Six Degrees of Separation No More:
Using Data Linkages to Improve the Quality of Cancer Registry and Study Data

David Harris
RTI Health Solutions, Research Triangle Park, NC, United States

ABSTRACT
Background: A data linkage is a process commonly used to enhance the quality of cancer registry data and one database also resides in a second database. There are two general types of linkages: deterministic (name-based) and probabilistic (statistical). Specific linkages are used to perform the linkages. For those cancer registries that don’t have a data linkage program, the Centers for Disease Control and Prevention (CDC) offers Link Plus free on their Web site.

OBJECTIVE: To explore the variety of reasons to link a database to a cancer registry file. The presentation will also illustrate the value of data linkages in increasing the quality of cancer registry and study data.

METHODS: The stated objectives will be achieved by offering real-life examples of the value of linking population-based cancer registry databases with other sources. Potential examples include linking a study cohort to a cancer registry database to determine cancer diagnoses and burden; assisting the cohort; using the linkage process to update the vital status and date of last contact for patients in the cancer registry or other database, and linking study data to cancer control and prevention programs; and using linkages for stingy surveillance studies.

RESULTS: The presenters will include results from data linkages between cancer registry files and other files, including linkages with public health surveillance data and other databases with cancer control data to evaluate program effectiveness, and with other databases to determine cancer burden in specific populations.

CONCLUSIONS: If used properly, data linkages can be an effective tool in increasing the quality of a cancer registry database. Linkages can lead to a better understanding of cancer burden in their study cohorts, help to determine if cancer screening efforts are effective, and allow cancer registry data to be used in novel ways.

EXAMPLES FROM STUDIES WHERE DATA LINKAGE WAS USED TO IMPROVE CANCER REGISTRY AND STUDY DATA

Using Data Linkages to Improve the Quality of Cancer Registry Data

• State cancer registries are graded on certified or national completeness, and current their data are.

• For example, the Surveillance, Epidemiology, and End Results (SEER) program sets criteria for following rules and categorizes the rates as unacceptable, acceptable, or stage for specific age categories.

• Thus the results for the SEER 9, 15, and 21 cancer registries from 1998-2000 and from 2000-2002 from the study performed in October 2008 (Example 5).

• Linkages with other sources can improve state of last contact and provide updated patient contact information.

• Accurate vital status and date of last contact are important for survival analysis and mortality rates.

• Linkages against own database are useful for identifying and removing duplicate cases.

Table 1: 2008 Data Set for Percentage of American Indians With Updated Follow-Up Data for American Indian- and Alaska Native-Identified Cancers 1998-2000 and Followed-up 2000-2002

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Total</th>
<th>2000-2002</th>
<th>Acceptable</th>
<th>Unacceptable</th>
</tr>
</thead>
<tbody>
<tr>
<td>20-64 years</td>
<td>100%</td>
<td>97%</td>
<td>97%</td>
<td>3%</td>
</tr>
<tr>
<td>65+ years</td>
<td>100%</td>
<td>99%</td>
<td>99%</td>
<td>3%</td>
</tr>
</tbody>
</table>

Example 1: Breast Cancer Screening Program Impact on Stage of Diagnosis

Data from a breast cancer screening program for undetermined women were linked to data from the California Cancer Registry to analyze stage at diagnosis for breast cancers during stage update (September 2004). (4)

• Overall, women in the program were diagnosed at a stage earlier than nonparticipants (Figure 1).

• However, the varying results from the linkage (data from both databases), where breast and mammogram and diagnosis were taken into consideration, the following results were found:

• Women diagnosed immediately after receiving a mammogram were diagnosed at statistically significantly longer stage from nonparticipants, driven the overall results.

• Women diagnosed were diagnosed one week later with both of the same rate as nonparticipants, a success for this program.

Figure 1. Stage at Diagnosis Among Women Receiving Mammograms Through a Screening Program for Underserved Women

Using Data Linkages to Improve Cancer Registry Data

• Investigators can link any type of study cohort or patient registry to one or more cancer registries to determine the incidence or burden of cancer among the cohort or registry, including changes over time.

• The Food and Drug Administration (FDA) Amendments Act of 2007 gives the FDA additional authority to require pharmaceutical companies to perform cancer registry linkages.

Linking to state cancer registries is an effective way of tracking the incidence of cancer in a cohort of patients over an extended period of time.

Example 2: Linking Two Types of Registries to Investigate Topism of Breast

• A linkage performed between cases from the San Francisco AIDS registry and the California Cancer Registry showed that the use of highly active antiretroviral therapy (HAART) or other treatments reduced the risk of breast carcinoma, systemic non-Hodgkin lymphoma (SNHL), and chronic myeloid leukemia (CMML), with no significant impact on AIDS patients.

Example 3: Linking a Cohort of Medications Used to State-Cancer Registries

• A current FDA-mandated study is using data linkages between multiple registries and a cohort of medication users over an extended period of time to ensure that the incidence of a specific type of cancer is not significantly higher among this cohort.

• Early results from this long-term study demonstrate the feasibility of using a standardized data linkage algorithm to help ensure consistent linkage results among the numerous state cancer registries participating in the study.

• Results can also be used to identify cases where follow-up is required for safety reasons.

CONCLUSION

If used properly, data linkages can be an effective tool in increasing the quality of a cancer registry’s data, allow registries to have a better understanding of their study populations, help to determine if cancer screening efforts are effective, and allow cancer registry data to be used in novel ways.

REFERENCES
4. Women receiving mammograms on a regular basis were diagnosed with breast cancer at statistically significantly later stage than January 1, 2003

CONTACT INFORMATION
David Harris, MPH
RTI Health Solutions
200 Park Offices Drive
Louisville, KY 40207
Phone: 502-443-7200
Fax: 502-443-7201
E-mail: david.harris@rti.org

Published in: 2011 NACCR Annual Conference
June 10-15, 2011
Los Angeles, CA