Deliverable D4

Guidelines for data sources and quality for RD Registries in Europe

Manuel Posada#, PhD, MD; Laura del Otero#, Ana Villaverde#, Pharm; Verónica Alonso#, PhD, Biol; Manuel Hens#, PhD, MD; and Ignacio Abaitua#, MD; María José Carroquino, PhD, Biol; Josep Jiménez**, PhD, MD, Luciano Vittozzi, PhD* and Domenica Taruscio, PhD, MD*.

# Institute of Rare Diseases Research (Instituto de Investigación de Enfermedades Raras - IIER), Instituto de Salud Carlos III (Instituto de Salud Carlos III - ISCIII); SpainRDR and Consortium for Biomedical Research in Rare Diseases (Centro de Investigación Biomédica en Red de Enfermedades Raras -CIBERER)

** Catalanian Health Service (Servei Català de la Salut), SpainRDR

* EPIRARE Project. Coordinating Team. National Center for Rare Diseases (NCMR).
Building consensus and synergies for the registration of rare diseases patients in Europe: the EPIRARE project
Leadership by Domenica Taruscio, CNMR, ISS

**Aims:** To agree on a Common data set, disease-specific data collection and data validation (Objective # 4) to conduct epidemiological research on rare diseases at the EU level.

**WP7.** Data Quality, Validation and Data Source Integration in Rare Disease Registries

**WP Leader:** Manuel Posada, IIER, ISCIII

**Deliverable:** Report on guidelines for data sources and quality of RD Registries in Europe: section Quality Criteria - Benefits from the assessment report by the EU Scientific Committees (WP4)

**Developing Timing:** From 9th to 24th months

**Authors:** Manuel Posada**, PhD, MD; Laura del Otero**, Ana Villaverde**, Pharm; Verónica Alonso**, PhD, Biol; Manuel Hens**, PhD, MD; and Ignacio Abaitua**, MD; María José Carroquino, PhD, Biol; Josep Jiménez**, PhD, MD, Luciano Vittozzi, PhD* and Domenica Taruscio, PhD*, MD.

**Affiliations:** # Institute of Rare Diseases Research (Instituto de Investigación de Enfermedades Raras - IIER), Instituto de Salud Carlos III (Instituto de Salud Carlos III - ISCIII); SpainRDR and Consortium for Biomedical Research in Rare Diseases (Centro de Investigación Biomédica en Red de Enfermedades Raras -CIBERER)

**Affiliations:** ** Catalonian Health Service (Servei Català de la Salut), SpainRDR

**Affiliations:** *EPIRARE Project. Coordinating Team. National Center for Rare Diseases (NCMR).

**Date:** Feb, 28th, 2014

**Citation:** Manuel Posada, Laura del Otero, Ana Villaverde, Verónica Alonso, Manuel Hens, Ignacio Abaitua, María José Carroquino, Josep Jiménez, & Luciano Vittozzi and Domenica Tarsucio. Data Quality, Validation and Data Source Integration in Rare Disease Registries. WP 7 deliverable EPIRARE project, 2014
# Contents/Outline

## Executive summary

5-9

## Section 1  Overview

10-29

### Chapter 1  Introduction

11-13

1.1 Purposes
1.2 Scope
1.3 Methods

### Chapter 2  Background

14-16

2.4 European actions in the field of RD registries
2.5 Other international actions in the field of RD registries
2.6 The EPIRARE project

### Chapter 3  Terminology and RD registry specificities

17-23

3.7 Registry definition
3.8 Quality concepts
3.9 Quality and Health Care Systems
3.10 Quality and Research
3.11 Internal and external validity

### Chapter 4  RD registries information

24-29

4.12 Literature review
4.13 Orphanet information
4.14 EPIRARE survey: Quality analysis

## Section 2  Quality topics

30-52

### Chapter 5  Guidelines for data sources and quality of RD registries in Europe

31-39

5.15 Quality of RD registries
5.16 RD registry tasks involving quality assurance
5.17 Design phase
5.18 Geographical, target and study population definitions and temporal settings
5.19 Design
5.20 Case-definition and Case-inclusion/case-exclusion criteria
5.21 Sources of cases
5.22 List of co-variables, data dictionary and codification
5.23 Data stored, access and safety
5.24 Operating manual including standardized procedures and report forms.
5.25 Pilot Phase
5.26 Training
5.27 Pilot testing
5.28 Registry Activities
5.29 Data collection
5.30 Case ascertainment
5.31 Standardized data report form (DRF)
5.32 Data Completeness
5.33 Data delivery and ethical issues
5.34 Reporting
5.35 Patient involvement
5.36 Assessment Phase
5.37 External assessment

Chapter 6  Addressing systematization of quality procedures  40-46
6.38 The development of a quality assurance plan
6.39 Quality control measures
6.40 Data security and confidentiality
6.41 Follow-up
6.42 Timeliness
6.43 Reporting
6.44 Coordination
6.45 Quality assurance in children registries: A special situation

Chapter 7  Quality Indicators  47-52
7.46 Types of quality indicators
7.47 Surveillance of RD registry activities
7.48 Case ascertainment
7.49 Analysis of data completeness
7.50 Consistency
7.51 Timelines
7.52 Data security and confidentiality
7.53 Validity
7.54 Summary

Section 3  Quality assurance plan for the European platform for rare
diseases registries and databases  53-65

Chapter 8  EPIRARE commitments and RDR platform  54-59
8.54 Quality Assurance and the RD registries Platform: General
principles
8.55 RD registry’s QAP vs RDR platform’s QAP
8.56 RDR platform and Quality Assurance implications
8.57 Services to be provided by the RDR platform
8.58 Public Health Area
8.59 Research-development Area
8.60 Quality Assurance Plan Characteristics
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.61</td>
<td>Quality assurance of the platform management processes</td>
<td></td>
</tr>
<tr>
<td>8.62</td>
<td>Governance and Quality</td>
<td></td>
</tr>
<tr>
<td>8.63</td>
<td>Ethics and Quality</td>
<td></td>
</tr>
<tr>
<td><strong>Chapter 9</strong></td>
<td>Quality and scientific validity and reliability of the platform outcomes</td>
<td>60-65</td>
</tr>
<tr>
<td>9.64</td>
<td>Manual of Procedures</td>
<td></td>
</tr>
<tr>
<td>9.65</td>
<td>Data quality</td>
<td></td>
</tr>
<tr>
<td>9.66</td>
<td>Outcomes Quality</td>
<td></td>
</tr>
<tr>
<td>9.66.1</td>
<td>Process indicators</td>
<td></td>
</tr>
<tr>
<td>9.66.2</td>
<td>Monitoring indicators</td>
<td></td>
</tr>
<tr>
<td>9.66.3</td>
<td>Outcome indicators</td>
<td></td>
</tr>
<tr>
<td>9.67</td>
<td>Quality assurance (user satisfaction assessment) of tools and services, including promotion of registration and networking</td>
<td></td>
</tr>
<tr>
<td>9.68</td>
<td>Identify improvement opportunities.</td>
<td></td>
</tr>
<tr>
<td>9.69</td>
<td>Tools and Resources Quality</td>
<td></td>
</tr>
<tr>
<td>9.70</td>
<td>Summary and conclusions</td>
<td></td>
</tr>
</tbody>
</table>

**References** 66-76

**List of abbreviations** 77-78
Overview of the documents produced by EPIRARE

D1.1 The current context of activities and policies on RD (WP7, WP8)

D1.2 Survey on the expectations and needs of patients (WP5)

D1.3 Survey on the situation and needs of registries (all WPs)

D1.4 Statistical Analysis of the EPIRARE survey data (Parts 1 and 2) (WP6)

D1.5 Data mining on the EPIRARE survey data (WP6)

D3 Proposed Aims, Scope, Governance and Sustainability options for a European Platform for Rare Disease Registries (WP5)

D5 Developing a European Platform for Rare Disease Registries (WP8)

D4 Guidelines for data sources and quality for RD Registries in Europe (WP7)

D9.1.1 Overview of the scientific literature on Common Data Elements in the Rare Diseases Registries setting (WP6)

D9.1 Report on the survey on Common Data Elements (WP8)

D9.2 Analysis of the EPIRARE survey on registries data elements (WP6)

D9.3 Proposal for a Platform set of Common Data Elements (WP8)

D2.1 The Legal and Ethical Framework of EU Rare Disease Policies (WP4)

D2.2 Epidemiological and public health considerations for the EPIRARE briefing document on RD and data protection. (WP7)

D2.3 Amendments to the Draft Regulations on General Personal Data Protection (WP4, WP5, WP8)

D2.4 Briefing to the European Parliament (WP4, WP5, WP8)

Disclaimer

The contents of this document is in the sole responsibility of the Authors; The Executive Agency for Health and Consumers is not responsible for any use that may be made of the information contained herein.
Executive summary
Executive summary

Rare Diseases Registries are important epidemiological tools for health policy makers and researchers working in the field of low prevalence diseases. The quality of procedures used when a rare diseases (RD) registry is defined and also during the first steps of their development sets the basis its success and it is at the same time the best way to guarantee the long-term sustainability. Therefore the quality of RD registries is one of the key questions, to be assured and designed for during the first steps of their design.

To address quality, it is first necessary to separate between these two related concepts: quality and outcomes´ validity. What is the framework of these two terms?, where are they at play? Where do they play a role?, and what are their main aims?.

- Quality applies to planning of a either registry or platform of registries and procedures associated to their development. They can also apply to the periodical control and assessment of those procedures assuring the reliability of those procedures through the time of the registry
- Validity applies to the final outcomes or results. Internal validity, which is the only possible achievement to be gotten in only one observational study, should be checked by the presence of biases and, if so their impact in the final results
- While quality functions in both types of activities (registries and platform of registries), validity mainly works at the registry level. Although a platform of registries have to provide valid data, they should be based in the validity of the data provided by their registry partners
- Quality is aimed to standardization while validity is aimed to make inferences to the target populations from where the study sample belongs.
- Quality does not provide clues by itself to modify clinical practice criteria and policy making decisions but health outcomes validity come from RD registries are usually important sources of information which can be translated to the health care system

The second group of definitions comprises those terms related with the quality. In this way, these are the definitions to be considered

- Data quality. Features and characteristics of a data set, that bearing on its ability to satisfy the needs that result from the intended use of the data
- Quality Assurance (QA). Activities undertaken before data collection to ensure that the data are of the highest possible quality at the time of collection
- Quality Control (QC). Activities undertaken during and after data collection aimed at identifying and correcting sources of data errors
- Quality Assessment (Qass). Process of quality evaluation of the consolidated database
- Quality Results (QR). Value or set of values resulting from applying a data quality measures
Quality Indicators (QInd). List of quality measurements

The following recommendations are made separately for registries and for platform of registries because as explained, they have to be differentiated.

Recommendations for the quality of RD registries

Addressed to the design and implementation

1. As registries are in the framework of the epidemiological observational study designs, their development should bear in mind all steps which apply to these types of studies, from their original definition of aims and scopes to the presentation of the final results achieved
2. The starting point has to be a good definition of aims and scopes and thereafter the bases for the rest of methods to be applied have to be set up, including quality related topics
3. A detailed and comprehensive list of actions has to be defined, and associated procedures should be clearly described from the beginning of the process of the RD registry construction
4. A Manual of Procedures containing all actions, procedures, and registries associated to the procedures (i.e.: incidental questions, backup dates, partners list and their capacities, etc) are two fundamental documents of the quality assurance plan
5. The Manual of Procedures should also contain a list of Common Data Elements including their characteristics. From them, a minimum data set from the whole CDE should be stated as mandatory. They will be the fundamental elements for exploring the outcomes validity.
6. A case standardized form should be provided to the partners and data have to be curated before being included in the database
7. Standardized Operating Processes have to be developed for each one of the procedures included in the registry
8. The use of some of the current quality guidelines for publishing results (i.e.: STROBE; STRENGTH; PRISMA, etc) is highly recommendable to communicate registries results

Addressed to the quality control

1. A quality assurance plan has to be designed, which should include quality criteria, quality indicators, quality control mechanisms, quality assessment processes and quality results
2. The list of quality of indicators would contain topics related with the process, the monitoring and outcomes of the registry (i.e.: Surveillance of RD registry activities, Case ascertainment, Analysis of data completeness, Consistency, Timelines, Data security and confidentiality and Validity)
3. Some assessment phases have to be included during the life of the registry at different periods. At each of these phases, a revision should be made of the development of a quality assurance plan and the quality control measures, indicators, data security and confidentiality, timeliness, reporting, coordination. The assessment phases should also include some external assessment

**Addressed to the cases and their data**

1. Inclusion and exclusion case criteria as well as target population should be defined
2. Analysis of sources of information and their capacity of providing valid information should be explored
3. Case selection and case ascertainment are the two most important questions to minimize the selection bias
4. Control of duplicates and minimizing the mistakes in the interpretation and diagnosis are important clues for the quality
5. Errors in coding, data entry, data transformation, data consistency across sites and over time and Intentional errors should be care using electronic forms, personal acting as a data curators, external audits, among some other mechanisms
6. Reliability and data accuracy have to be frequently explored
7. Data completeness must be checked

**RD Registries Platform recommendations**

**General considerations**

1. Registry Description, Sponsor and Conditions of Access, Registry Design, Eligibility for enrollment, Common Data Element Groups, Manuals of procedures, Progress Report could vary in size but not in the general principles, which are essentially the same, regarding the registry development
2. Use of an indicators’ checklist is also similar although with different scopes
3. Governance process and ethical issues involve much more complexity in platforms
4. Management IT tools system is more complete and complicated because it is necessary to add procedures for each one of their components
5. The development of Standard Operating Procedures (SOP) show also a high level of complexity because they should define the relationships among several type of stakeholders and aims
6. A problems solution system and mechanisms for making decisions should be also implemented but aimed to supra aims instead specific aims, which are addressed in a registry

**Quality assurance plan**
1. The type of data in the platform comes from the metadata of each participant RD registry. Mechanisms to aggregate information data and methods to de-identify subjects should be part of this quality assurance plan.

2. The type of providers has to define the registries owner, the European Reference Network scheme and the national partners.

3. Outcomes and outputs have to be addressed to provide epidemiological information on RD (including natural history of RD), source of patients for research studies, tools for surveillance of drug side effects and cost-effectiveness analysis.

4. Governance rules, ELSI criteria and coordination of activities among partners are important questions to be taken into account for a RDR platform.
Section 1  Overview

Chapter 1   Introduction
1.1       Purposes
1.2       Scope
1.3       Methods
Chapter 1. Introduction

1.1 Purposes

This report seeks to ensure background knowledge in the area of data quality and validity in RD-registering activities, as fundamental criteria for further development RD-registries. Usually, several different data sources of information are used. This renders usefulness of the registry to a variety of purposes. Because of this diversity, it is important to consider the internal and external validity of both primary and secondary sources, as well as the completeness of information. Accordingly, this report will address the following aspects: design and operational aspects (such as data elements\(^1\) and data sources, communication and processing); registry data quality and findings; completeness of all variables; case ascertainment; inclusion/exclusion criteria; case-definitions; coding and classification; monitoring; procedures; and health outcomes\(^2\).

This document will be also useful to support the necessary data-source integration in a common Europe-wide repository. The opportunities afforded by the collaboration of patients and their organisations in terms of increasing the completeness of case registration is also considered in this review, which includes a critical review of the quality of the papers that provide evidence of this contribution.

1.2 Scope

This report seeks to provide high level criteria for data-source and data-quality guidelines for RD registries in Europe. It reviews existing literature on this subject, and compiles the feedback of experts and stakeholders the own experiences of the EPIRARE partners.

This document also includes a set of recommendations to be useful for epidemiologists, clinic researchers, patient organizations, public health policy makers and in general governments and funding agencies. The ultimate goal is to support health planners in their decisions with respect to building and funding RD registries as one of the ways to provide the best quality level of health services and interventions for those living with a rare disease.

1.3 Methods

Most of the data and information used to develop this report were collected through a literature review, personal and external expertise and engagement of stakeholders. The

---

\(^1\) **Data Element (DE)**: any named unit of data used to record information in a registry or database. It is characterized by a name, a definition, representation terms and the set, range and/or format of values.

\(^2\) **Outcomes**: something that follows from an action, situation; result; consequence. It refers to the effect produced by some action or cause
content of the following RD registries guidelines documents have also been reviewed for relevance to the purposes of this report:


A previous literature review was also carried out and reported in April, 2012 as a previous EPIRARE project commitment Posada et al, 2012). A summary of results of this literature review report is included in this report in the current situation section.
Section 1  Overview

Chapter 2  Background
2.4 European actions in the field of RD registries
2.5 Other international actions in the field of RD registries
2.6 The EPIRARE project
Chapter 2. Background

2.4 European actions in the field of RD registries

The European Community (EC) acknowledges the relevance of registries as key instruments for developing rare disease (RD) centred clinical research, enhancing patient care and health planning, and improving social, economic and quality-of-life outcomes. In its RD Communication, the European Commission stressed the strategic importance of RD registries (Commission of the European Communities, 2008; Council recommendation on an action in the field of rare diseases, 2009). Indeed, the extent of RDs among the population must be known, reliable data on the epidemiology of such diseases must be made available, and there must be registers of these diseases to inform health authorities about the population burden that they entail, as a whole and individually. In the same vein, the EC stated, "Areas to be supported by the MS [Member States] and the European Commission include: quality standards, including development of strategies and tools for periodical monitoring of the quality of databases and for database upkeep; a minimum common set of data to be collected for epidemiological and public health purposes; attention to user-friendliness, transparency and connectivity of databases; intellectual property, communication between databases/registries (genetic, more generically diagnostic, clinical, surveillance-driven, etc.). Importance should be given to linking international (European) databases to national and/or regional databases, when existing" (Commission of the European Communities, 2008; Council recommendation on an action in the field of rare diseases, 2009). It is noteworthy that "quality of databases" is listed as a major topic, along with some other aspects closely related to the quality of information.

The European Commission has funded through both the European Agency for Health and Consumers and DG Research several projects, surveillance projects and European Reference Networks which included the development of RD registries of specific diseases or related group of RD diseases.

2.5 Other international actions in the field of RD registries

A large number of organisations and researchers around the world maintain RD registries, of which no information is publicly available to researchers or health-policy makers on the uniformity of the data or accepted standards governing their collection, organisation and accessibility, among other issues. In some cases, this state of affairs results in wasteful duplication, and in others, in under-use of information furnished by patients or research participants. This raises doubts on the benefits for the population and the long term sustainability of the registry. Fortunately, the National Institutes of Health (NIH) has also launched important initiatives to tackle all these problems, such as the Global Rare Diseases Patient Registry and Data Repository3 - GRDR (GRDR, 2012; Yaffa, 2010).

3 Data Repository: it refers to the platform component dedicated to the data collection from the participating registries and databases. It may also imply some ability to retrieve information from it.
2.6 The EPIRARE project

One of the main important actions in the RD this field funded by the European Agency Health and Consumers (EAHC), is “The European Platform for Rare Disease Registries-EPIRARE” (EPIRARE, 2012), which main mission is to provide RD methods and guides for European researchers and policy makers. The EPIRARE is also aimed at agreeing on a common RD data set, disease-specific data collection and data validation, among other goals. The task of defining a common data set is independent of the rare disease registered, and furnishes information consistent with the agreed scope of the registry and/or underlying platform. It is also useful for public health actions.

Taking this general aim into account, the remit of the EPIRARE work package 7 (WP7) is to assess: i) the need for implementing an external quality-control scheme to monitor platform registration-activity compliance with inclusion criteria, and define the principles for performing activity; ii) the need for transnational pooling of data-collections on RD patients, in order to ensure sufficient statistical power for clinical and public health research aimed at providing high-quality care to RD patients; iii) the necessary data-source integration and completeness of reporting, so as to obtain reliable results with respect to different aims and data analysis and ways of assessing this. The WP7's main deliverable is entitled, "Guidelines for data sources and quality of RD Registries in Europe: Quality Criteria section" and will be delivered at month 24.
Section 1  Overview

Chapter 3  Terminology and RD registry specificities

3.7  Registry definition
3.8  Quality concepts
3.9  Quality and Health Care Systems
3.10 Quality and Research
3.11 Internal and external validity
Chapter 3. Terminology and RD registry specificities

3.7 Registry definition

A disease registry is a collection of persons having the disease of interest and belonging to a defined population. Its main objectives are usually: health surveillance and disease control; performing research into etiology; describing the natural history of disease; assessing clinical effectiveness and/or cost-effectiveness; assessing safety or harm of interventions; and measuring or improving quality of care (Gliklich, 2010, Gliklich, 2007). Moreover, in the RD field other important aims can be added to the above list including monitoring prevalence and incidence and providing an inventory of patients for clinical research (RDTF, 2011).

More simple, the WHO define a registry as “a file of documents containing uniform information about individual persons, collected in a systematic and comprehensive way, in order to serve a pre-determined scientific, clinical or policy purpose” (RDTF, 2011).

Whatever definition is selected, what is clear is that a registry involves a systematic and organized process of collecting observational data with long-term aims and sustainable activities. This is important because a registry shares methods with different types of observational studies. A registry follows similar organizational steps to observational studies but generally combining different study designs through time such as case series and cross-sectional study methods, case-control and cohort study designs. This is more evident when different sources of information are used or depending on specific objectives. In fact, experimental studies – like clinical trials – are one of the most important studies to be linked with a subset of patients, and are often included in a registry. In summary, a registry is the best way of pooling data to achieve a sufficient sample size for epidemiological and/or clinical research. Indeed, registries serve as a recruitment tool for the launch of studies focusing on disease aetiology, pathogenesis, diagnosis or therapy (Posada de la Paz, 2010; RDTF, 2011; Richesson, 2010). A second way in which a registry can help is in cases where the information collected includes biological specimens or links to specimen data (Rubinstein, 2010; Richesson, 2010). In brief, registries are important elements in a strategy targeted at accelerating research into rare diseases and the development of new drugs.

Traditionally, a registry is considered as an important tool for the Health Information Systems (HIS) at the National Health Service (NHS). In fact, registries are defined according to how their target populations are defined, and have traditionally been classified as population- (Lotus Mallbris, 2007; Gunnar Juliusson, 2012) or hospital-based (Drolet, 2008; Brian, 2008). The former seeks to estimate disease burden using incidence, mortality, trends and prevalence measurements. Nested case-control and case-cohort designs can be developed within the context of this type of registry, which affords the most powerful tool for health and social care, public health decision-making and surveillance and etiological research (van Walraven, 2010). Hospital-based registries focus on patient follow-up and the undertaking of clinical studies involving areas such as prognosis and clinical trials (Carla, 2011; Anna Sárkózy, 2008; Maurizio
Luisetti, 2010). More recently, a new conceptual framework, the so-called patient registry (patient-specific RD outcome registry) (DEcIDE Center, 2011, Forrest, 2011, Mehta, 2010, RDTF, 2011, Richesson, 2010, Wrobel, 2009; Samir Gupta, 2011) has been comprehensively discussed in the literature. A patient registry is more than a mere list of patients with a particular condition: it involves the systematic collection of uniform information for a specific purpose(s). Such registries are basically fed with specialized clinical information provided by systematic follow-up of patients affected by a given disease. Depending on their specific aims some of these patient registries can receive specific names such as diseases registry (etiological research, natural history, etc), patients registry (specific points of the natural history, prognosis and treatments assessment), post-marketing registry (drug safety and long-term effectiveness), products registry (important for cost-effectiveness assessment of intervention) (Richard Gliklich and Michelle Leavy. Patient registries and rare diseases. Applied Clinical Trials 2011: 20(3)

3.8 Quality concepts

The meaning commonly ascribed to the term "quality" is degree of excellence, as in, "a quality product". Yet, one could also define this term as fitness for the purpose. According to the International Standards Organization (ISO) definition (http://www.iso.org-2012), there are several terms used in association with the word "quality". The first of these is Quality Assurance (QA), a term applied where the intention is to design planned, systematic activities that will lend trust and confidence to the final product: in the case of registry activities, quality assurance aims to assure that the data will be in fact collected in accordance with the plan previously agreed and that the data, which will be stored in the registry database, will meet the requisite standards of quality, which are generally defined based on the intended purposes (Gliklich et al, 2007). The other important term is Quality Control (QC) which refers to the joint observation techniques and activities that are used to fulfill requirements for quality. In general, QA defines the standards to be followed in order to meet the requirements, whereas QC ensures that these defined standards are followed at every step (EMA, 2012).

Therefore, the quality process comprises several terms as QA, QC and three other important tools such as Quality Indicators (QInd), a term used to define a list of quality measurements within a quality process Quality Assessment (QAss), which refers to the process of quality evaluation, and Quality Results (QR), which refers to the value or set of values resulting from applying a data quality measure.

3.9 Quality and Health Care Systems

In national health systems, however, the term, "quality" tends to be reserved for health care outcomes and is usually included in the assessment of some certification process (Ulmer et al, 2010). The quality health care comprises several topics such as effectiveness, safety, patient/family centered, timeliness, access and efficiency. They are focused on three main aims of the health care: Prevention, treatment and long-term management. At the same time, these should be based on two global dimensions: the
added value provided by the interventions and ethics rules (AHRQ, 2011).

As the designated aims of rare disease registries include that of improving both the health planning and socio-health care of the registry's target population, the term, "quality", should be used in this field, not purely as a criterion for checking procedures, but also as a goal for assessing pre-defined final outcomes. In other words, the final aim of a RD registry should be to provide the best health care to patients and families throughout the improvement of the knowledge of the natural history of the diseases. This final objective clearly states that the quality process should be focused on health outcomes and not only in the procedures themselves.

3.10 Quality and Research

For many years, however, other terms, including "validity", "reliability" and "agreement", have been widely used, both to check those final health outcomes and in the sphere of health research in general. Furthermore, "accuracy" and "completeness" are also concepts used in the registry field. All these terms create a certain degree of confusion as regards the ultimate aim of a registration process.

3.11 Internal and external validity

The study of validity refers to absence of bias and is closely related with the absence of bias in the measurements of the main variables. Exposures and outcomes as well as other co-variables and confounding factors are considered main variables of clinical and epidemiological studies. To have a study free of bias is known as internal validity and it should be the main purposes of our work. In some literature internal validity is associated with accuracy of measurements. On the other hand, external validity refers to the generalization of our results to other target populations. In fact, the external validity is not a question for some types of specific study. It should be based on consensus after several studies are carried out in different populations and settings can show similar results – consistency⁴ – and also they are in agreement with the rest of existing knowledge about this topic – coherence – among some other criteria of causality (Coughlin, 2011). Some authors distinguish between two types of external validity, (i) whether the results of a study or trial are valid for patients other than those in the original study population but in a similar setting that the original study (‘external validity’); (ii) whether the results are valid for patients to whom they are generalizable but who are in a different setting than the original study population (‘applicability’) (Rothwell, 2010).

There are also important differences between the following two related concepts: agreement and reliability. Agreement points to the question, whether diagnoses, scores, or judgments are identical or similar or the degree to which they differ. In this situation,

---

⁴ Consistency: Close conformity between the findings in different samples, strata, or populations, or at different times or in different circumstances, or in studies conducted by different methods or investigators. In reference to RD indicators, it would mean that the information provided is comparable across time and places.
the absolute degree of measurement error is of interest. The term reliability is generally used to refer to the reproducibility of a measure, i.e. how consistently a measurement can be repeated on the same subjects and at under identical conditions. Lack of reliability may arise from divergences between observers or instruments of measurement or instability of the attribute being measured, or variability in the information sources. Reliability is typically defined as the ratio of variability between scores of the same subjects (Kottner, 2011). Reliability can be assessed as **Intramethod reliability** which measures the reproducibility of an instrument, either applied in the same manner to the same subjects at two or more points in time (**test–retest reliability**) or applied by two or more data collectors to the same subjects (**interrater reliability**). A third method is the **Intermethod reliability** which refers to a measure of the ability of two different instruments which measure the same variable to yield similar results on the same subjects. In general, this method is used as a gold standard and in this case, it would be method for the validation the measurement of a specific variable (White et al 2008).

The achievement of some level of accuracy is crucial for the assurance the study validity and it should be applicable to whatever type of research studies but mainly addressed to those using human subjects. The following issues are important points to be taken into account to deal with the internal validity study,

- Study aims and hypothesis with their background and bases
- Definition of the target population, settings and study population
- Study design
- Case definition and inclusion/exclusion criteria
- List of set of variables including sources and definitions
- Standardized data report form
- Pilot study phase description (not always is needed)
- Data collection methods
- Data interpretation/abstraction
- Data codification
- Data stored
- Data analysis
- Ethical issues
- Study limitations and bias analysis
- Standardized data reporting, including discussion of results and their relationships with other published studies.

In the latest years, several groups and organizations have been working in the standardization of study methods and reporting. As there are several types of study design, different methods have been suggested and proposed for checking studies quality and their results throughout the standardization of final reports and papers for publication (Sanderson, 2007). Among all these tools found in the literature to test the quality of studies, it is considered that some of them (listed below) can be applied as standards for registries and or at least for some specific questions concerning registries,
• STROBE - Strengthening the Reporting of Observational Studies in Epidemiology (Vandenbroucke, 2007)
• GRRAS - Guidelines for Reporting Reliability and Agreement Studies (Kottnera, 2011)
• STROBE-ME - STrengthening the Reporting of OBservational studies in Epidemiology (Gallo et al, 2011)
• STREGA - STrengthening the REporting of Genetic Association Studies (Little, 2009)
• GRIPS - Strengthening the reporting of genetic risk prediction studies (Janssens, 2011)
• CONSORT - Consolidated Standards of Reporting Trials (Schulz et al, 2010)
• PRISMA - Preferred Reporting Items for Systematic Reviews and Meta-Analyses (Moher, 2009)
• MOOSE - Meta-analysis Of Observational Studies in Epidemiology (Stroup, 2002)
• STARD - Standards for Reporting of Diagnostic Accuracy (Bossuyt, 2003)
• REMARK - REporting recommendations for tumor MARKer prognostic studies (McShane, 2006)
• GRACE - Good Research for Comparative Effectiveness (Dreyer et al, 2010)

Additionally, some basic criteria have been proposed as general criteria for validation this kind of tools.

Domains and criteria for evaluating each tool’s content (Sanderson, 2007)

• Methods for selecting study participants: Source population and inclusion/exclusion criteria
• Methods for measuring main exposure and outcome variables: Measurement methods
• Design-specific sources of bias (excluding confounding): Methods outlined to deal with any design-specific issues such as recall bias, interviewer bias, biased loss to follow or blinding
• Methods to control confounding: Design and/or analytical methods
• Statistical methods (excluding control of confounding): Use of statistics for primary analysis of effect
• Conflict of interest: Declarations of conflict of interest or identification of funding sources

The above suggested standard procedures and rules were convened mainly for reporting phases according with their study aims. However, they should also be considered in the study design phase, so that also study design and methodology fit with their corresponding standards.
Despite of above standards for high quality observational research and reporting exist, they are not consistently adopted, in part because of their complexity and the difficulty of including all components listed in published articles (Laine, 2012). A combination of both, creating a registry of observational studies, similar to a clinical trials registry, and a more detailed methodology about prospective statistical analysis plan (SAP) available in published papers have been proposed for the improvement of quality and accuracy in the observational studies (Laine, 2012, Williams, 2010).

Registries also aim at research objectives. They need to have standardized information to enable the development of existing knowledge to facilitate a better health policies decisions in health interventions (health care planning strategies and treatments). This is why, there are several aspects affecting the final quality. Indeed, we can say that quality does not result automatically from valid information in the registry by itself but that a quality process must be included some quality control of the data validity and procedures validity providing a final results validity. Therefore, registries cannot be limited their activities to collect standardized information but to promote good research results and hence they should use also external standards use for observational study designs.
Section 1  Overview

Chapter 4  RD registries information
4.12  Literature review
4.13  Orphanet information
4.14  EPIRARE survey: Quality analysis
Chapter 4. RD registries information

4.12 Literature review

The following literature search strategy was used in PubMed:

Search (((((control, quality[MeSH Major Topic]) OR data quality[MeSH Major Topic]) OR validity and reliability[MeSH Major Topic]) OR validity epidemiology[MeSH Major Topic]) AND registry[MeSH Major Topic])

Standardised terms were selected from among those included in the list of MeSH Major Terms. Documents eligible for inclusion were defined as those published from 1 December 2004 to November 2013 (ten years filter), including all major reports drawn up by agencies and policy papers.

Relevant details were extracted from each document (Table 1). Each document underwent analysis by epidemiologists (i.e., the authors of this report), who selected those papers that described methods which were in line with the main aims of this report. Specifically, only papers that addressed validity and reliability assessment and data-quality methods or principles were considered for analysis purposes.

Table 1. Search locations

<table>
<thead>
<tr>
<th>Type of location</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electronic database</td>
<td>Medline (PubMed)</td>
</tr>
<tr>
<td>General search engines</td>
<td>Google</td>
</tr>
<tr>
<td></td>
<td>As well as using the general search terms, we also searched for relevant</td>
</tr>
<tr>
<td></td>
<td>organisations, including patients and professional associations</td>
</tr>
<tr>
<td>Specific professional organisations</td>
<td>National health technological assessment agency websites</td>
</tr>
<tr>
<td>Councils and committees</td>
<td>European Union Committee of Experts on Rare Diseases (EUCERD); Rare</td>
</tr>
<tr>
<td></td>
<td>Disease Task Force (RDTF); IIER Ethics Committee; EPPOSI; California</td>
</tr>
<tr>
<td></td>
<td>HealthCare Foundation</td>
</tr>
<tr>
<td>Specific sites</td>
<td><a href="http://www.iso.org">www.iso.org</a>; Centre for Review and Dissemination</td>
</tr>
<tr>
<td></td>
<td><a href="http://www.crd.york.ac.uk/crdweb/">http://www.crd.york.ac.uk/crdweb/</a></td>
</tr>
</tbody>
</table>

The electronic database search retrieved 130 papers. Of these, 80 were excluded based on reading the title and/or abstract, leaving 50 for review.
In line with the definitions outlined above, procedures should be separated from outcomes. Procedures, which also include methods, have to use standard quality criteria, while outcomes have to be assessed from the standpoint of the validity and reliability of the results.

As an epidemiological tool, a registry can be viewed as an observational study design and/or research platform. A registry requires two things for its development, namely, systematic activities and quality control over such activities (von EE, 2008). Nevertheless, the quality of the process itself does not always have an impact on the final outcomes and their validity or utility to health systems in the same way (Zurriaga, 2006). To understand the relationships between quality and validity, tasks included in registry activities are listed below.

Building a registry calls for the implementation of several tasks which, though equally important, will not affect the final goals in the same way. These tasks can be broken down into those which are addressed at defining the study design, objectives and methods, and others which are addressed at systematising all procedures. One of the first steps is to define clear aims. The declaration of aims -including any hypotheses if present- should be a standard feature of all scientific action and reporting. The recording of aims and hypotheses in a registry prior to collection of data would assure readers that these had indeed been defined a priori (Desbiens, 2008).

4.13 Orphanet information

The Orphanet Report Series - Disease Registries in Europe - 2013, describes the state of RD registries declared in the Orphanet website. According to this source of information,
588 RD have been identified. 62 are from Europe, 35 international, 423 national, 65 regional, 3 undefined). Most of them are supported and funded by academic institutions and very few are sustained by either patient organizations or industry.

_Geographical coverage of rare disease registries registered in the Orphanet database (January 2013)_

In a previous report of the same series, the medical areas covered by all those registries were

_Medical areas with patient registries, 2011._
4.14 EPIRARE survey: Quality analysis

A survey of RD registries was recently conducted by the EPIRARE. This survey aimed to know the main activities, quality status and needs of existing RD registries in the EU to help identify options to develop and support the EPIRARE platform. The sampling frame was constituted by the Orphanet registry database and the Member States national focal points involved in the EPIRARE project. The questionnaire was made up of 12 thematic areas, with more than 60 different questions but only 9 of them related to the quality analysis of registry participants. A total of 272 registries provided answers to this survey but selecting criteria in terms of incompleteness of survey answer and lack of consistency among related variables lead us to eliminate 52 of the registries. Finally, 220 were included in the global survey analysis (See Deliverable 1.3.- Survey on the Activities and needs of existing RD registries).

The majority of registry participants in this survey had national coverage (61,4%) or regional coverage (16,8%). Considering all European and International registries together accounted for nearly 18% of the total. The design of these registries were population based (56,4%) and hospital based (23,6%). Other important survey results were,

- The most common aims of the registries was “Epidemiological research” (70,4%), “Clinical research” (60,9), “Natural history of the disease” (60,4%) and “Disease surveillance” (55,4%). The majority of them declared having very different aims from pharmacovigilance (53,2%).

- Case definition was available in 87,7%, and the inclusion and exclusion criteria were standardised in 78,6% of the registries.

- Almost the totality of registries collected data on the diagnosis (95%), while clinical and genetic data was collected in 86,8% and 72,3% respectively.
- The most common data sources were clinical units (83.1%), patients and families (48.1%), clinical genetics units (43.6%) and laboratories/central services (43.1%). Only 6.8% had been taken information from other secondary databases or administrative registries. Commonly, data entry process were developed by the registry’s staff (51.8%), or directly online by data providers (41.8%), albeit, patients collaboration were cited in 7.2% of the respondents.

- A highly variability in the disease coding classification used among registries was observed. The most common “coding system” used was indeed “No coding system, just the disease name” (34.5%), followed by the ICD10 (25.9%) and by own coding system (23.6%).

- The quasi totality of registries declared making periodical data updating (95%). However a non negligible part of registries do not collect the date of the patient’s death (26.4%).

- The main method used to avoid data entry mistakes were the automatic control (44.5%) while 22.2% of registries declared to not have specific methods for data entry avoiding mistakes. 87.7% of the registries had been developed methods to control of duplicates cases.

More specific and concrete questions about quality showed the following figures,

- Only 36% of the registries had a set of quality indicators.

- 48.6% of registries checked for reliability, 46.4% for the external validity and 58.2% for internal validity. Also 43.6% performed periodically quality/test surveys.

- The main needs declared related with quality of the registries were: to assess the quality of the data (41.8%); to link the registries with other registries (32.7%), biobanking and bioinformatics activities (22.2%) and to improve the training of the register’s staff (11.8%)
# Section 2 Quality topics

## Chapter 5 Guidelines for data sources and quality of RD registries in Europe

<table>
<thead>
<tr>
<th>Section</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.15</td>
<td>Quality of RD registries</td>
</tr>
<tr>
<td>5.16</td>
<td>RD registry tasks involving quality assurance</td>
</tr>
<tr>
<td>5.17</td>
<td>Design phase</td>
</tr>
<tr>
<td>5.18</td>
<td>Geographical, target and study population definitions and temporal settings</td>
</tr>
<tr>
<td>5.19</td>
<td>Design</td>
</tr>
<tr>
<td>5.20</td>
<td>Case-definition and Case-inclusion/case-exclusion criteria</td>
</tr>
<tr>
<td>5.21</td>
<td>Sources of cases</td>
</tr>
<tr>
<td>5.22</td>
<td>List of co-variables, data dictionary and codification</td>
</tr>
<tr>
<td>5.23</td>
<td>Data stored, access and safety</td>
</tr>
<tr>
<td>5.24</td>
<td>Operating manual including standardized procedures and report forms.</td>
</tr>
<tr>
<td>5.25</td>
<td>Pilot Phase</td>
</tr>
<tr>
<td>5.26</td>
<td>Training</td>
</tr>
<tr>
<td>5.27</td>
<td>Pilot testing</td>
</tr>
<tr>
<td>5.28</td>
<td>Registry Activities</td>
</tr>
<tr>
<td>5.29</td>
<td>Data collection</td>
</tr>
<tr>
<td>5.30</td>
<td>Case ascertainment</td>
</tr>
<tr>
<td>5.31</td>
<td>Standardized data report form (DRF)</td>
</tr>
<tr>
<td>5.32</td>
<td>Data Completeness</td>
</tr>
<tr>
<td>5.33</td>
<td>Data delivery and ethical issues</td>
</tr>
<tr>
<td>5.34</td>
<td>Reporting</td>
</tr>
<tr>
<td>5.35</td>
<td>Patient involvement</td>
</tr>
<tr>
<td>5.36</td>
<td>Assessment Phase</td>
</tr>
<tr>
<td>5.37</td>
<td>External assessment</td>
</tr>
</tbody>
</table>
Chapter 5. Guidelines for data sources and quality of RD registries in Europe

5.15 Quality of RD registries

Over the past years the number of medical registries has increased sharply. Their value strongly depends on the quality of the data contained in the registry (Arts et al 2002). Quality of registries is a global concept involving from activities starting from the aims and design to the reporting and the external assessment phases. A comprehensive list of topics involved in a whole RD registry process includes,

- Background and fundamental bases of the registry
- General and specific aims
- Geographical and long-term settings
- Target and study populations
- Design
- Case-definition
- Case-inclusion and -exclusion criteria
- Sources of cases
- List of co-variables
- Data model and data dictionary
- Identification of disease codes and co-variables codification
- Statistical analysis plan
- Operating manual including standardized procedures and report forms
- Training
- Pilot testing
- Data collection
- Data stored, access and safety
- Data delivery
- Quality control procedures
- Reporting
- Governance
- Patient involvement
- Long-term funding
- Ethical and legal issues
- External assessment

All these phases are also involved in the process of quality assurance which also includes issues such as governance, ethics and confidentiality rules, patient involvement and funding and sustainability issues (Gliklich et al, 2010; RDTF, 2011; EUCERD/EMA WORKSHOP REPORT, 2011; Richesson et al, 2010; DEcIDE Center, 2011). At the same time, all these phases are closely related although there is not needed to consider a hierarchical order among them, except for those items going naturally linked (Richard Gliklich, 2012). However, given that there are other EPIRARE work
packages dealing with major RD registries topics, for the purposes of this report, only specific features closely related to quality assurance, quality control and quality indicators will be commented. For other purposes and items, see the general RD registries reports cited in previous pages of this report.

5.16 RD registry tasks involving quality assurance

From the above comprehensive list of topics, we will focus our attention in those related with the quality and validity. We have classified them in: design, piloting, registry activities and assessment.

5.17 Design phase

5.18 Geographical, target and study population definitions and temporal settings

Regularly, the geographical and target population definitions are usually defined by the aims. The target population can be international, national, local, hospital based. From the quality assurance point of view, this topic could be a problem if researchers want to make extrapolations to the whole population from those living at some point in their lives in the area of the registry or if only current residents are considered from that geographical area. Methods for population-based registry have been defined to prevent the bias regarding the inferential process. At the same time, if we are considering a registry including specific patient outcomes – as a result of some intervention – the geographical setting definition become less concrete because patient interventions are not limited to a specific area. If this type of bias wants to be avoided, the target population should be equal to the population included in the registry.

Regarding the temporal setting, it is clear that a registry is not a study limited in the time. Generally, a registry is considered as a long-term and sustainable action, or in some cases the end can be pre-defined if the outcome has been correctly already evaluated or if the registry aims are clearly unachievable under the current circumstances (several factors such as funding, efficiency, lack of collaboration, etc).

In summary, these topics are very important if valid results have to be further applied to a very well defined population and therefore, the quality assurance process must take them into account checking if their definitions are feasible and according with aims, methods and resources.

5.19 Design

Globally considered, the design of a registry comprises many steps. Specifically, the design phase usually refers to the epidemiological observational methods used for the case ascertainment. This design can follow some of the traditional study designs already described in the literature such as case series, prospective cohort of patients or even a case-control study design, if a control population is needed for the aims of the registry. As a registry is not usually formed by a closed population, except if the cohort is defined by some exposure such as some external intervention, cross-sectional methods
are usually also used in combination with other basic ways of case ascertainment.

However, it is not always simple to define the observational method using the traditional epidemiologic terms. As an example, a simple method for case ascertainment consists in a single declaration – case by case – made from different sources by curators or even physicians not directly involved in the registry. In this case, the design can be considered as an opened cohort or simply a case series of patients under some specific diagnosis. In many other circumstances, primary databases are used to identify new cases and downloaded into the registry database. Indeed, the main question to be solved from the perspective of the registry design definition is if some other questions apart from the diagnose-related are included in the case definition and, and also if all cases, a randomized sample or just a convenience sample with the same diagnostic characteristics are considered in the registry. To answer these questions will lead to the researchers to a correct registry design definition. Quality will be linked in different forms with all of these types of designs and not all are able to guarantee accuracy data and inferences.

5.20 Case-definition and Case-inclusion/case-exclusion criteria

These points are important when building a registry. They should define what types of cases are going to be included (i.e.: diagnosed patients; specific patients undertaken to some intervention; specific patients diagnosed but with some specific features of that disease, etc). A correct definition, including criteria for inclusion/exclusion, is an important part of a correct quality process. These rules definitions will form the whole population of the registry and will facilitate the quality control procedures as well as the validity analysis.

5.21 Sources of cases

Once the above terms have been defined, analysis of the sources of information analysis is a crucial step for a quality case ascertainment process. If this analysis is not correctly made before the start the registry, it may occur that only a subset of mild-severe patients, whom will be better identified than slight cases, will be included in the registry and selection bias could be stated in the final results.

5.22 List of co-variables, data dictionary and codification

Other variables such as outcomes (if they have been defined), socio-demographic, clinical data, prognosis factors, follow-up information as well as some other considered of interest by researchers have to be listed, described and their procedures of ascertainment guaranteed. A data dictionary containing data structure, data meaning, data sources and data coding for storing and statistical analyses have to be included in this document. This is the only way to assure the quality of all information which should be included in the registry.

5.23 Data stored, access and safety
Data have to be saved and stored for a long time because it is the basic principle of a registry. Electronic forms that allow researchers to type the data directly into the computer are usually the best way for the data entry process. These types of electronic forms facilitate the control of the data entry reducing the mistakes inherent to this process. In some other situations, a data report form (DRF) hard paper is used in as a primary step although it should be typed later on by their own registrars or by the research leaders. A third alternative consists in some form of electronic data already standardized but containing several cases (Excel, Access, ASCII files, etc) which is globally uploaded into the central server where the registry is stored.

Access to the registry has to be controlled by the administrator. Some measures of security access should be implemented in the central system and a security document should be elaborated, clearly establishing policies and criteria for confidentiality and person rights.

5.24 Operating manual including standardized procedures and report forms

This manual should collect all steps, procedures and questions that may arise in the complex task of building a registry. The manual is the main document, which should contain all types of explanations about the registry functioning and solve any hypothetical question. Procedures are important parts of this manual because they describe all type of activities regarding data collection, cleaning, storing, monitoring, reviewing and reporting. Each procedure has to be clearly written, the workflow very well defined and the involvement of each person clearly stated. Difficulties and troubles during the process have to be prevented and alternate rule options including responsible person for making final decisions should be included in each procedure.

A quality assurance process is based in a manual. The quality control process will check if all things stated in that manual is followed and respected at all times. The manual closes the design phase and lead us to the operational work.

5.25 Pilot Phase

5.26 Training

One important step that must be developed before starting the pilot study is the training. Educational materials as well as standardized procedures have to be taught and present to curators and in general to all participants in the registry. This training phase is basic for the success of pilot study. At the same time, trainings insisting in specific weak points must be repeated if necessary along the life registry course. Documents used during training have to be available all time for the partners and officers because they can facilitate the solution of doubts and critical unspecific questions. A comprehensive training design would comprise the development of a training manual with standardized training examples, and individual or group sessions where problems and doubts can be clarified. A double-review of the first cases registered where abstractions methods and
codification can be assessed. On-site visits are also useful because they allow to partners to ask directly to the registry staff and clarify certain constrains. At the same time, an on-site visit shows added values to the partner because this engagement can be renewed and reinforced. Periodically double-review is also recommended based on certain times or number of cases registered by each partner. Conference calls can be also useful for having a permanent link between local and central sites (Lisa et al, 2003).

5.27 Pilot testing

After finishing the description of steps summarized above mainly focused on building a comprehensive list of designing topics and scientific policies rules, a pilot test is usually the next phase. All procedures and rules are checked and tested before the start of the project. Piloting phase can be also launched for specific but main procedures as well as for some subset of diseases. Constrains and limits are evidenced, solved and written in a new version of the procedures manual. Quality testing is sensible to this pilot phase because it can be improved and future mistakes and gaps in the quality assurance process be avoided.

5.28 Registry Activities

In this phase, the real world of a registry life is starting and if all procedures and rules have been carefully designed and care, all steps should be working without difficulties. However, it is not rare that in many occasions procedures do not work according to predicted plans. In that case, this phase will find out certain types of problems which they have to be solved by a governance committee.

5.29 Data collection

Data have to be collected in accordance with the procedures previously defined by the aims and purposes of the registry. This step comprises several activities which have to be adequately connected and standardized. The list of activities is collecting information and cleaning data, data storage process and data monitoring and reviewing. In this step two major actions should be highlighted: case ascertainment and data completeness.

5.30 Case ascertainment

Special attention is called for when it comes to defining the case-ascertainment procedure. Also known as a “sampling bias” this process is in charge of consists in the identification, verification and case selection from the sources of information. Clinical centers and records are the main source of cases to be included in the registry, though patient organisations are also an important source of cases in the RD field. Nonetheless, case reporting is not a simple process because it can be done directly by professionals - generally physicians- by patients or their guardians, if the patient in question is a minor or a person with some intellectual disability (The Society of Teachers of Family Medicine. 2011), or by patient organisations (Gliklich, 2011; Specialized Healthcare Alliance, 2011). In addition, albeit less frequently, this information can be directly
downloaded, using some electronic means, from clinical records or secondary case sources (i.e., hospital discharge registry, mortality registry, other). These types of sources -patient, clinician, electronic medical records (Dan Belletti, 2010; Jane Metzger, 2004)- and secondary sources are subject to different types of errors, ranging from the validity of the diagnosis to the completeness of the case report form. If the registry is based on all these sources, strict rules and periodic quality assessment is recommended. In contrast, if the registry is based on just one of these sources, special care must be devoted to the main source of errors associated with the selection process.

Nowadays, clinical records continue to be the main source of case data but they are not sufficiently systematised or standardised. This is important particularly in those registries where retrospective cases are recruited. But, if only new diagnosis cases are accepted, the registry should define a protocol that has to be accepted by the professionals involved in these tasks, and researchers have to check that such protocols are being correctly applied. Full protocols remain the most comprehensive source of methodological information and should be made publicly available (Reveiz, 2010). Yet a protocol will not necessarily influence the quality of clinical records, the lack of standardization of this important clinical document is still one of the main challenges. If, on the other hand, the registry protocol requires patients to be examined directly by researchers and the information extracted directly from them or through some analysis/test, quality levels, including that of diagnosis, can be adequately guaranteed.

Easier recruitment may perpetuate potential selection biases which do not allow for the possibility of assessing the real representativeness of the study population recruited to the registry (Hegedus, 2010). One of the methods for resolving this potential bias is to enroll patients from a wider population base, through primary and social care (Iliffe, 2011; Eigil, 2011). The number of missing patients in the database could be minimized by a concerted effort targeted at meticulously controlling the data and then comparing them against data housed in other national registries (Wille-Jorgensen, 2011).

5.31 Standardized data report form (SDRF)

After a case has been identified and the disease already codified following a previous and harmonized criterion of the coding list used by all researchers involved in the registry, the list of selected variables should be collected and coding following the rules declared in the data dictionary. This sheet of information is namely Data report Form (DRF) or Case Report Form. Both terms can be used as synonymous but in this report we use DRF to separate this report from those regarding to case report (case report declaration form a source). If paper form is used, it should be stored for future audit mechanisms and checking of errors. If the DRF is electronic, the software has to be programmed taking into account possible ranges of values and internal consistence with regard other variables (i.e.: a women coding with a X-linked disease; old ages do not allow for diseases with high mortality rate during the first years of life, etc). Data validations rules which refer to mechanisms (electronic or manual) to be implemented to be sure that information collected and stored are within the expected range of values for each specific variable included in the dataset.
5.32 Data Completeness

It is one of the main topics where registries fail because of researchers want to collect many variables at different times including repeated measurements and follow-up. Finally, if the pilot testing has not been very well developed, most of these variables agreed during the designing phase cannot be achieved or they are not very well standardized it will lead the registry to some potential risks of bias affecting to case selection and also to the case information. In many registries a Common Data Elements\(^5\) with a Minimum Common Data (MCD) set are defined giving the opportunity of collecting in a complete way the MCD but not all Common Data Elements (CDE). This strategy can be associated to some quality control mechanisms that do not accept new cases which do not have full MCD filled out. But they allow new cases to be included in the registry with some other secondary variable which is not filled out.

The registry completeness could be biased due to the identification cases because of high level of suspicion but not yet diagnosed cases have been included in the registry. The second important point affecting to the diagnosis is the uncertainty. In some registries such as cancer registries, a certain level of uncertainty cases is allowed. This is due the fact that some cancers are not histologically analyzed due to the natural history of this disease or they are identified in the index mortality registry with some lack of certainty in the diagnosis. Underreporting due to systematic mistakes in some rules of clinical procedures in the source of information is a third problem which share similarities with the two above mentioned. In all these situations, only a high degree of completeness in case-finding procedures will ensure incidence rates and survival proportions are close to their true value (Parkin et al 2009).

Some other rare diseases registries share similar difficulties to rare cancer registries. Congenital anomalies have in some cases very complex and their classification and final full diagnosis are not always correct. Underreporting and underdiagnosed are very familiar questions to be corrected in these types of registries. Autoimmune rare diseases and genetic diseases have the same problem because their complexity, lack of specific biomarkers or they are not accessible for all cases, and the difficulties for the standardization of their procedures and criteria.

5.33 Data delivery and ethical issues

In several occasions, a local registry is participating in geographically broader strategy such as national or international. In these situations where data may be transferred from the local site to a central site, procedures for this transfer have to be stated because of final quality and validity criteria. In principle, there are no differences between criteria

---

\(^5\) Common Data Element (CDE): any data element which is used in the data repository to collect information regarding the registry and database features and data. It is disease-unspecific and is adopted by all or groups of registries and databases participating in the platform.
methods used for each local registrar regarding the central site in a small setting or in a large geographical setting. Quality standards and procedures are the same but responsibilities can be substantially different. Both sites and stages (central and peripheral) may carry systematic or random bias in several or their activities and each of them has to ensure the correct functioning in their own areas of responsibilities and create efficient systems and procedures to detect these types of bias. Errors in the procedures designed, in the DRF, insufficient CDE and MCD definitions, software tools among some other are crucial criteria which have to revised, piloted and checked at each time.

Confidentiality and ethics are important issues to be considered in data delivery process. Each country and even supranational scenarios such as the European Union have strict legal rules and ethics procedures relating to data delivery either among partners or between central and local sites. Registries and ethics criteria constitute another EPIRARE important deliverable and this is why this topic is not dealing with in deeper in this specific deliverable – although it is considering as fundamental in the quality assurance process.

5.34 Reporting

The final aim of all types of registries is to provide information to the stakeholders including policy makers, physicians and patient organizations in order to update the information about the disease course and to suggest some evidenced intervention criteria. There are many types of reports which registries can standardize for their own aims and purposes. As we have indicated in one previous section of this report, several standardized reporting guidelines are currently available for different purposes and outcomes. Registries can report descriptive data, even in understandable bulletins for the media and patients but they should always care the quality and validity of the information. The reputation of a registry is closely associated to the validity of information that can provide through, bulletins, reports, papers and policy documents that disseminate registry information. These dissemination elements must fit with some of the standards already existing for them. One of the frequent mistakes incurred by registries consists on the dissemination of new data under the assumption that there are not other possible counterparts who can argue in the opposite site about results. They know that replication by other groups are not viable because they have the control of the whole information of that specific disease(s). The only way to limit these effects in the registry reporting is combine ethics rules with evidence base analysis of their results. Limits in the generalization of some types of results waiting for further confirmations of the same results that can give consistency to their results are basic strategies that credit scientifically to the registry.

---

6 Scenario: a predicted sequence of events, possible - hypothetical situations. But in the RD Platform context is: The initial scenario of the EPIRARE project assumes that the existing registries and databases are mainly spontaneous and addressing specific research goals (as resulting from the EPIRARE survey) and that the platform has no regulatory power; therefore, it cannot force existing registries to modify their operation to comply with centrally defined rules and data elements.
5.35 Patient involvement

Patients are important in registry developments. They can invite all stakeholders and in particular policy makers and regulators to create the conditions to allow the creation of disease registries. Patients often wish to be involved and can concretely contribute to the definition of the content and specific purposes of the registry (evaluation of quality of life - a more restrictive term exclusively used to describe specific measurements of how patients or healthy people feel about their own lives - and quality of care), governance (definition of best practices), contributing to the general reporting for patients, mass media and general public. They can give feedback for several other registry topics and provide an important help to the researchers in the preparation of specific information for patients (EUCERD/EMA WORKSHOP REPORT, 2001).

Patient involvement is crucial in patient registries where the informed consent is absolutely necessary by law to be included in the database. They can be passively engaged in the registry or they can be actively participating in some other nested studies in the registry (i.e.: observational and/or experimental studies, samples donations, etc). Conversely, this voluntary engagement has to be carefully organized because voluntary symptoms declarations always express a wider range of health concerns than do non-volunteers, and differences in the level of co-operation between ethnic groups should be taken into account, specifically where genetic information is one of the outcomes (Bishop, 2011). Mild and severe clinical forms can be included in the registry as a non-desirable of this voluntary cooperation. Hence, to provide an accurate estimate of the generalizability of studies conducted using genetics registries, the method of recruitment – and in particular voluntary recruitment - should be examined when interpreting and analyzing results (Henrikson, 2007). In addition, recruitment through volunteers does not prevent a pronounced, healthy volunteer effect on mortality analysis (Burnell, 2011). This is why further assessment of strategies for successful recruitment of minority participants in epidemiological studies and also in registries is warranted.

5.36 Assessment Phase

5.37 External assessment

External assessment of both procedures and quality control strategies is highly recommended because deviances from the main rules are easier to detect and correct though training or simply with slight modifications of the workflow. There are different methods to provide external assessment to the registry. The most common is to account with an advisory board (committee) constituted by experts in some of the registry topics, who are in charge of evaluating actions and results periodically. It is also possible to hire some external audit company, which provide credit and certify the registry quality. It is recommendable to have some internal mechanisms for the evaluation of the quality and validity of the registry and to harmonize these control internal rules with the external requirements.
Section 2  Quality topics

Chapter 6  Addressing systematization of quality procedures

6.38  The development of a quality assurance plan

6.39  Quality control measures

6.40  Data security and confidentiality

6.41  Follow-up

6.42  Timeliness

6.43  Reporting

6.44  Coordination

6.45  Quality assurance in children registries: A special situation
Chapter 6. Addressing systematization of quality procedures

After making a brief review of some of the main quality-related registry topics, it is important to answer other questions like, how to design a quality assurance plan, what kind of rules should be implemented for testing the quality control of that plan, and what are quality indicators which must be defined.

6.38 The development of a quality assurance plan

A quality assurance plan is defined taking into account all aspects already mentioned in this report and others in which have been discussed in other EPIRARE deliverables. Specific points of each one of the topics must be considered and rules and criteria very well established, defined, piloted and stated in the corresponding procedures. All together form the manual of procedures which included all aspects related with the quality, such as the quality plan itself, the quality control mechanisms, the quality indicators list and the program for a periodically activity quality assessment. Therefore, documentation is the main clue for a quality assurance plan and all documents constitute the Manual of Procedures. The lack of data quality assurance could produce random errors, while the lack of quality of certain procedures, including case ascertainment, would lead to systematic errors. Changes in question formats may significantly change the quality of register data. Compared to open-ended questions, check-boxes seem to improve quality. Both types of errors can be dangerous and have an impact on outcome validity. This is why all activities have an important role and none may be rejected or dismissed. In a broader sense, quality assurance should comprise all the tasks correctly documented listed below.

List of documents contained in the Manuals of procedures

- **Policy rules and governance:** Regulatory documents, including fundamentals, specific aims and governance criteria, registry owner, funding, sustainability plan …
- **IT tools document:** Description of all IT tools related questions such as server system, software supporting database, software versions, back-ups protocols, incidences and solutions …
- **Security document:** All rules controlling access, data safety, confidentiality, patient and professional rights, data delivery…
- **Ethic rules:** List of ethic criteria and rules that guarantee patient confidentiality, informed consent, data owner and data delivery. Conflict of interest are also considered under the scope of this point
- **List of procedures for each topic:** List of documents describing the workflow of each topic(s) with alternative plans for possible unforeseen contingences and, at the same time, they should care data validity, data completeness, control of data errors and actions to be taken for these problems. Central and local site procedures can be different because of respectively responsibilities.
- **Manual training:** Specific documents for each type of participant profile which helpful instructions for developing their works
• *Instructions for database users:* A manual where users can find how to proceed to access to the system and use the software according with their profiles.

• *Data dictionary:* Contains all variables, definition, data model, relationships with the others, coding as well as tables and their relationships.

• *Data Report Form:* The information sheet with the Common Data Elements to be searched for each individual case and which is included in the database later.

• *Catalogues:* Auxiliary tables such as center participants, diagnosed tests, medical proofs…

• *Classifications:* List of diseases to be registered with the code system defined by the researches. If more than one list is proposed, the relationships among them would be provided for harmonizing this crucial quality question

• *Personnel functions and tasks:* Documents describing each activity and competency for each individual person involved in registry activities

• *Checklists:* Containing short list of activities to be revised and marked after they have been developed by workers

• *Nested research study protocols:* Including protocols already defined as well as protocols for future proposals

• *Quality assessment:* Document containing all procedures, control mechanisms and methods for the correct evaluation of the registry and their frequency.

The Quality Assurance Plan could be a separate part of the Manual of Procedures or it should be contented in the own manual and their respectively parts, although the main sections of this plan have to be stated in the quality assessment report. The question of who has the responsibility for the data quality assurance has to be also clearly stated. Moreover, all above questions and their specific details have to be stated in very well written documents which are accessible electronically and/or paper hard copies for all participants and actors of the registry.

6.39 **Quality control measures**

The Manual of Procedures should include the quality control mechanisms and the tools to be used for these types of control of activities. Registry quality control refers to all activities but especially to those related with both data and procedures. The following main types of data errors should be identified: 1) Problems in case selection and case ascertainment; duplicates, mistakes in the interpretation and diagnosis; 2) errors in coding, data entry, data transformation, accuracy, 3) Data consistency across sites and over time and 4) Intentional errors.

Cases are identified from multiple *sources*, where a source is defined as any location where a case was reported. Using the various sources, cases are matched to identify duplicates across sources. The sources are grouped according to “modes” of ascertainment (Richesson et al, 2010). Data sources are important windows for bias and their quality should be analyzed and known before the start of the registry activities. Primary and secondary sources are the best of sources because they provide actual information from patients, clinicians or from the clinical records. The quality of the clinical records can vary across sites and also between electronic and paper forms. The
best tool for avoiding variability in this source of information is to collect prospective information previously standardized throughout appropriately trained professionals. Retrospective information always has problems and recall bias, uncertainty of either clinical features description and/or interpretation and lack of information are the main difficulties inherent to the collection of this type of information. In any case, data abstraction including codification and interpretation of clinical findings is also a problem even for prospective information. Training and periodical assessment of this procedure are the only possible tools for assuring the quality control of these important data collection process. However, data accuracy should be analyzed by comparing the registry data to the gold standard (g.s.) data, at least for the main variables of case definition, exposure and outcomes. The rest of the variables can undergo different types of quality control such as automated or manual cleaning, automated coding control and/or using specific queries which test the consistency among variables (i.e.: gender, age, ethnic group, disease…), as well data tracking throughout the pathway from the origin of the information to the database where they are finally stored.

Data in registries can also vary in quality and completeness, mostly due to time limitations on the part of physicians. Personal incentives can be used but they are not the best mechanism to control this constrain. On-site audits along with professional promotion, giving them opportunities for being authors in papers, conferences as well as to create an actual engagement network where each participant can consider their contribution as important, are better solutions to promote the completeness of the registry.

**6.40 Data security and confidentiality** are also important topics for the quality control. A security plan, control of Access rights with Electronic signatures under some periodic security assessment process are good tools for improving and controlling the quality of this area of activity of whatever registry.

**6.41 Follow up.** A patient follow-up is one of the main tasks of many RD registries. There are different objectives which can be susceptible to a follow-up such as mortality surveillance, prognosis factors assessment, evaluation of interventions, cost-effectiveness analysis, and description of the natural history of the disease, among some others. Follow-up can introduce some bias in the information even though the procedure is very well described and also controlled. External factors such as changes in diagnosis criteria lead us to modify our own case definition criterion and also codification rules which should be carefully adopted and harmonized with the past information. The Lost in follow-up process also introduce certain type of selection bias and can lead us to erroneous assertions about disease prognosis or treatment efficacy. If these losses are not random, the final effect in our outcome could be varied and the knowledge and analysis of the type of patient lost (severe or slight disease) is the most important point for a correct analysis of our outcomes. The same happens in the case that some efficacy treatment cures or limits the disease. If that is the case, patient can refuse to be followed for much more time and their lack of participation would create troubles in the results interpretation.
6.42 Timeliness. It is not rare that data registries have some trade-off between timely data and the extent to which they are complete and accurate. This delay is known as a timeliness of registry and refers to the time which is needed to consider that data is ready and useful for reporting and communication. This topic has not the same importance in all registries. Population-based registries and also patient outcome registries assessing side treatment effects have must speed all type of completeness and validity data in order to report immediately (hours, days) important public health and/or patients concerns which demand some policy decision (i.e: preventive actions; withdrawing an orphan drug for dangerous side effects, etc). In some other occasions, this delay is not critical and reporting can be delayed until data have been checked and validated. This is usually seen in patient registries where the natural history of diseases is the only aim under scope. However, timeliness is an important question in the registry quality control process. The reason why a registry is set up is because it can produce information useful to develop new actions and measures in either public health perspectives – population affected – or for individual patients. In fact, a registry is a tool to either count or characterize health or disease characteristics population, with the intent to extrapolate those results back to a larger or different population (Richesson et al, 2010). This information is considered as important and fundamental for the people under their scope, and delays or gaps in their dissemination is considered as a fail of the registry. Data completeness and data validity should be under strict surveillance and monitoring within the quality control process of the registry.

6.43 Reporting is closely linked with the above topic. There are many ways of dissemination of the information used by registries. Websites, guides, papers, bulletins and some other which modern communication technologies offer are means for communicating the information. Official reports addressed to the health authorities, however, are the most important type of reports to be used. As we commented before, reporting should be carefully planned among some other registering activities as well as includes in the quality assurance plan. Quality control measures of this topic should include standard reports, control of missing data from the patient database, patient involvement and provide accessible reports easier to read by the patients.

The observational nature of registries needs that analyses and reports be undertaken with the appropriate consideration of potential biases and confounders, using standard epidemiologic methodology. A quality control process attempts to assess the presence and potential influence of these biases and provides criteria and standardized rules for preventing their influence in the final results. Although there are recent documents and good guides for building RD registries, there are currently no standards for developing registry programs, information systems, or data collection instruments universally accepted. This is why, observational studies guidelines already described in the literature and summarized in the page XX of this report are useful tools that helped us and improving several sections and topics of the quality control of a registry.

---

7 Timeliness: means that information is made available to decision makers before it loses its ability to influence decisions (25). In reference to RD indicators it would mean that the information provided by the indicator is on time to assist with effective decision making.
RD registries collect information also coming from genetic diseases. These types of diseases require information on genetic factors (presence of genes, deletions, specific mutations and/or even description of the whole gene sequence) for a clear identification and diagnosis of the diseases and/or for a genotype-phenotype correlation analysis, among some others. Therefore, a correct genetic testing made by a certified lab is an important key for both the case clinical management and the case classification into the registry. As other basic variables, valid information of the genetic background has to be periodically tested and improved using some gold standard or the data entry should not be permitted if it has not been provided by lab with an external quality certification. Some difficulties can arise during this process because most of genetic tests are not useful for diagnosis but just for research, which means that there are no standards for a good validation of them. In some other cases, the lack of accessibility to that test due to the fact that it is not considered in the reimbursement criteria of the health service or the insurance company, so it does not allow to the researchers to have complete data from all of our cases. This is why, in many occasions phenotypes are the only criteria for registering, and genetic information acts as co-variable of the case but not as a classificatory variable. In addition, the modern technology which allows high speed in the gene sequence analysis and provides the whole information of the exome of a case at lower costs, is introducing new challenges in research but also in registering activities because their validity and the potential misuse of the incidental findings not previously included in the aims of the registry. These, and some other future matters will have to be faced by RD registries taking into account new methods and applying the traditional ones to these new scenarios.

6.44 Coordination. Quality control of coordination activities is also important for most of the RD registries. RD are defined by their own prevalence and registries are recognized worldwide as a best tool for collecting sufficient number of cases for research purposes and for the improvement of the disease knowledge. This is why, coordination – generally between central and local sites – is becoming an important role in the registry design and its development. A clear definition of different competencies, workflows and responsibilities is recommended. Meetings, conferences, and site visits already commented above, are tools which can facilitate the quality of this coordination. In general, all topics are applied to this coordination but differences in the activities to be carried out by each of them have to be clearly stated in the Manual of Procedures. A lack of control of the coordination could be very dangerous for data completeness and validity.

6.45 Quality assurance in children registries: A special situation

It is very frequent that RD affect children because many of them are congenital diseases (inherited or not) or diseases diagnosed during the infancy. It is said that more than 80% of RD are seen in the earlier ages. A quality assurance plan in RD registries where children are the main source of cases should not be different to other registries. Determining clinical effectiveness and cost effectiveness of drugs, devices or procedures, monitoring safety of specific products and treatments, diagnostic methods,
quality of care and quality of life measurements, high risk group and survival are among some other topics to be treated in many RD registries, but they become to more critical when children are the cases to be registered (Simon et al, 2011). The use of such registries is particularly beneficial in areas of very small patient populations, such as those affected by ultra-rare diseases which are very usual in children with rare genetic diseases.

Many of the RD described lead child to severe mental and physical disability. Mental delay and developmental disorders also often include clinical features to be both included as variables and to face on as constrains in these types of registries. In addition to being minors there may be other types of disabilities which may condition their capacity for making decisions affecting their personal lives (Katherine et al, 2010). These types of registries have to be designed taking into account mainly strict ethic rules regarding to informed consent, participation in other nested studies and parent authorization while they are minor. If other mental disability is not present in the diseases, all these authorizations have to be confirmed by the patient once he/she reaches. Obviously, if there is some mental disability, the capacity of making decisions has to be carefully considered, and comply with legal rules. Case ascertainment is a problem for many of these RD because the lack of correct diagnosis and visibility. Many families are not interested in showing and exposure to the child to some unknown circumstances and they prefer to have the child at home without few contacts with the health system. Parents’ education and explain carefully the advantages of the registry are important topics in this field.

This is why, strengths and limitations in general cognitive competencies, self-awareness and understanding of health and illness concepts impact the reliability and accuracy of child RD registries beside other registry activities such as ethic and governance which have to be adapted to this circumstance.
Section 2  Quality topics

Chapter 7  Quality Indicators
7.46 Types of quality indicators
7.47 Surveillance of RD registry activities
7.48 Case ascertainment
7.49 Analysis of data completeness
7.50 Consistency
7.51 Timelines
7.52 Data security and confidentiality
7.53 Validity
7.54 Summary
Chapter 7. Quality Indicators

As with any other quality assessment procedure, some ways of measuring are required and these are based on indicators. The concept of indicators has already a long trajectory in the public health field. The development and use of indicators are integral parts of planning and designing health services, as they are management tools for health care services and health systems (20). Indicators show important clues for managing the quality of health information systems development because they standardize the information and enable making readable and valid conclusions for new public health decisions. At the same time, they are the basic instruments for monitoring the quality of the information system itself. This report is not intended to be an in-depth analysis of quality indicators: instead it briefly summarizes the properties of these measures. Like other types of health and monitoring indicators (EUROPLAN, 2012), indicators have to be accurate, valid, reliable, sufficiently sensitive to detect changes, and predictable. At the same time, they should be acceptable, simple, and flexible enough to be adapted to new modifications.

A RD registry is a health information system because it provides new knowledge about disease burden, natural history of the diseases and possible ways of new interventions that can lead to change the patient prognosis.

In the RD quality assurance plan the indicators selection is important due to the scarcity of resources. The relationship between each one of the indicators selected and their real potential usefulness for achieving their final RD registry objective must be considered (24). Prioritization is focused on utilizing those indicators that have the capacity of pointing to changes in the trend of actions implemented, enabling comparison of the current period with a previous period. If this is not possible, the system at least has to provide accurate information about the current state of the actions and activities implemented in order to consider if the registry is progressing according their aims and quality assurance plan.

The following criteria should be carefully considered for selecting and defining the list of indicators (AHRQ, 2011)

- Provide assessment of quality and disparity
- Provide baselines to track progress
- Identify information gaps
- Emphasize interdependence of quality and disparities
- Promote awareness and change

7.46 Types of quality indicators

7.47 Surveillance of RD registry activities. All written procedures should be followed and partners should adhere to them. A registry of incidents in the achievement of each procedure should be carefully developed because they are useful for detecting difficulties in the design but also for the identification of actions that have not been accomplished. Indicators should be able to detect these kind of problems during the
building phase of the registry, but also periodically during lifetime of the registry.

7.48 **Case ascertainment** can be monitored using the expected number of cases for the target population and the difference with regard to the actual number of cases registered. Appropriate indicators can be focused on some standardized measures of this number of cases such as incidence rate, prevalence by age and gender strata, temporal trend, differences between these same data from one year to another, etc and they are basic figures to be explored. Special attention should be made to the different sources of information because in some cases the problem can be located in one specific source. This question can be also checked in by comparing figures of the registry being developed with some other estimators provided by similar registries with a longer trajectory.

7.49 **Analysis of data completeness**. Data completeness should be monitored for all variables. Further analysis using secondary variables can lead us to make some mistakes in our conclusions. Data completeness can hide selection bias particularly if some deep statistical analysis is developed. In these situations, statistical modeling uses cases with full information for all variables included in the model. Many researchers are not aware about this type of constrains, and they consider their results as reliable for the whole cases included in the data set. Statistical packages eliminate rows of the data set when they found some missing value and some selection bias can be introduced in the conclusions. The proportion of registered cases with unknown values for various data items is an indicator of data quality. These types of gaps can result from problems with:

- data collection system
- access to necessary source documents
- item abstraction process
- coding rules
- rate of diagnostic case verification
- rate of duplicate cases detected
- relationship between mortality cases detected in the national index of death and case already registered

The rate of diagnosis and coding errors, and the rate of completeness of the main and secondary variables across sites are some of the most important indicators. Improvement in case reporting and coding is essential for accurate assessment.

7.50 **Consistency**. Data editing may be performed after the addition of new records interactively, or as a batch process before records are sent to the central registry, or both, and can be carried out at several points in the registration process after any changes have been made. This type of quality indicators requires of a previous development of rules which tested the presence of incompatible information within the same individual case. Each RD registry needs its own criteria because each disease shows its own features, for example, in some cases gender or age is not feasible for the disease registered; unexpected symptoms useful for establishing differential diagnosis among diseases sharing some clinical features; lab information, …
7.51 Timelines reflects the delay between any two (or more) steps in surveillance system. The timeliness of the system can be assessed by the ability of the system to take appropriate actions based in the urgency of the problem and the nature of the public health response. There are four points in time in the surveillance process are most often considered when measuring timeliness: a) time of onset of the disease or occurrence of an injury, b) time of diagnosis, c) time of receipt of the report of case by public health agency responsible for control activities, and d) time of implementation of control activities. Usually one of the first two points in time (a or b) is used as the starting point, and each of the other two points (c or d) is used as an endpoint. Timeliness is usually measured in days or weeks, but in hospital settings it might be measured in hours; for diseases that do not necessitate an immediate response, it might be measured in months or even years.

7.52 Data security and confidentiality can be monitored using indicators to check data access, number of failed trials, data downloaded, and characteristic of persons accessing the system, number of patients with informed consent, number of patients accessed after being registered for further purposes, … (Wilson et al, 2006). Understanding patient perspectives is also a key to developing balanced policies that protect patients' privacy and facilitate their opportunities to make autonomous decisions about participating in registry-linked research (Beskow, 2005). At the same time, registry policies on patient education and physician involvement can have an important impact on researchers' ability to conduct nested patient registry studies. It is a truth that registries are a valuable resource for recruiting participants for public health-oriented research, though such recruitment raises potentially competing concerns about patient privacy (Yaffa, 2011; de Abajo, 2008) and participant accrual. This is why policies for research contact with patients should avoid substantial variations. However, given that one of the final successes of a registry will be measured by the proportion of people from it who are recruited for research projects, and its impact on overall accrual to studies (Iliffe, 2011), methods of approaching patients are very important. Indicators related with patient and physician education such as number of training sessions as well as the creation of ways of patient expression (i.e.: patient involvement in registry committees) could help staff registry to explore the level of patient participation and therefore it can be one of the mechanisms to control quality of the confidentiality and privacy.

7.53 Validity. Indicators about registry and reliability registries are very important topics. These are main characteristics of whatever indicator itself but in this case indicators have to measure the own registry results. The only way to assess the information validity is to measure the size of bias. Each variable can be affected by some bias but the cost for measuring the potential bias of all variables included in a registry could be very high. To take some cases record and checking all information on data collected against primary source, could be one of the mechanisms to check if some bias in some of the variables collected have been introduced in the registry. Records random sample or random samples for specific list of variables can be drawn and then
tested against their original data. Monitoring of these types of assessment procedures can be easily changed into indicators (i.e.: number of variables including mistakes by number of variables checked). Abstraction is one of the sources of either systematic or random errors in variables. Special care should be given to diagnosis and outcomes variables. Errors in the diagnosis should be checked looking for clinical records, after consultation with experts, based on specific lab diagnostic tools and coding. Outcomes are also susceptible to be checked by the above methods because many of them are clinical features, functional imagines or lab tests. Indicators are then built using the number of mistakes identified in each one of variables under surveillance divide by the total of them revised.

7.54 Summary

Registries are usually voluntary programs, and the data quality depends on the motivation and commitment of health care professionals who collaborate in this process providing data from their own patients. Although success depends to a large extent on the resources available, much can be achieved by establishing a culture of integrity and enthusiasm among study staff which is conveyed to participants at all times (Golding, 2009). It is well known that the most restrictive strategy for registry staff is the need to obtain physician permission and contact patients with an opt-in approach. There are also mandatory registries organized under the scope of public health institutions, which demand specific cases reporting from all professionals working in that public health organization. In general, mandatory registries aim to know the disease dimension, distribution and possible determinants while patient registries are focused on specific interventions development or to improve the knowledge of the natural history of the disease.

Registries are considered in many places as a health information system because it provides information for the health status and follow-up of some or all population. Precisely, one concern about studies on healthcare information is that the data might be incomplete or inaccurate, leading to unsafe conclusions and hence undermining the credibility of the research (Alison et al, 2004). The epidemiologic usefulness of a registry increases the longer it has been in existence, often meaning that data collection, documentation, and quality control activities be conducted for many years before a register becomes fully productive for epidemiological purposes (Richesson et al, 2010) [46].

Registry policies should use balanced recruitment approaches, designed, in each case, to protect patient privacy and encourage beneficial research (Beskow, 2006a). The most efficient way is probably to involve patients' physicians, thereby affording an opportunity for physician input (Beskow, 2006b). Even so, registry exclusion criteria significantly alter the apparent severity of injury and resource utilization. The use of divergent exclusion criteria in the analysis of registry data may be misleading, at least in some disease registries where severity could play a role in the registration process (Bergerson, 2006). Quality assessment must look to these specific issues because patients are the reason why registries are created, and the validity of final outcomes is
the reason for carrying out research.

Quality is an understandable and desirable term but a complex concept. Quality assessment is the process that facilitates analysis of the achievement of different tasks and procedures. Though not a guarantee per se of the validity of the outcomes, it contributes to the analysis of possible mistakes, erroneous interpretations and potential bias. The process comprises data validation, data completeness, validity and reliability analysis, and global assessment of the registry (cost-effectiveness, potential benefits, etc.) (D. Max Parkin, 2009; Freddie Bray, 2009). The auditing process should be rigorous, robust and reflect a high degree of accuracy of data collected by participants (Andrianopoulos, 2011).

A final aim of this concept would be closely related with the accuracy term which means that our outcome or product fits our aim fully. In a simple definition, it could be read as the “susceptibility to provide a biased outcome”. In some registry stages there are aspects which do not need to be tested for bias because apparently they would not be directly linked with the presence of this kind of troubles. Therefore, they should be candidate activities to be included in a quality assurance plan. However, examples such as the use of an appropriate informed consent - which is not a priori undertaken a bias -, could lead us to some bias in the final outcome throughout their erroneous use. The same could happen with standards procedures for coding and case ascertainment. They look for quality and systematic activities but quality does not necessarily would restrict the bias. Hence, quality is not far from validity and the latest is not a minor step of the quality process. Conversely, the validity of a registry is the final goal and all activities have to undergo quality control measurements while looking at the validity and/or the interaction of this specific activity with the final outcome of the registry.

RD registry methods do not consist in simply a common platform and Common Data Elements, but also methods and criteria designed to achieve the best and most comprehensive information of supplying to policy makers, researchers, patients and their families (RDTF, 2011). In some other words, there are constrains and barriers a registry must be overcome, such as:

- insufficient utilization of information systems by clinicians and policy makers;
- limited application of statistical routines in health reports;
- inadequacy of software available in the public domain;
- insufficient use of medical records, due to increasing privacy concerns; and,
- lack of standardized approaches for secure data transmission.
Section 3  Quality assurance plan for the European platform for rare diseases registries and databases

<table>
<thead>
<tr>
<th>Chapter 8</th>
<th>EPIRARE commitments and RDR platform</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.54</td>
<td>Quality Assurance and the RD registries Platform: General principles</td>
</tr>
<tr>
<td>8.55</td>
<td>RD registry’s QAP vs RDR platform’s QAP</td>
</tr>
<tr>
<td>8.56</td>
<td>RDR platform and Quality Assurance implications</td>
</tr>
<tr>
<td>8.57</td>
<td>Services to be provided by the RDR platform</td>
</tr>
<tr>
<td>8.58</td>
<td>Public Health Area</td>
</tr>
<tr>
<td>8.59</td>
<td>Research-development Area</td>
</tr>
<tr>
<td>8.60</td>
<td>Quality Assurance Plan Characteristics</td>
</tr>
<tr>
<td>8.61</td>
<td>Quality assurance of the platform management processes</td>
</tr>
<tr>
<td>8.62</td>
<td>Governance and Quality</td>
</tr>
<tr>
<td>8.63</td>
<td>Ethics and Quality</td>
</tr>
</tbody>
</table>
Chapter 8. EPIRARE commitments and RDR platform

One of the main important actions in the RD field is to set up a European Platform for Rare Disease Registries (RDR). The project aims at preparing the ground for developing a “horizontal” multi-disease platform as a public, institutional undertaking which takes into account the interests of a number of stakeholders and cares for its long-term sustainability. In particular, EPIRARE tasks for developing RDR platform criteria are based on:

- Analysis of the current needs of EU Registries on RD and platform facilities
- To agree on the Register and Platform Scope, Governance and long-term sustainability
- Identification and characterization of services and facilities of a European Platform of EU RD Registries
- To design standardized information for policy-makers and researchers
- To agree on a Common Data Set, disease-specific data collection and data validation
- To identify tools and other facilities supporting the operation of the platform users
- To develop criteria and principles for a platform quality assurance program and validity methods of registered data.

8.54 Quality Assurance and the RD registries Platform: General principles

In the scope of the (RDR) Platform it is also essential to develop and maintain an organizational structure, procedures, processes and resources to implement quality management known under the name of Quality Assurance System (QAS).

A RDR Platform would ideally be limited to those registries characterized by having “high-quality” data and research practices. Registries involved should fulfill a minimum prerequisite of quality criteria and the adoption of the elements of the minimum data set with their definitions. In fact, to assure the platform quality, a manual of procedures having data repository structure and its criteria, tools and resources to be used by stakeholders, visitors and registries responsible should be developed. Nevertheless, the RDR platform intends to attract the majority of RD databases, in order to promote the use of their data, their sustainability.

To this aim, the platform should act as an external auditor with respect to the registries establishing basic elements to be fulfilled by the registries, indicators measurements, procedures, quality reports and analysis of quality by audits.

However, this QAS is clearly depending on the platform model and services to be included. If many services and complexities are included, difficulties to assure the quality of the whole platform will grow.

Furthermore, an external assessment is highly recommended, because deviances from the main rules are easier to detect and correct through training or simply with slight modifications of the workflow. There are different methods to provide external assessment. The most common is to account with an advisory board (committee) or with external audits, which provide credit and certify all activities and procedures. Therefore, to assure an external quality control procedures and targets in the platform, the QAS should be design in accordance with the requirements of the ISO 9000 family of standards.

Finally, a Multi-National Network based platform scenario would be related to the development of national multi-disease, multi-purpose registries. This potential scenario makes more difficult the quality management of the platform, because the QAP should be developed with the
participation of several types of stakeholders and taken into consideration all their needs and expectations, which affect directly to the quality.

In summary, a RDR platform is a project which necessarily needs to be developed by phases and in a close collaboration with stakeholders whom adapt the platform to the real registries life in parallel with the implementation of quality control measurements.

8.55 RD registry’s QAP vs RDR platform’s QAP

A QAP shows similarities and also differences between rules and procedures applied to a RDR platform or just to one RD registry. However, a QAP for a RDR platform involves much more complexity than one RD registry because of an overall RDR platform has different and broader aims which involve health policy makers, and future research networking activities.

The most important affects to some differences in complexity and/or extension of the activity.

- Registry Description, Sponsor and Conditions of Access, Registry Design, Eligibility for enrollment, Common Data Element Groups, Manuals of procedures, Progress Report could vary in the size but not in the general principles, which are essentially the same.
- Use of a indicators checklist is also similar although whit different scopes.
- Fulfill the requirements and stakeholders needs are common aims either a registry or a RDR platform.
- Governance process and ethical issues involve much more complexity.
- Management IT tools system is also more complete and complicated because it is necessary to add procedures for each one of their components.
- The development of SOP show also a high level of complexity because they should define the relationships among several type of stakeholders and aims.
- A problems solution system and mechanisms for making decisions should be also implemented but aimed to supra aims instead specific aims, which are addressed in a registry.

However, the main difference between a RDR platform and a RD registry consist on the former should develop several other services apart from the creation and sustaining of the main repository of RD registries. Other services aimed to help those registries in the process of achieving the full quality standards should be also developed. Services and tools such as reports and information as well as communication and networking facilities should be important issues to be treated in the platform. Those registries which do not comply the minimum quality criteria cannot provide data to the platform repository but they can receive assistance from the platform to improve its quality assurance and then provide data.

The special section of data repository for ad hoc studies can be also used for health technology assessment purposes such as case recruitment for new clinical trials and cost-effectiveness analysis for orphan drug post marketing authorization. In this case the platform should be ready and their QAP available for providing safeguards for protection of commercial interests and other types of conflict of interests. This QAS should allow acceptance of studies for regulatory purposes.

8.56 RDR platform and Quality Assurance implications

There are some different possibilities for building a pan European RDR platform but they would cover different quality strategies. Platform strategies should include,
About type of data
   a) Metadata of each RD registry
   b) Aggregated information data
   c) De-identify data

About type of providers
   a) Registries owner
   b) European Reference Network scheme
   c) National partners

About outcomes an outputs
   a) As source of epidemiological information on RD (including natural history of RD)
   b) As source of patients for research studies
   c) As source of surveillance for side drug effects, cost-effectiveness analysis, etc

In addition, funding criteria should be an important question to taken into account. Sustainability is the main concern related to this question followed by the data protection and criteria for the inter-relationships between private and public sectors.

8.57 Services to be provided by the RDR platform

The EU Platform for Rare Disease Registries should make possible the pulling and analysis of rare disease patient data, in the fastest and most efficient way, in full respect and protection of patient rights and needs. The RD platform should be based on a website platform that could contain the following components:
   - Data repository to receive data from registries and databases addressing different aims, populations and diseases
   - Support ad hoc collaborative projects
   - Access to ad hoc data extractions
   - Tools and resources of using registries
   - Predefined outputs for funding organizations and society
   - Mechanisms for the promotion of registration and networking

Quality of specific services to be provided by the platform should be are according to different area of interest:

8.58 Public Health Area (including Orphan Medical Products monitoring - OMP) which is applied for EU and Member States levels. Services should valid and reliable, developed in parallel with the EU and MS RD public health policies priorities, showing existing geographic inequalities in health care as well as areas where health and social interventions are needed, integrated with other services (Biobanks, Neonatal screening, exemption procedures, social services) and providing capacity building for the assessment of appropriate use of health technologies, clinical added value and relative effectiveness of OMP, among some others

8.59 Research-development Area

Facilitate the access to more extended data bases for either information or as source of voluntary cases for new research studies, provide links with other research initiatives (biobanks, quality assessment networks, certified laboratories), promoting the networking among experts and centers, translational research and OMP assessment.
However, services have necessarily to be based on types of data providers whom in combination with sponsors define the RDR platform model. The three possible data providers configure the following RDR platform models: Free Based RDR; European Reference Network; Comprehensive National Strategy. A Free Based RDR platform model is one of the possible scenarios. Patient registries and databases born mainly based on particular initiatives from physicians and researchers and address specific research goals. Therefore, the platform should not force the existing registries to modify their operational capacity but to comply with defined rules and data elements. In this scenario, the prerequisites to set out for the platform could be:

- To guarantee the improvement in the exploitation of currently existing data with minimum impact on existing registries and databases. The existing registries or databases would be requested to adapt some matters to platform tools and resources.
- The data collected by the platform are the basis for elaborating outputs of interest to granted organizations and also to the underlying registries and databases.
- To attempt to provide, as far as feasible, multiple outputs fulfilling the interests and needs of the different stakeholders. These outputs should be satisfied, fulfilled and complied with the quality assessments and procedure of the platform.
- The adaptation of requested by the platform to the registries or databases cannot involve changes in their main aims, scientific competence and studied population. Therefore, the platform should be sufficiently flexible to receive and elaborate data from registries and databases, but also, should be ensured standardized procedures and considered a list of quality control measures.

In this scenario, the participation in the platform is granted to registries and databases complying with the data quality criteria established in the Platform Quality Assurance plan. The registries and databases participating in the platform would communicate data using common data elements definitions. They would have conditional access to platform data and get tools and resources for the improvement of their activities.

8.60 Quality Assurance Plan Characteristics

Quality components are classified as either "basic elements of good practice", which can be viewed as a basic checklist that should be considered for all registries, or as "potential enhancements to good practice" that may strengthen the information value in particular circumstances. In general, a QAP for the RDR Platform plan should include both of strategies although this plan cloud implement in several phases. Whatever kind of data provider is selected, each registry to be included in the platform either metadata or de-identified data should achieve several principles of quality. Taking into account their basic elements, registries should have:

- Description: identification and description information
- Classification and Purpose: information about the type of registry and its intended purpose
- Governance: holders as well as some external experts should be involved
- Sponsor: information about the sponsor, collaborators
- Conditions of Access and related contact information
- Ownership: The platform should recognize the ownership of the achievements reached by the participating registries
- Design: information and references to the Registry design, including the protocol definition
- Eligibility: criteria for patient enrollment in the registry.
- Common Data Elements
Quality Procedures: information about the quality procedures being conducted for the registry
Progress Report: information associated with the registry including growth of the registry and any relevant references to available progress reports.
Other related Information: links to related publications, citations, and other relevant information

Moreover, the achievement of some level of accuracy is crucial for the assurance the registries validity and it should be applicable to whatever type registry platform - potential enhancements to translate into best practices. The following issues are important points to be taken into account to deal with the internal validity of the registries:

- Aims and hypothesis with their background and bases
- Definition of the target population, settings and study population
- Design
- Case definition and inclusion/exclusion criteria
- List of set of variables including sources and definitions
- Standardized data report form
- Pilot registry phase description (not always is needed)
- Data collection methods
- Data interpretation/abstraction
- Data codification
- Data stored
- Data analysis
- Ethical issues
- Limitations and bias analysis
- Standardized data reporting

In addition, to ensure the implementation of the Quality Assurance Plan as an external quality procedure (ISO 9000 Standards)\(^8\) the general documentation for the all procedures of the platform must to be compose of: Documented statements of a quality policy and quality objectives; A quality manual; Documented procedures; documents to ensure the effective planning, operation and control of its processes and Activities records.

### 8.61 Quality assurance of the platform management processes

### 8.62 Governance and quality

Within the quality assurance process, it should be develop standardized criteria for the quality

---

\(^8\) The ISO 9001:2008 sets out the criteria for a quality management system and, is based on a number of quality management principles that should be followed in order to lead the Rare Disease Registry Platform towards improved performance:
- **Customer focus**: organizations depends on their customers and therefore should understand current and future customer needs, should meet customer requirements and strive to exceed customer expectations.
- **Leadership**: leaders establish unity of purpose and direction of the organization. They should create and maintain the internal environment in which people can become fully involved in achieving the organization's objectives.
- **Involvement of people**: people at all levels are the essence of an organization and their full involvement enables their abilities to be used for the organization's benefit.
- **Process approach**: a desired result is achieved more efficiently when activities and related resources are managed as a process.
- **System approach to management**: identifying, understanding and managing interrelated processes as a system contributes to the organization's effectiveness and efficiency in achieving its objectives.
- **Continual improvement**: should be a permanent objective of the organization.
- **Factual approach to decision making**: effective decisions are based on the analysis of data and information.
- **Mutually beneficial supplier relationships**: an organization and its suppliers are interdependent and a mutually beneficial relationship enhances the ability of both to create value.
control and governance of the platform, and moreover, it should be implement a protocol to guarantee quality in overall process and in all the service of the platform.

Platform products and outputs can be affected very strongly by the way in which these outcomes are delivered. The primary roles involved in the RDR platform should aim to ensure the quality governance:

- Registry platform Holders, who would provide information regarding their registries.
- Registry platform Users, who would search and find information regarding registries included in the platform
- Registry Platform Reviewers, who would ensure the listed registry information was accurate, consistent, and of high quality
- Registry Platform Administrators, who would handle the maintenance and operation of the platform, and support the needs of the preceding roles

Furthermore, transparent and trustworthy operation is a clue for the success of the platform: all stakeholders should be represented in the governing bodies of the platform.

In general, the overall coordinating mechanisms of the platform are:

- import data from patient registries and databases using the Common Data Elements previously agreed
- use the collected information to find subsets of patients who may be potential candidates to participate either clinical trial or observational studies.
- participating registries will retain their data ownership and control and they may collect additional data beyond the CDEs core
- researchers can also use the information to report about the latest research findings to those families already registered.

The governance mechanism should aim at a distribution of responsibilities and tasks between the platform level and the registry level, so that it can assure the achievement of the platform aims (data quality, registry data interoperability, promotion of registration, predetermined platform outputs and sustainability) while maintaining flexibility of registry aims, reduction of costs for registry establishment and maintenance, research freedom, registry ownership and acknowledgement of intellectual property and of contribution to registry data. It is also necessary that the approval process for running a registry within the platform is transparent, not too time consuming and bureaucratic, to avoid creation of further barriers to the establishment of registries.

8.63 Ethics and quality

Ethical and legal considerations should guide the development and use of the RDR platform, and therefore, any registry included. The research purpose of a registry, the status of its developer, and the extent to which registry data are individually identifiable largely determine applicable regulatory requirements. Other important ethical and legal concerns include transparency of activities, oversight, and data ownership and conflict of interest. Also, confidentiality and ethics are important issues to be considered in data delivery process. Each country and even supranational scenarios such as the European Union have strict legal rules and ethics procedures relating to data delivery either among partners or between central and local sites. Therefore, it should be developed a list of ethic criteria and rules that guarantee patient confidentiality, informed consent, data owner and data delivery as an important piece of the QAP.
Section 3  Quality assurance plan for the European platform for rare diseases registries and databases

Chapter 9  Quality and scientific validity and reliability of the platform outcomes

9.64 Manual of Procedures
9.65 Data quality
9.66 Outcomes Quality
9.66.1 Process indicators
9.66.2 Monitoring indicators
9.66.3 Outcome indicators
9.67 Quality assurance (user satisfaction assessment) of tools and services, including promotion of registration and networking
9.68 Identify improvement opportunities.
9.69 Tools and Resources Quality
9.70 Summary and conclusions
Chapter 9. Quality and scientific validity and reliability of the platform outcomes

9.64 Manual of Procedures

The Manual of Procedures should include all aspects related with the quality, such as the quality plan itself, the quality control mechanisms, the quality indicators list and the program for a periodically activity quality assessment. Therefore, this documentation is the main clue for a quality assurance plan and all documents of procedure constitute the Manual of Procedures.

Therefore the Manuals of procedures should also contain the following list of documents as it was stated in the first section of this report:

- Policy rules and governance: Regulatory documents, including fundamentals, specific aims and governance criteria, registry owner, funding, sustainability plan …
- IT tools document: Description of all IT tools related questions such as server system, software supporting database, software versions, back-ups protocols, incidences and solutions.
- Security document: All rules controlling access, data safety, confidentiality, patient and professional rights, data delivery…
- Ethic rules: List of ethic criteria and rules that guarantee patient confidentiality, informed consent, data owner and data delivery. Conflict of interest are also considered under the scope of this point
- List of procedures for each topic: List of documents describing the workflow of each topic(s) with alternative plans for possible unforeseen contingences and, at the same time, they should care data validity, data completeness, control of data errors and actions to be taken for these problems. Central and local site procedures can be different because of respectively responsibilities.
- Manual training: Specific documents for each type of participant profile which helpful instructions for developing their works
- Instructions for database users: A manual where users can find how to proceed to access the system and use the software according with their profiles.
- Data dictionary: Contains all variables, definition, data model, relationships with the others, coding as well as tables and their relationships.
- Data Report Form: The information sheet with the Common Data Elements to be searched for each individual case and which is included in the database later.
- Catalogues: Auxiliary tables such as center participants, diagnosed tests, medical proofs…
- Classifications: List of diseases to be registered with the code system defined by the researches. If more than one list is proposed, the relationships among them would be provided for harmonizing this crucial quality question
- Personnel functions and tasks: Documents describing each activity and competency for each individual person involved in registry activities
- Checklists: Containing short list of activities to be revised and marked after they have been developed by workers
- Nested research study protocols: Including protocols already defined as well as protocols for future proposals
- Quality assessment: Document containing all procedures, control mechanisms and methods for the correct evaluation of the registry and their frequency.

The Quality Assurance Plan could be a separate part of the Manual of Procedures or it should be contented in the own manual and their respectively parts, although the main sections of this plan
have to be stated in the quality assessment report. The question of who has the responsibility for the data quality assurance has to be also clearly stated. Moreover, all above questions and their specific details have to be stated in very well written documents which are accessible electronically and/or paper hard copies for all participants and actors of the registry.

9.65 Data quality

The requirements for data quality assurance should be defined by the platform owners but achieved by each registry when is incepted or added in the platform. Data quality assurance should affirm that the data are collected in accordance with established procedures and that they meet requisite standards (completeness, validity of accuracy, coding system, security and confidentiality). All quality activities related with the data process have an important role in the platform development. Standardization of data elements and also outcome measures would result in reducing efforts in developing RD registry - added to the platform - and increasing opportunities to link and compare data across all the registries.

Furthermore, registries may be very useful vehicles for providing clinically relevant real-world information, even when they meet relatively few of the basic elements of good practice. In many cases, some data are better than no data, and even registries that fall short of including all the basic elements of good registry practice may still provide valuable insights about real-world medical and consumer practices and disease etiology. Evaluations of the quality of each registry in the platform must therefore be done with respect to the context-specific purpose of the registry, must take into account both the internal and external validity of the data, and should be tempered by considerations of cost and feasibility.

Therefore, the platform:

- should develop a set of Common Data Elements (CDEs) and their definitions in a standardized and meaningful manner, which are necessary for the predefined outputs of the platform. CDEs are necessary to ensure that data in the repository are defined in the same way, using the same standards, and the same vocabularies. Use of CDEs will facilitate the standardization of data entry and allows for harmonization, sharing and exchange of information across registries and diseases and facilitates analyses and studies within a specific disease and across multiple diseases. Participating registries should comply with the predefined data sets, while remaining free to record any other variable for the aims in which they are interested.
- should determine which Data Elements are absolutely necessary and which are desirable but not essential.
- should develop a guide of use of CDE to give indications of how to handle definition and other coding problems or inclusion criteria
- would define a general training manual

9 Requirements for registries should be found in the document: “Report on guidelines for data sources and quality of RD Registries in Europe: section Quality Criteria - Benefits from the assessment report by the EU Scientific Committees”

10 **Feasibility**: is defined as the viability, practicability, or workability of a task, program or intervention. In reference to information systems and specifically to RD indicators, feasibility would refer to the viability of collecting, measuring and recording the indicator. Effective indicators are based on data that is easy to access or that can be measured directly at the setting
would carry out general staff training sessions to raise attention on specific points of interest in order to fulfill quality of the platform (e.g. missing data or out-of-range data, data consistency, data validity, items for follow-up, ……)

should establish an internal and external assessments (analysis) of quality and validity of data and perform periodic data quality reports

could develop a quality data criteria checklist to help in the quality data assessment

Critical factors in the quality of the data include mainly how data elements are structured and define, how personnel are trained, and how data problems are handled (missing, out-of-range, or logically inconsistent values). Moreover, the selection of data elements requires balancing such factors as their importance for the integrity of the platform of registries and for the analysis of primary outcomes, their reliability, etc… Participating registries should comply with one or more of the predefined data sets, while remaining free to record any other variable for the aims in which they are interested. Overall, choice of data elements should be guided by validity, integration with platform components and each registry’s purpose. Finally, the selection of data elements requires balancing such factors as their importance for the integrity and aims of the platform, the registries aims and for the analysis of primary outcomes, their reliability, etc…

9.66 Outcomes quality

In the process of RDR platform development is essential to standardize all outcomes and to facilitate access to that information for all the stakeholders interested in the RDR. Nevertheless, while begin the platform activities, it is useful to use indicators measurements which could serve for monitoring the platform activity and also its quality and impact.

Otherwise, registries are undertaken for many purposes, ranging from scientific understanding of patient outcomes to inform policy decisions and some of them, could serve for multiple purposes and change their roles over time. Moreover, it should ensure to provide outputs that fulfill the needs of all involved stakeholders. Although there are limitations in whatever type of the assessment of outcomes quality, it should take into account differences between research quality (which pertains to the scientific process) and evidence quality (which pertains to the findings emanating from the research process). On multiples occasions, it is not practical or feasible to achieve all of the basic elements of quality assurance, but it is useful to consider these quality characteristics in planning and evaluating the registry platform.

A preliminary list of indicators that can be useful for assessing the quality of the RDR platform outputs could be as follow,

9.66.1 Process indicators

- Suitable (Appropriate) and clarity in the definition and development of the objectives of the participating registries.
- Existing a quality management system certified in participating registries.
- Existing a quality management system certified in the platform level.
- Existence of a coordination system among registries included in the platform.
- Providing clear, operational and standardized definitions of aims and common data elements.
- Existence of standard instructions for use of the platform and all of their components
- Suitability of use of a glossary of terms (CDE) used by the participating registries
- Entry and exit criteria for the participating registries clearly established
- Existence of a communication plan within the components of the platform
9.66.2 Monitoring indicators

- Use of validated scales and tests to measure data and outcomes
- Completeness of information of each registry is assessed and described
- Existing regular update of tools and resources
- Number of research studies already promoted through the use of the platform information
- Results for all the main objectives of the registry platform are reported

9.66.3 Outcome indicators

- Periodic development of statistical data analysis
- Incidences, prevalence and other epidemiological descriptors reports in the platform level (stratified by RD group? ??)
- Number of registries participating in the platform (stratified by country or region)

This list should be adapted to the final platform model once it is decided, taking into account stakeholders opinions – including patient organizations – and data providers.

9.77 Quality assurance (user satisfaction assessment) of tools and services, including promotion of registration and networking

9.68 Identify improvement opportunities

Other resources/tool offer and perform by the RDR Platform should be provided a quality satisfaction evaluation for the participating registries and other user of the platform. Related to this, it can be implemented for example satisfaction questionnaire, a kind of suggestion box or other tool to receive, analyze and feedback some comments or recommendations from the users and providers of the platform.

9.69 Tools and Resources Quality

A RDR platform can also offers some registry-related resources such as promotion of networking, training activities, facilities for setting up a new registry, among some others. The platform should also care this sort of resources and applies quality criteria for them. Indeed, good standardize procedures establishing how to request each of those tools, how to wait for receiving them and who is responsible their provisions are main topics affecting their quality. A registry of the tools provided or downloaded and their impact are also criteria to be implemented to deal with the effectiveness of these types of interventions.

9.70 Summary and conclusions

To build a RDR platform will be implemented in one of the JRC institutes in Ispra, Italy, following the decision made by the EC. This platform is challenging for rare diseases research and health and social planning and it can be dramatically change the reality of the RD patients and their family. The correct knowledge of these large group of low prevalent diseases is still a pending matter and not only because their mechanisms involved nor etiology, which are very important, but because registries are observational tools which provide knowledge by themselves for several purposes and
aims ranging from promotion research to facilitation of the health and social plans. Quality procedures have to be developed in the RDR platform mainly based on external audit and certification but also quality should look for the impact of their activities in the real life of people affected through valid and reliable results which could facilitate translational research and best clinical practices among some of their main aims.
References
References


EU Clinical Trials Register. Data quality. 10 March 2011. Available at: https://www.clinicaltrialsregister.eu/dataquality.htm

EPIRARE. 2012 European Platform for Rare Diseases Registries. Available at: http://www.epirare.eu/


GRDR. 2012 Global Rare Diseases Registry. Available at: http://www.grdr.info/.


Matthew Bellgard, Christophe Beroud, Kay Parkinson, Tess Harris, Segolene Ayme, Gareth Baynam, Tarun Weeramanthri, Hugh Dawkins, Adam Hunter. 2013. Dispelling


Roehr, B. 2012. FDA given new powers over data reporting to national clinical trials registry. BMJ 345:e6629.


List of abbreviations
List of abbreviations

CDE   Common Data Elements
DRF   Data Report Form
EAHC  European Agency Health and Consumers
EPIRARE The European Platform for Rare Disease Registries
ELSI  Ethic, Legal and Social Implications
EMA   European Medicine Agency
EPPOSI European Platform for Patients' Organisations, Science & Industry
EUCERD European Union Committee of Experts on Rare Diseases
EUROPLAN The European Project for Rare Diseases National Plans Development
GRDR  Global Rare Diseases Patient Registry and Data Repository
IIER  Institute of Rare Diseases Research
ISO   International Organization for Standardization
IT    Informatics Tools
MCD   Minimum Common Data
MeSH  Medical Subject Headings
NIH   National Institutes of Health
QA    Quality Assurance
Qass  Quality Assessment
QAS   Quality Assurance System
QAP   Quality Assurance Plan
QInd  Quality Indicators
QC    Quality Control
QR    Quality Results
SOP   Standard Operating Procedures
RD    Rare diseases
RDTF  Rare Diseases Task Force
RDR   Rare Diseases Registries
SDRF  Standardized Data Report Form
WHO   World Health Organization
WP    Work package