Neurological Registry Quality Control and Quality Assurance

James Marriott[1], Vanessa K. Noonan[2], Elizabeth Donner[3], Mark Lowerison[4], Darren Lam[4], Lundy Day[4], Janet Warner[4], Eric E. Smith[4], Jean K. Mah[4], Paula de Robles[4], Nathalie Jette[4,5], Megan Johnston[4], Tamara Pringsheim[4], Lawrence Korngut[4]

This section of the guideline discusses procedures and best practices around quality control and quality assurance. In developing this section of the guideline we reviewed available literature and best practice; consulted with registry and disease experts; and derived consensus recommendations.

Quality, as defined by the International Standards Organization (ISO) in standard ISO 8402:1994, is the "totality of characteristics of an entity that bear on its ability to satisfy stated and implied needs."[194] In the context of registries, this means that registry data characteristics must altogether satisfy the intended and implied needs of the registry purpose. For example, if the purpose of your registry is to study all female adults of child-bearing age with epilepsy; then your registry data must consist only of female adults of child-bearing age who have a diagnosis of epilepsy. It is important to note that quality and registry purpose are inherently related. Registry creators will therefore need to define what quality means for their specific purpose(s).

While quality control and quality assurance are related concepts it is important to understand that they are different. Quality assurance (QA) is the process that maintains a desired level of quality.[195] QA is a proactive process done in advance of obtaining an outcome. Examples of QA activities might include audits, training, procedure documentation, selection of quality tools etc. Quality control (QC) is the assessment of whether an outcome meets quality expectations.[195] QC is a reactive process done once an outcome has been obtained. Examples of QC activities might include testing a product sample to determine if it meets requirements; or conducting a site inspection visit. Useful registries must have good quality data.[196,197]

RELEVANT LITERATURE

Quality Attributes

Without high standards for capturing data in registries, data quality can be compromised. The absence of high quality data in a registry may limit its use and generalizability. Arts et al[196] conducted a literature review on the subject of registry quality between 1990 and 2000. The two most frequently cited attributes that determine registry usability were accuracy and completeness. Based on an amalgamation of the definitions discussed in the literature these attributes were defined as[196]:

Accuracy – the extent to which registry data represent the truth

Completeness – the extent to which all necessary registry data has been entered

The above definitions were further supported in additional literature.[197]

Types and Causes of Data Errors

Many types and causes of error can be identified. The literature reviewed by Arts et al[196] divided data errors into three types: interpretation errors, documentation errors, and coding errors. Causes for these errors fall into two classifications: systematic and random.[197] Systematic data errors might be caused by computer programming errors; poor data dictionary definitions; inadequate or poor training; or data collection methodology violations or errors.[196,197] Random errors might be caused by incorrect data transcription (e.g. typing error), incorrect data collection (e.g. source documentation is illegible), or data are incorrectly entered into the data field (e.g. correct data in wrong location).[196,197] The most frequently cited errors in the literature reviewed by Arts et al[196] were inaccurate data transcription and computer programming errors. The average error rate found in the literature, accounting for both systematic and random data errors on Case Report Forms (CRFs) is 976 errors per 10,000 fields.[197] Overall, this is an error rate of approximately 1%.

Table 11: QA Activities

QA Action	Activity
Prevention	• Select and train adequately motivated personnel. • Design a data collection protocol including standardized definitions for data fields and guidelines for data collection method(s).
Detection	• Routinely monitor data and compare with the original data source; this could include centralized audits or site audits. • Utilize automated field parameters to detect errors within known value ranges. • Consider using dual entry or visual check methods to reduce random errors.
Action	• Correct identified errors • Identify and remedy root causes of errors

Case Report Forms

The literature reviewed by Arts et al[196] discussed a number of factors that would influence data quality on CRFs. Quality can be improved by the use of closed rather than open-ended questions on the CRF; and collecting data promptly from the original data source whenever possible or having data entered by a clinician if the original source is not available. Additional literature suggests direct connection between electronic medical records and CRFs to reduce transcription errors and capitalizing on the functionality of online CRFs[197] with easy to use fields and self-explained fields (e.g. pop up help).

Quality Control

Arts et al[196] found that the two main mechanisms discussed in the literature for registry QC were completeness checks and site visits.

Quality Assurance

Arts et al[196] found a number of quality assurance activities that should be undertaken in high quality registries. (see Table 11 above)

Other validation studies have also reiterated or reported additional points for maintaining and improving data quality in registries (Table 12). A review of the literature suggests that a number of strategies can be utilized. These include having motivated, well-trained, up-to-date, and accountable staff,[102,196,198-211] user-friendly data collection forms,[104,196,200] clear data collection methods,[202,212] clear objective definitions,[102,104,196,201,203-205,211], uniform data collection methods across sites,[93,201,205,209,212] a minimum set of necessary data elements in the registry,[104,196] drop down menus (as opposed to free text fields),[196] a system for automated data checks (e.g. software algorithms),[102,196,202,208,213] and an integrated delivery system between medical facilities for sharing patient records and information.[214] In addition, drafting and evaluating data

Table 12: Strategies to Maintain and Improve Data Quality in Registries

Strategy	Article(s)
Having motivated, well-trained, up-to-date, and accountable staff	102,196,198,211
Having user-friendly data collection forms	104,196,200
Having clear data collection methods	202,212
Having clear objective definitions	102,104,196,201,203-205,211
Having uniform data collection methods across sites	93,201,205,209,212
Having a minimum set of necessary data items	104,196
Having drop down menus	196
Having a system for automated data checks (e.g. software algorithms)	102,196,202,208,213
Having a minimum set of necessary data elements in the registry	104,196
Having an integrated delivery system between medical facilities for sharing patient records and information	214
Drafting and evaluating data collection protocols	196,215
Routine monitoring of data	102,196,197,199,202-204,208,215,216
Limiting the number of steps when collecting registry data	102,104,201,205,217
Collaboration and communication between staff, sites, and the registry	196,199,206,212,218-220
Providing constant feedback to participating sites for quality control	196,200
Comparing data with external sources to ensure complete case ascertainment	104
Collecting data in a time a location sensitive manner from those directly involved in the patients' care	196
Mandatory reporting	204,221

collection protocols,[196,215] routine monitoring of data,[102,196,197,199,202-204,208,215,216] limiting the number of steps when collecting registry data (e.g. electronic forms can directly collect data to reduce entry error affiliated with paper forms),[102,104,201,205,217] collaboration and communication between staff, sites, and the registry,[196,199,206,212,218-220] providing constant feedback to participating sites for quality control,[196,200] comparing data with external sources to ensure complete case ascertainment,[104] collecting data in a time and location sensitive manner from those directly involved in the patients' care,[196] and mandatory reporting[204,221] can all aid in maintaining and ensuring that registry data is of high quality.

OTHER CONSIDERATIONS

Quality Assurance Considerations – Data Collection

Appropriate and accurate data collection is inherently linked to the success of a registry so it is of paramount importance that the ultimate goals of the registry be reflected in the development of data collection procedures. This involves keeping the "big picture" in mind from the outset. To that end, case report forms (CRFs) should be designed to only collect the minimum amount of information needed for the registry. It is important to note that the minimum amount of information needed for the registry may not just be the minimum dataset. Sometimes it may be necessary to collect items beyond the minimum dataset needed for research in order to engage key stakeholders, or provide the registry with long-term funding sustainability. However, collecting the minimum amount of data serves multiple purposes. It reduces the burden on the front-end individuals (clinicians, researchers, data abstracters), minimizing the chance for survey fatigue to lead to errors or neglect of the registry. Interrelated to this, the data being collected should be seen as valid and important to the front line individuals (e.g.; patients of physicians) completing CRFs to keep them engaged and enthusiastic about the registry itself.

The CRFs should be designed to allow accurate data collection of uniform quality. This involves the creation of a data dictionary to explicitly define each variable being collected and the precise range of values allowed for all variables. For example, formatting of dates should be standardized throughout the CRFs and registry databases. These technical considerations need to be addressed early in the registry and CRF process. CRFs should also be constructed with a logical flow for the benefit of those entering the data. In registries which will be utilizing various primary data sources (egg; patient questionnaires, health care databases), CRFs should be identified so the appropriate data is collected from the best source for that data to help ensure data quality. Alternatively, if different sources are going to be used, methods for achieving consensus for data points needs to be identified a priori and mechanisms built-in to the registry to resolve discrepancies among data sources.

CRFs should also be designed with the needs of the content provider in mind. In the case of clinicians, as mentioned above, this means ensuring that registry data entry process is fluid and not overly time-consuming. When patients will be contributing to data collection, additional considerations are also needed. For example, limited dexterity might make typing or pencil/paper data collection time-consuming or impossible in some populations. Similarly, those with hearing deficits may find it challenging to complete a phone interview. It is therefore

essential for the modality of data collection be tailored to suit the needs of the population providing the content. If a diverse patient population with varying abilities is involved, the same information may need to be captured through different CRFs. In that case, it is important track the mode of data collection utilized for each dataset. This can be used to evaluate registry quality despite different data collection modalities and can allow for comparison between the methods.

Some registries will be collecting information from pre-existing datasets. These might include governmental health-care databases or office/hospital based electronic medical records (EMRs). Similar to when data is being collected initially from a patient or a clinician, it is important that the registry not become overloaded with extraneous data from the pre-existing dataset that does not suit the purpose and goals of the registry. Related to this issue, some administrative datasets may contain information that may comprise confidentially or privacy and should be stripped from the dataset used to construct the registry.

In some cases, it may be desirable to compare the registry data with an external database. For example, this may be needed to ensure the validity of the registry. If this is planned for a specific registry, it is important to consider data linkage up front during the CRF and registry design. It should be noted that creating database linkages may increase the burden of data collection. Nevertheless, this may be justified given the purpose of the considered linkage. More information on Data Linkage and Validation can be found in other sections of this document. Overall, it is paramount that the design of the data collection instruments and the registry as a whole be focused on collecting the least data amount of data necessary and collecting this through the easiest modality possible. It is also important that the tasks of data validation and data cleanup be delegated to those running the registry and not the front-line personnel providing registry content.

Data Cleaning

Data cleaning refers to the process in which errors in the registry's dataset are identified and corrected. Errors can include incorrect data (such as out-of-range values), absent data (missing values), duplicate entries or contradictory, mutually-exclusive data entered into different fields. As with other aspects of registry production, it is important that data cleaning be considered upfront in the design of the registry. This involves the production of a "data management manual" that explicitly details how data will be queried and the steps that will be taken to resolve data conflicts. The data dictionary and data validation rules will need to be specified. Data cleaning methodology might include automatic periodic query reports, automatic data cleaning algorithms and manual data cleaning and query reports.

As with other aspects of registry initiation, it would be important to include data cleaning in any pilot testing phase to ensure that the data cleaning procedures are adequate.

When anomalies or errors are found in data cleaning routines, it is important that problematic data is (at least initially) retained with the registry without being removed until the uncertainty is addressed. The data cleaning routines should go back to the original source of the data whenever possible. This will minimize the potential for bias to be introduced by a third party causing further errors by incorrectly assuming what value the

anomalous data should take; (i.e.: a registry technician thinking that they know "what that person meant"). Furthermore, the registry needs to have a mechanism to track and record when data changes are made which can be analyzed in the future as needed. This applies not only to data cleaning, but also in situations where incomplete records are later appended.

If multiple data entry methods are being used to populate the database, the data cleaning routines may need to take this into account. While this could be used as a way to help ensure data accuracy (eg: if different data sources contain the same information), the data management manual will have to explicitly detail how to resolve conflicts that arise because of different values in different data sources.

Quality Control

In addition to data cleaning, more comprehensive auditing of the registry contents may be required. As with all aspects of registry design and implementation, audits will vary in scope, frequency and location (i.e.: either onsite at data collection points or remotely) depending on the requirements of a specific registry and the funding constraints. Audits are important to ensure that training of data collectors is adequate and that data is complete and consistent. Audits can be conducted on a random basis or alternatively, for-cause audits can be created it there is a concern for example about a particular data collection site or a particular data field. The procedure for future planned audits should be incorporated into the registry design so that data collection methods are audit friendly. Furthermore, the registry should be structured such that a certain level of change within the collected data elements automatically triggers an audit. Such an audit would help to identify the root cause(s) of the change; is it because of simple error, or is there a systemic issue.

Quality Plan

Registries should document their quality management practices into a quality management plan. Quality management plans will address how, when and where quality activities will take place and who is responsible for them. Two key best practices in constructing quality plans include flowcharts and checklists. Flowcharts are diagrams that show the flow of data through the registry process and identify points at which quality breakdowns may occur. Checklists can be used to help to control quality and to communicate components of quality within a given process. The quality management plan can either be integrated into the data management plan, or developed as a separate standalone document.

RECOMMENDATIONS

✓ Consider the big picture when designing data collection procedures with quality in mind. Especially consider reducing data collection burden and enhancing participant engagement.

✓ Determine the best source for data being collected. If multiple sources are to be used ensure a methodology to address discrepancies is in place. Address uncertainties produced by vague data directly at the source.

✓ Consider the population from which data is being collected and tailor the data collection modality(ies) to suit the population's needs.

✓ Reconcile registry data with external data to ensure external validity.

✓ Ensure data collection process can be modified if quality assessments or pilot testing detect the need.

✓ Do not change data based on assumptions. This could distort the data or introduce bias.

✓ Develop a data cleaning plan during the registry design stage. Ensure the plan addresses any data linkage needs.

✓ Track when data are changed and the manner in which they are changed. Consider keeping multiple versions of a record instead of overwriting a single record.

✓ Define clear triggers for auditing. Audits should be random and for cause. Data should be collected in a manner that facilitates auditing.

✓ Develop a quality management plan. Ensure quality acceptance criteria and acceptable range (if applicable) are documented in the data management plan or quality plan.