

## REPORT OF THE AUTOMATED TUMOR LINKAGE WORK GROUP OF ROC, OCTOBER 2005

### Automated Tumor Linkage Work Group Roster

Mary Jane King, MPH CTR, Co-Chair, Massachusetts Cancer Registry  
Jennifer Seiffert, MLIS, CTR, Co-Chair, TRW & RegistryPlus Software  
Peggy Adamo, RHIT, CTR, NCI/SEER  
Deborah Bringman, MPH, University of California, Irvine  
Susan J. Capron, Chicago, IL 60647  
Diane Kirsch, RHIA, CTR, Pennsylvania Cancer Registry  
Gary Levin, Florida Cancer Data System  
Kamini Milnes, Cancer Care Ontario  
Jerri Linn Phillips, MA, CTR, American College of Surgeons  
Wendy Scharber, RHIT, CTR, Minnesota Cancer Surveillance System  
Colleen Sherman, RHIA, CTR, NYS Cancer Registry  
Judy Williams, RHIT, CTR, IMPAC Medical Systems, Inc.

The Record Consolidation Subcommittee of ROC began as an ad hoc committee to investigate central cancer registry strategies for patient and tumor record linkage and consolidation. The goal of the subcommittee has always been to discover and make available accurate, reliable strategies to conduct the linkage and consolidation processes. Its first published report<sup>1</sup> presented a survey of several central registries and their linkage and consolidation strategies, as well as the results from the first tumor record linkage and consolidation test ever conducted by NAACCR. The report also established a consistent set of terms and definitions for these processes and for cancer data reports that are still in use today.

The subcommittee's second report<sup>2</sup> was a feasibility study preliminary to creating and conducting a larger, more complex test to measure the impact of different methods on the consistency and accuracy of determining the number of patients and tumors.

A new, larger test was created and is currently available on the NAACCR web site in an ICD-O-2 version and an ICD-O-3 version. The subcommittee's third report included the findings of this Consolidation Test.<sup>3</sup>

Presently, the Record Consolidation Subcommittee is carrying forward its work on automated tumor linkage through the Automated Tumor Linkage Work Group of ROC

---

<sup>1</sup> NAACCR, *Report of the Record Consolidation Committee, 1999*, Preface and Appendices A-D, [http://www.naacr.org/index.asp?Col\\_SectionKey=7&Col\\_ContentID=33](http://www.naacr.org/index.asp?Col_SectionKey=7&Col_ContentID=33)

<sup>2</sup> NAACCR, *Report of the Record Consolidation Committee to the NAACCR Board of Directors, March 1, 2000*, [http://www.naacr.org/index.asp?Col\\_SectionKey=7&Col\\_ContentID=33](http://www.naacr.org/index.asp?Col_SectionKey=7&Col_ContentID=33)

<sup>3</sup> NAACCR, *Report of the Record Consolidation Committee, 2003*. (Braun, JE, Ross FE, Capron SJ, Creation of a Record Consolidation Test File: Report to the NAACCR Board, December 2003) [http://www.naacr.org/index.asp?Col\\_SectionKey=7&Col\\_ContentID=33](http://www.naacr.org/index.asp?Col_SectionKey=7&Col_ContentID=33)

## **Objective**

This past year the Tumor Linkage Work Group set itself the objective of developing best practices for automating the determination of which records represent the same tumor using the SEER multiple primary rules as the standard. The results of this project are intended to serve to facilitate automation of the tumor linkage process by central cancer registries.

## **Methods**

The group sought to identify what tumor linkage routines existed in registries, particularly, automated linkage systems. Seven presentations were solicited by the group from various sources, mainly population-based registries, but also a specialty registry, and a software vendor. Five of these gave presentations to the group. The process, while informal, included written documentation and an extensive question and answer period via teleconferences with the system representatives. The representatives were drawn from the full spectrum of automated cancer registry management. The systems were selected by convenience sampling, but after thorough networking, the group believed that it covered the scope of automated systems. Four of the systems presented used automated tumor linkage routines; one linked cases manually.

The group used the information provided for each system to empirically create a list of criteria by which to compare the way in which each system handled tumor linkage. Importantly, this evaluation included review of site-pair and histology linking logic.

## **Tumor Linkage Systems Identified**

The following tumor linkage systems were identified. They are shown in the order that corresponds to the companion document, the “Tumor Linkage Comparison Chart”.

- Minnesota’s MN-PATRL
- Pennsylvania’s and CDC’s TLC Plus (Registry Plus)
- Florida’s FCDS
- National Cancer Data Base (NCDB)
- SEER DMS (not presented, no automated linkage at this time)
- California’s Eureka (no automated tumor linkage at this time)
- IMPAC Precis/Central (not presented).

## **Results**

### **1. Comparison Criteria and Notable Individual System Comparisons**

Comparison criteria were developed as the five systems were being presented, as it became clear what characteristics affected their linkage procedures. These included which site and histology groups were considered the same, what time periods (for synchronous tumors) were considered the same, what flexibility the system had, how portable the system was, the registry’s size, and reporting sources. Twelve criteria were decided upon and appear here along with some specific examples. Interim results of this comparison were presented at the NAACCR annual meeting in 2005. A complete list of system and registry comparisons based on these criteria is available in the Tumor Linkage Comparison Chart handed out at NAACCR 05. **This chart is available now on the NAACCR web site as [ATL\\_Registry\\_Comparison.xls](#).**

## **Criteria**

Criterion 1. Characteristics of the central registry using the system may explain the specific needs that the linkage logic was designed to address. For example, (a) the average total number of records received by the system annually, (b) the average total number of consolidated tumors (cases) annually, (c) the types of reporting sources and proportion of cases reported from each source type, (d) whether both analytic and non-analytic cases are reported.

Criterion 2. The proportion of tumors having more than one report can be handled automatically by the system.

Criterion 3. Which sites are considered the same, and how that is encoded. Minnesota, Pennsylvania, Florida and the NCDB use a site-pairs table. Pennsylvania also includes some “equivalences” outside of their table. In addition to evaluating location/site of the tumor, an automated tumor linkage system needs to include evaluation of diagnosis date, laterality and histology before determining whether an incoming record represents a new primary. The site-pairs table lists pairs of ICD-O-3 topography codes that can be considered the same in an automated tumor linkage system. Additional data items, such as diagnostic confirmation, number of lesions or facility identification number, may also be useful in the decision-making process.

Criterion 4. Which histologies are considered the same and how that is encoded.

Criterion 5. The time period used to define the same versus separate cancers, and whether that unit is measured in days or months. Is the interval for comparison within one facility different from that used for comparisons among different facilities? Florida does not use a timing rule, Minnesota and Pennsylvania consider greater than one year as automatically separate primaries, and Pennsylvania considers two months or less to automatically be the same primary. Both Minnesota and Pennsylvania review other time intervals manually. The NCDB considers greater than 365 days to automatically be separate primaries, otherwise NCDB considers them to be the same.

Criterion 6. Determine whether the system compares an incoming record to a consolidated record or to the individual source records/abstracts. Minnesota and Pennsylvania compare incoming records against consolidated tumor record. Florida and the NCDB compare incoming records against unduplicated source records.

Criterion 7. The point in the process at which problem cases are resolved (when does manual review occur). Does this happen before final disposition of each incoming record or later, in a “clean-up” process. The Minnesota system sets automatic flags, which are cleared as further admissions arrive and resolve the problem. The cases are added to the incidence database with these flags up. As the year is closed, any remaining flags are reviewed and cleared. The

Pennsylvania system resolves questionable cases during linkage, before final disposition into the incidence database. Florida and the NCDB do not manually review cases.

Criterion 8. Determine how portable each system was. Characteristics that affect portability include information/variables structured in tables, programming language, standardized definitions and logic, modular components, amount/quality of documentation, open source/free programming, widely-used platform/operating system. Minnesota's tables are fully portable. Pennsylvania and Florida have (e.g., site, histology) tables that are potentially portable. Minnesota and Pennsylvania use SEER rules. No system presented was fully portable. One system was planned to be fully portable by 2007.

Criterion 9. Determine how flexible or customizable each system is, depending largely on characteristics elaborated in item 8 (above). Minnesota, Pennsylvania and Florida are self-described as very flexible and customizable. Pennsylvania has user-interfaces for changes. Minnesota and Florida systems can only be changed by programmers.

Criterion 10. Determine if each system provides metrics on its behavior and results. Minnesota provides information on the confidence level of the tumor linkage result; Pennsylvania and Florida systems also provide various metrics. Criterion 11. Determine if the owner registry performs audits of the automated system's results and decisions. Minnesota, Florida and the NCDB perform independent audits of automated results. Pennsylvania does not because they review questionable cases before final disposition.

Criterion 12. Identify what information the owner registry has on the acceptability of its automated tumor linkage to CTR users or to consumers of the results. In Minnesota, it took four years for the CTRs to trust the established system; a comparison of manual tumor linkage results by the CTRS with the automated results validated the system's accuracy and modified the mindset. Currently, the CTR's and user-epidemiologists trust the results and availability of information regarding system performance. CTR staff in Pennsylvania trust their system's results. After implementation, they saw and appreciated the significant reduction in time needed to perform annual file cleanup. Florida reported that staff and field CTRs, the Florida Department of Health, and researchers fully trust FCDS system tumor linkages. California is not currently using an automated linkage system for new cases due to their quality assurance staff preferring manual decisions. California is planning to initiate an "expert systems" project that will create linkage strategies to use for new cases.

## **2. Site Pairs Table**

Regarding Criterion 3, a Site-Pairs Table was adopted based on the table used by Minnesota, but including sites paired by all five of the systems presents. **This chart is available now on the NAACCR web site as [ATL\\_SitePairs\\_Table.xls](#).** This table is color

coded to indicate degree of agreement on any particular pair, for example, pairs that are considered matches by all five are blue, and pairs that are considered matches by only Minnesota and Florida are pink. This table is annotated with comments explaining the logic of individual differences.

Additionally each pair was reviewed by the group to achieve consensus on which recommended pairs should be considered the same. Agreement was based on conformity with current SEER multiple primary rules, and/or the predictability of the pair resolving into a match agreeing with SEER rules. The degree of consensus is expressed as three groupings:

Group 1: pairs that correspond to the written SEER rules for same site.

Group 2: pairs that are a reasonable extension of the SEER rules, because they represent a specific and general category [in ICD-O3] but differ at the 2 or 3 character level. Example: Gum, NOS C039 and Mouth NOS C069.

Group 3: pairs that experience has shown are likely the same but may be handled differently in different registries because of different coding practices, different mix of data sources included in the registries, differing levels of automation, and differing levels of review of automated decisions. Example: Transverse Colon, C184, and Back/Flank/Trunk, NOS, C767.

Group 4: Pairs that meet the written SEER rules for same site unless histology = 8720-8790.

Group S\*: The workgroup did not consider these to be a routine match. The "S\*" pairs are used in Minnesota to ensure sarcomas are considered the same primary tumor, even if registrars miss-assign the site. These pairs are critical if site is the first data item processed by automated tumor linkage.

The committee believes that the site pairs table as presented represents the consensus of all automated tumor linkage systems studied. The table has been prepared in a way to facilitate its incorporation into other automated systems that might be developed. [A third document assists with the site-pairs table: ATL\\_SitePairs\\_ExecutiveSummary.doc. This document is also available on the NAACCR web site](#)

### **3. Histology Pairs Table**

Since the multiple primary and histology rules are due to change in 2007, the committee decided against expending time creating an accompanying histology table. While the site pairs table will remain useful, the histology table would not. Instead the committee is currently devoting its time to the creation of a histology table using the new rules.

### **Conclusions**

Tumor linkage is the process of using defined criteria to determine whether source records for the same patient refer to the same tumor. In a central registry, automated

tumor linkage can greatly reduce the amount of human intervention required in the record linkage process. Common to each automated linkage system reviewed is a mechanism for comparing specific data items (diagnosis date, primary site laterality, histology, and behavior at a minimum) and, based upon the degree of agreement between reported values, allowing the computer to complete routine tumor linkage decisions. Although registry characteristics and data management philosophies have influenced the manner in which automated systems were implemented, this Work Group advocates the development of accepted definitions for which values should be considered the same in all automated tumor linkage systems. The Site Pairs Table is the first product toward this goal.

### **Future Plans**

During calendar year 2006, the group will work on histology groupings tables (analogous to site-pairs table) as final new histology documents become available. The group will then evaluate the feasibility of, and propose approaches to the development a fully portable tumor linkage module.