

Linkage Between Neurological Registry Data and Administrative Data

Nathalie Jette^{1,2}, Kristin Atwood³, Mark Hamilton⁴, Rachel Hayward^{4,5}, Lundy Day⁴, Theo Mobach⁴, Colleen Maxwell⁶, Claire Marie Fortin⁷, Guillermo Fiebelkorn⁴, Karen Barlow⁴, Michael Shevell⁸, Moira K. Kapral⁹, Steve Casha^{2,4}, Megan Johnston⁴, Samuel Wiebe², Lawrence Korngut², Tamara Pringsheim⁴

Can J Neurol Sci. 2013; 40: Suppl. 2 - S32-S34

This section of the guideline discusses considerations when planning and designing a registry that will require linkage to administrative data. Administrative data may include hospitalization and surgical or other procedure data; physician billing data; vital statistics data (e.g. births, deaths); prescription and other pharmacy data; long-term care services and admissions; and other data collected by provincial and territorial governments, health authorities and hospitals or care sites for administrative purposes.

Linkage of registry data to administrative data may create unique challenges and considerations during registry design. For example, linking to administrative data may require patient identifiers such as a personal health number that might not otherwise be collected. Linkage may present schedule challenges or constraints with respect to registry launch, operation, and data analysis. And finally, linkage of administrative data to registry data may impact the data fields collected and research ethics board (REB) approval required for the registry.

Clinical registries often capture a wealth of clinical information, and as a result help address and answer key health related questions. However linkage of registry data to other data sources is sometimes necessary to achieve a particular registry's pre-determined objectives. For example, a large number of patients with a particular neurological condition may self-register to a particular registry, and agree for their data to be linked to administrative data. Linkage to administrative data may allow for confirmation of a diagnosis that may not have been possible solely by self-report, enhancing the validity and quality of the registry data.

Although administrative health data were originally used solely for "administrative purposes", they have become a rich source of health data for research and surveillance purposes. They are many advantages to administrative data, in that they are often population-based capturing nearly every contact with the health care system, they are often cost effective, they can allow

investigators to follow people over time, they are not affected by selection or recall biases and they can be used to study rare outcomes. However, the amount of data can at times be overwhelming and lack some of the rich clinical information which is often captured in clinical registries. Data quality, as with any other data sources can also be an issue. Regardless, administrative health data are used widely in health care for quality improvement, surveillance, and to study health services, morbidity and to study a variety of outcomes, including mortality.¹⁰⁹

In preparing this section of the guideline we reviewed available literature and consulted with registry, disease, administrative data, legal, ethics and privacy experts to derive consensus recommendations.

RELEVANT LITERATURE

Unfortunately there was a paucity of literature addressing the issues surrounding linkage between registry data and administrative data. Furthermore, many of the identified articles from the literature review pertaining to this topic were not relevant in the Canadian context. Although the Agency for Healthcare Research and Quality (AHRQ) manual provided rich information on how to develop and implement registries, the information regarding data linkage was not Canadian specific.⁵

Policy & Legislation

There is an ethical and legal obligation to protect patient privacy when collecting health data, whether from a single source or from multiple linked sources. There is a need for methods and formal approaches to ensure that individual identifiable information is protected. There were no peer reviewed articles discussing ethics regulations and privacy legislation requirements in each of the Canadian provinces and territories. It is well known that these differ from one province

From the ¹Institute for Public Health, University of Calgary, Calgary, Alberta; ²Hotchkiss Brain Institute, University of Calgary, Calgary, Alberta; ³BC Ministry of Health, Victoria, British Columbia; ⁴University of Calgary, Calgary, Alberta; ⁵Office of the Information and Privacy Commissioner of Alberta, Edmonton, Alberta; ⁶University of Waterloo, Waterloo, Ontario; ⁷Canadian Institutes of Health Information, Toronto, Ontario; ⁸McGill University, Montreal, Quebec; ⁹University of Toronto, Toronto, Ontario.

FINAL REVISIONS SUBMITTED JANUARY 28, 2013.

Correspondence to: Lawrence Korngut, Clinical Neurosciences, 480060, 4th Floor Administration, South Health Campus, 4448 Front Street SE, Calgary, Alberta, T3M 1M4, Canada. Email: Lawrence.korngut@albertahealthservices.ca.

to another. In some provinces, ethics approval is only required for research questions and privacy impact assessments are only required for clinical care or quality assurance purposes. Details of the relevant policy and legislation by province can be found in the Ethical and Legal considerations section of this document.

OTHER CONSIDERATIONS

Linkable Data Sources

Although the focus of this section of the guideline is on linkage of registry data to administrative data, it is important to consider all possible linkable data when developing a registry, as similar issues arise whether or not these linkable data are administrative in nature or not. Sources of data beyond administrative data include clinical databases, survey and census data (e.g. national health surveys), imaging data, electronic medical records, laboratory data, and biological specimens.

Significant gaps in the literature review were identified in this area.

Technical Considerations

Although there was a lot of information in the published literature regarding how to enhance data quality and data collection for registry, there was a paucity of articles discussing the technical aspects of data linkage in a Canadian context. There was a good chapter in the AHRQ manual discussing linking registry data, but much of the information provided was only relevant to American researchers,⁵ with a major emphasis on the Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy Rule.

Linking patient data from multiple sources can increase the quality and completeness of data collection and assist in tracking patients who are lost to follow up. A group of researchers created and evaluated the feasibility of electronic linkage of the North Carolina Emergency Medical Services (EMS) Data System with the North Carolina Stroke Care Collaborative Registry.¹¹⁰ This system matched de-identified data from a prospective registry to EMS data using hospital name, arrival date, time, age and sex. The system was validated in three registry hospitals manually using patient names. Results of the study generated 63% probable patient matches and 89% of these matches were verified as true matches. One limitation to data verification and linkage was the quality of EMS data. However, linking EMS records electronically to a stroke registry was feasible and led to a large number of valid matches. They concluded that data linkage was a useful tool for registries to collect patient information from various sources and enhance coordinated systems of care. However, linkage may not be possible when databases are coded differently or when data collection methods and privacy laws might place limitations on using identifiable data.

Obtaining Linkage Data

Computer-assisted record linkage dates back to the 1950s.⁵ There are several data linking methods, most of which rely on the use of unique identifiers. The most commonly used method is the so-called “deterministic” method where unique identifiers are used in each of the databases of interest.⁵ In Canada, ideal unique identifiers used for health data linkage include:

- i. Personal health care number (PHN) that is a unique identifier given to all Canadian citizens with provincial health care coverage.
- ii. Last name
- iii. First name
- iv. Date of birth
- v. Postal code

The most unique challenges not only in Canada, but internationally relates to obtaining linkage data. In some Canadian provinces, administrative health data can take up to three years to be released to researchers, despite following all of the proper ethical and legislative processes.

Data Protection

One of the most important aspects of the data linkage process includes the processes in place to ensure ongoing data protection. Here, we do not discuss in details the methods used to mitigate the risk of re-identification. These are discussed in some detail in the AHRQ manual.⁵ We however emphasize the need to involve data linkage, privacy and legislative experts who have familiarity with these processes early on in the registry inception to ensure risk mitigation is in place.

Cost

If data linkage is being considered, the cost for data linkage must be addressed during the planning stages of the registry. Cost considerations include: jurisdictions coverage; length of review and time to acquire data; and ongoing cost to maintain access to data. Costs for administrative data vary by data type and by province. Careful research into the required jurisdiction and type of data should be conducted during the planning phases. Jurisdictions also have substantially different lengths of time for review of administrative data requests and even following approval, acquisition of actual data files may involve further time delays. Consider the costs of these time implications against the registry plan. Finally, if the registry project plans ongoing linkage activities the sustainability costs must be addressed. For example, are there annual fees; ongoing security needs; or other aspects to this activity that will impact the project budget?

RECOMMENDATIONS

- ✓ Ensure data linkage is necessary, feasible, and ethically sound during the registry planning and design phase. For example, relevant data exists and can be obtained in a financially sound and time appropriate manner.
- ✓ Involve administrative data experts and data custodians early on if data linkage needs are identified.

- ✓ Carefully examine jurisdictions that will need to provide administrative data and examine costs; identify data custodians; research application requirements; and identify projected time to obtain data.
- ✓ Incorporate data linkage permission into participant informed consent at the outset.
- ✓ Determine the data fields needed to provide adequate linkage. Use a minimum dataset approach.
- ✓ Consider establishing reciprocal data sharing agreements and educating people about the benefits and value of data linkage in order to overcome challenges in obtaining linkable data.
- ✓ Registries containing linked data or data linkage references must have appropriate security protection in place which may include: password protection; levels of access by user; suppression and encryption. Additionally, database systems should be regularly backed up.
- ✓ Establish a desired timeline and linkage plan considering the time required to obtain data and how often it should be updated.
- ✓ Prepare policies and procedures around linked data including how data will be stored and when and how it will be destroyed. In some cases these aspects may be partially dictated by the data custodian.