| Version 01 | February 2019

Procedural document:
Epidemiology of rare diseases in Orphanet

Prevalence, incidence and number of published cases or families

www.orpha.net www.orphadata.org

Inserm
La science pour la santé
From science to health

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
MINISTÈRE DES SOLIDARITÉS ET DE LA SANTÉ

Co-funded by
the Health Programme
of the European Union

I. Introduction

1.1. Purpose/objectives

Orphanet carries out a systematic survey of literature in order to estimate the prevalence and incidence of rare disorders (RDs). These data are of particular interest for public health, and research and development purposes, amongst others.

This procedural document aims to describe the collection, analysis and validation of epidemiological data related to RDs integrated into the Orphanet database.

1.2. Disclaimer

- This publication is part of the OrphaNetWork Direct Grant (831390), which has received funding from the European Union's Health Programme (2014-2020).

- The content of this publication represents the views of the author only and is his/her sole responsibility; it cannot be considered to reflect the views of the European Commission and/or the Consumers, Health, Agriculture and Food Executive Agency or any other body of the European Union. The European Commission and the Agency do not accept any responsibility for use that may be made of the information it contains.
- Epidemiological data on rare disorders is scarce. In most cases, patients are so rare that studies cannot be carried out using solid methodologies. It is possible that prevalence is overestimated in some cases as epidemiological studies are generally based on hospital data in regions with a higher prevalence. Furthermore, clinical presentation may overlap with common comorbidities.
- Orphanet collects raw data with a variable level of scientific evidence ranging from meta-analysis to case reports/series.
- The validity and exactitude of raw data sources is taken for granted and have not been verified. However, confusion between terms such as incidence and prevalence and/or birth prevalence due to the interchangeable use of these terms in certain sources is corrected as much as possible when detected.
- The arithmetical average values do not take into account the heterogeneous nature of the methodologies employed by the studies considered in the literature survey.
- The prevalence and incidence data published are only estimations and cannot be considered to be absolutely correct.

1.3. Range of application

Orphanet registers rare disorders affecting not more than five in 10 thousand persons in the general population, established by the European Regulation on orphan medicinal products (Regulation (EC) No 141/2000 of the European Parliament and of the Council of 16 December 1999 on orphan medicinal products). Collected epidemiological data allows the rarity of a disease to be assessed. Only epidemiological data corresponding to the European definition of a RD is included in the Orphanet database.

Epidemiological data are managed by an epidemiologist who closely collaborates with the Orphanet information scientists in charge of the inventory, classifications, scientific annotations and encyclopaedia of rare disorders. The epidemiologist is also in continuous interaction with experts in the rare disease field, identified through their publications and their activity related to the given disorder/group of disorders such as research projects, clinical trials, expert centers, and dedicated networks. They are consulted in order to validate averaged and extrapolated data.

1.4. References

1.4.1. Procedural documents

- Orphanet inventory of rare diseases
- Orphanet standard operating procedures

1.4.2. Information sources

- Summaries of opinion and Press releases from EMA (European Medicine Agency) and **FDA** (Food and Drug Administration)
- European Medicines Agency Points to Consider on the Calculation and Reporting of the Prevalence of a Condition for Orphan Designation. 2002. [Accessed January 10, 2019].
- Medical registries/networks (herein referred to as "specialised resources"):
    - Rarecarenet: Incidence of rare cancers in Europe
    - IARC (International Agency for Research on Cancer): Incidence statistics of cancers worldwide
    - EUROCAT (European surveillance of congenital anomalies): Prevalence of congenital anomalies recorded in Europe
    - ICBDSR (International Clearinghouse for Birth Defects Surveillance and Research) : prevalence of congenital anomalies recorded in certain parts of the world
    - MONICA (Multinational MONItoring of trends and determinants in CArdiovascular disease): List cardiovascular diseases in certain parts of the world
    - RAMEDIS (Rare Metabolic Diseases Database) : Metabolic diseases
- National/international health institutes and agencies :
    - World health organization (WHO- Global Health Observatory Data Repository), Worldwide
    - Center of Disease Control, USA
    - Institut de Veille Sanitaire (French Institute of Health Surveillance), France
    - National Cancer Institute, France
- PubMed – NCBI: Bibliographical search engine
- Rare disease experts identified by Orphanet

1.4.3. Sources of demographic data (Population estimates by age and sex)

- The World Bank, World wide
- Eurostat, Europe
- Census, United States
- Insee, France
- Office for National Statistics, United Kingdom

1.5. Definitions

1.5.1. Diseases related definitions

Disorders are clinical entities including diseases, syndromes, anomalies and particular clinical situations. Disorders are clinically homogeneous entities described in at least two independent individuals, confirming that the clinical signs are not associated by fortuity.

Rare disorder (RD): defined according to the European legislation defining a prevalence threshold of not more than 5 affected persons per 10'000 (Regulation (EC) N°141/2000 of the European Parliament and of the Council of 16 December 1999 on orphan medicinal products)

However, for disorders such as cancers that have a sub-acute clinical course, incidence is used to define rare cancers with the threshold of 6/100,000/year, as defined by RARECARE (Surveillance of Rare Cancers in Europe project).

1.5.2. Epidemiological related definitions

Point prevalence: Number of cases scaled up to the general population at a given time.

Birth prevalence[1]: Number of cases observed at birth relative to the number of children born alive at a given moment.

Annual incidence: Number of newly diagnosed cases in a population in one year.

Lifetime prevalence: Number of cases presenting or having presented the clinical entity during their lifetime scaled up to the general population.

Cases/families: Number of cases or family (ies) published in the literature.

1.5.3. General process related definitions

Expert: Experts mentioned in this procedural document are the health professionals identified by Orphanet as leaders in the medical field for a rare disorder or a group of rare disorders, based on a set of criteria including: number of publications; leadership in expert centres; leadership in expert networks; leadership in research, including clinical trials.

Geographical area: Finite list of terms composed of the countries present in the ISO 3166-1 alpha-2 norm and 9 large geographical areas (for more details on the latter please refer to annex 1 of this document).

[1] Birth prevalence and birth incidence are similar concepts. Birth prevalence is methodologically more accurate (http://www.ncbi.nlm.nih.gov/pubmed/16240384).

Validated status: Information that has been checked and is accepted. A data element is validated when identified from its original source (e.g. registry, institute, study) or if an expert confirms its accuracy.

Medical Subject Headings (MeSH): Hierarchically-organised terminology that uses comprehensive controlled vocabulary for indexing and cataloging of journal articles and books in the life sciences. It serves as a thesaurus that facilitates searches in the literature.

Medline: Database that contains more than 25 million references to journal articles in life sciences with a concentration on biomedicine, all indexed with the NLM Medical Subject Headings (MesH).

National Library of Medicine (NLM): Largest electronic biomedical library in the world. It maintains and makes available a vast print collection and produces electronic information resources on a wide range of topics.

Pubmed: Free search engine accessing primarily the Medline database of references and abstracts on life sciences and biomedical topics.

Source: The person or document that supplies the information. (e.g. expert, registries, institutes, books, medical literature on Medline).

1.6. Filing and updates

This procedure is updated at least annually and as often as necessary by the epidemiologist in charge of the epidemiological data. The most up-to-date version is available on the Orphanet website:https://www.orpha.net/orphacom/cahiers/docs/GB/Epidemiology_in_Orphanet_R1_Ann_Epi_EP_05.pdf

II. Epidemiological data characteristics

Orphanet aims to collect epidemiological data on every rare clinical entity included in its Nomenclature. This data can describe groups of disorders, disorders or subtypes of disorders.

Even if point prevalence is recognised as the most appropriate epidemiological indicator for RD as it provides a measurement of the disease's burden, the scarcity of data and the diversity of RDs has led Orphanet to collect various epidemiological indicators: point prevalence, annual incidence, birth prevalence, lifetime prevalence and cases and families reported in the literature. These indicators are registered either as numerical estimates or pre-defined ranges <1/1,000,000, 1–9/1,000,000, 1–9/100,000, 1–5/10,000, 6–9/10,000, >1/1,000 (see Figure 1).

Data is further characterised by a geographical area (country, continent or worldwide) or a particular population if relevant (e.g. Ashkenazi Jews).

The sources from which each epidemiological feature is derived is also made available.

A 'validation status' also informs users on the reliability of the data. A data is validated when identified from the original source (peer-reviewed studies, specialised reports, national or international registries) or if an expert confirms its accuracy. 'Not yet validated' status means that the original source cannot be identified and that there has been no validation from an expert.

Finally, for some RDs, no epidemiological data can be retrieved, even after a comprehensive search. Orphanet delivers this information as « unknown » data for a geographical area. Indeed, for a given disorder, an epidemiological feature can be available in one specific European country, but be « unknown » in Europe.

As such, « Unknown » data is a meaningful information, different from « not yet documented », which means that no documentation has yet been carried out.

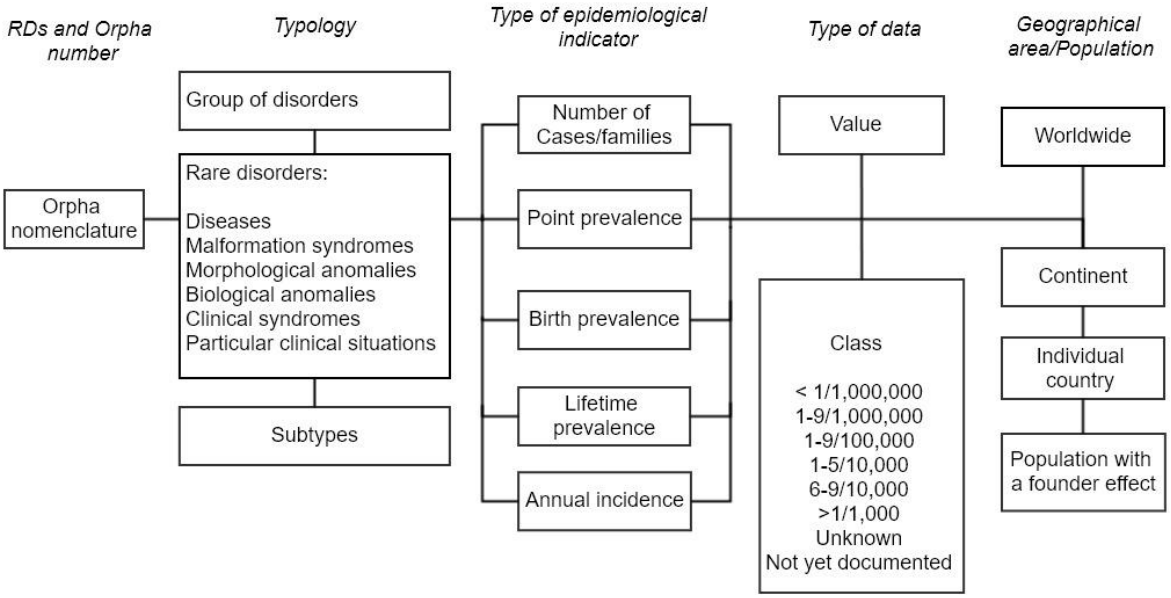

Figure 1 : Representation of Epidemiological data in the Orphanet database

II. Methodology

2.1 Flowchart

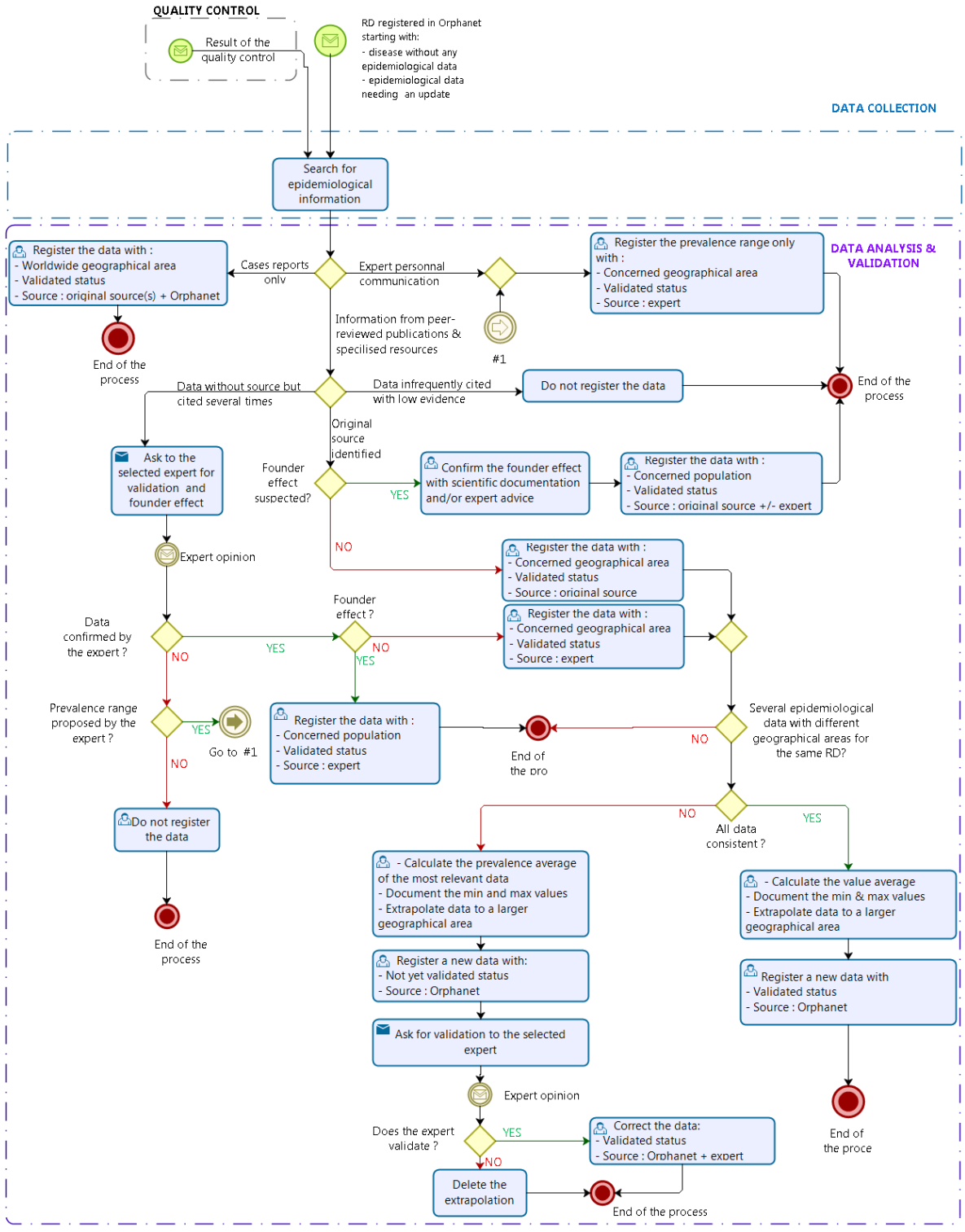

Figure 2 : Workflow of a epidemiological data in Orphanet. This workflow has been powered by Bizagi

2.2 Description

An epidemiological data search is performed for every rare disorder registered in the Orphanet database, but priority is given to:

- Rare disorders for which epidemiological data has not yet been documented;
- Rare disorders for which epidemiological data have been documented but needs updating according to the most recent study (or studies) or according to quality control results.

2.2.1 *Data collection*

As much as possible, data collection is based on population-based studies, meta-analysis and population surveys. They are identified through a continuous and systematic peer-reviewed literature survey on RDs, including publications with the highest impact factor in their speciality (i.e. Lancet neurology). The complete list of peer-reviewed journals surveyed can be found in Annex 2 of this document. Specific queries complete this systematic survey. There are carried out on the NLM (National Library of Medicine) website through the PubMed database using the following algorithm: {[Disease names [MeSH] AND Epidemiology [MeSH: NoExp] OR Incidence [Title/abstract] OR Prevalence [Title/abstract] OR Epidemiology [Title/abstract]}. [2]

Furthermore, some valuable specialised resources including reports, registries, databases or websites are also monitored to increase epidemiological data collection on RDs. It includes for example: the public summary of opinion on orphan designation published by the European Medicines Agency's Committee for Orphan Medicinal Products; EUROCAT for congenital anomalies in Europe; the World Health Organization (Global Health Observatory Data Repository).

Finally, for RDs where an epidemiological estimate cannot be easily inferred from peer-reviewed publications or specialised resources, Orphanet calls on expert advice to provide additional information.

For a complete list of documentation sources, see the reference section of this document.

2.2.2 *Data analysis*

The curation process is schematised in figure 2 and explained in detail in the following sections. Three different scenarii are applied depending on the characteristics of the epidemiological data analysed.

2.2.2.1 Data from peer-reviewed publications and specialised resources

Information from peer-reviewed publications or specialised resources are systematically examined. Orphanet also studies data frequently cited and collectively admitted in the literature. However, data infrequently cited with a low level of evidence are never retained.

[2] This allows extracting all the records where a disease name has been indexed as a MeSH term and where you can find either incidence, prevalence or epidemiology in the title or abstract. The 'NoExp' term is used to reduce the noise as much as possible

By default, Orphanet keeps but does not rely on data frequently cited and collectively admitted without an identified original source. An expert is systematically contacted when the original source is not clearly identified. However, if the original source(s), whether it be a peer-reviewed publication or a specialised resource, is/are identified, the documentation process continues without requesting futher expert advice.

At this step, epidemiological data are registered in the database according to the following rules:

If a founder effect is made evident (either by scientific documentation and/or by expert advice), data is registered for the concerned population and not for a geographical area. However, if the disorder's aetiology is not associated with a founder effect, the data is registered according to the geographical area (worldwide, continent, country) on which the selected study is based.

If the original source is clearly identified or if an expert confirms that the frequently cited data is trustworthy (even without an identified original publication), then the data obtains a "validated" status. If available, the reference of the original source is added to the dataset as well as the name of the expert having taken part in the validation. Otherwise, the data obtains a "not yet validated" status. When an expert invalidates a data that is commonly cited but that lacks a clearly identifiable source, then a number of options are possible: only a prevalence range is attributed; the "unknown" value for the specific geographical area is attributed; when no range can be attributed, the data is not included for the given geographical area if the disease is already sufficiently documented in the database.

No further investigation is carried out for data related to a founder effect and/or for frequently cited data not confirmed by an expert.

Specific cases:

- Some resources, such as registries, publish their data repeatedly over time. As the same methodology is applied, the average value is registered with the smallest and the highest value, e.g., Annual incidence of renal agenesis in France available from 2008 to 2012. The mean value within this period is collected for France.
- When several studies with different methodologies provide various data for the same disorder in the same geographical area, the most recent data with the most reliable methodology is registered, .e.g., Orphanet detected two point prevalences for primary Sjogren's syndrome in Greece, one published in 2005 based on a questionnaire or clinical examination (population survey), and another based on a medical record search or physicians' personal registries (population-based) published in 2006. Orphanet only registered the 2006 value for Greece.
- When a study is undertaken in a single part of a country and is related to a disorder without known differences among ethnicities, then the data is extrapolated to the whole country. The publication as well as Orphanet (as it assumes responsibility for this extrapolation) are added as sources, .e.g., point prevalence for Alexander disease in Scotland is extrapolated to the United Kingdom.
- When a study is undertaken in a sub-population, the data is extrapolated to the general population of the specific geographical area. The publication as well as Orphanet (as it assumes this extrapolation) are added as sources, e.g., point prevalence on X-linked diseases originally reported in the male population is divided by two and considered as extrapolated to the general population.

In order to expand as much as possible the available knowledge, further investigations are carried out for trustworthy data (with a "validated" status) registered for the same disorder in different geographical areas. The main objective is to extrapolate to a larger geographical area

for disorders with no differences amongst ethnicities. To do so, the consistency of the data registered in Orphanet is collectively analysed and only the most accurate data sources that meet a certain number of quality criteria[3] are kept for further calculation. An average value is calculated and extrapolated to the closest larger geographical area, e.g., point prevalence value from Sweden (7/100,000), and Denmark (31/100,000) are registered for Mucopolysaccharidosis type 4. The average value (19/100,000) is then registered with the minimum value (7/100,000) and the maximum value (31/100,000), and then extrapolated to Europe.

If there is an overall consistency amongst all the original studies, then the extrapolated data is validated and Orphanet is added as a source as it assumes responsibility for its reliability. If some of the original sources have been discarded for calculation due to their lower quality level, the extrapolated data is validated only if confirmed by an expert. Then, the expert is added to Orphanet as a source.

2.2.2.2 Data from case reports

For extremely rare disorders for which there is no population-based study, the number of cases or families is collected from published case report(s), case serie(s) or recent review of the literature with a worldwide coverage. The total number of cases or families reported is calculated. All the publications are added as sources, with a "validated" status. The lowest point prevalence range (<1/1,000,000) is assigned to the disorder, with a worldwide coverage and Orphanet is also added as a source as it assumes responsibility for its reliability.

2.2.2.3 Data from personal communication only

When there are no peer-reviewed publications or specialised resources enabling epidemiological data collection for a given rare disorder, an epidemiological range (including the "unknown" data) can be assigned on expert advice. The data is considered reliable, so a "validated" status is added and the expert is noted as a source.

2.2.3 *Calculation of point prevalence value from other epidemiological indicators*

As point prevalence is the most reliable epidemiological indicator for RDs, Orphanet tries to estimate a resulting point prevalence value, when not available, from birth prevalence and incidence value according to the following method:

- For congenital disorders with a well-known life expectancy:

Point prevalence = birth prevalence x (patient life expectancy/general population life expectancy).

- For disorders with a known mean duration:

Point prevalence = incidence x disease mean duration

[3] Study design based on population, study with case ascertainment established using the most recent internationally accepted diagnostic criteria, and case finding method including administrative databases or hospital medical records.

The resulting point prevalence estimates are validated only if confirmed by an expert. The expert is added to Orphanet as the source.

2.2.4 *Quality control*

Post-release quality control is performed on a regular basis to ensure the coherence between the most up-to-date summary information published in the Orphanet encyclopaedia and the corresponding rare disorder epidemiological data.

Systematic data checks and error detection strategies are carried out regularly to ensure consistency, e.g, the "validated" status is associated to the original source, the lowest point prevalence range (<1/1,000,000) is assigned to disorders for which there are only case or family reports.

2.3 Data availability

The worldwide prevalence range is displayed on the Orphanet website on each specific disorder page. If no worldwide prevalence range is available, then European values are given.

The entire set of epidemiological data: low-high and average values for point prevalence, prevalence at birth and annual incidence, point prevalence ranges, as well as cases or families described in the literature, are available on the Orphadata website as free datasets (CC BY 4.0). The geographical area or specific population category is given for each datum.

Point prevalence, prevalence at birth, annual incidence figures and cases and families reported in the literature are published in two Orphanet Report Series and are updated every six months.

This information feeds into the epidemiology section of the abstracts displayed on the website.

III. Annexes

3.1 Alphabetical list of the large geographical areas registered in Orphanet

- Africa
- Eastern Mediterranean Asia
- Europe
- Latin America
- North America
- Oceania
- South East Asia
- Western Asia
- Worldwide

3.2 Alphabetical list of peer reviewed journal consulted for the epidemiological data curation in Orphanet

American Journal of Human Genetics	Journal of Investigative Dermatology
American Journal of Medical Genetics	Journal of Medical Genetics
Annals of Internal Medicine	Journal of Neuromuscular Diseases
Annals of Neurology	Journal of Rare Diseases: Research & Treatment
Annals of the Rheumatic Diseases	Journal of Rare Disorders: Diagno & Therapy
Archives de Pédiatrie	Journal of the American Medical Association
Arthritis and Rheumatology	Journal of the American Society of Nephrology
Arthritis Care & Research	Lancet
Blood	Lancet Infectious Diseases
BMC Medicine	Lancet Neurology
Bone	Lancet Oncology
Brain	Molecular Genetics and Metabolism
British Journal of Haematology	Molecular Syndromology
Bulletin du Cancer	Molecular therapy
Cell Reports	Muscle & Nerve
Cell Stem Cell	Nature
Circulation	Nature Genetics
Clinical Genetics	Nature Medicine
Cochrane Reviews (Cochrane Database of Systematic Reviews)	Nature Neuroscience
Current Rheumathology Reports	Nature Reviews Clinical Oncology
Dermatologic Clinics	Nature Reviews Endocrinology
Diabetes	Nature Reviews Immunology
European Heart Journal	Nature Reviews Nephrology
European Journal of Haematology	Nature Reviews Rheumathology
European Journal of Human Genetics	Neurology
European Journal of Internal Medicine	Neuromuscular Disorders
European Journal of Medical Genetics	Orphan Drugs: Research and Reviews
Expert Opin Orphan Drugs	Orphanet Journal of Rare Diseases
Familial Cancer	Pediatric Research
Gastroenterology	Pediatrics
Gene Therapy	Plos Genetics
Genet Med	Proceedings of the National Academy of Sciences
Genome Medicine	Progress in Retinal and Eye Research
Gut	Rare Diseases (Taylor & Francis)

Hepatology	Revue de Médecine Interne
Human Genetics	Science
Human Molecular Genetics	Science Translational Medicine
Human Mutation	Stem Cell Reports
International Journal of Rare Diseases & Orphan Drugs	Stem cells
Intractable & Rare Diseases Research	Stem Cells Translational Medicine
JAMA Neurology	The Journal of Rare Disorders
Journal of Allergy and Clinical Immunology	The New England Journal of Medecine
Journal of Clinical Endocrinology and Metabolism	Translational Science of Rare Diseases
Journal of Clinical Investigation	

For any questions or comments, please contact us: contact.orphanet@inserm.fr

Editor of this procedural document: Stéphanie Nguengang Wakap and Annie Olry - This procedural document has been approved by : Ana Rath - Quality control : Charlotte Gueydan

Identification code of the document: [code identification de la procedure]. Version of the document : 01

The correct form when quoting this document is:

« Procedural document on Epidemiology of rare disease in Orphanet (Prevalence, incidence and number of published cases or families), Orphanet, February 2019, Version 01

https://www.orpha.net/orphacom/cahiers/docs/GB/Epidemiology_in_Orphanet_R1_Ann_Epi_EP_05.pdf