ERICA

European Rare Disease Research Coordination and Support Action

Program title
Horizon 2020

Start date of project
01 March 2021

Duration
48 months

Project number
964908

Project full title
the European Rare dIsease Coordination and Support Action

Project acronym
ERICA

Project coordinator
Prof. A.M. Pereira, Leiden University Medical Center

Deliverable N°: D2.3
Title: Recommendations on registry data collection

WP N° and Title: **WP2 – Data Collection, Integration and Sharing**
Lead beneficiary: **UK-HD**
Type: **Report**
Dissemination level: **Confidential, only for members of the consortium**

Background

Rare disease (RD) registries aim to improve patient care through sharing of data, facilitate the planning of appropriate clinical trials, and support healthcare management [Kodra et al., 2018]. There has been a recent increase in the number of rare diseases registries in Europe, with more than 750 rare disease registries existing to date. However, these registries are often diverse in nature, funding, objectives and consequently non-interoperable. To overcome these limitations, there is a need to develop guidance for minimum standards necessary to maintain a high-quality and sustainable registry [Kodra et al., 2018]. This includes, but is not limited to, guidelines on establishing an efficient data collection framework, a suitable governance, making the collected data Findable, Accessible, Interoperable and Reusable (FAIR) and developing documentation. The objective of this report is to provide recommendations on RD registry data collection, based on the literature and discussions held during meetings of the ERICA WP2 EWGs on Data Collection and Data Quality.

By definition, a patient registry is "*an organised system that uses observational study methods to collect uniform data (clinical and other) to evaluated specified outcomes for a population defined by a particular disease, condition or exposure and that serves one or more predetermined scientific, clinical, or policy purposes*" [Glicklich et al., 2014]. Four primary use cases are typically identified: (1) describing the natural history of the diseases, (2) determining the clinical effectiveness of treatments, (3) assessing the safety of treatments and (4) evaluating or improving the quality of care [Boulanger et al., 2020]. It is therefore crucial for the data collection framework to be carefully designed to meet the registries' objectives in an efficient way.

Four different aspects of efficiency in terms of data collection are commonly identified, namely: (1) time and costs, (2) meaningful collection of data with respect to the objectives of the registry, (3) medium- and long-term sustainability of the data collection process, and (4) quality of the data collected. It is also crucial to integrate directly into the data collection framework the considerations related to the data protection regulation in force in the country of collection.

These different aspects require, prior to the beginning of data collection, to decide upon:

- the data elements to be collected, balancing the need for a broad dataset with the burden associated with data collection;
- the data elements to be collected should be clearly defined to ensure consistency in interpretation across the participating sites;
- if follow-up data are to be collected, it is necessary to decide not only on how often data will be collected but also how to encourage retention and minimise lost-to-follow up patients;
- the operational aspects related to data collection at the local care sites;
- how the data will be deleted in case patients withdraw their consent;

Respective recommendations for the ERN registries are described hereafter.

Objectives of the registry and general design considerations

The development and the use of a registry dataset is efficient if and only if there is a **clear definition of the objectives** the registry must reach and tracking of the outcomes is carefully planned [Viviani et al., 2014]. This is also necessary to evaluate the feasibility of success within the time frame of the available funding. Objectives can be reviewed and updated at a later stage when the data collection process is running smoothly.

In addition, it is necessary in early stage of the registry development to define unambiguously the **inclusion criteria** for the patients, which will ensure the uniformity of the population under study [Viviani et al., 2014]. This is typically based on a general agreement between expert clinicians involved in

the registry development. While these criteria can evolve in time, due to the discovery of new diseases, advances in knowledge and diagnostic capabilities, it is key to keep track of such changes.

Finally, the definition of **Key Performances Indicators (KPIs)** in early stage of the registry development allows to monitor if the registry reaches its objectives or not.

Choice of the data elements to be collected

During the planning of the registry, it is necessary to identify which aspects of the disease(s) will be recorded and identify the appropriate **data elements** to collect. In general, all the data elements collected in a registry should support one or several purposes of the registry or address a specific issue or need. In practice, it is advised to define first the analysis that will be conducted on the registry data and determine in a second step which data elements should be collected to fulfil these purposes. It is strongly advised at this stage of the development to consider the identified data elements and assess the **feasibility** of collecting them (i.e., are the data elements routinely collected in clinical practice, what is the cost of collecting them, etc.). One crucial point at that stage is to limit enthusiasm and the number of "nice to know" data elements to be collected, as it often leads to a substantial overestimation of the amount of information necessary to record [Viviani et al., 2014]. Collecting a larger number of data elements than strictly necessary is not desirable, as this automatically leads to a drastic increase of the burden of data collection (increase in complexity of the data collection system, increase in time and costs to collect the data). The choice of the data elements to be collected in a registry is typically made by consensus between expert clinicians, based on their knowledge and the existing literature.

Finally, it is reminded that the ERNs registries are required to collect the Common Data Elements (CDEs) developed by the Joint Research Centre (https://eu-rd-platform.jrc.ec.europa.eu/set-of-common-data-elements_en).

Definition of the data elements to be collected

Precise and explicit definitions of the data elements to be collected and how to collect them (i.e., methodology, unit of measurement) is key to ensure consistency across the collected data. It is strongly advised to develop a clear **data dictionary**, with explicit definitions of the data elements being collected and to provide adequate training to all the participating care sites joining the registry. To reduce ambiguities, it is also strongly recommended to use ontologies where possible (e.g., HPO codes for phenotypes, HGVS codes for genetic mutations, etc.) and to drastically limit free text fields.

Designing an efficient electronic data capture system

To reach its aims, a registry must set up an efficient electronic data capture system that coordinates data collection, data quality control, some level of data analysis and reporting [Viviani et al., 2014]. The electronic data capture system developed for the registry must be comprehensible to the end-user, be as automatised as possible, and the process of filling out data into the registry should fit into the normal practices of the healthcare provider.

Although currently unfeasible due to care sites regulations, if the option to automatically transfer data from the hospital Electronic Health Records (EHR) becomes available in the future, the design of the registry electronic data capture system should integrate the requirements for such an automatic transfer.

To improve data quality upon collection, it is strongly advised to implement **automatic logic checks** to prevent technically and/or biologically unplausible data from being saved into the database (e.g., a BMI < 10 is impossible, and systolic must be higher than diastolic blood pressure; saving of corresponding

data values should not be allowed). Missing data entries should be limited by maximizing the number of compulsory data fields.

Operational aspects of data collection at the local care sites

The successful implementation and long-term sustainability of a registry is strictly connected to a number of operational and logistical aspects. During the development stage of the registry, it is crucial to plan how the **data collection process will be integrated** into the local care sites' daily routine.

It is advised to identify, at each local care site, a limited number of **dedicated team members** (typically one to three) that will be responsible for the data entries. These responsible people should be provided with adequate training to develop an **efficient routine for entering the data** into the registry electronic data capture system.

It is also important for the local care sites to plan and implement a **routine for informing new patients** about the registry and request their consent. A routine should also be set up for contacting prevalent patients to inform them about any change in the registry or, if they gave consent, of any clinical project or research study related to their condition.

To ensure high data quality, it is strongly recommended that the central management team regularly analyses the quality of the data provided by each local care site and **send queries/requests to the local site teams** regarding due/overdue/pending or erroneous data entries.

A **routine for processing of queries/requests** from the central management team should also be set up by the local care sites.

Finally, identifying and securing the **budget required to initiate and maintain on-site data collection activities** is a key point to success. For long-term sustainability, registries should seek funding from multiple and complementary sources [Kodra et al., 2018].

Data protection legislation considerations

It is necessary to ensure that the patient identity and its personal data is **protected with state-of-the-art measures** that are compliant with the data protection legislation in force in the country of collection [Viviani et al., 2014]. Only pseudonymised data should be collected so that only the treating physician is able to link a specific pseudonym to the patient.

The data should be stored on a secured server, and access to the data or to the electronic data capture system limited to authorised users only.

It is also necessary to plan in the early stages of registry development standardised procedures for the removal of consent, ensuring that the data from a patient who withdraw their consent is deleted from the registry as promptly as possible (maximum within 30 days). This includes but is not limited to: deleting the record(s) of the corresponding patient from the database and its backups, deleting the patient pseudonym from the pseudonymisation tool used by the registry, and indicating in the patient file at the local care site that the patient does not wish anymore to be followed in the registry.

Summary/Conclusions

The setting up and the maintenance of a registry is difficult and challenging, and experience from past registries show how the development of an efficient data collection framework is crucial for success. Not only the technical part, which was discussed in this report, is crucial, but it is also necessary to

dedicate enough funding to collect, check and analyse the data. Finally, it is strongly advised, upon the development of a registry, to pilot the recording of few variables and test the applicability of the system designed before a full-scale deployment.

The deliverable is on schedule.

References

Boulanger, V., Schlemmer, M., Rossov, S., Seebald, A., & Gavin, P. (2020). Establishing patient registries for rare diseases: rationale and challenges. *Pharmaceutical Medicine*, *34*(3), 185-190.

Gliklich, R. E., Dreyer, N. A., & Leavy, M. B. (Eds.). (2014). Registries for evaluating patient outcomes: a user's guide. *Effective Health Care Program.*

Kodra, Y., Weinbach, J., Posada-De-La-Paz, M., Coi, A., Lemonnier, S. L., Van Enckevort, D., ... & Taruscio, D. (2018). Recommendations for improving the quality of rare disease registries. *International journal of environmental research and public health*, *15*(8), 1644.

Viviani, L., Zolin, A., Mehta, A., & Olesen, H. V. (2014). The European Cystic Fibrosis Society Patient Registry: valuable lessons learned on how to sustain a disease registry. *Orphanet journal of rare diseases*, *9*(1), 1-14.