Information for CINA Deluxe Data Investigators

1. Eligibility for data access

- You must be a member of the North American Association of Central Cancer Registries (NAACCR), or an employee of a member organization (http://www.naaccr.org/Membership/MembershipDirectory.aspx) to request access to CINA Deluxe data files. [The information on NAACCR membership can be found at http://www.naaccr.org/Membership/BecomeaMember.aspx]

- If you are not a NAACCR member, contact a NAACCR registry member or other member organization and discuss your proposal idea to determine if it is feasible based on available CINA Deluxe data. **The principal investigator for any CINA Deluxe application must be a NAACCR member**. Researchers who are not a NAACCR member may serve as a co-investigator and/or apply for Individual membership.

- You must be familiar with SEER*Stat (http://seer.cancer.gov/seerstat/index.html) and cancer statistics to ensure accurate data analysis.

2. CINA Deluxe data

2.1. Source of data

Data in the CINA Deluxe data files are from the NAACCR Call for Data. Population-based central cancer registries in the United States and Canada submit cancer incidence data, from1995 to the most current year, to NAACCR in December each year for its Call for Data.

2.2. Data year and geographic coverage and data quality

CINA Deluxe data files include data only from registries that meet all the CINA Deluxe inclusion criteria (see the table below) for at least three consecutive years.

Criteria	CINA Deluxe Inclusion Criteria
Completeness	≥90%
% Passing EDITS	100%
Death-Certificate-Only Cases	<=5%
Timeliness	Received by December due date (within 23 months of the close of the diagnosis year)

Duplicate Reports	<=2/1,000 cases
Missing Sex, Age, County	<=3%
Missing Race	<=5%

Data quality and completeness vary by registry and year of diagnosis. You are advised to assess the quality and completeness of the data before proceeding with data analysis. Researchers may have to limit data analysis to a subset of registries or to certain diagnosis years because of missing or low quality data.

2.3. Timeliness

Registries are required to submit data for the NAACCR Call for Data within 23 months of close of a diagnosis year (i.e., December 1, 2011, for all cases diagnosed between 1995 and 2009). The CINA Deluxe data files are available for use about 28 months after the close of the diagnosis year (i.e., May 15, 2012, for all cases diagnosed between 1995 and 2009).

2.4. Data elements

NAACCR collects a limited data set for certification, publication and evaluation. Three types of files are currently available for researchers.

- **CINA Deluxe Standard files** include the **v**ariables in Table 1 (http://www.naaccr.org/Research/CINADeluxe.aspx)
- **CINA Deluxe files with single year of age** include the **v**ariables in Table 1 plus single year of age
- **CINA Deluxe files with county data** include the **v**ariables in Table 1 plus the county and county-based socioeconomic measures.

You are advised to review the variable list to ensure that data elements that you are going to use are available in the CINA Deluxe data files. Further information regarding values and codes in the CINA Deluxe data may be found in the NAACCR Data Standards and Data Dictionary, Volume II (http://www.naaccr.org/StandardsandRegistryOperations/VolumeII.aspx).

If you still have questions after reviewing these materials, please send an email to durc@naaccr.org.

3. Application requesting access to CINA Deluxe data files.

You can find the application packet on
http://www.naaccr.org/Research/CINADeluxe.aspx

The application must include:

- Study proposal using the CINA deluxe proposal template
- Biographical sketches for all investigators
- Data Confidentiality Agreement for NAACCR Researchers form signed and dated by every person who will access the data

Incomplete applications will not be considered. Please submit your application to Royale Anne Hinds rahinds@naaccr.org at NAACCR.

4. Application review process

4.1. Review process

- Upon receipt of a complete application, the DURC chair distributes the application packet to DURC members and designates two reviewers (primary and secondary). An employee of Information Management System, Inc. (IMS), the company that houses the CINA Deluxe data, will review the proposal for the adequacy of technical detail.

- The reviewers will have two or more weeks to review the application and then present reviews during the next DURC monthly conference call.

- Based on reviewers' comments and the committee's discussion, the DURC will make one of the three decisions: Approved, Approved Subject to Revisions, or Reject.

- The primary and secondary reviewers finalize their reviews and submit written comments to the DURC chair within a week after the DURC conference call**.**

- The DURC chair notifies the applicant of the decision and provides reviewers' comments via email and copies Betsy Kohler and Royale Anne Hinds. The notification will include the information required for the NAACCR IRB application.

4.2. Review criteria

The DUR committee examines the proposal for

- Experience of investigators in analyzing cancer surveillance data
- Clarity of research questions and objectives
- Availability of data items to address stated research questions
- Appropriateness of analytic approach

4.3. NAACCR IRB and consent process

- You are required to complete and submit all forms for the IRB

- NAACCR IRB panel reviews the application for compliance with human subjects protection.

- Following the IRB approval, NAACCR will email all registries whose data are eligible for your project and ask them to consent to the use of its data in the study. Registries are usually given 14 days to respond.

- **Note:** NAACCR requires active consent when the proposed project requires area identifier (e.g., county code) or area-based measure (e.g., percent of persons in the county living below the poverty level), linkage with other datasets, or presentation of state-specific results with state name.

- If a registry does not respond to a request, NAACCR assumes the registry consents for projects requesting passive consent and does not consent for projects requesting active consent.

4.4. Length of the application process

- About a month - DURC review process
- About a month - NAACCR IRB review
- About 14 days - registry consent process
- About 14 days - IMS prepares the data files in SEER*Stat

You should be able to gain access to the data in SEER*Stat in about 3 months.

4.5. Access to CINA Deluxe data through SEER*Stat

- You will receive an email notification from Information Management Services, Inc. (IMS) when access to your NAACCR CINA Deluxe File is available.

- The data files will contain data from consenting registries only.

5. Confidentiality and responsibilities

- You (the researcher) must formally agree and acknowledge that patient confidentiality is of the utmost importance in using the data and in presenting or publishing research results. Please refer to the "Data Confidentiality Agreement for NAACCR Researchers with Access to CINA Deluxe" for further details. (http://www.naaccr.org/Research/CINADeluxe.aspx)

- Your agreement with NAACCR states specifically that the data must only be used for the purpose defined in the approved proposal. Any other use violates this agreement.

- Access to the data file will be terminated at the conclusion of the project.

6. Publication and review instructions

Upon completion of the study, NAACCR reserves the right to review findings and ensure compliance with privacy conditions.

6.1. Final product

- As a condition of release, a written report or a manuscript of the project's findings (positive or negative) must be completed and submitted to Royale Anne Hinds (rahinds@naaccr.org) at NAACCR within 12 months of the award notification.

- A brief description of the data quality findings shall be in the report.

- The report should not exceed 6,000 words, or a manuscript may be submitted.

6.2. Approval of NAACCR Scientific Editorial Board

- Any manuscript resulting from this effort must be submitted to the NAACCR Scientific Editorial Board (bkohler@naaccr.org) for review and approval before being submitted for publication. The review process takes about 6 weeks.

Note: Abstracts or presentations at meetings don't have to be approved. However, please email a courtesy copy to durc@naaccr.org.

6.3. Acknowledgment

Using the following wording, please acknowledge NAACCR and the use of its CINA Deluxe Data files in publications or presentations for the projects using the CINA deluxe data.

These data are based on the NAACCR December 20xx__ data submission. Support for cancer registries is provided by the state, province or territory in which the registry is located. In the U.S., registries also participate in the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program or the Centers for Disease Control and Prevention's National Program of Cancer Registries (NPCR) or both. In Canada, all registries submit data to the Canadian Cancer Registry maintained by Statistics Canada.

7. Special notes regarding CINA Deluxe data

CINA Deluxe data are complex and sometimes difficult to analyze. Please see the following comments about selected data items.

7.1. Race and Hispanic Ethnicity Variables

- **NAACCR Incidence CINA Analytic File, for Expanded Races**. Researchers should use this file for all analyses that include race as one of an analytic variables regardless of race included. This file encompasses the maximum number of registries with data available for studies.

- **NAACCR CIUSHL Analytic File:** This file is the source for all analyses that include Hispanic ethnicity as an analytic variable. The NAACCR Hispanic/Latino Identification Algorithm, version 2 (NHIA v2) is a method used by all cancer registries in the United States to enhance identification of persons of Latino heritage diagnosed with cancer. Information describing the NHIA v2 algorithm is available on NAACCR website (http://www.naaccr.org/DataandPublications/EpiReports.aspx). All investigators should be cautious in their interpretation of these data because in areas of small Latino populations, methods to indirectly identify Latinos can overestimate the ethnicity-specific counts of cancer cases. Also, even small errors in population estimates of Latinos can affect the magnitude of the cancer rates.

In the NAACCR CIUSHL Analytic file, race is combined with ethnicity to produce the following population categories: Hispanic (all races combined), non-Hispanic white, and non-Hispanic black.

7.2. County Information

- Some states are prohibited from submitting county identifiers to NAACCR. The standard CINA Deluxe file contains a rural-urban indicator rather than county identifiers.

- In 2002, Colorado created a new county, Broomfield County. Thus, the 2000 and 2001Census estimates did not contain data for Broomfield County. Population characteristics for Broomfield County were created for 2002 and later in order to produce accurate county-specific population counts by race, Hispanic origin, age and sex. These changes do not affect analyses of state-level data.

7.3. Urban Rural Continuum 1993 and 2000

The Urban Rural Continuum variable is used as a measure of urbanization on a county level. Both the 1993 and 2000 variables are available on the file. Please see the following website for more information on the codes and their definition: http://www.ers.usda.gov/Briefing/Rurality/RuralUrbCon/code93.htm

7.4. Impact of Hurricane Katrina

Hurricane Katrina severely affected cancer surveillance in 2005 in Alabama, Louisiana, Mississippi, and Texas. Case counts in CINA Deluxe for these states are for the entire year. However, age-adjusted rates are calculated only on the number of cases diagnosed between January 1, 2005, and June 30, 2005. To include case data for the entire year for these 4 states, the box "Cases in Limited-Use Database" located within the "Selection" tab" must be unchecked.

7.5. Duplication of Incidence Cases in CINA Deluxe

The CINA Deluxe file contains data from all registries, including those that are population-based but not state-based (e.g., the metropolitan areas of Atlanta, Detroit, Greater Bay Area, Los Angeles, and Seattle). This means that for the states encompassing these metropolitan areas, case counts in the file are duplicated. In order to avoid double

counting of cases, researchers must exclude overlapping metropolitan areas when running analyses that are not stratified by registry.

7.6. Analyses involving time trends (caution)

It is very important that any analyses involving time trends include only data from registries that have data for all years in the time interval under investigation. It has been demonstrated that registries that do not meet this criterion may have different rates and trends so that inclusion of their partial data can skew overall trend estimates and lead to biases in the data and inaccurate conclusions.

7.7. Stage of disease at diagnosis

- **Summary Stage 1977 and SEER Summary Stage 2000 (U.S. Only)**

NAACCR has evaluated the impact of the coding changes in summary stage 2000 compared with summary stage 1977. SEER Summary Stage 1977 is applied to the 1995 to 2000 cases, and SEER Summary Stage 2000 is applied to the 2001-2003 cases. This report is available on NAACCR web site:
http://www.naaccr.org/DataandPublications/DataQuality.aspx

- **Collaborative Stage Derived Summary Stage 2000 (2004 and forward)**

The Derived Summary Stage 2000 is the coding schema used for all cases diagnosed 2004 and after. It is calculated using data items from the Collaborative Stage System. NAACCR has evaluated the comparability of the Derived Summary Stage 2000 and the Directly Coded Summary Stage 2000 using data from cancer registries. The report is available on http://www.ncbi.nlm.nih.gov/pubmed/21688742

7.8. Using dates of treatment

Two variables are available as standards for the date of first treatment, one used by the SEER program and the other by the Commission on Cancer (CoC). SEER's Date of Initial Treatment is listed as a month and year. The second is the CoC standard, Date of First Course of Treatment. Registries that are part of the SEER program are required to use the SEER standard. Registries in the NPCR program have the option to use either standard. Thus, both variables need to be used with caution since the two standard setters may applied different meanings to the variable. Investigators using these variables must

refer to NAACCR Standards for Cancer Registries, Volume II: Data Standards and Data Dictionary. This document can be found on NAACCR web site at: http://www.naaccr.org/standards.

7.9. For Canadian Registry Research

- Canadian registries do not collect information on race.

- The Rural Urban Continuum codes are not available for Canada.

- Canadian Registries have implemented Collaborative Stage Derived Summary Stage 2000, effective for cases diagnosed 2004 and later. The data are available on the file as submitted by provinces; however, these data are incomplete. Summary Stage 1977 and SEER Summary Stage 2000 are not available. Canadian registries do not follow NAACCR standards for collection of type of reporting source, and this data element is not standard among Canadian registries.