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

Prepared for the 138th Annual Meeting of the American Public Health Association
Denver, CO • November 6-10, 2010
David Harris, MPH • RTI Health Solutions
David Harris

The following personal financial relationships with commercial interests relevant to this presentation existed during the past 12 months:

No relationships to disclose
Purpose of This Presentation

• **Provide a brief background on the following:**
  - State cancer registries
  - Data linkages

• **Describe reasons for using data linkages:**
  - For evaluating the effectiveness of cancer control and prevention programs
  - For improving the quality of cancer registry data
  - For research purposes:
    • Cohort studies
    • Registry studies
    • Postmarketing drug approval safety surveillance
Background: Cancer Registries in the United States

- Every state in the US has a central cancer registry
- Cancer reporting to state registries is mandatory
- One goal of population-based cancer registries in the US is to collect complete, timely, and high-quality data for cancer control and research efforts
Data linkages are used to determine if persons in one database also reside in a second database.

There are two general types of data linkages:

- **Deterministic** (rules-based) linkages require exact matches to establish variable agreement:
  - Strength: Less time- and labor-intensive than probabilistic linkages
  - Weakness: Require accurate, discriminating variables

- **Probabilistic** linkages use underlying statistical methods to identify possible matches:
  - Strength: More effective than deterministic linkages when variables are missing or incorrect
  - Weakness: Can be very labor-intensive to review possible matches; require specialized technical expertise to implement
Evaluating the Effectiveness of Cancer Control and Prevention Programs

- **Data from a cancer control and prevention program can be linked to a cancer registry to assist in evaluating the program**

- **Example:** Data from a breast cancer screening program for underserved women was linked to a state cancer registry to analyze stage at diagnosis for breast cancers vs. stage among nonparticipants¹
  - Using variables from both databases, the study found that the program was being used in two distinct ways:
    - Women suspected of having breast cancer were sent for a confirmatory mammogram (diagnosed at later stage)
    - Women were getting screened on a regular basis (diagnosed at earlier stage)

Example to Demonstrate How Data Linkage Can Be Used to Evaluate the Effectiveness of a Cancer Control Program

Breast cancer screening program participants overall more likely to be diagnosed at late stage vs. nonparticipants¹
Adjusted odds ratio (OR)
1.2 (1.1-1.3)

Diagnosis 0-1 month after screening
Adjusted OR 1.8 (1.6-2.1) (vs. nonparticipants)

Diagnosis 1-10 months after screening
Adjusted OR 0.8 (0.7-1.0) (vs. nonparticipants)

Diagnosis after history of regular mammograms
Adjusted OR 0.9 (0.6-1.1) (vs. nonparticipants)

Improving the Quality of Cancer Registry Data

- **State cancer registries are graded or certified on how complete, accurate, and current the data are**
  - For example, Surveillance, Epidemiology, and End Results (SEER) sets criteria for acceptable follow-up (date of last contact) rates for tumors for 3 age categories plus all ages combined

- **Linkages with other sources can improve date of last contact as well as provide updated patient contact information for IRB approved follow-up**

- **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**
An Example of How Data Linkages Can Improve Cancer Registry Data

• A study performed in October 2004 analyzed the effect of data linkages in improving date of last contact (follow-up rate) for the Cancer Registry of Greater California (CRGC) SEER Region¹

• The following sources, among others, were linked to the CRGC database to obtain an updated date of last contact:
  – Social Security Death Master File and National Death Index
  – Centers for Medicare and Medicaid Services (CMS) database
  – California Voter Registration
  – National Change of Address
  – California Department of Motor Vehicles

The study included cases diagnosed in 2000-2001 with a date of last contact in 2003

- As of October 1, 2004 (before the linkages), only 88% of all cases diagnosed in 2000-2001 contained a date of last contact into 2003—an unacceptable rate
- As of October 19, 2004 (after the linkages), the date of last contact into 2003 for all ages combined had increased to an acceptable rate of 95%

Research: Tracking Outcomes in Cohorts

- Investigators can link a study cohort to one or more state cancer registries to determine whether cohort members have been diagnosed with cancer.

- There is an increased interest in analyzing the association between cancer and environmental exposures, such as with the World Trade Center and Camp Lejeune.
  - Linkages with cohorts from these events can be used to track outcomes.
Research: Linking Different Types of Registries to Determine Burden

- **Other registries (e.g., HIV/AIDS, organ transplant) can link to cancer registries to determine the burden of cancer among individuals in their registries, including changes over time**

- **Researchers can use data from the linkage to investigate topics of interest specific to their registry population, such as:**
  - Has the incidence of Kaposi sarcoma and non-Hodgkin’s lymphoma (NHL) decreased over time in HIV-positive populations?
  - Do patients who have undergone a transplant have a higher incidence of solid organ tumors than the general population?
  - Do patients receiving AIDS medications have a better outcome after diagnosis with cancer than patients not receiving medication?
• Pharmaceutical companies are sometimes required to track the safety of new medications by more active means than just spontaneous adverse event reports

• Linking a cohort of medication users to state cancer registries can help determine the incidence of cancer among the cohort

• The results of linkages with state cancer registries can be used to identify cases where follow-up is required for safety reasons
Conclusion

• **Data linkages can be useful for the following:**
  - Evaluating the effectiveness of a cancer control and prevention program
  - Improving the quality and timeliness of cancer registry data items such as date of last contact, vital status, and contact information for cancer registry patients
  - Determining disease prevalence and burden among a study cohort
  - Helping to assess the long-term safety of a new drug
Contact Information

David Harris, MPH
(919) 541-7493
dharris@rti.org

RTI Health Solutions
Research Triangle Park, NC, US
Ann Arbor, MI, US
Barcelona, Spain
Lund, Sweden
Manchester, UK
Sheffield, UK
Waltham, MA, US

www.rtihs.org