Neurological Registry Data Collection Methods and Configuration

Eric E. Smith[1], Janet Warner[2], Megan Johnston[2], Kristin Atwood[3], Ruth Hall[4], Jean K. Mah[2], Colleen Maxwell[5], Claire Marie Fortin[6], Mark Lowerison[2], Moira K. Kapral[4], Vanessa K. Noonan[7], Ted Pfister[2], Gail MacKean[2], Lisa Casselman[2], Tamara Pringsheim[2], Nathalie Jette[1,8], Lawrence Korngut[1]

The purpose of this section of the document is to identify issues related to data collection and registry configuration. When designing a disease registry, it is important to consider the registry's purpose and target population as this will influence the type of data, source(s) of data, and the manner in which it is collected. A data dictionary defining the specific data elements to be collected is key to ensuring registry data quality. Compliance of physicians and patients who provide registry data is instrumental to data collection and should be addressed early. Additionally, it is important to consider if the registry will be linked to other databases. Finally, it is important to address procedures for making changes in the registry and to establish what types of documentation are necessary.

In preparation of this section, we reviewed the literature, scholarly sources, and consulted with medical experts and registry/database specialists on the topics mentioned above.

RELEVANT LITERATURE

In preparing the information below 139 full text articles were reviewed.

Conventional and Innovative Roles for Registries

Most registries aimed to serve conventional registry roles. As examples, Byrne et al used a registry to examine the natural history of Pompe disease[68]; three registries pursued quality improvement (stroke care, cardiac catheterization and management of childhood diabetes[69-71]); disease subgroup characterization was facilitated for pituitary and lung tumors,[72,73] post-market device monitoring for cardiovascular stents,[74,75] as well as monitoring of treatment outcomes and safety follow up for cardiac transplant patients receiving everolimus, general cardiac care, biological agents for arthritis, giant intracranial aneurysms, and TPA for ischemic stroke.[76-80] Other registries aimed to serve innovative roles:
• examining the impact of a diagnostic test on patient management (i.e. PET scanning on management of cancer patients),[81]
• clinical instrument development (i.e. management of hospitalized patients with heart failure),[82]
• linkage of clinical data to a DNA bank for patients with congenital heart disease,[83]
• imaging information capturing MRI data from stroke patients,[84]
• curation of data for genetic linkage analysis in systemic lupus erythematosus,[85]
• monitoring of adverse events and medical errors in surgical patients,[86,87]
• adherence to clinical best practice guidelines (i.e. lack of lymph node dissection for penile cancer),[88]
• serving as a ureteral stent removal reminder system to Urologists,[89] and
• identification of management not consistent with best available evidence (i.e. not targeting evidence-based INR for prevention of venous thromboembolism.[90]

Registry Configuration

The Agency for Healthcare Research and Quality (AHRQ) users' guide[5] was the most comprehensive document identified to guide the process of registry development. The authors illustrate current uses for patient registries and how they may play critical roles in providing high quality evidence in circumstances where randomized trials cannot be conducted or may not generate generalizable results. In particular, the guide focused on patient outcomes including studies of natural history, effectiveness determination, measuring or monitoring safety and harm, and measuring quality. Registries can be designed as product, health services, or disease registries, or combinations thereof.

The AHRQ guide[5] provides suggested steps in planning a registry. These begin with articulation of the purpose, determination that registry design is the appropriate methodology for the purpose, identification of key stakeholders, feasibility assessment, building of the team, establishment of governance and oversight, scope of data, as well as defining of

From the [1]Hotchkiss Brain Institute, University of Calgary, Calgary, Alberta; [2]University of Calgary, Calgary, Alberta; [3]BC Ministry of Health, Victoria, British Columbia; [4]University of Toronto, Toronto, Ontario; [5]University of Waterloo, Waterloo, Ontario; [6]Canadian Institutes of Health Information, Toronto, Ontario; [7]Rick Hansen Institute, Vancouver, British Columbia; [8]Institute for Public Health, University of Calgary, Calgary, Alberta.
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.

the core dataset, patient outcomes and the target population. The guide also suggests that a clear protocol and project plan be developed and planning for study completion.

Is a Registry Appropriate?

The first step in designing a registry for obtaining information on neurological diseases is to determine whether a registry is the best means of obtaining the desired information.[91] Registries are useful tools for facilitating research, performing audits, facilitating policy decisions, and managing health care services and associated resources.[5,92] However, sometimes registries are not an appropriate means of obtaining information; for instance if the data necessary to answer a research question have already been collected, the data are of the quality needed to properly address the question and the data are accessible to the researchers.

Selecting a Registry Design

Once it has been determined that a registry is the best method of collecting data, there are a number of decisions regarding its configuration that must be made. First, it is important to consider what kind of registry is most desirable. There are several types of registry designs including clinic-based, community-based, online, patient self-registration that have a variety of purposes such as disease surveillance, quality improvement, natural history studies and longitudinal research.[5,36] It is important to clearly describe the registry's purpose[36] as well as the specific research questions the registry will purport to answer and specific, measurable objectives the registry will seek to accomplish[5] before defining data collection methods and forms. This will ensure that registry data collection is relevant in terms of the study objectives, that the data that gets collected can be used to its full potential, and that data collection is proportional to the resources that are available.[5,36,69,90,91,93]

Target Population

It is important to determine the registry's population. A population-based registry is one that represents all incidences of a given condition in a given population.[39] A province-wide registry accessing all potential participants is an example of a population-based registry.[39] Beghi et al emphasize the importance of strict adherence to population-based registry design when examining for disease risk factors, the importance of appropriate control selection when using registry participants in a case-control study, and validation of the quality of data registration.[94] Establishing a population-based registry will provide more complete and comprehensive information about those afflicted with the condition of focus in that population.[39] Although a population-based registry is the most desirable, the challenge is the reliance on voluntary consent. It is therefore possible that targeting a subset of a population for a registry may be more practical and sustainable.[39]

Regardless of whether or not the registry is population-based, it is essential to define the population to which the registry findings are intended to be applied, i.e. the target population.[5,91] For example, registries may choose to focus on patients with particular diseases, those with an exposure to a particular product or procedure or those who participated in a quality improvement project or other program.[5] The target population of the registry

will influence many aspects of registry planning and design, such as which sampling practices are most appropriate.[5]

Patient Recruitment

It is important to consider various challenges faced by patients (such as cognitive issues, mobility issues, etc.) and physicians/centres (such as limited time, limited staff, limited resources) when planning a recruitment model because addressing the needs of the people who will be providing data is instrumental to successful data collection. Using only a single mode of data collection may lead to biased sampling because patients with cognitive and/or mobility issues may find particular modes of data collection more challenging than others. Hence, using diverse recruitment practices - such as telephone, mail, in-person discussions during clinic visits, online recruitment and mobile applications - will decrease the likelihood of biased sampling.

One challenge with respect to rare disease research is that one must sample from centres in multiple jurisdictions in order to get an appropriate sample size. However, under-representation is likely to occur in those centres which lack the resources to participate in registry recruitment. Diverse recruitment practices are one way to resolve this issue. Additionally, in Canada, different jurisdictions have different regulations with respect to privacy and research ethics: these are discussed in more detail in the Ethical and Legal Considerations section of this document.

Incentives which conform to the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS-2)[7] may also increase compliance with data collection requirements. For example, providing patients and physicians with regular correspondence in the form of individualized reports may serve as an incentive for participation because those involved in registries tend to want to have access to the data they are providing and tend to want to know what is being done with the data that they are providing.[95] Additionally, it is recommended that clinicians be guaranteed open access to their own registry data as it can be used for clinical studies[95] and may facilitate clinic note dictation. In addition to providing useful information, this will likely increase transparency about how the registry is using the data it collects.

With respect to using registries to facilitate study recruitment, it is recommended that passive recruitment such as notifying patients of existing research be used as opposed to active recruitment such as marketing for and promoting other studies.

Concerns about privacy may influence patients' willingness to participate in registries.[48,96] In order to address these concerns, it is important to consider how data will be stored, who will have access and what security measures will be taken. Furthermore, it is important to address how patients' privacy will be protected during the informed consent process.[5] More specific information about data storage and privacy considerations can be found in the Data Storage and Curation section of this document as well as the Ethical and Legal Considerations section of this document.

For more information about patient recruitment, consult the Patient Recruitment section of this document.

Data Collection Sources and Methodology

Registry data can be obtained from such sources as patients, clinicians, paper medical records, electronic medical records,

administrative sources, other registries, national disease organizations, laboratory data and physician billing data.[5]

Full text review of 34 documents was conducted and revealed that there are a number of potential sources of data and methods of data collection for registries.

Types of Registries

Physician Driven Registries

Physician driven registries have great potential to gather large amounts of clinical and demographic information,[93] but time constraints on physicians make it challenging for them to be able to gather large amounts of patient data for registries[93] Recruitment of patients for a registry by a physician is one of the most successful recruitment strategies because the direct involvement of a patient's physician in a registry is a key factor influencing participation.[5] However, to avoid data collection fatigue physician driven registries must have unambiguous datasets which are not a tax on physician time.[93] Clearly defining and documenting expectations of clinical professionals involved in recruitment and making use of technology to automate data entry and reminders are measures that can be taken in order to reduce the physician's burden. Efficient workflows that align with clinical process will maximize data quality.[97]

Patient Driven Registries

Patient driven registries can provide access to large patient populations in a cost-effective manner,[93] and readily cross geographic boundaries. However, these registries may not gather uniformly high quality data due to the high potential of errors in diagnosis and other key data points when physician review of collected data does not occur.[93] While it is possible to create successful patient driven registries with accurate diagnoses,[93] in general physician driven registries are more likely to produce datasets with limited bias and registries that retain patient interest and commitment. One concern with any registry methodology is the potential for patient populations to be biased through recruitment methods (selection bias). This concern can be partially addressed by stratifying registry data to represent geographic distribution, and then sub-sampling across the registry for study purposes.[98]

Periodic reassessment of registry participants in either physician or patient driven registries has the potential to provide rich longitudinal data which would also be beneficial for examining outcomes and facilitating research.[83,85]

Approaches to Data Collection

Data can be abstracted from patient records by a person other than the clinician who interacts with the patient.[5] Sometimes the person who abstracts the data from the record will "code" the data onto the case report form (the form that contains the data elements the registry intends to collect from its patients). Coding consists of replacing a text diagnosis in a chart with a standardized code: these codes are usually defined in a data dictionary.[5] If data linkage is being considered, it is important to ensure that data elements allowing linkage are compatible with the linkage data source.

Web-based Registries

Web-based registries collect retrospective data over the internet from patients or clinics and transmit the data to a central repository. The data are manually entered by the patient healthcare provider or delegated research assistant.[99] While there is a general perception that web-based registries improve the speed of data entry, one study found only a ten second difference between paper-based data collection and web-based data collection per patient over a total of one initial entry and one follow-up visit and initial data entry collection was actually longer by eight second versus paper-based data collection.[100] Additionally one study compared online registry data collection to a previous paper-based methodology in the same discipline and found that it increased participation by 42%.[101]

Electronic Chart – Based Registries

As a new concept,[99] electronic chart based registries can enroll patients in real time based on chart data such as International Classification of Diseases (ICD) codes and populate registry fields through automatic download of relevant chart data.

When comparing the two data collection methods above, a key benefit of the electronic chart-based registries is the elimination of manual data entry errors.[99] However, electronic chart-based registries rely on compatibility with electronic charting systems (e.g. versioning etc.) and may require periodic updates or reconfiguration. Clinical follow-up may also be problematic if patients do not return to the primary hospital although this issue also exists with other types of registries.[99]

When selecting a modality with respect to registry type, consideration should be based on the availability of patient data through a given modality and the likelihood of registry success considering comprehensive factors from data collection efficiency to overall cost efficiency. What may be appropriate and successful with one patient population in any particular country may be starkly inappropriate in another patient population or country and evidence to support or refute any particular choice is likely to be found in the literature. Evidence in the literature is clear that registry usefulness is far more impacted by the overall quality of the data present in the registry, not the method of data collection.[101]

Data Elements and Data Dictionary

It is essential for registries to clearly define which data elements are to be collected, how they are to be collected and ultimately to collect these data elements in a uniform way.[5] It is also essential to clearly describe and document guidelines for data abstraction and coding[5,102] and for those in charge of abstraction and coding to be properly trained in order to minimize the probability of errors.[36] Using a paper or electronic case report form or formatted list of elements as well as producing a manual which clearly defines the data elements, how the data elements are interpreted, acceptable parameters and logical rules for data elements are recommended practices for encouraging uniform data collection.[5] Finally, it is important for the data entry process to be standardized and user-friendly.[36,39,40,103,104]

Linkage of data among registries can be facilitated by using common data elements.[5] More information about common data elements suggested for Canadian neurological registries can be found in Part 3 of this document.

Data Linkage

Registries can be linked to one another or to various data sources. Planning of new registries must consider existing registries and possible linkages or overlap in patient recruitment. New registries may also expand the target population of existing registries through linkage. Since one disease registry is typically not representative of the entire population with that disease, linking registries to one another can provide more representative information across a disease population. It is best to consider data linkage from the outset rather than attempting to link data after registries have been developed, because data definitions and formats developed separately are often not standardized across databases; translation between systems would be required and transferring data from one database to another could potentially lead to errors.[5] Hence, the process of linking registries is facilitated if registries are consistent in the data elements they collect and the manner in which they are collected. It is for this reason that the use of common data elements is recommended[5,93]. Increasingly, networks of registries are emerging to facilitate collaboration and planning of large studies such as the Orphanet rare disease database and meta-registry.[105]

Furthermore, appropriate permissions for data linkage should be sought a priori.[5] Additionally, it is important to be transparent about the sources from which data are being obtained and to consider establishing reciprocal data sharing agreements. With respect to data linkage, it is essential to consider who owns the data and who is responsible to maintain privacy during the inception stage.[5] More information can be found in the Data Linkage section of this document; more information about privacy considerations can be found in the Ethical and Legal Considerations section of this document.

Data Quality and Management

Given the potential usages of registry data, the data should be complete and accurate.[36,40] In the planning and design stages of a registry, it is important to consider issues related to data quality. For example, in a multi-disease registry, using and clearly stating standardized disease definitions, familiarizing participants with these disease definitions as well as using standardized sampling techniques are all recommended practices in order to promote optimal data quality.[7,90] More specifically, it is essential to work to maximize internal and external validity as well as generalizability.[5] Additionally, it is important to consider possible forms of biases and work to minimize bias within the registry.[5] Although bias cannot be eliminated, having a documented understanding of what biases exist and how such effects can be managed will be helpful, particularly in reporting outcomes.

Additional issues which can have a deleterious impact on registry data quality include: missing data, invalid entries, erroneous entries and inconsistent data.[5] It is recommended that registries have a manual which addresses how to assess and ameliorate these issues.[5] Possible ways to resolve these issues with data include re-checking the case report form, interviewing the patient and examining an alternate source of patient information.[5] It is important to perform database queries or reviews designed to screen for problems in the database.[5] The date, time and results of all reviews of the database should be documented. Additionally, for good registry management, it is essential to track all data received, all information entered into the database, and all data cleaning practices that are implemented.[5] It is recommended that a member of the registry's staff should have the role of quality assurance[36] i.e. someone who regularly assesses data items for accuracy, completeness and relevance.[36,39,40]

It is recommended that epidemiologists, statisticians and other database specialists be consulted throughout all stages of the registry in order to ensure that it is designed in a manner that maximizes the potential for gathering high quality data.[92] Registry data should be as comprehensive as possible while also being simple enough to reduce data collection burden.[10,36,69,91,104,106] It is hence recommended that the registry balances the need for parsimony (for instance, collecting the minimum amount of variables that it needs in order to answer its study questions and accomplish its objectives) with a reasonable anticipation of future needs that may require additional data not immediately necessary for the study's initial objectives.[36,94] In some instances, it may not be feasible to assess all aspects of the data initially, but future plans could include more detailed analyses that warrant the collection of additional information. However, these plans should be fully articulated at the onset of the registry design so that all data elements can be rationalized. Inclusion and exclusion criteria must also be clearly defined and the rationale for these criteria should also be clearly documented.[5,94] Five additional characteristics of a high quality registry database are as follows:[92,99]

1. The registry must be representative of its target population
2. Data must be complete and accurate
3. Data validation procedures should be used to assess data accuracy
4. Variables must be explicitly defined
5. There must be independence of observation of outcomes

Given that the quality of a registry's data is related to the abilities of the registry's staff, proper training of data collection staff is essential.[5,91] This training could take the form of an initial training session during on-boarding followed by regular continuing education sessions.[99] Database training environments, videos, and webinars are all useful training tools which are especially beneficial for registries with multiple centres as they can be used in a remote training situation. Additionally registries should have a manual of operations in which data collection staff members are well versed. When multiple sites are involved, it can also be beneficial to have regular meetings (teleconferences with abstractors/sites) or site visits to discuss progress, review procedures and resolve any issues that may arise.[83]

Prior to the launch of the registry, it is recommended that pilot tests be conducted. Pilot tests will allow registries to detect and resolve data collection issues which will detrimentally impact the successful implementation of the registry.[5,107] It is

helpful to directly involve staff who will be responsible for collecting data in the piloting process in order to receive feedback on considerations such as how user-friendly data collection methods are.

More specific information about how to address these issues of data quality can be found in the Quality Control/Quality Assurance section of this document while more information about how to evaluate data quality post hoc can be found in the Registry Evaluation section of this document.

Data Analysis

Factors to consider when developing observational cohorts with respect to subsequent data analysis include participant (database) bias, missing data, and subject misclassification.[108] The frequency and manner in which registry data are analyzed also need to be considered in the design stages of the registry because the way in which data is collected will influence whether or not planned analyses are feasible. The anticipated size of the registry, and the duration of the registry will also influence the way in which data are collected and analyzed and are hence important to define in the early stages of a registry.[94]

RECOMMENDATIONS

✓ Establish clear objectives for the registry based on its purpose.
✓ Define your target population, and what will constitute an appropriate sample.
✓ Employ diverse recruitment methods in order to reduce selection bias. Consider your target patient population and your chosen recruitment strategy to identify potential challenges that may be present
✓ Present clinical staff with clear expectations and use technology to reduce their recruitment burden.
✓ Utilize regular reporting to increase registry transparency and participation by physicians/healthcare professionals and participants.
✓ Develop a training program for data collection staff. Provide them with aids and resources to maintain training on an ongoing basis. Consider using technology to facilitate remote training and reduce costs.
✓ Consider including healthy age/sex match controls in the registry to facilitate research.
✓ Thorough documentation is essential for registry success. To that end, it is important to clearly document the following aspects of the registry in the registry protocol and additional documentation[5]:

1. Purpose of Registry
2. Research Questions/Specific, Measurable Objectives
3. Inclusion/Exclusion criteria and rationale for these criteria
4. Target population and sampling methodology
5. Anticipated size and duration of the registry
6. Manner and Frequency of data collection and analysis
7. Data dictionaries and coding manuals as appropriate.
8. Sources of registry data
9. How to use the paper and/or electronic case report form, whether or not the case report form is to be retained/copied/archived.
10. Roles of registry personnel and corresponding job descriptions and necessary qualifications for each position
11. Recruitment/withdrawal procedures including copies of appropriate consent/withdrawal forms and how they should be retained/copied/archived.
12. Procedures for promoting and subsequently evaluating data quality.
13. How patient identification codes are assigned, how duplicate records are prevented
14. Procedures for access to data for research purposes (internal and external).
15. Data security measures and procedures in the event of a security breach.
16. Registry governance structure and roles.
17. Legal and ethical documentation such as: confidentiality agreements; data-sharing agreements and ethics certificates and submissions.
18. Data management policies and agreements governing data management (e.g. contractor agreements; database administrator position description etc).

✓ Define anticipated registry size and duration to assist with selection of data collection strategies.
✓ Conduct pilot data collection to evaluate training protocols and database function.
✓ Determine data linkage needs in advance and seek appropriate permissions.
✓ Utilize passive recruitment methods for research study recruitment within a registry population.