

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

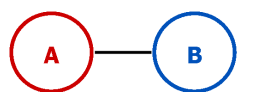
- (1) A keyword search with 'link' was conducted to identify various types of data linkages.
 - (2) An additional search using the criteria 'Cross-sectional Population Databases' and keyword term 'longitudinal' was conducted to identify DBs with records linked across survey periods (Figure 1).
- The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

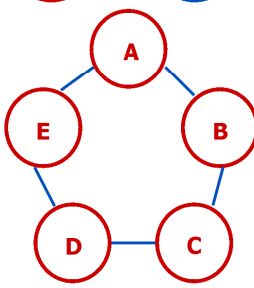
- The set of 109 DBs had the following non-exclusive characteristics: 81 (74%) DBs directly linked to another DB (Figure 2A), 19 (17%) had indirect linkage capabilities (Figure 2B), and 38 (35%) were formed through DB linkages (Figure 2C).

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)



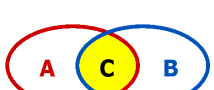
- #### A.2 Multiple Direct Linkage (Network of linkages across DBs 'A' through 'E')
- Norwegian national registers
 - Manitoba Population Health Research Data

B. Indirect Linkage (n=18)

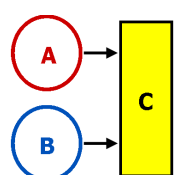


- (Linking DB 'A' to DB 'B' requires an extra step)
- Icelandic Cancer Registry needs approval prior to linkage of datasets to Cause of Death Registry

C. Formed by Linkage (n=38)



- #### C.1 Combination of Database Subsets (DB 'A' subset 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

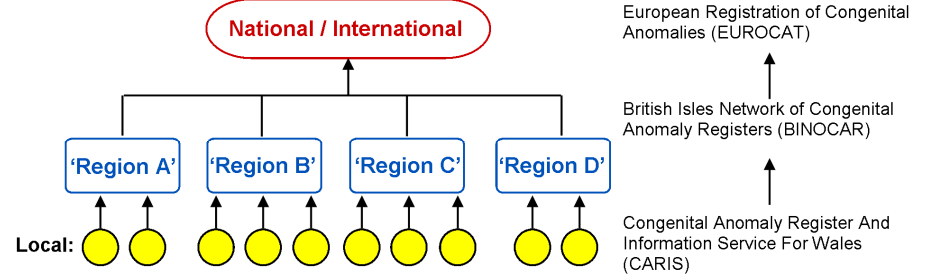
RESULTS – Part 2

The most common patterns included linkages by:

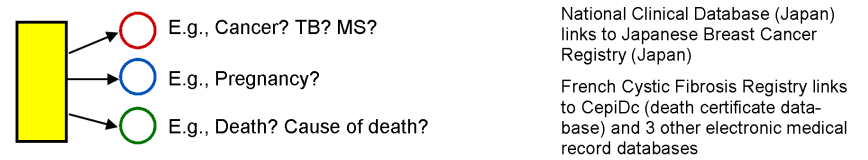
- Type of health services, e.g., prescription, diagnoses, and hospitalization data (66; 61%);
 - Region, e.g., national registers (65; 60%) (Figure 3A);
 - Health status (57; 52%) (Figure 3B);
 - Vital statistics (41; 38%) (Figure 3B); and/or
 - Civil / Administrative information, e.g., government administrative 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

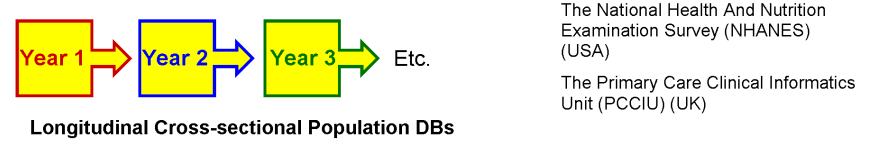
A. By Region (n=65)



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



C. Across survey years (n=9)



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

FIELD NAMES	Manitoba Population Health Research Data Repository: Health Insurance Registry (Canada)	National Death Index (NDI) (USA)	AGIS Health Database (AHD - Netherlands)
Region	Manitoba, Canada	United States	Central region of Netherlands
Brief Database Description	Manitoba Health maintains a current registry of all individuals registered to receive health services in Manitoba for administrative purposes, typically relevant for a short period of time and used by various claim-processing systems to verify eligibility requirements. The Manitoba Health Services Insurance Plan (MHSIP) is financed from general revenues of the Province of Manitoba and funds provided by the Government of Canada. Every family is assigned a family registration number, and every individual is assigned an unique encrypted Personal Health Identification Number (PHIN) by the Ministry of Health. From 1979 to 1984, individuals were specified using a combination of family registration, date of birth, and sex. Encrypted PHIN started being assigned in 1984 at which a person moved into Manitoba. It allows MCHP to analyze individual-level data over time and across databases without requiring names, addresses, or other identifying information.	The National Death Index (NDI) is a central computerized database composed of death record information on file in the state vital statistics office since 1979. NDI is compiled from computer files submitted to the National Center for Health Statistics (NCHS) by the states vital statistics office. NDI enables investigators conducting statistical studies to determine deaths of study subjects in the US. The investigator provides NCHS with identifying information on the subjects in statistical studies. NCHS runs this information against data in NDI and the investigator is sent a report showing which NDI records were possible matches with the investigator records. The investigator may then contact the state offices for copies of the relevant death certificates and/or obtain cause of death codes using the NDI Plus service.	The AGIS Health Database (AHD) is a registry of a health insurance company in the Netherlands. AHD contains demographic and health care consumption data of ~1.3 million inhabitants since 1998. It also contains information on characteristics of health care providers (HCP). Although AHD is not completely representative of the entire Dutch population, it does represent the urbanized area of the Netherlands. Data are preserved on health care procedures by all contracted HCPs such as general practitioners, specialists, physiotherapists, pharmacists, midwives and hospitals, in both primary (consultations, referrals, prescriptions) and secondary care (consultations, prescriptions and Diagnostic Treatment Codes). A number of personal characteristics are documented, such as date of birth and gender. As all prescriptions must be registered electronically in this database before costs are reimbursed, AHD is an accurate and detailed registration of pharmaceutical prescriptions and all other health care services of GPs and specialists.
Database Type	Registry (Longitudinal Population-Based Registry)	Other (Population Database - with dates of death, states of death, death certificate numbers, and cause of death codes only)	Longitudinal Population Database (Drug and Diagnosis Data, Outpatient and Inpatient)
Database Source	Manitoba Health registration files. All identifiers known to Manitoba Health since 1970 are reviewed and a single unique encrypted PHIN is assigned to each individual. This enables it possible to construct individual histories over the entire period of the database and to obtain information from other databases containing encrypted PHIN.	Death Certificates	Medical Insurance Claims (Computerized Prescriptions)
Population Type	General Population (Insured - All Manitoba residents registered with the Manitoba Health Insurance Plan. A small proportion of non-Manitoba residents receiving health services in Manitoba may also be present in the registry.)	Decedents (Decedents in the US, 1979 onward)	General Population (The Dutch population is insured mandatory by the new Health Insurance Act from 2006. The AHD population is representative of the urbanized central region of the Netherlands.)
Database Population Size (Range)	1-5 Million (Just under 2 million records representing active and inactive registrants)	50 - 100 Million (Currently over 70 million records are in the database since 1979. About 2.5 million records are added annually.)	1 - 5 Million (1.5 Million)
Death Recorded	Yes (Date of death is also recorded)	Yes (Information includes date of death, state of death, death certificate number, and cause of death code)	Yes (Stopping the insurance is recorded by the date and reason; death is one of the reasons. Transfer to another insurance company is also a reason, but there is no follow up.)
Other Demographic Data	Yes (Data fields for registration, birth, entry into province, migration in/out of province, and death provide the needed follow-up information to track residents for longitudinal and intergenerational analyses. Other data elements include encrypted PHIN, demographic characteristics, and family composition data fields. Residential postal codes are recorded semi-annually making it possible to track residential mobility, while marital status is reconstructed from family registration information.)	Yes (Marital status, race, sex, first and last name, middle initial, father's last name (surname), date of birth, social security number (SSN is access-restricted), state of death, state of birth, state of residence, date of death, age at death. These data are not released. The NDI only confirms whether these data elements agree with the same data elements provided by our NDI clients.)	Yes (Zip code, social status, education level, citizen administration number, medical specialty assigned, zip code and name of hospital used)
Diagnoses Coded	ICD-10 CA (Prior to April 1, 2004 the ICD-9-CM coding system was used)	ICD-9 (prior to 1999) & ICD-10 (after 1999) These mortality codes are used for underlying and multiple causes of death. NDI also provides the state and date of death, and the corresponding death certificate number for each possible NDI record match generated. The NDI client can also purchase copies of death certificates from the appropriate state vital statistics offices to obtain cause of death data. Not all states record cause of death codes on their certificates.	Yes (Currently, AHD includes disease classification codes only for hospital care, not for general practitioners (as they do not submit clinical information). These codes, called 'DSC-codes', were introduced in the Netherlands for the reimbursement of hospital services. This system does require an appropriate diagnosis of each patient involved.)
Physical Examination Findings	Yes (Via linkage)	N/A	Yes (Data are preserved on health care procedures by all contracted health care providers such as general practitioners, specialists, physiotherapists, pharmacists, midwives and hospitals including type and date of GP consultations and of chronic disease management programmes run by practice nurses in primary care.)
Cancer Data	Yes (Via linkage)	Yes (If diagnosed as cause of death)	Yes (Only if treatment has been carried out. The DBC code includes a diagnosis and treatment regime, but not the stage and pathology.)
Procedure Data	Yes	N/A	Yes (Information on surgical procedures is collected)
Laboratory Information	Yes (Via linkage)	N/A	No
Drug Data	Yes (Via linkage)	Yes (Only if listed as one of the causes of death on the death certificate. Drug name is usually not included.)	Yes (Prescription only AHD contains medical data on all insured patients including computerized lists of all medication prescribed by GPs. All doses and prices are standard and prescriptions (12 million to date) are electronically verified before being saved in the database which is updated weekly. Note: AHD does not include drugs delivered through hospital pharmacies and OTC medication.)
Type of Cost Data	The cost list provides a standard cost for each type of case that receives care in a hospital in Manitoba. This includes care provided to inpatients and for outpatient surgeries.	N/A	Yes, AHD records payments for the provision of all medical care to its insured patients, including prescription costs
Linkage to Other Databases	Yes (Use of a consistent set of identifiers (with identification numbers of both patients and physicians encrypted to ensure confidentiality) permits researchers to build histories of individuals across files. For example, individuals who are discharged from hospital can be linked to the medical claims file in order to determine if adverse events are being treated in physicians' offices.)	Yes (NDI is permitted to be linked to any database containing records of study subjects, survey participants, or employees, as long as the linkage is solely for statistical purposes in health and medical research)	Yes (To overcome the lack of clinical information, AHD was linked to other databases such as those of regional (e.g. hospital, psychiatric and primary care databases), or national (cancer databases, death registries) databases. Currently, a virtual network (the so-called Mondriaan Project) links several databases including AHD on a structural basis. To protect privacy during this linkage procedure a trust-trusted party is involved, that uses a so-called chance-linking procedure to identify patients on the basis of three individual characteristics.)
Brief Description of Linkage Capabilities	The anonymized data can be linked with other databases in the Repository housed at MCHP using an encrypted Personal Health Identification Number (PHIN)	NDI is linked to most NCHS surveys	In the near future this network of permanently linked databases will be able to efficiently provide complex datasets for research purposes.
Database Contact Data	Jack Rach - Communications Officer (Full contact information on B.R.I.D.G.E.)	Lillian Ingster, Ph.D. - National Death Index Director (Full contact information on B.R.I.D.G.E.)	Hugo M. Smeets - Julius Center for Health Sciences and Primary Care (Full contact information on B.R.I.D.G.E.)

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

Presented at the 28th International Conference on Pharmacoepidemiology & Therapeutic Risk Management,
Barcelona, Spain. August 25, 2012.



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