

**NAACCR Town Meeting
Registry Operation and Data Tools
December 8, 2006**

1:00 pm Eastern; 12:00 pm Central; 11:00 am Mountain; 10:00 am Pacific

Present – Representatives from the following registries and organizations:

27 U.S.: Alabama Statewide Cancer Registry, Arizona Cancer Registry, California Cancer Registry, Los Angeles Cancer Surveillance Program, Northern California Cancer Center, Centers for Disease Control and Prevention, Colorado Central Cancer Registry, Delaware Cancer Registry, Georgia Comprehensive Cancer Registry, Cancer Data Registry of Idaho, Illinois State Cancer Registry, Indiana State Cancer Registry, Information Management Services, Inc., Louisiana Tumor Registry, Missouri Cancer Registry, National Cancer Institute, National Cancer Registrars Association, New Jersey State Cancer Registry, New York State Cancer Registry, North Carolina Central Cancer Registry, North Dakota Cancer Registry, Oregon State Cancer Registry, Pennsylvania Cancer Registry, Texas Cancer Registry, Utah Cancer Registry, Virginia Cancer Registry, Washington State Cancer Registry

3 NAACCR Staff: Dr. Holly Howe, Moderator – Executive Director, Royale Anne Hinds – Assistant to the Executive Director, and Shannon Vann – Program Manager of Education and Training

1. Introduction and Background **Holly Howe**

The NAACCR Board wants to explore a new activity tool, the NAACCR Interstate Duplicate Case Identification System, and gauge the interest of the membership. We will also discuss updates on data tools that were discussed in September of 2006.

2. NAACCR Interstate Duplicate Case Identification System **Holly Howe**

The potential and impact of incidence cases that are registered in more than two registries has been discussed for a long time. California has made huge efforts to eliminate duplication within their regions from their systems. Even after this substantial effort, they were surprised to find out they had an 8 percent duplication rate within their state [see amended statement at the end of this section]. There are rates in the NAACCR data which are higher than the SEER rates. SEER registries don't share borders within their regions like the NAACCR registries do. The NAACCR Board has worked with IMS to develop a protocol for a new service that NAACCR would like to offer to members. NAACCR needs to find out if there is enough interest to support the activity through revenue from members.

NAACCR would ask all registries to submit an identified confidential file to IMS. The data would be going back to the base year of 1995. Within two months, IMS would process the data and create a linkage program between the registries that submitted their data. The reports would be sent by IMS to NAACCR, who would forward the reports to the individual registries. The reports would tell the registries the following things:

1. The names that are appearing as a duplicate on your list and another registry's list. NAACCR would develop guidelines to assist the registries in determining which registry the case belongs to.

2. Identify by count, the number of cases that are reported as another state's resident, but did not appear on the resident state's file. If the states have a case sharing agreement they can determine who the case belongs to. If they do not have a case sharing agreement, they would need to develop one so they can share the missing cases.

NAACCR would need to create a leadership group to determine the linkage criteria and decide how to resolve the duplicates and determine in which registry they belong. NAACCR would like to develop and implement this system in 2007. A certain number of registries are needed to cover the development costs. There would be a higher fee the first year of use. Subsequent years would have a reduced fee of approximately 75 percent of the first year's fee. Each year, the data would be submitted from 1995 forward and would reside at IMS for approximately two months. All confidential data would be destroyed once the reports are generated. NAACCR is willing to work with the states to get IRB approval if needed. The cost would be dependent upon the size of the registry.

There is a description of the program and survey on the NAACCR website for members to indicate their interest level in the NAACCR Interstate Duplicate Case Identification System. See attached.

An amendment to comments made about California duplicates follows:

Holly Howe mentioned Lilia's presentation where we found 8% Duplicates. Upon clarification by Lilia's staff, they knew about most of those duplicates (and the states will know about most of them too) - as these were our regional case-share cases. When the regions put their data into Eureka they put in all of the records they had, including those that they had already case-shared with other regions. I believe that most of the 8% were these case-share cases based on some linkages that were done before switching to Eureka. After determining what was known and not previously known to each regional registry, it was found that each region had only 0.5-0.9% possible duplicates with another region.

3. Group discussion of Proposed System

Open Forum

Brenda Edwards asked:

- Will the output of cases from the tool go back to all registries involved? Yes, the cases would be identified by name and the resident registry would be identified.
- Will the summary table with overall counts be broken down by cancer type? At this point no, because of confidentiality. The registries would only be informed of which registry has the case and the registries would work it out between themselves. A summary report would come to NAACCR for the purpose of assessing the impact on rates.
- Is this tool already developed or to be developed? The tool is not developed yet. We don't have dollars to develop it, which is why we need to know the membership's interest in the tool.
- Do you have a study design or strategy that will incorporate the performance and validation of this tool? IMS and the Board members that worked on this project say that we have very good matching programs. The linkage and match criteria would be developed by a small group. The protocol would be available after the first of the year.

Melanie Williams from Texas asked:

- Will the registries be sharing the same person or the same tumor? The tumor is the case. In the linkage process they would identify the person that is being reported in two states. The registries would need to determine in which state that incidence case belongs. Whether a person matches or the tumor would need to be determined by the advisory group. It could be either or both.

Jeannette Jackson-Thompson from Missouri commented that their state is concerned about using this system even if it were free because of strict confidentiality issues. They do not want confidential data going to a non-government agency. Holly agreed that each registry would need to decide if this is worth spending their resources on this project. The end result is a better data set which most registries would want. Once the process takes place if the results are showing that there are not many duplicates then the process can be put to rest, but there could be a substantial amount of duplication.

Stacey Carson from Idaho is interested in the service and asked:

- Can the registry decide once they know who else will participate because if their surrounding states aren't going to participate, it would not be very effective? Holly will let everyone know which states want to participate by the results from the survey.

Jan Snodgrass from Illinois is concerned about some state's ability to reconcile duplicates. Many states do not clean up their duplicates and sometimes files come in from other states and Illinois finds duplicates within the other state's file. Also, there is a difference in the application of multiple primary rules which may impact the rates as well.

Rosemary Dibble from Utah asked:

- If this will be valid if some registries participate and others do not. Won't it reduce the rates in some states, but not in others? Holly replied it is too early to know the impact. It is possible that in the future, NAACCR would require de-duplication to be included in the combined rates.

Betsy Kohler from New Jersey said their registry is always surprised at the number of summer residents of New Jersey that reside in Florida during the winter and they expect are counted in both areas.

Hannah Weir from CC commented that the NPCR - NDI application allows registries to contact the state in which their resident died after the link is made. The two registries would determine whether it was a duplicate case or not. Not all registries have applied to the NPCR – NDI, so Hannah encouraged registries to proceed with NAACCR's process.

Vivien Chen from Louisiana is interested in the system. This could be a big step for the surveillance community to produce accurate data. IMS could guarantee the confidentiality and integrity of the data not being compromised.

- Would each state need to submit their out of state cases along with their resident cases? Yes.

A participant asked:

- Is it would be possible to get CDC Funding through grants? Hannah said that CDC would like to extend an opportunity to work with NAACCR on this. CDC has a certificate of confidentiality that covers the personal identifier.

Holly explained that the idea of the system was generated from a sub-group of the NAACCR Board. It would be a service that NAACCR would like to offer to its members and a way for NAACCR to generate revenue to build financial stability. Further, they felt it may be a good time because many registries are in the fifth year of NPCR Cooperative Agreements and may have unexpended dollars that could be put toward this service.

4. Survey of Registry Interest in the System **Holly Howe**

The survey was made be available on the NAACCR website shortly after the meeting.

5. Revisiting Other Data Tools **Holly Howe**

Updates on other data tools that were mentioned at the September Town Meeting:

- a. Archive 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? NAACCR has had an offer from Rocky Feuer at NCI to develop delay adjustment models for the nation. Currently, there is only one for the SEER program. We are beginning to use time trends more often and feel that it makes the interpretation of the trends more correct. NAACCR would need to change the agreement with registries so that we could archive the data rather than destroying it one year after the Call for Data is completed.

Comments and questions include:

Jim Martin from Virginia thinks the idea is interesting and could be very revealing and asked:

- What changes may be necessary? Holly responded that NAACCR does not believe this is a confidential file, although different state laws define it differently.

Illinois would not have a problem with the archive of data submissions if the county identifier is not archived. They asked:

- Will the county information be archived? Brenda responded that it is used in special projects, but needed to check with staff if it is used in delay adjustment. Holly said the county identifier could be stripped from the archive file.

Melanie Williams in Texas would be willing to have state-level data archived, but not the county identifier.

Washington State feels this is a worthwhile project and would be comfortable down to the county level.

Holly asked that people email her with comments or questions. She will look into the possibility of excluding the county identifier so that more registries could participate.

- b. How would registries feel about the CINA Deluxe data file being made available to non-NAACCR members?

Kurt Weiss from Nebraska asked:

- Is there is a fee charged to the non-NAACCR researchers? Yes, there would be a charge to them. Nebraska is in favor of doing this.

Illinois asked:

- Would the non-NAACCR researchers sign the same agreements that the registries sign? Holly responded that there would not be a change in the process. It would just be opening the data to other uses. Illinois would be in favor of doing this as long as it was not a county-level release. Holly pointed out that registries have the ability to say “No” during the consent process.

Brenda talked about the importance of county level data being used for research. The county level data are relevant and extremely useful as part of a calculation in an analysis even though the county data are not published.

Mel Lehnherr from Illinois asked:

- Was the county level data being used for the time delay adjustment? Brenda does not believe that it is being used for the time delay adjustment. As far as Brenda knows the county level data are not being published by ACS for Facts & Figures.

Vivien Chen from Louisiana asked:

- Are there criteria for the non-NAACCR researchers to be given access to the data file? Holly reported that a proposal would have to be submitted to be reviewed and approved by the Research Proposal Review Subcommittee. The subcommittee will need to decide if additional criteria should be required to secure that they are a capable researcher. There are some requirements in place currently. NAACCR will need to have written guidelines so that all researchers and data requestors would be treated the same.

- c. The data from CINA Volumes I through VI are available in SAS Format. The file can be downloaded from the NAACCR website.
- d. CINA + Online available in SEER*Stat Format. This format would be 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 errors and confidence intervals. Researcher agreements, usernames, and passwords will be required for access.

Kurt Weiss from Nebraska asked:

- Can registries opt-in or opt-out of the SEER*Stat version? Maria Schymura and Kurt think the Confidentiality Subcommittee and the DURC decided that registries should have the opportunity to opt-out of the SEER*Stat version because there are differences in the data. Based on this decision, registries will need to consent to having their data included.
- e. A SAS program available on the NAACCR website that links the percent poverty for census tracts with the census tract geocodes for registry cases. This program is now available for registries to download from the NAACCR website. The output will give registries the percent poverty in the census tract and it will also aggregate those percentages into four major poverty groups. We will be including the grouped output

variable in the 2008 Call for Data. It can be run now to work through any bugs that may come up later. Email Holly at hhowe@naaccr.org if you have trouble locating the program on the NAACCR website.

6. Closing Remarks

Holly Howe

Please complete the survey on the NAACCR website about interest in the NAACCR Interstate Duplicate Case Identification System. There will be another NAACCR Committee Town Meeting on December 21, 2006 on the committee activities of the Registry Operations and Education Committees.

NAACCR Interstate Duplicate Case Identification System

Definitions

- "Primary Registry" designates the registry for which the reports are being generated.
- "Matching Registries" designates all other registries included in the analysis

Submission of Data from Registries

- Registries will submit their data in NAACCR format
- Registries will submit a data file for all diagnosis years, 1995-2004, including all cases not within their catchment area. The file will include only the variables required to determine matches (e.g., name, social security number, date of birth, etc)
- Registries will submit their data for analysis at a designated time. The schedule for this submission will be set by NAACCR to allow sufficient time for the processing of the data, generation of the registry reports, and resolution of items identified in the analyses by the registries prior to the annual data submission for CINA.

Data Analysis

- One analysis will be generated:
 - A. *Duplicate Report*: List of cases in the Primary Registry which designates exact, possible, and probable matches in data from the Matching Registries

Methodology and Reporting – Duplicate Report

- The data from the Primary Registry will be matched against all other Matching Registries
- Cases from the Matching Registries data files that are possible duplicates of cases within the Primary Registry data file will be identified.
- A case report will be prepared of cases that had a match with another state. The report will designate the case by match category (i.e. exact, possible) and the state(s)/registry that caused a match.
- A Summary Report (including a data file for further analysis) of the number of duplicate cases by the three categories will be generated. The report will also include other relevant summary statistics, such as total number of cases, appropriate percentages, categories of largest and least duplication (by age, cancer site, etc) The Summary Report will be stratified by Diagnosis Year, Matching Registry, Sex, Primary Site, etc.

Definition of a Match

- The definition of a match will be determined by an advisory panel.

Data Retention

- Data from registries will be retained at IMS in a secure location only for the duration of the analysis. All data will be destroyed following the completion of analysis and no backup copies will be retained.
- NAACCR and the Registry will be notified of the data destruction
- Summary reports of the analysis results will be maintained for analytic purposes.

Data Security

- Data will be maintained in a secure location on the IMS network
- No IMS staff will be allowed access to the data without prior authorization

Data Reporting and Usage

- The summary reports will be forwarded to NAACCR for review and distribution to the registry
- The NAACCR office will forward the results to each participating registry
- The NAACCR office will retain the results for the purpose of producing a national summary report for duplicates. Additional reports may be written, such as a report on the burden of cancer.

Data Sharing Agreements

- A data sharing agreement will be developed for this protocol based on similar agreements already established.

Fees for Analysis

- Two levels of fees will be utilized. An *initiation* fee for the first year of participation by a registry, and a lower *maintenance* fee for subsequent years of participation by a registry.
- The initiation fees charged per registry will be set such that the development costs of the system will be covered, based on an anticipated number of registries participating in the first year of service. Initiation fees collected in the following years of service will be used to enhance the system in areas to be defined later.
- The maintenance fees charged per registry will be set such that the processing and administrative costs of the analysis will be covered, based on an anticipated number of registries participating.

Future Enhancements

- Possible future enhancements for the system include analyses related to sequence numbers and multiple primaries.

NEEDS ASSESSMENT FOR THE
NAACCR INTERSTATE CASE DUPLICATION IDENTIFICATION PROGRAM

Name: _____
Registry: _____
Country: _____

A summary description of the program is attached. The details will need to be worked out with an advisory committee, and this will be done in early 2007. If you have any questions about the program, please contact Holly Howe at 217.698.0800 ext 2 or at hhowe@naaccr.org.

1. Would you be interested in participating in this new program? Yes no
2. As you understand the program today, would you be able to submit a file with case identifiers to participate in the program?
No – under no circumstances
Yes – under any circumstances
Yes, but only if _____

NAACCR will need to charge a fee for this service. Although the scale is not yet finalized, it will be higher in the first year of participation (regardless of when you actually start participating) and then a lower maintenance annual fee. The fee will depend on the size of the registry and we estimate that it will be:

Large registries (60,000 cases+) \$6000 year one; \$4500 annually thereafter
Medium registries (20,000-59,999 cases) \$4000 year one; \$3000 annually thereafter
Small registries (<20000 cases) \$2000 year one; \$1500 thereafter

3. Would you be willing to pay this fee for the service?
Yes no Not applicable (can't participate)
4. If you are interested, would you be able to participate in a May 2007 due date for file submissions?
Yes
No
If not, what is the first year that you could participate? _____
Not applicable

Thanks for your response!