethnic populations (e.g., black men have higher prostate cancer rates than other racial or ethnic groups), and thus when comparing cancer rates across geographic areas, the racial and ethnic makeup of that area should be considered; and (3) the population burden of cancer in a geographic area is determined by the number of cases diagnosed and the number of cancer deaths, not by the age-adjusted rate; therefore, a relatively high or low cancer age-adjusted rate may not be a reflection of the true cancer burden within that geographic area.

United States Cancer Statistics: 2002 Incidence and Mortality provides a basis for states and researchers to describe the variability in cancer incidence and death rates across different populations and to identify certain populations for evidence-based cancer control measures. We will continue to ensure data from all NPCR and SEER registries are of the highest quality. Since the publication of our first report covering diagnosis year 1999, additional registries have contributed data each year, resulting in increased coverage of the U.S. population with every new report. We expect that future reports will include high-quality data from the few remaining state registries, thus allowing a more comprehensive description of the cancer burden across racial, ethnic, and geographic populations in the United States.

Attachment 6

NPCR-CSS Data Release Policy

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

Data Release Policy

October 2005

Cancer Surveillance Branch
Division of Cancer Prevention and Control
NCCDPHP, CDC, Mailstop K-53
Atlanta, Georgia 30341-3717
Phone: 770-488-4783
E-mail: cancerinfo@cdc.gov (specify "NPCR-CSS" in subject line)

Table of Contents

Appendices

Acknowledgement

The Cancer Surveillance Branch, Division of Cancer Prevention and Control, NCCDPHP, CDC, wishes to acknowledge and thank the members of the state-based NPCR-CSS Scientific Workgroup for their contributions to the concepts included in this document and for their review and comment on drafts. In 2004 and 2005, the members of the Small Data Release Group played an important role in the review and release of the on-line USCS expanded data set and the USCS restricted-access data set. Appendix A lists the members of both groups.

National Program of Cancer Registries
Cancer Surveillance System
Data Release Policy
October 2005

I. Introduction

This document describes the format and content of data that the National Program of Cancer Registries' Cancer Surveillance System (NPCR-CSS) will release or share. This multi-year policy updates the September 2004 NPCR-CSS *Data Release Policy*. This policy applies to data submitted to Centers for Disease Control and Prevention (CDC) for the January 2006 data submission and for all future data submissions until a new policy is provided.

The NPCR-CSS Privacy Steward (Joseph Rogers) 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 than 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 states will be notified and will 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 states 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 in accordance 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 the NPCR-CSS data (Appendices B, C, and D).[4] Special care is needed even for 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 or from laboratory, 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 USC 242m(d)) (Appendices B and C). 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 USC 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.

III. Overview of the Data

In 1992, Congress established the National Program of Cancer Registries (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 planning and evaluation of cancer control interventions and for research. NPCR adopted reporting requirements and definitions consistent with the National Cancer Institute's (NCI's) Surveillance, Epidemiology, and End Results Program (SEER);[11,12] required the use of uniform data items and codes and record layouts as defined by 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. Currently, 45 states, three territories, and the District of Columbia are funded by NPCR (Appendix E).[15,16] NPCR-funded state registries collect data on patient demographics, primary tumor site, morphology, stage of disease at diagnosis, and first course of treatment. In addition, NPCR 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, and three territories. 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 state cancer registry receives case reports from facilities and practitioners throughout the state, processes them according to standard data management procedures.[14] Personal identifiers including the patient's name, social security number, and street address are removed 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 D). For more information on the NPCR-CSS data, see "Technical Notes" in *United States Cancer Statistics: 2002 Incidence and Mortality,*[16] which is updated and published annually. NPCR's Cancer Surveillance System 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 CFR 46) (http://ohrp.osophs.dhhs.gov/nsearch_t.htm).

State cancer registries and federal agencies routinely publish cancer incidence data between 22 and 25 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 states have preliminary data available as soon as 12 months after the close of each diagnosis year. Even after the publication of official statistics, state cancer registries (and CDC and NCI) continue to update and re-publish data with new information incorporated. Thus, 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 the file was prepared. Changes in state cancer registry incidence data that occur more than 22 to 25 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. Requests for State or Local Data and Notification of States

The Freedom of Information Act (FOIA) (http://www.cdc.gov/od/foia/foi/htm) 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 only Federal agencies, and covers only records in the possession and control of those agencies at the time of the FOIA request (except in certain narrow instances involving grantee-held data). Because State based data becomes 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:

1) 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
2) Exemption 6, which exempts from disclosure personnel and medical files and similar files the disclosure of which would constitute a clearly unwarranted invasion of personal privacy.

In general, non-FOIA requests to CDC from the public, the 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 1992 Cancer Registries Act, Public Law 102-515; [4] (2) the relative infrequency of data submission to federal agencies assures that the state health department or its designated cancer registry will have the most complete, accurate, and up-to-date information; and (3) the state 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 data sets that are released in accordance with practices described in this document. At this time, it is anticipated that two kinds of data sets will be released: *public-use data sets* and *restricted-access data sets* (see definitions below).

CDC staff members or contractors perform analyses of NPCR program data as needed, including assessment of the completeness, timeliness, and quality of cancer incidence data; and analyses of the cancer burden 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. In compliance with the 308(d) Assurance of Confidentiality, CDC employees and contractors are required to handle the information in accordance with principles outlined in the CDC Staff Manual on Confidentiality and to follow the specific procedures documented in the NPCR-CSS Confidentiality/Security Statement (Appendices B, C, and D).

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 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 the National Program of Cancer Registries (NPCR) and submitted to CDC in the (insert submission date) NPCR-Cancer Surveillance System data submission.

V. Public-Use Data Sets

For purposes of this policy, public-use data sets (PUDS) are defined as data sets that are comprised of aggregated data (i.e., not individual case-specific data or microdata) that have been modified as needed, according to accepted procedures, to block breaches of confidentiality and prevent disclosure of the patient's identity or the patient's confidential information.[2, 5-10] A combination of confidentiality protection measures is employed for each PUDS (Table 1). PUDS will not contain information that is identifiable or potentially identifiable according to currently accepted procedures for reducing disclosure risk.[2, 5-10] One of the proposed PUDS that will be available as a Web-based query system will have a database behind a CDC firewall that is case-specific microdata; however, even for this database users will be able to access only aggregate counts and rates with all confidentiality protections built in. 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 9) will caution users against inappropriate use of the data or inappropriate disclosure of information. PUDS will be released as delimited ASCII files and/or on a Web-based query system, or possibly through other vehicles (Table 1). States will have an opportunity to review their state's data before each PUDS is released and have adequate time to notify CDC if they see a problem with the data. As a convenience to NPCR state registries, states may request from CDC a copy of their complete state-specific analytic database that is used to create each PUDS. The following three PUDS are anticipated:

- *United States Cancer Statistics (USCS)* data set
- *USCS* expanded data set
- *USCS* county cancer incidence data set.

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 will be as described in the technical notes for *USCS* unless noted in separate documentation that accompanies the data set. While data from NCI's SEER program is included in the *USCS* data set, these data are not currently included in all releases described in this document. However, based on the underlying principles for coordinated nationwide cancer control as stated in the CDC and NCI Surveillance Memorandum of Understanding, the goal is to have all datasets include data from both NPCR and SEER.

Separate documentation may accompany each PUDS which will describe its unique features; for example, 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. *United States Cancer Statistics (USCS)* Data Set

The *USCS* data set is a PUDS that will contain the same aggregate counts and rates for incidence and mortality published annually (Table 1). The PUDS is an HTML edition of *USCS*. Tables of male and female combined counts and rates will be in the HTML edition in addition to the published sex-specific tables. Users will be able to download the data in ASCII format for use in other applications.

B. *USCS* Expanded Data Set

The *USCS* expanded data set is a PUDS that displays the aggregate counts and rates published yearly in the *USCS,* plus additional aggregate values created from the same analytic file containing finer breakdowns of counts and rates based on selected variables (Table 1). This PUDS will be available as a Web-based query system and will have a database behind a CDC firewall that is case-specific microdata; however, users will be able to access only aggregate counts and rates with all confidentiality protections built in. Because these PUDS will present data in more detail than is presented in *USCS*, states will have the option to notify NPCR if they prefer not to have their state data included. It is anticipated that the 2004 version of this PUDS may be available in late 2005. The first version of this file will contain data from the 1999-2002 diagnosis years. It is anticipated that diagnosis years before 1999 will be added when resources permit and when data quality for those years has been tested and found to meet the specified criteria.

C. *USCS* County Cancer Incidence Public Use Data Set

The *USCS* county cancer incidence data set is a PUDS that will consist of aggregate cancer incidence counts and rates for selected counties in the United States (Table 1). This PUDS will be available as an ASCII file. Because this PUDS will present data at a sub-state geographic level, states will have the option to notify NPCR if they prefer not to have their state data included. A limited version of this data set was released in 2004 to a small number of users, including the American Cancer Society for its Community Assessment "E-Tool" CD-ROM, NCI for the State Cancer Profiles project (www.statecancerprofiles.cancer.gov), and the U.S. Department of Health and Human Services' Women's Health Initiative Project. Future versions may contain more detail about cancer at the county level. Unless states are notified well in advance, this data set will always contain a cell suppression threshold that is at least as high as that used in 2004 with counts and rates averaged over at least three years. When states have accumulated more consecutive years of high quality data, we will publish county data averaged over five years. This next version of this PUDS, based on the January

2005 data submission to CDC, will be available in late Fall 2005. It is anticipated that diagnosis years before 1997 will be added when resources permit and when data quality for those years has been tested and found to meet the specified criteria.

D. 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:
>
> - 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.

VI. Restricted-Access Data Sets

For purposes of this policy, restricted-access data sets are defined as versions of the full NPCR-CSS analytic data set, 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. Restricted-access data sets will not contain personal identifiers such as a patient's name, street address, or social security number as this information is not transmitted by state cancer registries to CDC. However, they may contain information that is potentially identifiable especially when linked with other data sets, such as the occurrence of a rare cancer in a person of a certain age or racial or ethnic group. A list of the variables included in the data set are in Appendix G. Restricted-access data sets will be released only after CDC authenticates the requestor's identity and the requestor either signs a detailed data sharing agreement or signs a less detailed data sharing agreement and submits confirmation of IRB approval by an institution with appropriate assurance from the U.S. Department of Health and Human Services' Office for Human Research Protections (http://ohrp.osophs.dhhs.gov/nsearch_t.htm). Because restricted-access data sets may potentially contain identifiable information, states will have the option to notify NPCR if they prefer not to have their state data included in a restricted access data set.

To define and develop the datasets, CDC staff members are working collaboratively with a subset (Small Data Release Group) of the NPCR-CSS Scientific Workgroup. The resources needed for preparation, data quality testing, releasing, logging [2] and monitoring the release of restricted data sets have been and will continue to be substantial. Thus, release of restricted-access data sets will occur in over a multi-year period. The first release to ACS, NCI and members of the Small Data Release Group is scheduled for November 2005. In November 2006, the data will be available to ACS, NCI, and all NPCR programs. In the third year, if all goes as planned, the data will be available to researchers who meet the criteria jointly established by CDC and NPCR programs. This restricted-access data set will be available as an ASCII file. User documentation will be provided and will be facilitated by the availability of a NAACCR data dictionary for every diagnosis year since 1994.[13] Detailed data sharing agreements, and procedures for user authentication and for logging and monitoring of data releases, will also need to be developed. The recently completed *CDC/ATSDR/CSTE Data Release Guidelines for Re-Release of State Data* described in detail the recommended contents of such a data sharing agreement and provided guidance for user authentication and logging and monitoring of data releases.[2] CDC staff members and partners who are working together to plan for the release of these data sets will review these CDC recommendations and other relevant materials and develop data sharing agreements that meet the requirements of state and federal agencies. A similar approach will be used for developing methods for user authentication, logging of data releases, and compliance monitoring.

VII. Emergency and Provisional Data Releases

It is not anticipated that CDC will ever 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 and C).

Provisional data and draft data tables will be shared with CDC employees and contractors, NPCR states, and other partners as needed 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.

VIII. 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 a restricted access data set under certain controlled conditions.[1] These controlled conditions may include a CDC-controlled data center such as the data center established at 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 this policy is in place. Release of data under controlled conditions will be considered as part of discussions with partners toward development of restricted access files, and a determination will be made as to whether such releases of data will be considered in the future for NPCR-CSS data.

IX. 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 Statisitics; 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 Statisitcal Association; 2005. Available at http://www.amstat.org/comm/cmtepc/index.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://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. *United States Cancer Statistics: 2002 Incidence and Mortality, Technical Notes.* Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention *and* National Cancer Institute; 2005. Available at http://www.cdc.gov/cancer/npcr/uscs.

17. Method for estimating case completeness is available in: Wu XC, Hotes JL, Fulton PJ, Cormier M, Correa CN, McLaughlin CC, Kosary C, Howe HL, Chen VW, editors. *Cancer in North America, 1995-1999, vol. I: Incidence.* Springfield, IL: North American Association of Central Cancer Registries; 2002. Available at http://www.naaccr.org/Stats/CINAPubs.html.
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 1. Comparison of NPCR-CSS Data Sets

	Public-Use			Restricted-Access
	USCS	*USCS* Expanded	*USCS* County	*USCS* Restricted Access File (RAF)
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 individual record level data
Mode of Access	Web-based query system and downloadable ASCII file tables	Web-based query system	Flat ASCII file and Web-based query system and separate brief text documentation; in future years may be more than one flat ASCII file; www.statecancerprofiles.cancer.gov	Flat ASCII file
Web Address or Contact Information	*USCS* website (www.cdc.gov/cancer/npcr/uscs)	CDC Wonder (http://wonder.cdc.gov)	Request from cancerinfo@cdc.gov (specify "NPCR-CSS Data Release" in subject line) or by phone (770-488-4783)	Request from cancerinfo@cdc.gov (specify "NPCR-CSS Data Release" in subject line) or by phone (770-488-4783)
Contains Identifiable or Potentially Identifiable Information	No	No	No	Yes
Registry Eligibility Criteria for Data Completeness and Quality	USCS publication criteria	USCS publication criteria; data meet criteria for inclusion of the additional variables	USCS publication criteria; data meet criteria for passing single-field and inter-field edits for county at diagnosis	USCS publication criteria
When Available	Updated November 2005	2005	2006	2005

Table 1. Comparison of NPCR-CSS Data Sets, *continued*

	Public-Use			Restricted-Access
	USCS	*USCS* Expanded	*USCS* County	*USCS* RAF
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	Single year of data from 1999-2002 (additional diagnosis years to be added in future years)	Single year of data from 1999-2002 (additional diagnosis years to be added in future years)	Aggregated years of data from 1999-2001, 2000-2002, 1999-2003** (additional diagnosis years to be added in future years)	Single year of data from 1999-2002 (additional diagnosis years to be added in future years)
Cancer Sites	All invasive cancers *in situ* female breast	All invasive cancers *in situ* female breast	Current: All 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. (The following sites will be added in 2006 if they meet data quality standards: non-Hodgkin lymphoma, liver and intrahepatic bile duct, pancreas, esophagus, and childhood cancers (all sites <20 and <15 of age)**	All invasive and *in situ* cancers

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

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

Table 1. Comparison of NPCR-CSS Data Sets, *continued*

	Variables Included			
	Public-Use			Restricted-Access
	USCS	*USCS* Expanded	*USCS* County	*USCS* RAF
Geographic Levels	All areas combined; U.S. census region and division; NPCR and SEER state or territory; SEER metropolitan area	All areas combined; NPCR state or territory; MSA for cities of 500,000 or more* (additional levels may be added)**	NPCR state; county* (additional levels may be added in 2006; i.e., Health Service Regions)**	NPCR state or territory*
Race	All races; white; black; Asian/Pacific Islander; American Indian/Alaska Native (subpopulations may be added)**	All races; white; black; Asian/Pacific Islander;** American Indian/Alaska Native** (subpopulations may be added)**	All races; white; black (additional groups or subpopulations may be added)**	All races collected by NAACCR
Ethnicity (Hispanic)	Yes	Yes	No (may be added)**	Yes
Age Groups	For adult cancers, all sites combined only all geographic areas combined only: standard 5-year age groups.** For childhood cancers, all geographic areas combined only: 0-4; 5-9; 10-14; 15-19; 0-14; 0-19. (additional groups may be added)**	For adult cancers: standard 5-year age groups that can be combined by the user. (additional groups may be added)**	No (may be added)**	Standard 5-year age groups
Summary Stage	No	As a future enhancement; when comparability between SEER Summary Stage 1977 and 2000 are resolved, stage may be added for female breast, colon and rectum; prostate; lung and bronchus; and cervix uteri (additional sites may be added).**	No	Yes
Histology	No	ICCC for childhood cancers (per SEER monograph)	No	Yes

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

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

Table 1. Comparison of NPCR-CSS Data Sets, *continued*

Confidentiality Protection/Disclosure Limitation Measures Employed				
	Public-Use			Restricted-Access
	USCS	*USCS* Expanded	*USCS* County	*USCS* RAF
No Direct Identifiers	Yes	Yes	Yes	Yes
Aggregation	Yes	Yes	Yes	No
Limited Number of Variables	Yes	Yes	Yes	Yes
Grouping/Collapsing of Variables or Response Codes	Yes	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 3 years of data	No	No	Yes	No
Cell Suppression	Counts and rates: count of less than 16	Counts and rates: count of less than 16	Counts and rates: 3-5 year total count of less than 16**	No
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

** Pending further data quality investigation and discussion with the NPCR-CSS Scientific Workgroup

NPCR-CSS Scientific Workgroup Members

Lyn Almon, MSPH, Centers for Disease Control and Prevention
Mark E. Allen, MS, California Cancer Registry
Virginia C. Andrews, MSPH, South Carolina Central Cancer Registry
Cheryll Cardinez, MSPH, Centers for Disease Control and Prevention
Vivien W. Chen, PhD, Louisiana Tumor Registry
Catherine N. Correa, PhD, MPH, Louisiana Tumor Registry
Temeika Fairley, PhD, Centers for Disease Control and Prevention
Carol Friedman, DO, Centers for Disease Control and Prevention
Susan T. Gershman, MS, MPH, PhD, CTR, Massachusetts Cancer Registry
Georgette G. Haydu, MS, Ohio Cancer Incidence Surveillance System
Holly L. Howe, PhD, North American Association of Central Cancer Registries
Jeannette Jackson-Thompson, PhD, MSPH, Missouri Cancer Registry
Missy Jamison, MPH, Centers for Disease Control and Prevention
Alison T. Johnson, CTR, Vermont Cancer Registry
Amy Kahn, MS, CTR, New York State Cancer Registry
Karen L. Knight, MS, North Carolina Central Cancer Registry
Betsy A. Kohler, MPH, CTR, New Jersey State Cancer Registry
Sue Min Lai, PhD, MS, MBA, Kansas Cancer Registry
Jill A. MacKinnon, CTR, Florida Cancer Data System
Howard J. Martin, PhD, Virginia Cancer Registry
Stacey Neloms, MPH, Maryland Cancer Registry
Xiaoling Niu, MS, New Jersey State Cancer Registry
Patricia Reese, CTR, Wyoming Central Tumor Registry
Maria J. Schymura, PhD, New York State Cancer Registry
Tiefu Shen, MD, PhD, Illinois State Cancer Registry
Laura Stephenson, BA, Wisconsin Cancer Reporting System
Thomas C. Tucker, PhD, MPH, Kentucky Cancer Registry
Lydia Voti, MS, Florida Cancer Data System
Hannah Weir, PhD, Centers for Disease Control and Prevention
Nancy S. Weiss, PhD, Texas Cancer Registry
Melanie Williams, PhD, Texas Cancer Registry
Brian D. Wright, BS, Pennsylvania Cancer Registry
William E. Wright, PhD, California Cancer Registry
Xiao Cheng Wu, MD, MPH, CTR, Louisiana Tumor Registry

NPCR-CSS Small Data Release Group

Mark E. Allen, MS, California Cancer Registry
Pam Agovino, New Jersey State Cancer Registry
Vivian W. Chen, PhD, Louisiana Tumor Registry
Catherine N. Correa, PhD, Louisiana Tumor Registry
Georgette G. Haydu, MS, Ohio Cancer Incidence Surveillance System
Amy Kahn, MS, CTR, New York State Cancer Registry
Karen L. Knight, MS, North Carolina Central Cancer Registry
Sue Min Lai, PhD, MS, MBA, Kansas Cancer Registry
Jill A. MacKinnon, CTR, Florida Cancer Data System
Xiaoling Niu, MS, New Jersey State Cancer Registry
Maria J. Schymura, PhD, New York State Cancer Registry
Jeannette Jackson-Thompson, PhD, MSPH, Missouri Cancer Registry

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 Prevention and Control (CDC), an agency of the United States Department of Health and Human Services, and Macro International Inc., a contractor of the CDC. The information to be received by the 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. Once a year in January, CDC will request 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 USC 242k) with an assurance that it will be held in strict confidence in accordance with Section 308(d) of the PHS Act (42 USC 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 the 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, aside from NCCDPHP 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 and organizations (e.g., the North American Association of Central Cancer Registries, the American Cancer Society, the National Cancer Institute) will be required to sign a detailed data release agreement to have access to restricted release data.

Appendix C

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 provide a degree of protection for personally identifiable data, the Public Health Service Act, Section 308(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 (then a separate agency) to grant 308(d) confidentiality protection in 1983. Section 308 (d) of the Public Health Service Act (42 U.S.C. 242 m(d)) 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 fifty projects have received 308(d) protection since CDC received this authority, and currently there are only approximately 25 active projects with 308(d) confidentiality assurances. As a testament of 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 staff of CDC, or its 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. 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.

b. 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.

c. 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.

d. An IRB Review Status Statement verifying NPCR-CSS's exemption from CDC Institutional Review Board (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 not necessary.)

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 five years to maintain the assurance.

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 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 Freedom of Information Act (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 non-release 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, Macro International, Inc. Data that are released to external researchers are done so in accordance with the Data Use Agreement (copy attached) 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), the American Cancer Society (ACS), and the 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, a detailed data use agreement must be signed by the requesting party. Information that could lead to the identification of cancer patients, through either 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. NPCR does not plan to release individual record level data submitted in 2003.

9. Are there penalties for violating the confidentiality assurance?

NPCR employees and contractors at Macro International, Inc. 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 Macro International, Inc. 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 ORC Macro. To ensure the security and confidentiality of project data, the following provisions have been incorporated into the ORC 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 ORC Macro's Bethesda office with a guard on duty in the lobby 24 hours a day. Elevator and stairwell access is controlled by card key. The server resides on its own local area network (LAN) behind ORC Macro's firewall.

- Access to the NPCR-CSS server is limited to authorized ORC Macro project staff (see below). It is password protected on its own security domain. No one, including nonproject staff at ORC Macro, is allowed access to the NPCR-CSS data.

- All ORC 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 ORC 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 when no longer needed.

- A comprehensive security plan has been developed by ORC Macro's security team. The security team consists of June Bray, Managing Director; Kevin Zhang, Project Director; Leo Shen, Data Manager and Security Officer; David Radune, Database Administrator; and Gretchen Stanton, LAN and WAN Security Steward. All project staff receive annual security awareness training covering security procedures. The ORC Macro project security team oversees operations to prevent unauthorized disclosure of the NPCR-CSS data.

- Periodic (currently quarterly, but no less than once a year) review and update of ORC 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.

ORC Macro
Authorized Project Staff

Staff Member	Position
June Bray, Ph.D.	Managing Director
Kevin Zhang, Ph.D.	Project Director
Leo Shen, M.B.A.	Data Manager/Security Officer
David Radune, B.S.	Database Administrator
Gretchen Stanton, M.S.	LAN and WAN Security Steward
Qiming He, Ph.D.	QA Coordinator/Programmer Analyst
Yuan Ren, Ph.D.	Statistical Programmer/Sr. Analyst
Jonathan Stanger, M.P.A.	SQL Programmer
Shaobin Xu, M.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 2005
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

Appendix E

State, Metropolitan Area, or Territory	First Diagnosis Year for Which Cancer Cases Were Reportable to NPCR or SEER*	Federal Funding Source 2005
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
Palau	1999	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
Utah	1973	SEER
Vermont	1996	NPCR
Virgin Islands (U.S.)	1999	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

National Program of Cancer Registries
Cancer Surveillance System
Data Sharing Agreement

It is of the utmost importance to insure the confidentiality of individuals diagnosed with cancer when information about their cancer is entered into a data base for the purpose of establishing a research resource. In order to protect this 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 this data can only be used for the purpose for which it was obtained. In utilizing data on such individuals for research purposes, it is absolutely necessary to insure, 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, would be prosecuted to the full extent of the law.

The Division of Cancer Prevention and Control (DCPC) does all it can to assure 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 it may be possible in rare instances, through complex analysis and with outside information to ascertain from the dataset the identity of particular persons. Considerable harm could ensue if this were done.

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 data sets or any part of them to any person except with the written approval of DCPC;
3. I will not attempt to link nor permit others to link the data set with individually identifiable records from any other CDC or non-CDC data set;
4. I will not attempt to use the data sets or permit others to use them to learn the identity of any person or establishment included in any set; and
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.

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

- No figure, including totals, should be less than 6 in tabulations for substate geographic areas, unless it is a tabulation routinely published by DCPC.
- No data on an identifiable case should be derivable through subtraction or other calculation from the combination of tables in a given publication.
- No data should permit disclosure when used in combination with other known data.

Appendix F

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 USC 1001 and is punishable by a fine up to $10,000 or up to five years in prison.

Signature Date

Print or type name

Title

Organization

Mailing Address

Telephone

Fax

E-mail

Proposed use:

Please return completed form to:
Attention Mr. Joseph Rogers
Division of Cancer Prevention and Control
NCCDPHP
Centers for Disease Control and Prevention
4770 Buford Hwy, N.E., Mailstop K-53
Atlanta, GA 30341-3724
Phone: (770) 488-3026
Fax: (770) 488-4759

Description of the Restricted Access File (RAF)

The restricted access data set is individual case-specific data of the NPCR-CSS data set. The data items to be included in this data set are listed below.

NAACCR Name and Item Number

Patient Id Number [20] – recoded
Address at Diagnosis – State [80]
Race 1 [160]
Race 2 [161]
Race 3 [162]
Race 4 [163]
Race 5 [164]
Spanish/Hispanic Origin [190]
NHIA Derived Hispanic Origin [191]
Sex [220]
Age at Diagnosis [230]
Birth Date [240] – Year only
Sequence Number – Central [380]
Date of Diagnosis [390] – Month and year only
Primary Site [400]
Laterality [410]
Histology ICD-O-2 [420]
Behavior ICD-O-2 [430]
Grade [440]
Diagnostic Confirmation [490]
Type of Reporting Source [500]
Histologic Type ICD-O-3 [522]
Behavior Code ICD-O-3 [523]
SEER Summary Stage 2000 [759]
SEER Summary Stage 1977 [760]
Over-ride Age/Site/Morph [1990]
Over-ride SeqNo/DxConf [2000]
Over-ride Site/Lat/SeqNo [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]

Derived Fields

Primary Site Recode
Primary Site Recode including Mesothelioma and Kaposi's Sarcoma
Behavior Code for Analysis
SEER-modified ICCC Recode
Standard 5 Year Age Recode
Race Recode (white, black, Asian/Pacific Islander, American Indian/Alaska Native)