Assessing Healthcare Data Linkage Capabilities Using an Online Database Resource

Anokhi J Kapasi, Sharmila A Kamani, Earl L Goehring Jr, Bao Nguyen-Khoa, Varinder P Singh, Judith K Jones

DGI, LLC, Arlington, VA, USA   E-mail: info@bridgetodata.org

BACKGROUND

The scope of epidemiology research can be enhanced by linking different health & population datasets to allow broader analyses of diverse characteristics. It is therefore important to be aware of the variety of data linkages that can occur.

OBJECTIVE

To determine the types of database (DB) linkages possible within or across various healthcare databases.

METHODS

B.R.I.D.G.E TO DATA® (www.bridgetodata.org), an online resource with 184 population healthcare DB profiles worldwide (as of August 2, 2012), was utilized to identify DBs with data linkage capabilities.

Figure 1. B.R.I.D.G.E. TO DATA® Search Page

(1) A keyword search with 'link' was conducted to identify various types of data linkages.
(2) An additional search using the criteria 'Cross-sectional Population Databases' and keyword term 'longitudinal' was conducted to identify DBs with records linked across survey periods (**Figure 1**).

The two searches resulted in 141 unique DBs, 32 of which were excluded due to no data linkage capabilities. The remaining 109 DBs were reviewed for data linkage characteristics.

RESULTS – Part 1

• The set of 109 DBs had the following non-exclusive characteristics: 81 (74%) DBs directly linked to another DB (**Figure 2A**), 19 (17%) had indirect linkage capabilities (**Figure 2B**), and 38 (35%) were formed through DB linkages (**Figure 2C**).

Figure 2. Examples of Database Linkage Capabilities

A. Direct Linkage (n=81)
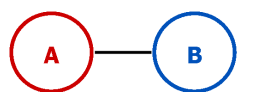
A.1 Direct Linkage (DB 'A' links to DB 'B')
- Korean Health Insurance Review Agency (HIRA) Database links to Korea Central Cancer Registry
- Multiple Risk Factor Intervention Trial (MRFIT) links to National Death Index (NDI)
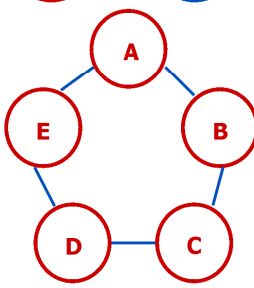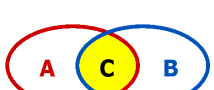
A.2 Multiple Direct Linkage (Network of linkages across DBs 'A' through 'E')
- Norwegian national registers
- Manitoba Population Health Research Data

B. Indirect Linkage (n=18)
(Linking DB 'A' to DB 'B' requires an extra step)
- Icelandic Cancer Registry needs approval prior to linkage of datasets to Cause of Death Registry

C. Formed by Linkage (n=38)

C.1 Combination of Database Subsets (DB 'A' sub-set links to DB 'B' subset to form new DB 'C')
- SEER - Medicare Database (USA) linkage of SEER cancer registries data, and the Medicare enrollment and claims files
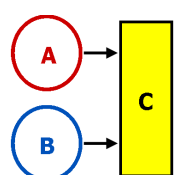
C.2 Merged Databases (DB 'A' merges with DB 'B' to form new DB 'C')
- North American Research Committee on Multiple Sclerosis (NARCOMS) Registry formed by multiple regional MS registries
- AIHW National Diabetes Register (Australia) formed by the National Diabetes Services Scheme database (NDSS) and the Australasian Paediatric Endocrine Group's (APEG) state and territory databases
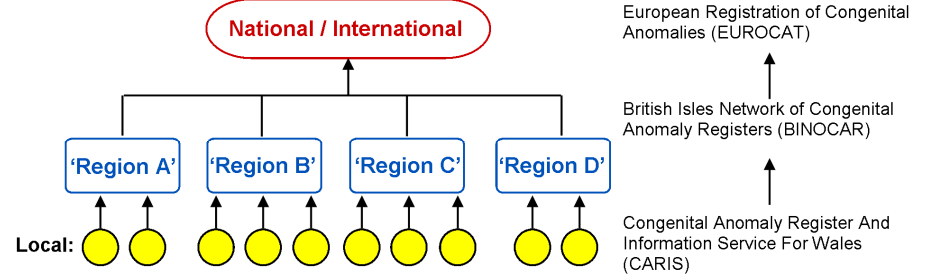
RESULTS – Part 2

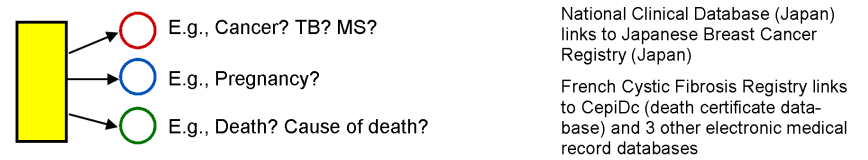The most common patterns included linkages by:
• Type of health services, e.g., prescription, diagnoses, and hospitalization data (66; 61%);
• Region, e.g., national registers (65; 60%) (**Figure 3A**);
• Health status (57; 52%) (**Figure 3B**);
• Vital statistics (41; 38%) (**Figure 3B**); and/or
• Civil / Administrative information, e.g., government administrative DBs (40; 37%).
• Some of the less common linkages were those by institution, practice type, across survey years/waves (**Figure 3C**), or study cohort.

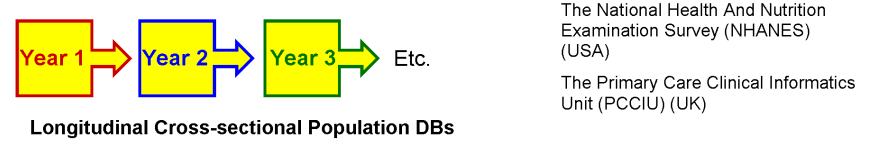### Figure 3. Examples of Types of Data Linkage Themes Across Healthcare Databases

A. By Region (n=65)

European Registration of Congenital Anomalies (EUROCAT)

British Isles Network of Congenital Anomaly Registers (BINOCAR)

Congenital Anomaly Register And Information Service For Wales (CARIS)

B. By Health Status (n=57) or Vital Status (n=41)

E.g., Cancer? TB? MS?
E.g., Pregnancy?
E.g., Death? Cause of death?

National Clinical Database (Japan) links to Japanese Breast Cancer Registry (Japan)

French Cystic Fibrosis Registry links to CepiDc (death certificate database) and 3 other electronic medical record databases

C. Across survey years (n=9)

Year 1 → Year 2 → Year 3 → Etc.

Longitudinal Cross-sectional Population DBs

The National Health And Nutrition Examination Survey (NHANES) (USA)

The Primary Care Clinical Informatics Unit (PCCIU) (UK)

Primary linkage methods were use of unique ID or probabilistic matching. Data elements obtainable via linkage varied, but frequently included data on vital status, cancer, hospitalizations, and prescriptions.

LIMITATIONS: This analysis was done using registries currently profiled within B.R.I.D.G.E. TO DATA®. More profiles of data sources are continually being added to this resource.

Each of the 75 data fields used in structured profiles in B.R.I.D.G.E. can be compared side-by-side to identify DBs with the most appropriate data elements captured within a DB or via linkage capabilities (**Table 1**).

Table 1. Excerpt from B.R.I.D.G.E. TO DATA® Comparing Data Elements in 3 Selected Databases with Linkage Capabilities

CONCLUSIONS

Through the use of a schematic notation system for mapping database linkages, this study highlights a growing number of databases with data linkage capabilities. Specifically, 59% (109/184) of the profiles on www.bridgetodata.org describe data linkages. While many linkages exist, the most frequent are to regional or health services DBs; common data elements obtained are vital status, cancer diagnoses, hospitalizations, and prescriptions.

Assessing Healthcare Data Linkage Capabilities Using an Online Database Resource

Anokhi J Kapasi, Sharmila A Kamani, Earl L Goehring Jr, Bao Nguyen-Khoa, Varinder P. Singh, , Judith K Jones

DGI, LLC, Arlington, VA, USA

DGI Center for Health Research and Education, LLC

1616 N Fort Myer Drive, Suite 1430 • Arlington, VA 22209 USA

www.bridgetodata.org   T: +1 703 894 0804   info@bridgetodata.org