

**NAACCR Town Meeting**  
**NAACCR Research Program: Growth and Directions**  
**September 7, 2006**  
**3:00 pm Eastern; 2 pm Central; 1 pm Mountain; 12 pm Pacific**

**Present** – Representatives from the following registries and organizations:

**1 Canada:** Alberta Cancer Registry

**20 U.S.:** American Cancer Society, Arizona Cancer Registry, Centers for Disease Control and Prevention, Cancer Data Registry of Idaho, Illinois State Cancer Registry, Indiana State Cancer Registry, State Health Registry of Iowa, Louisiana Tumor Registry, Michigan Cancer Surveillance Program, Minnesota Cancer Surveillance System, National Cancer Institute, National Program of Cancer Registries, Nebraska Cancer Registry, Nevada Statewide Cancer Registry, New Hampshire State Cancer Registry, New Jersey State Cancer Registry, North Dakota Cancer Registry, Tennessee Cancer Registry, Texas Cancer Registry, Virginia Cancer Registry, Susan Carozza

**4 NAACCR Staff:** Dr. Holly Howe, Moderator – Executive Director, Joellyn Ellison – Program Manager of Data Evaluation and Publication, Royale Anne Hinds – Assistant to the Executive Director, and Lori Havener – Program Manager of Standards

**1. NAACCR Submissions: Increasing Focus on  
Developing a World Class Research Database**

**Holly Howe**

This meeting is to update NAACCR members on activities that are in progress or are nearing completion for developing and enhancing the NAACCR Research Program. This is a global euphemism to cover all uses of the NAACCR Call for Data submissions. We want to enhance the data set to increase the use of registry data in research a major goal of Deapen's vision during his presidency and which Betsy Kohler has incorporated this into her Presidential vision as well. The NAACCR Board has defined strengthening the NAACCR Research program as a primary goal, with an endpoint being the creation of a world class multi-registry research database and with the CINA monograph and CINA+ Online being important products of the process, but not the only endpoint.

We recognize that the NAACCR Call for Data is a voluntary activity and we hope that you and your registry find value in participating. We believe in addition to the feedback received, that the data and research uses are not only good for the whole, they also are good for each registry to add to the list of their own uses. Different than 15 years ago when we started and talked about requirements, the onset of national programs has caused different interpretation of this word. Thus we will not be using the word required, unless you are submitting data that you want to be included in a particular activity of product for which there are requirements, like Registry Certification.

- a. This year's Call for Data was used by the American Cancer Society (ACS) to produce more accurate estimates of cancer incidence, which will be published in Cancer Facts & Figures. Linda Pickle is writing a paper on the application of her model and used by ACS to produce the estimates. The paper will be published in

- Ca: A Journal for Clinicians*. The 2007 Call for Data submission has included this use of the data as a primary use of the NAACCR submission.
- b. Given that we are more focused on the creation of a research database as an endpoint, NAACCR is asking that all cancer cases be included in their NAACCR submission if they are reportable in the years requested, that is 1995-200x. This means that those reported in ICD-0-2 but not ICD-0-3 would be included for the years they were reportable and vice versa. In addition, this would mean that benign brain tumors would be submitted for 2004 as they were reportable in that year. They would not be published in CINA, but they would be included in the CINA Deluxe research database.
  - c. The Call for Data 2007 submission is requesting that the output from the Great Circle Distance Program be included on the file. The Great Circle Distance Program is a SAS program that works with ArcInfo or MapQuest and was developed by the GIS Committee. The GCD from a patient residence to up to five hospital facilities can be calculated using this program. We would like to do some pilot studies with these data next year. The hospital facility information is proprietary and thus is not part of the program. This information must be inputted from the ArcInfo or MapQuest software, and the file contains the necessary latitude and longitude information used to compute distances. We have also been aware that without standards for order of multiple records on a patient, or a standard to associate a particular treatment with a particular facility, that researchers may not be able to identify these specific distances. However, we are aware of a myriad of health services questions that can be addressed using a tool that is less crude than a popular one of using concentric circles to measure distance.
  - d. NAACCR has submitted a proposal to the Komen Foundation with the assistance of the GIS Committee to develop a shortest route finder program that uses road networks. The program would compute network travel distance and time. NAACCR will find out next May if the three year project has been approved. Once developed, it will replace the GCD. We believe that this tool will have many uses for persons wanting to do health services research using registry data.
  - e. Committees and Work Groups have been working on enhancing the data that we have on race and ethnicity.
    - i. The NAACCR Hispanic/Latino Identification Algorithm v2 (NHIA v2) was released in 2005. Changes were made to the algorithm because it did not work well with data from states with a small Hispanic/Latino population. NHIA v2 was used in a feature of the Annual Report to the Nation which was released on September 6, 2006. The report is being published in *on* October 15, 2006. 90% of the US Hispanic/Latino population was used in the feature.
    - ii. Next year's feature in the Annual Report to the Nation will be on American Indian/Alaska Native populations. The I\_HS Linkage variable has been included in this year's Call for Data so the NAACCR data set will contain the information needed for the Annual Report to the Nation. NAACCR is asking registries to submit the I\_HS Linkage variable data that are available. We are aware that the US national programs are working with their affiliate registries about program requirements for conducting the linkage. NAACCR is not requiring linkage, but only that the linkage information be included in the NAACCR submission.
    - iii. NAACCR Asian/Pacific Islander Identification Algorithm (NAPIIA) is being completed for the six major Asian groups. The focus is on getting Asian, NOS cases into a more specific Asian group. The subcommittee responsible for the development is moving into a testing phase in registries with small

Asian/Pacific Islander populations. If you are interesting in doing testing please contact Frank Boscoe in New York or Holly Howe at NAACCR. The subcommittee is also looking at whether the algorithm can be expanded to include some Pacific Islander groups. This will be completed concurrent to the testing component of the Asian piece. The output from NAPIIA will be requested in next year's Call for Data.

- f. For the annual report to the nation, the NAACCR data were used to look at short term trends (fixed interval trends from 1995-2003). The data were from all registries that met high quality criteria, which covers about 70% of the US. When we presented rates for 1999-2003, the period represented 83% of the US population with the Latino section including 90% of the US Hispanic/Latino population. This was the first time that a substantial part of the report of the nation relied on data from NAACCR submissions. The authors remarked that it was a major achievement that for all but long-term trends, data – and high quality data – were available for describing cancer burden from the vast majority of the US.
- g. We have had requests that members want to be able to use NAACCR data in an electronic format to compare state rates with national/NAACCR or regional rates. This would not fall within a research request, but rather more like availability of currently public data in a format on which statistical testing could be easily done. We are developing two different electronic outputs of CINA data. A SAS dataset of rates, counts and standard errors from CINA will be available for download on the NAACCR website. This information is already public, so confidentiality agreements would not necessary. The second format would be to have a discretionary file in SEER\*Stat that would need to be requested and then made available through a client server environment that would include the information that is currently available on CINA+ Online, but also enable comparative analyses including standard errors and confidence intervals. CINA+ Online data enable age-specific and single year queries. It will suppress submission of small numbers. It will provide standard error and confidence interval, while preventing time outs and access limits experienced in CINA Plus Online. Researcher agreements, usernames, and passwords will be required for access.
- h. IMS is writing a SAS program for registries to run that would link the percent poverty for census tracts with the census tract geocodes for registry cases. The output would give registries the census tract percent poverty and it will also aggregate those percentages into four major poverty groups. This program will be available for registries to download from the NAACCR website. We would also like to include a request for the four-group output in the next Call for Data. If registries see confidentiality issues, please let Holly know.
- i. CINA Plus Online's dataset will be moved to a new software structure. The technology for the current query system is not current technology and we are facing increasing problems with operations. It is consistently the 3<sup>rd</sup> most frequent page hit on the NAACCR website. We have plans to move it to the query system developed by the University of Kentucky. We anticipate that the new site will be available in January 2007.

## **2. More Quality Assurance for Users**

- a. Registries are not looking at duplicates across state borders and we may be duplicating cases in national cancer rates. We have some observed higher rates in aggregated areas outside the SEER program than those reported in the SEER program, and one could wonder whether duplicates could be contributing. In California, a recent comparison of duplicates among their regions uncovered a duplicate rate of 8.7%. NAACCR is looking into the feasibility of providing a matching service to registries that would enable them to identify potential duplicates on their file and another registry's and also could identify instances of missed cases (those that are on another registry's file as a resident of the state in which your registry has jurisdiction). This project will be discussed more in the next year as feasibility issues are determined.
- b. The four treatment variables that were submitted during last year's Call for Data are being evaluated.
- c. DECC is looking at a queryable database of the specific completeness of data variables that would reside on the NAACCR website and that researchers could use to assess feasibility of specific research projects.

### 3. Member Discussion

Questions from Holly:

- a. A request from NCI was received that NAACCR begin to archive the data submissions so that they could be used for research purposes, such as delay-adjustment in reporting. Our agreement with registries is that we destroy the data after two years. What do registries think about NAACCR changing this policy?
- b. How would registries feel about this data file being available to non-NAACCR members? They would still go through the review processes, etc.

Questions from membership:

Why would NAACCR archive old data when the registry is already submitting cleaner cumulative data each year? NCI's request mentioned a primary use to estimate delay-adjustment factors on for registries, so delay adjusted trends would be available for all and not just the SEER program.

When using the Great Distance Circle program, will registries need to have latitude and longitude for facilities that have closed, opened, or moved? The Shortest Network Route Finder program has roads starting in 2001, and thus cases and facilities will be able to be matched, even if they have subsequently closed. Other than that if the facility is not in existence any more, then there will be a null value and distance will not be able to be computed.

Will NAACCR estimate the time and money needed to use this program? It takes more time to prepare the data file than to run the program, but it is too early to say how long it would take.

What is the current process for NAACCR researchers to access the data file? And what would the future process be for external researchers to access the data file? Notices are sent through the NAACCR Listserv about quarterly deadlines for CINA Deluxe applications. A proposal must be written and reviewed by a peer review panel before going on to the NAACCR IRB. Consent forms are sent to registries for their data to be

included in the specific project. Access is provided to the researcher. At this point in time, access to the file is not made available to non-member researchers.

### **Closing Remarks**

Holly and Board members would like to hear from registries their opinion, support, or concerns about the possibility of allowing non-member researchers to use NAACCR data and archiving NAACCR data submissions.