The Orphanet Nomenclature of rare diseases: a standard terminology for improved patient recognition and data interoperability

Caterina Lucano[1,*], David Lagorce[1], Annie Olry[1], Houda Ali[1], Valérie Lanneau[1], Mickael De Carvalho[1], Aysegul Dilsizoglu Senol[1], Marta Fructuoso[1], Emilie Gaillard[1,2], Marie-Cécile Gaillard[1], Seed Mihic[1], Mariana Tannoury[1], Florence Sauvage[1], Charlotte Rodwell[1], Sylvie Maiella[1], Marc Hanauer[1], Ana Rath[1,*]

[1]Inserm, US14-Orphanet, Paris, France

[2]French National Rare Disease Registry (BNDMR), Greater Paris University Hospitals (AP-HP), Paris, France

* Corresponding authors

E-mail: caterina.lucano@inserm.fr (CL)

E-mail: ana.rath@inserm.fr (AR)

Abstract

Although individually uncommon, rare diseases (RD) affect an estimated 300 million people worldwide. Establishing a public health approach to RD requires counting diseases and affected patients. However, RD are under-represented in medical terminologies, with only a small fraction of RD possessing specific and unambiguous codes, and those codes not being explicitly designated as rare.

To tackle the challenge of RD codification and interoperability, Orphanet has developed a nomenclature of RD that provides unique and time-stable disease identifiers (ORPHAcodes) and meets the gold standards for implementation in health information systems and systematic research collections. The Orphanet Nomenclature of RD is multilingual and versioned; its development and updates rely on standardized procedures, manual curation and expert validation, reflecting advancements in RD knowledge and clinical practice.

This work provides an overview of the Orphanet Nomenclature and classification system for RD, detailing its production, update and mapping methodology. As of July 2025, the Orphanet Nomenclature contains a total of 6,527 RD, multiclassified into 29 classifications, each corresponding to a medical domain, accurately representing the complex multisystemic nature of RD. Extensive qualified mappings ensure semantic interoperability across multiple terminologies. Overall, 97.4% of RD are mapped to at least one ICD-10 code (with only 6.3% exhibiting an exact equivalence), 71.7% are mapped to at least one ICD-11 code (15.3% with an exact equivalence) and 93.8% are mapped to SNOMED CT (all with an exact equivalence). Genetic diseases represent 72.2% of all RD, and 63.4% are mapped to at least one phenotypic OMIM number.

By addressing the underrepresentation of RD in medical terminologies, ORPHAcodes facilitate accurate patient identification, advance research and healthcare interoperability,

and help in shaping public health policies. The recognition of the Orphanet Nomenclature as the most appropriate terminology for RD clinical coding in Europe underscores its critical role in the global RD ecosystem.

Introduction

Cross-border data collection, sharing, and exploitation to support evidence-based policymaking, healthcare, and research are essential in the field of rare diseases (RD), as underscored by numerous international and European initiatives [1]. Although RD are individually uncommon, with over 6,000 different known RD, they collectively affect an estimated 300 million people worldwide [2]. The critical amount of data necessary to improve knowledge, care and treatment can thus only be achieved through cumulative data collection across different countries in a standardized manner [3,4].

The concept of RD first emerged in 1972 in Japan with the launch of the National Program on Rare and Intractable Diseases [5]. Since then, numerous definitions have been proposed, reflecting local legislations [4,6–11]. Most of these definitions use point prevalence as the common epidemiological indicator to define the rarity of a disease. However, the threshold for rarity varies geographically: less than one per 2,500 in Japan [5], less than 50 per 100,000 in Europe [12], or no more than 200,000 persons in the US, which translates to a prevalence of 86 per 100,000 at the time of the American Orphan Drug Act (1983) [13]. Furthermore, definitions can differ clinically, sometimes including restrictions such as “life-threatening” or “chronically debilitating” [6,12,14].

To address these definition inconsistencies, Rare Diseases International (RDI), in collaboration with a global panel of experts and the World Health Organization (WHO), has developed an Operational Description of Rare Diseases [15]. This work has defined a RD as a medical condition with a specific pattern of clinical signs, symptoms, and findings that affects fewer than or equal to 1 in 2,000 persons living in any World Health Organization-defined region of the world. The definition moreover emphasizes the need for a standardized approach to account for the unique challenges faced by individuals with RD, including long

74 diagnostic journeys, inadequate clinical management, and limited treatment options [15].
75 Despite this effort offering a path toward greater convergence, the continued lack of
76 harmonization among countries, combined with the limitations of the currently adopted
77 definitions, impacts research on RD mechanisms and therapies, diagnostic approaches, and
78 healthcare planning [1,3,8].

79 To take public health action on RD, the community needs to know how many people
80 are affected by these diseases (prevalence), how these conditions impact those affected, and
81 how to monitor their medical and societal consequences. Establishing a public health
82 approach to RD includes counting and measuring RD, which is essential to support actions that
83 reduce their impact on patients and society. This would improve health, quality of life, life
84 expectancy, disease management, and patient involvement in their communities and society
85 [2,4,16]. Indeed the United Nations Resolution on People Living with a Rare Disease
86 "encourages Member States and relevant United Nations agencies to collect, analyze, and
87 disseminate disaggregated data on persons living with a rare disease... to assess progress
88 towards the improvement of the status of persons living with a rare disease". This further
89 highlights that adequate coding of RD patient data is a prerequisite for an effective response
90 to this global public health priority [17].

91 However, RD are under-represented in medical terminologies used within health
92 systems, with only a small fraction of RD possessing specific and unambiguous codes [18–21].
93 This makes it difficult to trace patients with RD in health information systems at the national
94 level, preventing a robust understanding of the actual number of people affected by a RD.
95 Only a small proportion of RD have a specific code in ICD-10 [19–21]; the representation of RD
96 has increased in ICD-11 [20,22] and SNOMED CT [23], two other global medical terminology
97 standards, though it still remains below the number of known RD. Nevertheless, RD cannot be

98 easily identified in these generic systems, as they do not include any indication of the rarity of
99 a disease. Furthermore, different regions use diverse medical terminologies (e.g. the
100 International Classification of Diseases ICD or SNOMED CT) or diverse versions of the same
101 terminology (ICD local extensions, different ICD versions), hampering data sharing and
102 interoperability. Given the already scarce nature of RD data, this barrier to data collection
103 requires dedicated efforts to make RD recognizable in health information systems [24–26].

104 The availability of adequate codes for each RD is now a well-recognized need at
105 international level [24,26]. Improved codification for RD has been cited as a priority in the
106 2009 EU Council Recommendation on an action in the field of rare diseases [26], and more
107 recently by the Rare2030 recommendations [27]. Having codes for each RD would help
108 international, European, national, and local health authorities obtain better knowledge of
109 healthcare pathways and of their impact on both specialized (such as centers of expertise) and
110 general health care services as well as on budget planning for health and social services.
111 Improving the visibility of RD in healthcare information systems could help to define patient
112 pathways based on real-life data. A common standard for describing RD would also aid in
113 developing approaches to alleviate the substantial cumulative burden of RD [8,24,26].
114 Adequate coding is also necessary for identifying patients for clinical trials at both national
115 and transnational levels. Finally, a large proportion of RD patients remain undiagnosed despite
116 undergoing a full investigation, but this particular situation cannot be captured by the
117 codification systems in use.

118 To tackle the challenge of RD codification and interoperability, Orphanet has
119 developed and maintained a nomenclature of RD since 1997 and a classification of RD since
120 2006 [19,28]. In 2014, the Commission Expert Group on Rare Diseases adopted a
121 recommendation on improving RD codification [24]. In this document, Member States are

encouraged to assess the feasibility of implementing ORPHAcodes at a national level and to include RD codification into their national RD plans/strategies. In 2017, the EC Steering Group on Promotion and Prevention (SGPP) recognized ORPHAcodes as a best practice [29]. Since then, ORPHAcodes have been included in the set of common data elements for Rare Diseases Registration, released by the EU RD platform to enhance interoperability between registries [30]. The Orphanet Nomenclature of RD has also played a key role in implementing the EU 2011 Cross-Border Healthcare Directive as concerns RD. As a result, ORPHAcodes have been recommended for tracking RD diagnoses in the European Common Semantic Strategy (2019) [31] and the Guidelines on Patient Summary (Release 3.2, March 2022 and Release 3.4, November 2024) [32], and have been integrated into the Code Systems used in the MVC - My Health @ EU - eHealth Digital Service Infrastructure (eHDSI). The functional specifications developed by the X-eHealth project also include ORPHAcodes in the minimum dataset required for the electronic Patient Summary [32,33]. Finally, the WHO resolution "Rare diseases: a global health priority for equity and inclusion" adopted on the 24th of May 2025 urges Members States to commit to considering the implementation of ICD-11, and where appropriate, interoperable codification systems for RD such as the Orphanet Nomenclature of rare diseases, at their earliest possibility, and in accordance with their available resources, in order to enable the recording, reporting and monitoring of rare diseases at the national and international levels [34].

This paper presents an overview of the Orphanet Nomenclature and classification of RD, detailing its production and update methodology. It includes a clear and up-to-date count of RD based on the Orphanet knowledge base, as well as their distribution by medical domain, and information on mappings to other medical terminologies. Finally, this work highlights the

145 importance of maintaining a standardized and reliable RD nomenclature for clinical coding,

146 research, healthcare interoperability, and policy-making.

147 # Materials and Methods

148 ## The structure of the Orphanet Nomenclature and classification of RD

149 The Orphanet Nomenclature of RD is composed of clinical entities, each of which
150 includes several elements to ensure precise and unambiguous coding (Fig 1) [35]:

151 - ORPHAcode: a unique and time-stable numerical identifier automatically assigned in
152 ascending order to every entry registered in the Orphanet Nomenclature;
153 - Preferred term (or main name): the most widely accepted name for the diagnosis in
154 the literature and/or in the medical community;
155 - Synonym(s): additional terms perfectly equivalent to the preferred term, as many as
156 necessary according to the information found in the literature;
157 - Classification level: indicates the level of precision of the clinical entity, among three
158 possible granularity levels*:* Group of disorders, Disorder, or Subtype of disorder;
159 - Definition: a brief description of the major clinical characteristics that define the
160 diagnosis and differentiate it from other clinically related clinical entities.

161

162

163 **Fig 1. Technical elements composing the Orphanet Nomenclature of RD.**
164 Example of a clinical entity in the Orphanet Nomenclature of RD, defined by an ORPHAcode,
165 a preferred term, as many synonyms as needed, a classification level, and a textual definition.

166

167

168 In the absence of an international consensus on how to name diseases, preferred
169 terms and synonyms are assigned according to standard naming rules developed by Orphanet

170 to create a comprehensive RD terminology that aligns with current knowledge and clinical
171 practice [36]. These rules ensure consistency across different disorder groups and medical
172 domains, facilitating better integration and use of RD-related clinical information in databases
173 and healthcare systems.

174 The Orphanet Nomenclature of RD is produced in English and entirely translated into
175 eight languages (French, German, Spanish, Italian, Dutch, Portuguese, Polish and Czech) by
176 translation teams located in member countries of the Orphanet Network. All translations are
177 validated by RD expert medical doctors in each country. The English terminology serves as
178 the basis for the translation of concepts; however, the translation process considers the
179 particularities and local practices specific to every language, in particular with regard to the
180 number of synonyms [37].

181 The Orphanet Nomenclature of RD is structured in a classification system composed
182 by the three classification levels above-mentioned, namely Group of disorders, Disorder, and
183 Subtype of disorder (Fig 2). The classification level serves as the basis for parent-child
184 relationships between related clinical entities. These relationships structure the classification
185 according to successive criteria that are relevant in the clinical setting [35]. Additionally, each
186 clinical entity is attributed a typology, a characteristic that, within the classification level,
187 conveys the nosological definition of the clinical entity. For Group of disorders, typologies
188 include Clinical group and Category; for Disorders, they include Disease, Clinical syndrome,
189 Biological anomaly, Malformation syndrome, Morphological anomaly, and Particular clinical
190 situation in a disease or syndrome; and for Subtype of disorder, they include Clinical subtype,
191 Etiological subtype, and Histopathological subtype (Fig 2) [35]. While generally used
192 interchangeably to accommodate the widespread use of the term “rare diseases”, the terms
193 “disease” and “disorder” designate different concepts within the Orphanet Nomenclature of

194 RD. As previously explained, the term *disease* represents a narrower clinical concept, whereas
195 the term disorder is used to refer to the broader classification level [35].
196
197
198 **Fig 2. Classification levels and typologies of the Orphanet Nomenclature of RD.**
199 Every clinical entity in the Orphanet Nomenclature of RD is associated with a classification
200 level conveying the precision of the diagnosis (Group of disorder, Disorder, or Subtype of
201 disorder), and a unique typology within the classification level.
202
203
204 The Orphanet classification is also structured in 34 major classification groups,
205 individually referred to as classification hierarchy. Of these 34 classifications, 29 represent a
206 distinct medical domain, and 5 are ad-hoc classifications used for sorting purposes. These 5
207 hierarchies have been excluded from the analysis in this work, for the reasons outlined below:
208 - Orphanet classification of rare genetic disorders: this classification does not represent
209 a distinct medical domain but is rather a derived genetic version that mirrors all the
210 other clinical classification hierarchies;
211 - Orphanet classification of rare systemic or rheumatological disorders of childhood: this
212 classification is fully included into the Orphanet classification of rare systemic or
213 rheumatologic disorders;
214 - Orphanet classification of rare transplant-related disorders: this classification does not
215 represent a medical domain but is used to class disorders according to a surgical
216 intervention;

217 - Orphanet classification of teratologic disorders: this classification is fully included into
218 the Orphanet classification of rare developmental anomalies during embryogenesis;
219 - Orphanet classification of rare disorder without a determined diagnosis: this
220 classification does not represent a medical domain and is used to class one single
221 disorder, ORPHA: 616874 Rare disorder without a determined diagnosis after full
222 investigation [38]. This is a code that allows the identification of undiagnosed rare
223 disorder patients in healthcare information systems and that does not conceptually
224 fall under any of the other Orphanet classification hierarchies.

225 Every classification consists in an arborescence structure that starts with a unique ultimate
226 parent group at the top representing the broadest clinical concept/the medical domain (head
227 of classification), then branching out into as many intermediate groups as necessary with
228 increasing precision of clinical concepts, until reaching the first level of the confirmed
229 diagnosis, the disorder. For any given disorder, subtypes can be added when this level of
230 precision is required in clinical practice for appropriate management and follow-up.

231 Every clinical entity in the Orphanet Nomenclature of RD is introduced in as many
232 classifications as necessary according to major medical specialties involved in the
233 management of the disorder. This results in a multi-hierarchical and multi-parental
234 classification structure that reflects the multi-dimensionality and complexity of RD (Fig 3).
235 However, only one ultimate parent group is chosen as the entity's preferential classification
236 for linearization (preferential parent), according to the most severely affected body system,
237 the most determining involvement for the prognosis, and/or the medical specialty that will
238 most likely be relied on for the management of the disorder [39]. This process, named
239 linearization, prevents concepts ambiguity in the presence of multiple classifications and
240 enables coherent statistical analysis.

241

242

243 **Fig 3. Representation of a disorder in the Orphanet classification system.**

244 The disorder Atypical Hemolytic Syndrome (ORPHA:2134) is listed under four classifications.
245 It has two direct parent groups, each with broader parent groups, culminating at the top of
246 the hierarchy, where each ultimate parent group represents a medical domain. The Rare Renal
247 Disorder group (ORPHA:93626) was designated as the preferential parent of the disorder
248 Atypical Hemolytic Syndrome (ORPHA:2134). Note that all groups in the figure have additional
249 child entities not shown here.

250

251

252 Production and update process of the Orphanet Nomenclature of RD

253 The Orphanet Nomenclature of RD is produced and updated according to a
254 standardized and published procedure, in order to maintain a maximum level of quality of the
255 information, by a scientific team of information scientists at the Institut national de la santé
256 et de la recherche médicale (Inserm) US14 service unit, Paris, France. The information
257 provided in the Orphanet Nomenclature is sourced from up-to-date scientific and medical
258 literature (more than 2,000 papers are consulted yearly), manually curated, reviewed and
259 validated by Orphanet information scientists and RD medical experts, when pertinent [35,40].

260 The maintenance of the Orphanet Nomenclature of RD is ensured through three main
261 types of actions:

262 - the **inclusion** of a new clinical entity: to be included in the Orphanet Nomenclature, a
263 clinical entity must represent a clinically homogeneous phenotype with a prevalence
264 of no more than five in 10,000 people and be described in at least two individuals in

265 the literature [15]. Before inclusion, the clinical entity is compared with existing
266 entities to avoid overlaps and ensure classification consistency;
267 - the **modification** of an existing entity: a clinical entity in the Orphanet Nomenclature
268 of RD can be modified to improve or update its information, including the preferred
269 term, synonyms, classification level, typology or the definition. Clinical entities
270 described before the genetic era, and lacking publications since then, are labeled as
271 "Historic" and are kept in the Orphanet Nomenclature because their initial descriptions
272 are still available and carry a potential for further characterization;
273 - the **inactivation** of a clinical entity: a clinical entity in Orphanet can be inactivated due
274 to deprecation (it is now included in another concept), obsolescence (it has no valid
275 reason to exist, such as being an imprecise or duplicate concept), or for not meeting
276 the rare disorder prevalence criteria anymore.
277 To ensure continuity of information, especially in health information records and coding
278 systems, Orphanet provides a replacement active clinical entity for deprecated and obsolete
279 entities. This replacement code, or target code, must be used instead of the corresponding
280 deprecated clinical entity as it aligns with the evolution of knowledge in the literature. For
281 obsolete entities, the target often refers to a group of disorders, guiding users to explore the
282 relevant classification to identify the most appropriate replacement code based on the
283 patient's diagnosis. Entities inactivated as non-rare in Europe are not given a replacement
284 code, as they no longer fall under the RD scope.
285 An ORPHAcode is permanently linked to the same clinical concept and is never deleted
286 nor reused for a different entity, even if the original entity becomes inactive. In rare instances,
287 an inactive clinical entity may be reactivated after reexamination with updated knowledge

288 and reintroduced into the Orphanet Nomenclature and classification with all relevant
289 scientific information.
290
291
292 ## Mappings to other reference terminologies
293 The Orphanet Nomenclature of RD is mapped to international terminologies and
294 reference databases according to a standardized and published procedure, to enable
295 interoperability between different information systems [41]. Target terminologies include
296 ICD-10 [42], ICD-11 (MMS and Foundation) [43], SNOMED CT [44,45], OMIM [46], UMLS [47],
297 MeSH [48], MedDRA [49], GARD [50], and MONDO [51].
298 The mapping cardinality always goes from the ORPHAcode to a code of the target medical
299 terminology. All mappings are qualified by a proximity relationship that defines the nature of
300 the correspondence between the two concepts. The adopted proximity relationships are:
301 - E (Exact): the Orphanet entity designated by an ORPHAcode and the target code have
302 the same range of application, they describe the same pathological entity;
303 - BTNT (Broader Term to Narrow Term): the Orphanet entity designated by an
304 ORPHAcode has a broader range than the target code;
305 - NTBT (Narrow Term to Broader Term): the Orphanet entity designated by an
306 ORPHAcode has a narrower range than the target code.
307 Moreover, the mappings with the WHO International Classification of Diseases (ICD-10 and
308 ICD-11) [41] are further characterized by a specificity relationship, namely:
309 - Specific code: the ORPHAcode has its own code in the ICD-10/ICD-11 Mortality and
310 Morbidity Statistics (MMS);

- Index term: the Orphanet entity is listed in the ICD-10 alphabetical index/in the ICD-11 Foundation;
- Inclusion term (ICD-10 only): the Orphanet entity is included under an ICD-10 category in the tabulated list under the code but it has not its own code;
- Attributed code: the ICD-10/ ICD-11 MMS code is attributed by Orphanet because of the lack of matching terms.

All mappings are manually curated and validated by a medical doctor before publication, unless explicitly marked as *Not yet validated* to indicate pending medical review.

Data extraction and analysis

Data metrics from the Orphanet Nomenclature of RD were computed using the Orphanet Scientific Knowledge Files delivered on the Orphadata Science website [52]. These datasets, structured in an Extensible Markup Language (XML) file format, are derived from the Orphanet Knowledge base, which is also accessible through the Orphanet website [28]. They are delivered twice a year in July and December and all previous versions are archived in a dedicated GitHub repository [53]. For this study, we analyzed 4 datasets of the 2025 release:

- the rare diseases and alignments with terminologies and databases dataset (<iso>_product1.xml) in English, Czech, Dutch, French, German, Italian, Polish, Portuguese and Spanish languages; version 1.3.42 / 4.1.8 [2025-03-03] released 24 Jun 2025, available at https://sciences.orphadata.com/alignments/
- the linearization of rare diseases English dataset (en_product7.xml); version 1.3.42 / 4.1.8 [2025-03-03] released 24 Jun 2025, available at https://sciences.orphadata.com/linearisation/

335 - the genes associated with rare diseases English product (en_product6.xml); version
336 1.3.42 / 4.1.8 [2025-03-03] released 24 Jun 2025, available at
337 https://sciences.orphadata.com/genes/
338 - The classifications of rare diseases English datasets (en_product3_<x>.xml); version
339 1.3.42 / 4.1.8 [2025-03-03] released 24 Jun 2025, available at
340 https://sciences.orphadata.com/classifications/
341 Metrics were computed by using Python 3.8 scripts freely available in dedicated Jupyter
342 Notebooks (.ipynb files) shared in a Binder/Github-ready repository
343 (https://github.com/Orphanet/orphanet_nomenclature_binder). This repository provides a
344 publicly accessible, reproducible, and interactive computational environment that enables on-
345 demand calculation of metrics using Orphadata files. Eight scripts have been coded in order
346 to calculate the necessary metrics:
347 - 1_number_of_Orphanet_clinical_entities
348 - 2_production_of_the_Orphanet_RD_terminology_in_english_and_other_languages
349 - 3_proportion_of_genetic_versus_non_genetic_Orphanet_RD
350 - 4_number_of_Orphanet_RD_monoclassified_vs_multiclassified
351 - 5_distribution_of_Orphanet_RD_in_Orphanet_classifications
352 - 6_distribution_of_Orphanet_RD_by_preferential_parent
353 - 7_alignment_between_Orphanet_RD_and_international_terminologies
354 - 8_number_of_Orphanet_RD_with_definition_in_all_languages
355 Each script in the Jupiter Notebook has been annotated to clarify the calculation methodology.
356 The human readable Orphanet-SNOMED-CT mapping file (ORPHA-
357 SNOMEDCT_Mapping_File_production.xlsx); July 2024 version, released 15 Oct 2024,

358 available at: https://www.orphacode.org/pack-nomenclature/ was also used to report on the

359 Orphanet-SNOMED CT mappings.

Results

The Orphanet Nomenclature of RD

As of July 2025, the Orphanet Nomenclature of RD contains a total of 9,784 active clinical entities, including 6,527 disorders (among which is a specific disorder code for undiagnosed patients after full investigation [38]), 1,084 subtypes of disorders, and 2,173 groups of disorders. Among all active entities, 3.3% (325 disorders and 1 subtype) are considered historic, with no publications over the last 25 years. The distribution of all active clinical entities by their classification level and their typology is presented in Table 1 and in S1 File.

Table 1. Distribution of active clinical entities in the Orphanet Nomenclature of RD by classification level and typology.

	Number of active clinical entities	% on total clinical entities	% on each classification level
Disorders	**6,527**	66.7%	
Diseases	4,172		63.9%
Clinical syndromes	50		0.8%
Malformation syndromes	1,788		27.4%
Morphological anomalies	446		6.8%
Biological anomalies	11		0.2%
Particular clinical situations in a disease or syndrome	60		0.9%

Subtypes of disorders	**1,084**	11.1%	
Clinical subtypes	839		77.4%
Etiological subtypes	195		18.0%
Histopathological subtypes	50		4.6%
Groups of disorders	**2,173**	22.2%	
Clinical groups	450		20.7%
Categories	1,723		79.3%

Columns indicate the number of active clinical entities for each classification level and typology, their percentage relative to the total number of clinical entities (n=9,784), and the percentage each typology represents within its respective classification level.

Since Orphanet started tracking inactivations in 2013, 1,455 clinical entities have been inactivated and therefore excluded from the Orphanet Nomenclature of RD. Of these, 378 were groups of disorders, 809 were disorders and 268 were subtypes of disorders. The reason for inactivation is obsolescence for 68% of the cases (988 entities), deprecation for 22% (323 entities) and the fact that the disorder has been recognized as non-rare for the remaining 10% (144 entities).

Since 2020, an average of 123 ± 72 (mean ± SD) new disorders per year have been added to the Orphanet Nomenclature, while an average of 44 ± 18 disorders per year have been inactivated. A detailed report of the number of newly included, obsoleted, deprecated and non-rare disorders for each year from 2020 to 2025 is provided in Table 2.

Table 2. Yearly overview of the newly included and inactivated disorders in the Orphanet Nomenclature of RD (2020-2025).

	Newly included disorders	Obsoleted disorders	Deprecated disorders	Non rare disorders
2020	74	35	24	5
2021	53	8	16	1
2022	92	29	5	4
2023	114	21	8	5
2024	253	38	22	9
2025	152	13	19	2

Number of disorders newly included, obsoleted, deprecated, or identified as non-rare in the Orphanet Nomenclature of (RD), based on annual July dataset releases from 2020 to 2025.

Of the 9,784 active clinical entities in the Orphanet Nomenclature of RD, 6,311 (64.5%) carry at least one synonym in English, while the remaining 3,473 (35.5%) are not associated with any synonym. Of the 6,527 disorders, 4,805 (73.6%) carry at least one synonym in English, with a total number of synonyms at the disorder level amounting to 10,547. The minimum number of synonyms associated to a disorder is 1, and the maximum is 15, with a mean number of 2.2 synonyms for disorders with at least one synonym. The complete distribution of synonyms by classification level is reported in S2 File.

All preferred terms associated with active clinical entities registered in Orphanet Nomenclature of RD are translated in the following languages: French, German, Spanish, Italian, Dutch, Portuguese, Polish and Czech. Table 3 shows the rate of translation of preferred

terms for every language. The very minor delay in translation is due to the different timings and frequencies at which translations are processed by every translation team. Non translated preferred terms are nevertheless included in English, pending translation. Synonyms can be adapted in all translated languages according to the medical practice in the target language; the number of adapted synonyms for every translation language is shown in Table 3. The complete distribution of synonyms by classification level in all translated languages is reported in S2 File.

Table 3. Number of active entities with translated preferred term and adapted synonyms per the target language.

	FR	DE	ES	IT	NL	PT	PL	CS
Number of translated preferred terms	9,784	9,778	9,784	9,569	9,777	9,576	9,605	9,628
% of translated preferred terms	100%	99.9%	100%	97.8%	99.9%	97.9%	98.2%	98.4%
Number of adapted synonyms	13,000	12,854	12,912	11,234	14,546	5,858	12,804	14,298

Number and percentage (%) of active clinical entities with a translated preferred term available per language, relative to the total number of entities with English preferred terms (n=9,784). The table also shows the number of adapted synonyms per language. Data for French (FR), German (DE), Spanish (ES), Italian (IT), Dutch (NL), Portuguese (PT), Polish (PL), and Czech (CS) were manually reviewed to identify untranslated terms.

Additional translations are also available in Chinese (9,158 translated preferred terms, last update June 2020), Turkish (9,168 translated preferred terms, last update September 2021) and Ukrainian (9,612 translated preferred terms, last update July 2024), and are provided in separate files in the "Rare Diseases and Alignment with Terminologies and Databases" section of Orphadata Science [52].

Disorder definitions are produced in English for all disorders included in the Orphanet Nomenclature of RD and are subsequently translated into the eight translation languages: French, German, Spanish, Italian, Dutch, Portuguese, Polish and Czech. Table 4 presents the coverage of disorder definitions at the disorder level in each language. Currently, approximately 94% of rare disorders have a definition in English, while the rate of coverage in other languages depends on the priorities defined by each translating country.

Table 4. Number of disorders with an available definitions per language.

	EN	FR	DE	ES	IT	NL	PT	PL	CS
Number of disorders with a definition	6,116	5,941	4,572	6,095	4,592	6,068	1,186	1,372	17
% on the total number disorders	93.7%	91.0%	70.1%	93.4%	70.4%	92.9%	18.2%	21.0%	0.26%

Number and percentage (%) of disorders with an available definitions in English (EN) and translations into French (FR), German (DE), Spanish (ES), Italian (IT), Dutch (NL), Portuguese (PT), Polish (PL), and Czech (CS). Percentages are calculated based on the total number of disorders (n=6,527) in the Orphanet Nomenclature of RD.

444 ## The Orphanet classification of RD

445 The Orphanet classification of RD has a multi-parental structure: 4,415 (67.7%)
446 disorders are classed in more than one classification hierarchy while 2,106 (32.3%) are classed
447 in just one classification, independently of the number of parents that they have in this same
448 hierarchy. Of the 34 classification hierarchies in the Orphanet classification system, 29 were
449 included in the analysis. Five hierarchies not corresponding to distinct medical domains or
450 representing organizational or non-clinical groupings, were excluded as previously described
451 in the Methods section.

452 The majority of rare disorders are linearized into the rare developmental defect during
453 embryogenesis classification (32.3%, 2,109 disorders), closely followed by the rare neurologic
454 disorders classification (16.8%, 1,096 disorders), amounting together to almost 50% of all rare
455 disorders. The following most represented classifications are the rare neoplastic disorders at
456 7.6%, the rare inborn errors of metabolism at 6.1%, the rare skin disorders at 6%, the rare
457 bone disorders at 5.4%, the rare ophthalmic disorders at 3.7%, the rare immune disorders at
458 3.2%, the rare endocrine disorders at 3%, the rare hematologic disorders at 2.8%, the rare
459 infectious disorders at 2.7%, and the rare systemic or rheumatologic diseases at 2.5%. The
460 remaining classifications each include fewer than 100 linearized disorders each, collectively
461 representing the residual 8% of rare disorders (Figure 4 gray bars, detailed in S3 File).

462

463

464 **Fig 4. Distribution of rare disorders by medical domain according to their classification.** Grey
465 bars represent the number of disorders linearized according to their preferential parent
466 classification, while white bars indicate the distribution of multi-classified disorders, counted
467 as many times as they appear in the relevant classification hierarchies. Classifications

468 containing fewer than 100 disorders each are annotated with exact counts to enhance
469 readability. Six disorders were excluded from this analysis as they were either biological
470 anomalies without clinical manifestations or were linearized under the classifications of rare
471 genetic disorders or rare transplant-related disorders. ORPHA:616874 Rare disorder without
472 a determined diagnosis after full investigation was also excluded.
473
474
475 The majority of multiclassified rare disorders are again found in the rare
476 developmental defect during embryogenesis classification (2,793 disorders) and the rare
477 neurologic disorders classification (2,492 disorders). The following classifications contain
478 considerably less multiclassified disorders and are, in descending order: rare ophthalmic
479 disorders (868 disorders), rare skin disorders (835 disorders), rare bone disorders (793
480 disorders), rare neoplastic disorders (743 disorders), and rare inborn errors of metabolism
481 (704 disorders). The remaining hierarchies each contain less than 500 disorders, with only 3
482 hierarchies containing less than 100 (Figure 4 white bars, detailed in S3 File).
483
484
485 **Description of the mappings between the Orphanet Nomenclature of RD and other**
486 **reference terminologies**
487 Of the 6,527 active disorders registered in Orphanet, 6,355 (97.4%) are aligned to at
488 least one code ICD-10, with only 415 (6.4%) having an exact proximity relationship. The
489 remaining aligned codes are either present in the ICD-10 alphabetical index/are included in
490 the tabulated list under the ICD-10 code (655 disorders) or they have been attributed by
491 Orphanet to the ICD-10 code that corresponds to the closest entity according to Orphanet's

492 rules. The work on ICD-11 mapping has not been completed yet, however, as of July 2025,
493 4,683 disorders (71.8%) were aligned to at least one code in the ICD-11 MMS, with 958 (14.7%)
494 characterized by an exact proximity relationship and 3,013 (46.2%) present in the alphabetical
495 index/present in the ICD11 Foundation with an exact uri. The remaining mappings have been
496 attributed by Orphanet to the ICD-11 code corresponding to the closest entity. According to
497 the October 2024 release of the Orphanet-SNOMED CT mapping [54], 6,124 active disorders
498 (93.8%) are mapped to one SNOMED CT code, all with an exact proximity relationship (Table
499 5). Among all active disorders present in the Orphanet Nomenclature, 72.2% (n= 4,715) are
500 genetic disorders. A total of 4,141 disorders are mapped to at least one phenotypic OMIM
501 number and 3,319 mapped disorders have an exact proximity relationship (Table 5).
502
503
504 **Table 5. Number of active disorders mapped to reference terminologies**

A.	Number of active disorders with at least one mapping	% on total	Number of active disorders with an exact proximity relationship	% on total
ICD-10	6,355*	97.4%	415	6.4%
ICD-11	4,683*	71.8%	958	14.7%
SNOMED CT	6,124	93.8%	6,124	93.8%
B.	Number of active genetic disorders with at least one mapping	% on total	Number of active genetic disorders with an exact proximity relationship	% on total
OMIM	4,141	63.4%	3,319	50.9%

505 A) Number and percentage (%) of Orphanet active disorders mapped to ICD-10, ICD-11, and
506 SNOMED CT medical terminologies. Asterisks (*) indicate the inclusion in the total of mappings
507 attributed by Orphanet when the rare disorder was not present in the target terminology; B)
508 Number and percentage (%) of Orphanet active genetic disorders mapped to OMIM (Online
509 Mendelian Inheritance in Man). Percentages reflect the proportion relative to the total
510 number of disorders (n=6,527).

511 # Discussion

512 While collectively common, RD are numerous, clinically heterogeneous, and share
513 challenges directly related to their individual rarity: persons living with each RD are few and
514 dispersed, knowledge about each RD is limited and often inaccessible to non-specialist health
515 professionals, and expertise is scarce and dispersed as well. RD are hard to identify in medical
516 records since most are absent from medical terminologies in use in hospitals or not labeled as
517 rare, rendering them invisible in health systems. This lack of identification results in
518 inequalities in accessing a timely diagnosis and appropriate care, hinders data collection on
519 RD natural history and their personal and societal burden, and complicates patient
520 recruitment for clinical research and therapy development [4,8,16,55].

521 RD codification requires a nomenclature that meets the standards of a medical
522 terminology to ensure implementation in health information systems and in systematic
523 research collections, such as patient registries. Simply maintaining an inventory of RD is not
524 sufficient to address the codification challenge. A trustworthy and reliable medical
525 terminology must meet strict quality requirements: univocity and unambiguity of coded
526 concepts, unrestricted management of synonyms, consistency of semantic relationships
527 between concepts, multilingual support, accuracy of translations, correctness of semantic
528 alignments, and relevant coverage of identified use cases. Additionally, terminology standards
529 must adhere to rigorous and essential quality procedures, ensure sustainability and scalability
530 to support the rapid development of knowledge, and guarantee long-term stability,
531 preservation of the meaning of concepts from previous versions, and traceability of changes
532 [56]. Since 1997, Orphanet has maintained through its publicly available website an inventory
533 of RD to provide information on RD and their dedicated services across Europe [28]. To

534 address the gaps in the classification and codification of RD identified in the 2009 Council
535 Recommendation, the European Commission supported Orphanet, though a number of EU
536 funded projects, in developing a nomenclature of RD, which has evolved over time into the
537 only RD-specific standardized codification and classification system meeting the high-quality
538 standards of a medical terminology [19].

539 To meet the univocity and unambiguity criteria, Orphanet adopted a clinical definition
540 of RD that encompasses all RD, regardless of whether the etiology is known or not. This core
541 concept ensures a terminology that can be adapted to different use cases, i.e. health systems
542 with diverse diagnostic capabilities or different stages of the patient's diagnostic journey. A
543 RD in Orphanet is therefore defined as a clinical entity characterized by a set of homogeneous
544 phenotypic abnormalities, described in at least two independent cases in the medical
545 literature, and a consistent evolution, allowing for a definitive clinical diagnosis. The rarity of
546 a disease is defined according to the European prevalence threshold (not more than 1 in 2,000
547 persons in the European population). This choice was made for historical reasons, as Orphanet
548 was developed in Europe and co-funded by the European Commission, in particular for the
549 development of the Orphanet Nomenclature of RD. However, this definition is consistent with
550 those used in other geographical areas, which adopt a similar or a more restrictive cutoff, with
551 the exception of the United States, which includes diseases with a higher prevalence [2].
552 Another limitation of this definition is that it includes rare infectious diseases in Europe, even
553 if they are common in other parts of the world. To address this, some countries having
554 adopted the Orphanet Nomenclature of RD outside the European space, like Argentina, have
555 adapted this chapter to exclude infectious diseases that are common in their countries [57].
556 These limitations underline the need for consensus around a global operational definition of
557 RD, a goal the WHO has worked towards in collaboration with RDI, with the participation of a

558 multi-stakeholder, international group of experts including an Orphanet representative (A.R.)
559 [15]. The results of this work support Orphanet's adoption of a clinical definition of RD and
560 may lead in the future to an adaptation of the Orphanet RD lists to the different WHO regions.
561 In the meantime, Orphanet collaborates with non-European countries to help adapt the
562 Orphanet Nomenclature of RD to local needs, improving global RD interoperability.

563 A consensus on the definition of RD also allows to unambiguously count the number
564 of existing RD. Many enumerations of RD circulate, some derived from semi-automated
565 integrations of different databases and ontologies (most frequently Orphanet and OMIM),
566 sometimes resulting in estimates exceeding 10,000 RD [8,58,59]. However, these methods
567 aggregate different levels of granularity without clearly stating which RD definition is adopted.
568 While this overall figure aligns closely with Orphanet's total active clinical entities (9,784 as of
569 July 2025), it is essential to recognize that superimposing different levels of granularity and/or
570 different etiologies of the same RD can lead to overcounting the same disease multiple times.
571 Indeed, from a methodological point of view, it is necessary to define the object to be
572 enumerated in order to achieve a reliable result. In this respect, the Orphanet Nomenclature
573 of RD and its classification system are unique in distinguishing between groups, disorders (RD
574 as defined by the criteria above), and subtypes within its 9,784 active clinical entities. This
575 allows for a precise count of RD (i.e. disorder classification level), which, as of July 2025 totals
576 6,527. More than the numbers themselves, it is their interpretation and impact that matter.
577 For epidemiological and public health purposes, the count of clinically defined RD provides an
578 appropriate indicator for comparing countries and healthcare systems with varying
579 technological and diagnostic capacities. Conversely, for precision medicine or molecular
580 diagnosis, a more granular approach including subtypes or genetic variants may be needed.
581 Estimates exceeding 10,000 RD often reflect the growing knowledge of disease mechanisms

582 and treatable variants rather than an actual increase in the number of RD. However, using this
583 broader count in public health discussions without clarification can lead to misinterpretations,
584 potentially affecting policy decisions, budget allocation, and healthcare planning.

585 The field of RD is a rapidly evolving one: with the continuous advancement in genetic
586 knowledge and the growing interest in personalized medicine, new RD are described each
587 month. Additionally, the perimeter of previously identified RD can be refined, with medical
588 concepts being either merged or divided based on new discoveries or the examination of
589 larger numbers of patients. To ensure that the Orphanet Nomenclature and classification of
590 RD remains up to date with current scientific knowledge, Orphanet conducts a monthly
591 systematic review of the literature and continuously collaborates with medical experts,
592 particularly from European Reference Networks (ERNs) and international RD learned societies
593 [40,60]. All modifications to ORPHAcodes are performed after an extensive review of the
594 literature and are validated by the Orphanet medical and scientific committee, following
595 published standardized procedures [35]. Between 2020 and 2025, a total of 738 new RD were
596 included in the Orphanet Nomenclature of RD, averaging about 120 diseases per year. This
597 number reflects not only newly described RD, but also the outcomes of numerous
598 time-effective collaborative projects with experts, aimed at restructuring and updating the
599 classification to incorporate revised medical concepts. At the same time, Orphanet also
600 applies a strict inactivation policy to preserve the stability and univocity of ORPHAcodes.
601 Inactivated codes are never reused for different entities, and a substitution code is always
602 suggested to maintain traceability. While unusual, it is possible to reintroduce an inactivated
603 ORPHAcode when new evidence emerges that justifies a reconsideration [35]. This has
604 occurred in cases such as Systemic lupus erythematosus ORPHA:536 and Immunoglobulin A
605 nephropathy ORPHA:34145, which did not initially meet Orphanet's criteria for being

606 considered RD. Taken together, the integration of new RD and the number of inactivated ones
607 contribute to a small and relatively stable annual increase in the total number of RD, averaging
608 fewer than 80 diseases per year.

609 To further ensure the univocity and accessibility of concepts, Orphanet implements an
610 unrestricted management of synonyms, aiming to maximize semantic coverage across
611 different usages and data sources. Among the 6,527 active RD, 73.6% have at least one
612 synonym in English, totaling 10,547 terms. Synonyms must cover the exact concept perimeter
613 and are added as needed, without limitations in number [35]. The number of synonyms per
614 RD ranges from 1 to 15, with a mean of 2.2 for RD carrying at least one synonym. Additionally,
615 Orphanet provides a standardized, medically validated definition for each RD included in the
616 Orphanet Nomenclature. These definitions consist of short texts describing the clinical
617 features to unambiguously delineate the entity and distinguish it from other RD, and are
618 intended to support healthcare professionals in assigning the correct ORPHAcode to a patient
619 [35,61]. To further ensure international interoperability, the Orphanet Nomenclature and
620 classification of RD is also a multilingual resource. Translations are performed according to
621 established procedures that include a final manual medical validation [37]. The different
622 translations also include an unlimited number of synonyms, that are tailored according to the
623 relevance to each language. Taken together, these elements ensure semantic consistency and
624 facilitate ORPHAcode integration into health and research information systems, optimizing
625 their use for medical practitioners and coders across countries and contexts and facilitating
626 cross-border data comparison and sharing.

627 To ensure that the Orphanet Nomenclature of RD could effectively be used across
628 various use cases and settings, RD are embedded in a three-level classification system:
629 disorders, group of disorders and subtype of disorders. While the disorder level serves as the

630 primary reference level for data sharing among systems and countries, and for
631 epidemiological purposes, groups of disorders allow data aggregation for use cases analysis,
632 i.e. for public health and decision-making purposes, and subtypes of disorders allow to fine-
633 tune RD codification, at the clinical, etiological and histopathological level. This classification
634 structure in which each level and their respective relationships are clearly defined, responds
635 to the criterion of consistency of semantic relationships between concepts. This system is able
636 to respond to the needs of different stakeholders and use cases: RD groups can be used for
637 public health analysis, the disorder level serves as the aggregation level for RD counting and
638 epidemiological studies, and the more granular subtypes can be used for applications in
639 genetics, precision medicine, and other research and innovation purposes [62–68].
640 Furthermore, the Orphanet Nomenclature of RD is the only one providing a code for
641 undiagnosed RD patients who underwent all available diagnostic possibilities: this RD
642 population has been estimated to up to 50% of all RD patients, based on the experience of RD
643 expert centers [38]. Being able to code these patients makes it possible to bring them to a
644 research pathway and undiagnosed programmes, as recommended by IRDiRC's goal #1 [69–
645 71]. Moreover, this classification system can capture the multisystemic nature of RD, as
646 clinical entities are organized into 29 distinct medical domains and classed as many times as
647 needed according to their relevant domain of care. This is reflected in the distribution of RD
648 across classification hierarchies: the majority of RD fall within the rare developmental defect
649 during embryogenesis and rare neurologic disease classifications, together accounting for
650 nearly 50% of all RD. However, almost 70% of RD are classed in more than one classification,
651 reflecting their multisystem involvement. This approach ensures a more accurate
652 representation of the complex nature of RD and of their linked resources, allowing to better

653 orient patients and medical doctors towards appropriate information and specialized care
654 [62–64].

655 RD is a field in which the boundaries between diagnosis, healthcare, and research are
656 often blurred, therefore maintaining semantic interoperability between health and research
657 sectors is essential. Semantic interoperability is the ability of different systems to exchange
658 information so that each of them can interpret the meaning of the information correctly,
659 without altering it, and use it in conjunction with its own local data. Achieving this requires
660 the use of standardized and interoperable information, as well as the coding of data using
661 semantic resources, such as reference terminologies, value sets, and semantic alignments.
662 Semantic interoperability with other reference terminologies used in healthcare (ICD-10, ICD-
663 11, SNOMED CT), research (OMIM and others), or regulatory settings (MedDRA) is maintained
664 according to specific rules that comprise cardinality and conceptual equivalence information.
665 This allows for interoperability between settings using terminologies that are based on
666 different concept definitions (for instance OMIM, in which concepts are defined genetically
667 [46], or MeSH, in which concepts aggregate related terms useful to improve medical literature
668 research, or ICD-10, in which different disease concepts can relate to a same broad concept
669 code). As of July 2025, Orphanet has aligned 97.4% of its active RD to at least one ICD-10 code,
670 with 6.4% exhibiting an exact proximity relationship. The work on ICD-11 mapping is ongoing,
671 but by July 2025, 71.8% of Orphanet's RD were mapped to at least one ICD-11 MMS code,
672 with 14.7% showing an exact proximity relationship. Additionally, the October 2024 release of
673 the Orphanet-SNOMED CT mapping shows that 93.8% of Orphanet's active RD are mapped to
674 SNOMED CT, with all exhibiting an exact proximity relationship. Regarding genetic diseases,
675 Orphanet classifies 72.2% of its active RD as genetic (defined as either caused by a known gene
676 or having a clear familial inheritance pattern) with 4,141 RD mapped to at least one

677 phenotypic OMIM number. It is interesting to note that many genetic RD are mapped to
678 multiple OMIM numbers, reflecting the fact that Orphanet defines diseases as clinical entities,
679 whereas OMIM organizes entries around individual gene-phenotype relationships. Overall,
680 this extensive and accurate mapping system ensures consistency of transcoding across
681 multiple platforms, further enhancing the utility of the Orphanet Nomenclature of RD in
682 clinical, research, and regulatory contexts [65–68,72,73]. It is important to note that Orphanet
683 is currently the only producer of a scientifically validated nomenclature of RD. Other
684 terminologies are increasingly integrating Orphanet concepts as a source of RD information,
685 acknowledging the critical scientific contributions Orphanet has made to the RD community.
686 Outstanding collaborations between Orphanet, SNOMED International and WHO ensure that
687 the RD content of SNOMED CT and the ICD-11 continues to be improved by the inclusion of
688 RD concepts defined by Orphanet [22,23].

689 To support the effective and sustainable implementation in health and research
690 information systems, the Orphanet Nomenclature of RD (ORPHAcodes) is distributed under
691 an open license (CC-BY 4.0) on different platforms depending on use: Orphadata Science for
692 structured datasets intended for large-scale analyses [52], and Orphacodes platform for
693 coding resources [54]. Each release is versioned and accompanied by differential files to
694 ensure traceability of changes and continuity of coding, as per recommendation of the RD-
695 ACTION working group for routine maintenance of codification [74]. In addition, a suite of
696 technical tools is provided to facilitate implementation into diverse systems, promote
697 accurate patient codification, and ensure continuous support by the community. This
698 supporting infrastructure ecosystem, on top of the production and dissemination of the
699 Orphanet Nomenclature of RD, is essential to ensure evolutivity and respond to emerging use
700 cases in a rapidly evolving field.

Conclusion

Improved codification of RD has been recognized as a strategic priority within the European policy landscape and has been the subject of a series of official recommendations. Accurate codification is the foundation for accurately producing, collecting and sharing high-quality data essential to increase the knowledge in a field that has long remained largely invisible within health systems. Establishing a seamless data ecosystem for RD based on the Orphanet Nomenclature and classification of RD has been recommended as a key step toward improving the lives of people living with a rare disease [27]. Indeed, accurate codification of RD patients allows to produce data both for primary and secondary use. Primary use includes: recognition of the RD condition by non-expert health professionals, raising awareness on RD specificities ensuring best care practices at all stages of the patient's care pathway, and recognition of undiagnosed RD patients who might benefit from new discoveries, potentially leading to establishing a diagnosis. Secondary use of RD data includes: accurately measuring the prevalence of the entire RD population and monitoring public health interventions to inform evidence-based decision-making, production of real-world evidence for research and therapeutic development, and facilitating recruitment for clinical research.

The Orphanet Nomenclature and classification of RD is the only RD-specific, high-quality standard, interoperable medical terminology and classification system that helps close the visibility gap in health and research information systems, thereby enabling the consistent recognition of people living with a rare disease and facilitating cross-border and cross-sectoral data interoperability.

Acknowledgements:

We would like to thank the Orphanet translating members for their essential contribution to the translation of the Orphanet Nomenclature of RD, as well as the European Reference Networks (ERNs) and the Filières de santé maladies rares (FSMR) for their continued involvement in nomenclature and classification revisions, and in the production of textual content.

732 ## References

733 1. European Parliament: Directorate-General for Internal Policies of the Union, Kamphuis B,
734 De Jongh T, Bastiaanssen V. Tackling rare diseases – Challenges, opportunities and gaps
735 for action on rare diseases in the European Union. European Parliament; 2024. Available:
736 https://data.europa.eu/doi/10.2861/263904
737 2. Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, et al.
738 Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet
739 database. Eur J Hum Genet. 2020;28: 165–173. doi:10.1038/s41431-019-0508-0
740 3. European Parliament: Directorate-General for Parliamentary Research Services,
741 Schmiegelow K, Fehn K, Westman D, Clemente Pesudo I, Atanasovska D, et al.
742 Addressing challenges to European multi-country collaboration models for rare diseases.
743 European Parliament; 2024. Available: https://data.europa.eu/doi/10.2861/51755
744 4. Ferreira CR. The burden of rare diseases. Am J Med Genet A. 2019;179: 885–892.
745 doi:10.1002/ajmg.a.61124
746 5. Hayashi S, Umeda T. 35 years of Japanese policy on rare diseases. Lancet Lond Engl.
747 2008;372: 889–890. doi:10.1016/S0140-6736(08)61393-8
748 6. Richter T, Nestler-Parr S, Babela R, Khan ZM, Tesoro T, Molsen E, et al. Rare Disease
749 Terminology and Definitions-A Systematic Global Review: Report of the ISPOR Rare
750 Disease Special Interest Group. Value Health J Int Soc Pharmacoeconomics Outcomes
751 Res. 2015;18: 906–914. doi:10.1016/j.jval.2015.05.008
752 7. Khosla N, Valdez R. A compilation of national plans, policies and government actions for
753 rare diseases in 23 countries. Intractable Rare Dis Res. 2018;7: 213–222.
754 doi:10.5582/irdr.2018.01085

755 8. Haendel M, Vasilevsky N, Unni D, Bologa C, Harris N, Rehm H, et al. How many rare
756 diseases are there? Nat Rev Drug Discov. 2020;19: 77–78. doi:10.1038/d41573-019-
757 00180-y
758 9. Chan AYL, Chan VKY, Olsson S, Fan M, Jit M, Gong M, et al. Access and Unmet Needs of
759 Orphan Drugs in 194 Countries and 6 Areas: A Global Policy Review With Content
760 Analysis. Value Health. 2020;23: 1580–1591. doi:10.1016/j.jval.2020.06.020
761 10. Wainstock D, Katz A. Advancing rare disease policy in Latin America: a call to action.
762 Lancet Reg Health Am. 2023;18: 100434. doi:10.1016/j.lana.2023.100434
763 11. Abozaid GM, Kerr K, McKnight A, Al-Omar HA. Criteria to define rare diseases and orphan
764 drugs: a systematic review protocol. BMJ Open. 2022;12: e062126.
765 doi:10.1136/bmjopen-2022-062126
766 12. European Parliament and Council of the European Union. Regulation (EC) No 141/2000
767 of the European Parliament and of the Council of 16 December 1999 on orphan
768 medicinal products. Publications Office of the European Union; 2000 Jan. Available:
769 https://eur-lex.europa.eu/eli/reg/2000/141/oj/eng
770 13. United States Congress. Orphan Drug Act, 21 U.S.C. § 360bb(a)(2). United States
771 Government Publishing Office; 2023. Available:
772 https://www.govinfo.gov/app/details/USCODE-2023-title21/USCODE-2023-title21-
773 chap9-subchapV-partB-sec360bb
774 14. Abozaid GM, Kerr K, Alomary H, Al-Omar HA, McKnight A. Global insight into rare disease
775 and orphan drug definitions: a systematic literature review. BMJ Open. 2025;15:
776 e086527. doi:10.1136/bmjopen-2024-086527

15. Wang CM, Whiting AH, Rath A, Anido R, Ardigò D, Baynam G, et al. Operational description of rare diseases: a reference to improve the recognition and visibility of rare diseases. Orphanet J Rare Dis. 2024;19: 334. doi:10.1186/s13023-024-03322-7
16. Valdez R, Ouyang L, Bolen J. Public Health and Rare Diseases: Oxymoron No More. Prev Chronic Dis. 2016;13: E05. doi:10.5888/pcd13.150491
17. United Nations General Assembly. Addressing the challenges of persons living with a rare disease and their families: resolution. United Nations; 2021. Available: http://digitallibrary.un.org/record/3953832
18. Jetté N, Quan H, Hemmelgarn B, Drosler S, Maass C, Moskal L, et al. The development, evolution, and modifications of ICD-10: challenges to the international comparability of morbidity data. Med Care. 2010;48: 1105–1110. doi:10.1097/MLR.0b013e3181ef9d3e
19. Rath A, Olry A, Dhombres F, Brandt MM, Urbero B, Ayme S. Representation of rare diseases in health information systems: The orphanet approach to serve a wide range of end users. Hum Mutat. 2012;33: 803–808. doi:10.1002/humu.22078
20. Aymé S, Bellet B, Rath A. Rare diseases in ICD11: making rare diseases visible in health information systems through appropriate coding. Orphanet J Rare Dis. 2015;10: 35. doi:10.1186/s13023-015-0251-8
21. Aymé S, Rath A, Bellet B. WHO International Classification of Diseases (ICD) Revision Process: incorporating rare diseases into the classification scheme: state of art. Orphanet J Rare Dis. 2010;5: P1. doi:10.1186/1750-1172-5-S1-P1
22. World Health Organization. Rare diseases (frequently asked questions). In: WHO – International Classification of Diseases [Internet]. [cited 15 June 2025]. Available: https://www.who.int/standards/classifications/frequently-asked-questions/rare-diseases

23. SNOMED International. INSERM and SNOMED International release SNOMED CT to Orphanet map supporting representation and use. In: SNOMED International News [Internet]. 2021 [cited 15 June 2025]. Available: https://www.snomed.org/news/inserm-and-snomed-international-release-snomed-ct-to-orphanet-map-supporting-representation-and-use
24. Commission Expert Group on Rare Diseases (CEGRD), European Commission. Recommendation on ways to improve codification for rare diseases in health information systems. 2014 Nov. Available: https://health.ec.europa.eu/publications/recommendation-ways-improve-codification-rare-diseases-health-information-systems_en
25. Choquet R, Maaroufi M, de Carrara A, Messiaen C, Luigi E, Landais P. A methodology for a minimum data set for rare diseases to support national centers of excellence for healthcare and research. J Am Med Inform Assoc JAMIA. 2015;22: 76–85. doi:10.1136/amiajnl-2014-002794
26. Council of the European Union. Council Recommendation of 8 June 2009 on an action in the field of rare diseases. Publications Office of the European Union; 2009. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32009H0703%2802%29
27. Kole A, Hedley V, et al. Recommendations from the Rare 2030 Foresight Study: the future of rare diseases starts today. EURORDIS-Rare Diseases Europe (Rare2030 project); 2021 Feb. Available: https://download2.eurordis.org/rare2030/Rare2030_recommendations.pdf
28. Orphanet, Inserm. Orphanet – Knowledge on rare diseases and orphan drugs. In: Orphanet [Internet]. [cited 30 July 2025]. Available: http://www.orpha.net/

29. Directorate-General for Health and Food Safety, European Commission. Rare diseases. In: European Commission – Public Health [Internet]. 2025 [cited 15 June 2025]. Available: https://health.ec.europa.eu/rare-diseases-and-european-reference-networks/rare-diseases_en
30. European Commission – Joint Research Centre (JRC). Set of Common Data Elements for Rare Diseases Registration. In: EU RD Platform / European Commission Science Hub [Internet]. [cited 15 June 2025]. Available: https://eu-rd-platform.jrc.ec.europa.eu/set-of-common-data-elements_en
31. eHAction — Joint Action supporting the eHealth Network. Common Semantic Strategy for Health in the European Union (eHAction Deliverable D8.2.2). European Commission; 2019 May. Available: https://health.ec.europa.eu/document/download/92ac8823-19c4-4641-bf4d-5d9bf021a600_en
32. Directorate-General for Health and Food Safety. eHN Guideline on Patient Summary (Release 3.4). European Commission; 2024 Nov. Available: https://health.ec.europa.eu/system/files/2023-10/ehn_guidelines_patientsummary_en.pdf
33. X-eHealth Consortium. D5.6 – Refine PS functional specifications to account for eHN Guidelines and rare diseases. X-eHealth: eXchanging electronic Health Records in a commom framework; 2022. Available: https://www.x-ehealth.eu/wp-content/uploads/2022/11/D5.6-Refine-PS-functional-specifications-to-account-for-eHN-Guidelines-and-rare-diseases.pdf
34. Executive Board, World Health Organization. Rare diseases: a global health priority for equity and inclusion (EB156(15)). World Health Organization; 2025 Feb. Available: https://apps.who.int/gb/ebwha/pdf_files/EB156/B156_(15)-en.pdf

35. Orphanet, Inserm. Procedural document: Orphanet nomenclature and classification of rare diseases. 2023 June. Report No.: 05. Available: https://www.orpha.net/pdfs/orphacom/cahiers/docs/GB/eproc_disease_inventory_R1_Nom_Dis_EP_05.pdf
36. Orphanet, Inserm. Procedural document: Rare disease nomenclature in English. 2017 Apr. Report No.: 1. Available: https://www.orpha.net/pdfs/orphacom/cahiers/docs/GB/eproc_Disease_naming_rules_in_English_PR_R1_Nom_01.pdf
37. Orphanet, Inserm. Procedural document: Orphanet Rare Diseases Nomenclature Production in National Language. 2017 Sept. Report No.: 1. Available: https://www.orpha.net/pdfs/orphacom/cahiers/docs/GB/eproc_Rare_disease_Nomenclature_Production__national-language.pdf
38. Angin C, Mazzucato M, Weber S, Kirch K, Abdel Khalek W, Ali H, et al. Coding undiagnosed rare disease patients in health information systems: recommendations from the RD-CODE project. Orphanet J Rare Dis. 2024;19: 28. doi:10.1186/s13023-024-03030-2
39. Orphanet, Inserm. Procedural document: Linearisation rules for rare diseases. 2014 Sept. Available: https://www.orpha.net/pdfs/orphacom/cahiers/docs/GB/Orphanet_linearisation_rules.pdf
40. Orphanet, Inserm. Procedural document: Collaboration with networks of expertise for the revision of the Orphanet nomenclature and classification of rare diseases. 2024 Sept. Report No.: 1.1. Available:

871 https://www.orpha.net/pdfs/orphacom/cahiers/docs/GB/eproc_Collaboration_network
872 s_R1_Nom_Rev_Exp_EP_10.pdf
873 41. Orphanet, Inserm. Procedural document: Orphanet ICD-10 Coding Rules for Rare
874 Diseases. 2023 June. Report No.: 3. Available:
875 https://www.orpha.net/pdfs/orphacom/cahiers/docs/GB/Orphanet_ICD10_coding_rule
876 s_R1_Nom_ICD_EP_07.pdf
877 42. World Health Organization. International statistical classification of diseases and related
878 health problems. 10th revision (ICD-10). 2010 edition. Geneva: World Health
879 Organization; Available: https://icd.who.int/browse10/2010/en
880 43. World Health Organization. ICD-11: International Classification of Diseases, 11th
881 Revision. In: ICD-11 for Mortality and Morbidity Statistics [Internet]. [cited 15 June
882 2025]. Available: https://icd.who.int/en
883 44. U.S. National Library of Medicine. Overview of SNOMED CT. In: U.S. National Library of
884 Medicine – Health IT [Internet]. U.S. National Library of Medicine; 2014 [cited 25 July
885 2025]. Available: https://www.nlm.nih.gov/healthit/snomedct/snomed_overview.html
886 45. SNOMED International. What is SNOMED CT. In: SNOMED International [Internet]. [cited
887 25 July 2025]. Available: https://www.snomed.org/what-is-snomed-ct
888 46. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online
889 Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic
890 disorders. Nucleic Acids Res. 2015;43: D789–D798. doi:10.1093/nar/gku1205
891 47. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical
892 terminology. Nucleic Acids Res. 2004;32: D267–D270. doi:10.1093/nar/gkh061
893 48. Lipscomb CE. Medical Subject Headings (MeSH). Bull Med Libr Assoc. 2000;88: 265–266.

49. Brown EG, Wood L, Wood S. The Medical Dictionary for Regulatory Activities (MedDRA). Drug Saf. 1999;20: 109–117. doi:10.2165/00002018-199920020-00002
50. National Center for Advancing Translational Sciences (NCATS), National Institutes of Health. Genetic and Rare Diseases Information Center (GARD). In: Genetic and Rare Diseases Information Center [Internet]. [cited 15 June 2025]. Available: https://rarediseases.info.nih.gov/
51. Vasilevsky NA, Matentzoglu NA, Toro S, Flack JE, Hegde H, Unni DR, et al. Mondo: Unifying diseases for the world, by the world. medRxiv; 2022. p. 2022.04.13.22273750. doi:10.1101/2022.04.13.22273750
52. Orphanet, Inserm. Orphadata Science – Orphanet datasets. In: Science Orphadata [Internet]. 2025 [cited 30 July 2025]. Available: https://sciences.orphadata.com/
53. Orphanet, Inserm. Orphadata aggregated. In: GitHub [Internet]. 9 July 2025 [cited 25 July 2025]. Available: https://github.com/Orphanet/Orphadata_aggregated
54. Orphanet, Inserm. Orphanet Nomenclature Pack for Coding. In: ORPHAcodes [Internet]. 2025 [cited 25 July 2025]. Available: https://www.orphacode.org/pack-nomenclature/
55. Schieppati A, Henter J-I, Daina E, Aperia A. Why rare diseases are an important medical and social issue. The Lancet. 2008;371: 2039–2041. doi:10.1016/S0140-6736(08)60872-7
56. Direction générale de l'offre de soins, France numérique santé (e-Santé) – Ministry of Health, France. Référentiels d'intéropérabilité sémantique : mise en euvre de terminologies de référence pour le secteur santé-social en France. Ministère des Solidarités et de la Santé, France; 2015. Available: https://esante.gouv.fr/sites/default/files/media_entity/documents/20150504_etude_terminos_phase2_diagnostic.pdf

57. Ministerio de Salud de la Nación, Argentina. Listado de enfermedades poco frecuentes en Argentina. In: Argentina.gob.ar – Salud [Internet]. [cited 15 June 2025]. Available: https://www.argentina.gob.ar/salud/pocofrecuentes/listado
58. RARE-X Data Platform (RARE-X), Lamoreaux K, et al. The Power of Being Counted: A more accurate count of rare diseases and steps to getting counted. RARE-X Data Platform; 2022 May. Available: https://rare-x.org/wp-content/uploads/2022/05/be-counted-052722-WEB.pdf
59. Smith CIE, Bergman P, Hagey DW. Estimating the number of diseases – the concept of rare, ultra-rare, and hyper-rare. iScience. 2022;25: 104698. doi:10.1016/j.isci.2022.104698
60. European Commission (Directorate-General for Health and Food Safety). European Reference Networks. In: European Commission – Public Health [Internet]. [cited 25 July 2025]. Available: https://health.ec.europa.eu/rare-diseases-and-european-reference-networks/european-reference-networks_en
61. Orphanet, Inserm. Procedural document: Creation and update of disease summary texts for the Orphanet encyclopaedia for professionals. 2021 June. Report No.: 2. Available: https://www.orpha.net/pdfs/orphacom/cahiers/docs/GB/Orphanet_DiseaseSummaryText_R1_Prod_sum_EP_02.pdf
62. Mazzucato M, Pozza LVD, Facchin P, Angin C, Agius F, Cavero-Carbonell C, et al. ORPHAcodes use for the coding of rare diseases: comparison of the accuracy and cross country comparability. Orphanet J Rare Dis. 2023;18: 267. doi:10.1186/s13023-023-02864-6

939 63. Mazzucato M, Visonà Dalla Pozza L, Minichiello C, Toto E, Vianello A, Facchin P.
940 Estimating mortality in rare diseases using a population-based registry, 2002 through
941 2019. Orphanet J Rare Dis. 2023;18: 362. doi:10.1186/s13023-023-02944-7
942 64. Pichon T, Messiaen C, Soussand L, Angin C, Sandrin A, Elarouci N, et al. Overview of
943 patients' cohorts in the French National rare disease registry. Orphanet J Rare Dis.
944 2023;18: 176. doi:10.1186/s13023-023-02725-2
945 65. Gunne E, McGarvey C, Hamilton K, Treacy E, Lambert DM, Lynch SA. A retrospective
946 review of the contribution of rare diseases to paediatric mortality in Ireland. Orphanet J
947 Rare Dis. 2020;15: 311. doi:10.1186/s13023-020-01574-7
948 66. Rico J, Echevarría-González de Garibay LJ, García-López M, Guardiola-Vilarroig S,
949 Maceda-Roldán LA, Zurriaga Ó, et al. The interoperability between the Spanish version of
950 the International Classification of Diseases and ORPHAcodes: towards better
951 identification of rare diseases. Orphanet J Rare Dis. 2021;16: 121. doi:10.1186/s13023-
952 021-01763-y
953 67. Walker CE, Mahede T, Davis G, Miller LJ, Girschik J, Brameld K, et al. The collective
954 impact of rare diseases in Western Australia: an estimate using a population-based
955 cohort. Genet Med. 2017;19: 546–552. doi:10.1038/gim.2016.143
956 68. Scanlon P, Ridler G, Say G, Kellett M, Charlesworth J, Neil A, et al. Measuring the impact
957 of rare diseases in Tasmania, Australia. Orphanet J Rare Dis. 2024;19: 399.
958 doi:10.1186/s13023-024-03343-2
959 69. International Rare Diseases Research Consortium (IRDiRC). IRDiRC Vision & Goals. In:
960 IRDiRC (irdirc.org) [Internet]. [cited 15 June 2025]. Available: https://irdirc.org/about-
961 us/vision-goals/

962 70. Boycott KM, Rath A, Chong JX, Hartley T, Alkuraya FS, Baynam G, et al. International
963 Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. Am J Hum Genet.
964 2017;100: 695–705. doi:10.1016/j.ajhg.2017.04.003
965 71. Lochmüller H, Le Cam Y, Jonker AH, Lau LP, Baynam G, Kaufmann P, et al. 'IRDiRC
966 Recognized Resources': a new mechanism to support scientists to conduct efficient,
967 high-quality research for rare diseases. Eur J Hum Genet. 2017;25: 162–165.
968 doi:10.1038/ejhg.2016.137
969 72. Chiu ATG, Chung CCY, Wong WHS, Lee SL, Chung BHY. Healthcare burden of rare
970 diseases in Hong Kong – adopting ORPHAcodes in ICD-10 based healthcare
971 administrative datasets. Orphanet J Rare Dis. 2018;13: 147. doi:10.1186/s13023-018-
972 0892-5
973 73. Navarrete-Opazo AA, Singh M, Tisdale A, Cutillo CM, Garrison SR. Can you hear us now?
974 The impact of health-care utilization by rare disease patients in the United States. Genet
975 Med. 2021;23: 2194–2201. doi:10.1038/s41436-021-01241-7
976 74. Work Package 5, RD-ACTION Joint Action. Recommendation for routine maintenance of
977 codification resources for rare diseases (RD-ACTION Deliverable 5.5). European
978 Commission (Health Programme) / RD-ACTION; 2018 July. Available: https://www.rd-
979 action.eu/wp-content/uploads/2018/09/677024_DEL5.5_Recommendation-routine-
980 maintenance-codification.pdf

981

982

Supporting information captions

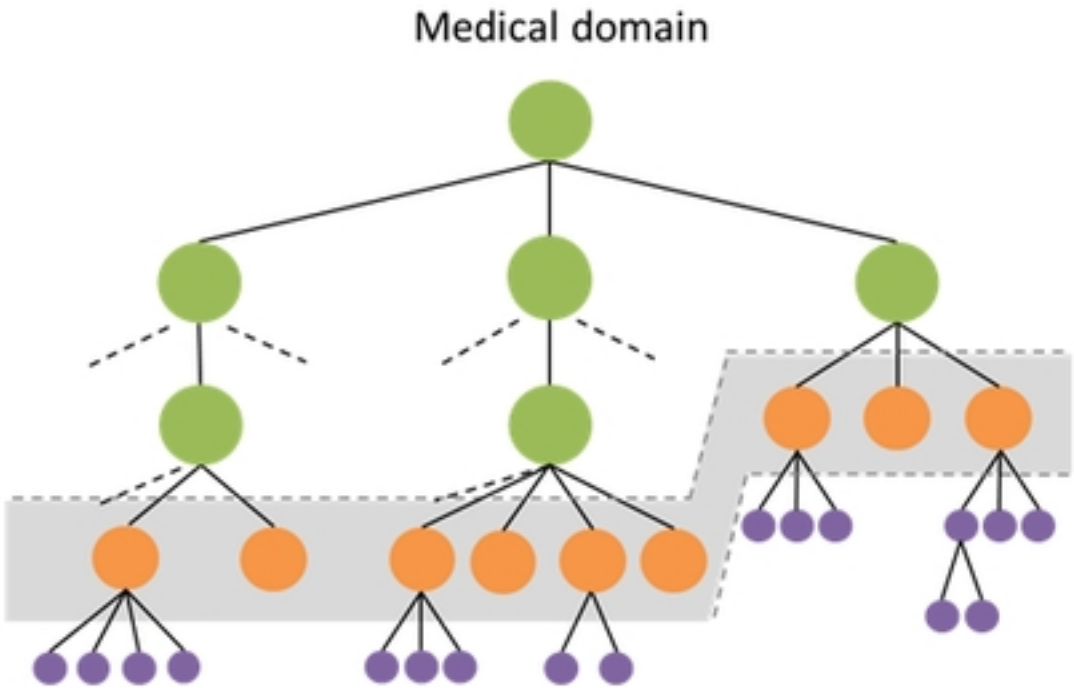
S1 File. Distribution of Orphanet clinical entities by classification level and typology. This file is the output of Notebook 1_number_of_Orphanet_clinical_entities.

S2 File. Production of the Orphanet RD terminology in English and translation languages. This file is the output of Notebook 2_production_of_the_Orphanet_RD_terminology_in_english_and_other_languages.

S3 File. Distribution of disorders in Orphanet classifications according to their preferential parent and multiclassified disorders. This file combines the outputs of Notebook 5_distribution_of_Orphanet_RD_in_Orphanet_classifications and Notebook 6_distribution_of_Orphanet_RD_by_preferential_parent.

ORPHAcode	Preferred term	Synonyms	Classification level
ORPHA:2134	Atypical hemolytic uremic syndrome	Atypical HUS aHUS	Disorder
Definition	A rare, genetic thrombotic microangiopathy due to dysregulation of the alternative complement pathway and characterized by the triad of hemolytic anemia, thrombocytopenia, and acute renal dysfunction.		

Figure 1



Group of disorders

- Clinical group
- Category

Disorder

- Disease
- Clinical syndrome
- Biological anomaly
- Malformation syndrome
- Morphological anomaly
- Particular clinical situation in a disease or syndrome

Subtype of disorder

- Clinical subtype
- Etiological subtype
- Histopathological subtype

Figure 2

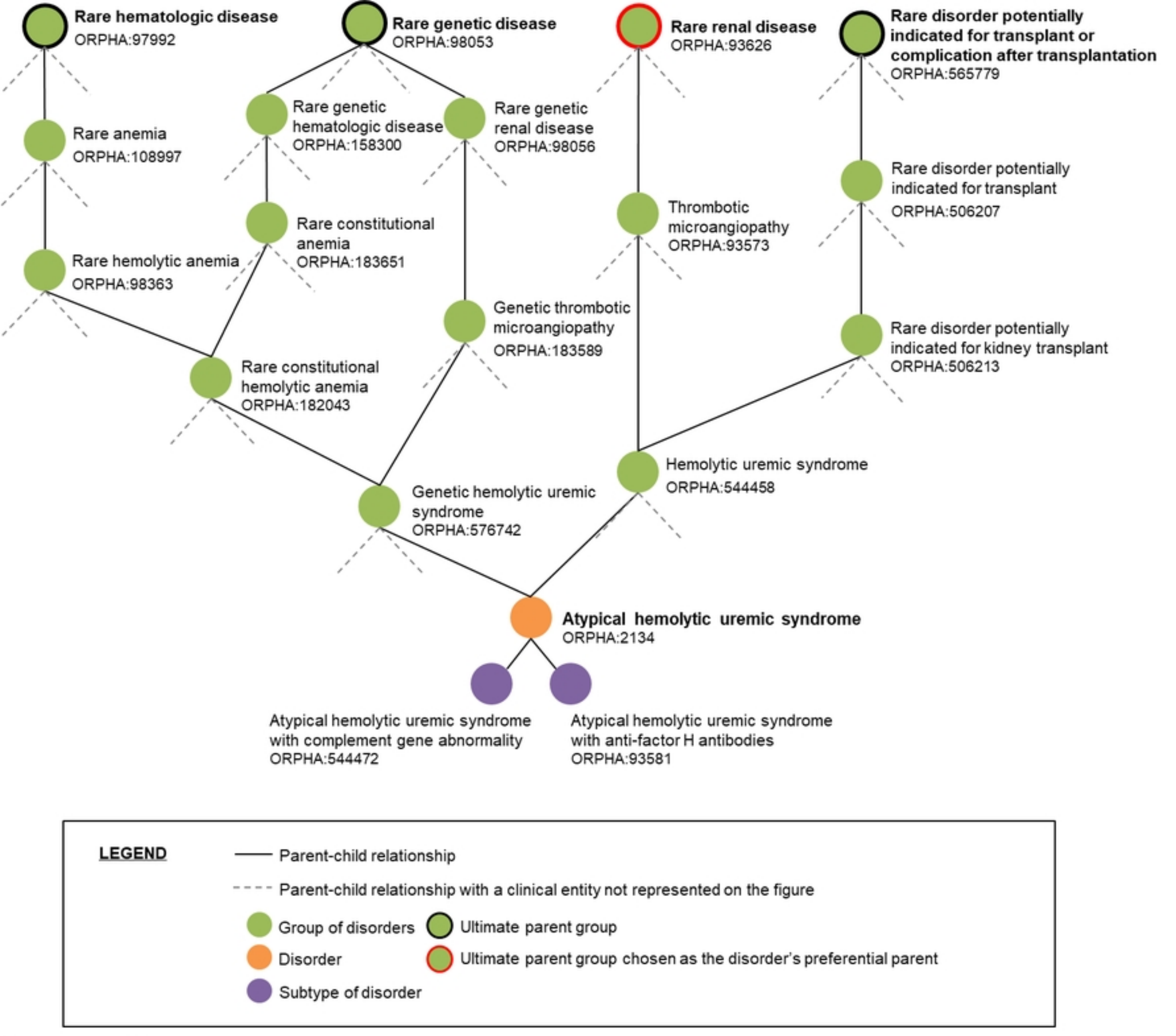

Rare hematologic disease
ORPHA:97992
Rare genetic disease
ORPHA:98053
Rare renal disease
ORPHA:93626
Rare disorder potentially indicated for transplant or complication after transplantation
ORPHA:565779
Rare anemia
ORPHA:108997
Rare genetic hematologic disease
ORPHA:158300
Rare genetic renal disease
ORPHA:98056
Rare disorder potentially indicated for transplant
ORPHA:506207
Thrombotic microangiopathy
ORPHA:93573
Rare constitutional anemia
ORPHA:183651
Rare hemolytic anemia
ORPHA:98363
Genetic thrombotic microangiopathy
ORPHA:183589
Rare disorder potentially indicated for kidney transplant
ORPHA:506213
Rare constitutional hemolytic anemia
ORPHA:182043
Hemolytic uremic syndrome
ORPHA:544458
Genetic hemolytic uremic syndrome
ORPHA:576742
Atypical hemolytic uremic syndrome
ORPHA:2134
Atypical hemolytic uremic syndrome with complement gene abnormality
ORPHA:544472
Atypical hemolytic uremic syndrome with anti-factor H antibodies
ORPHA:93581
LEGEND
Parent-child relationship
Parent-child relationship with a clinical entity not represented on the figure
Group of disorders
Ultimate parent group
Disorder
Ultimate parent group chosen as the disorder's preferential parent
Subtype of disorder

Figure 3

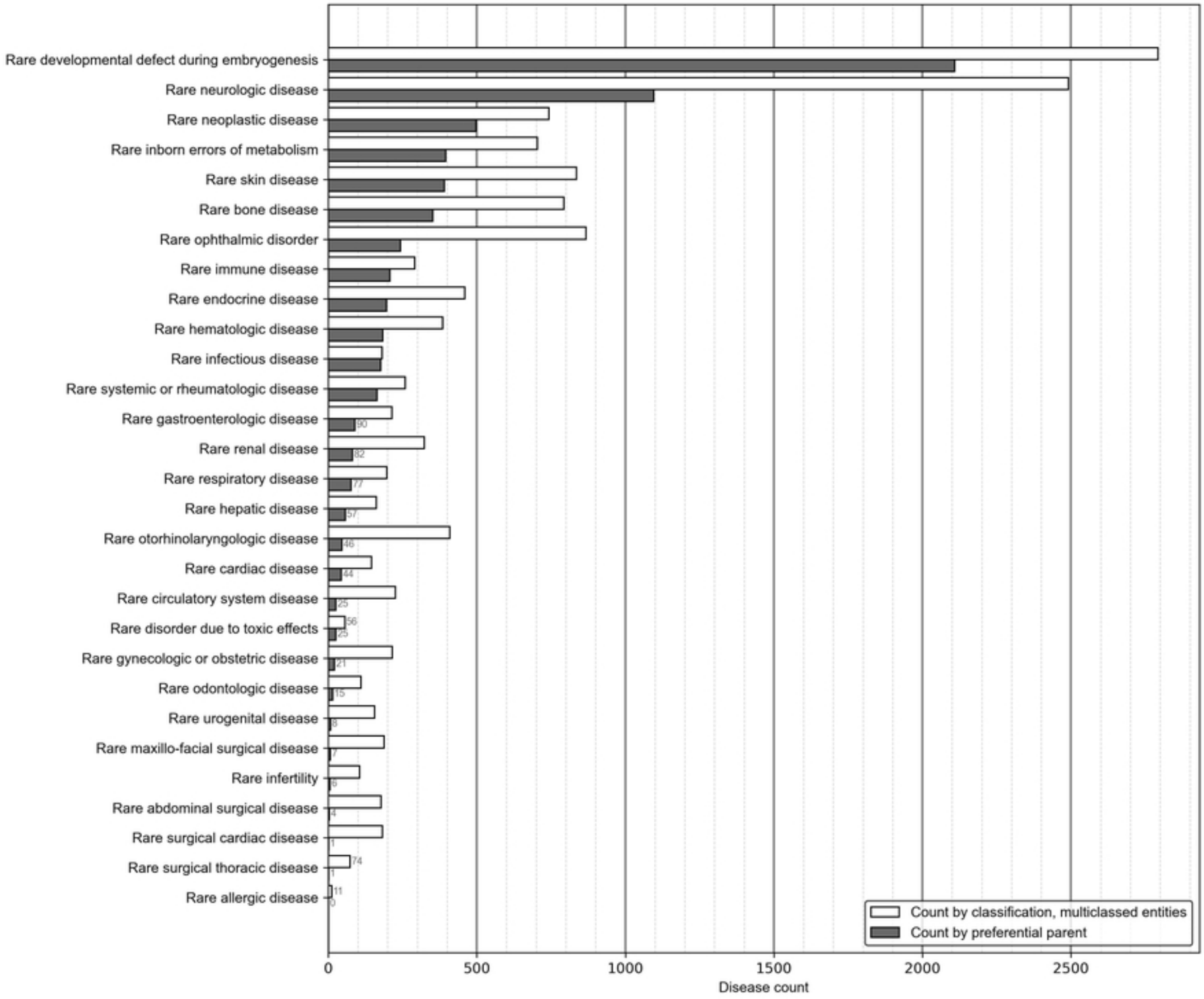

Rare developmental defect during embryogenesis
Rare neurologic disease
Rare neoplastic disease
Rare inborn errors of metabolism
Rare skin disease
Rare bone disease
Rare ophthalmic disorder
Rare immune disease
Rare endocrine disease
Rare hematologic disease
Rare infectious disease
Rare systemic or rheumatologic disease
Rare gastroenterologic disease
Rare renal disease
Rare respiratory disease
Rare hepatic disease
Rare otorhinolaryngologic disease
Rare cardiac disease
Rare circulatory system disease
Rare disorder due to toxic effects
Rare gynecologic or obstetric disease
Rare odontologic disease
Rare urogenital disease
Rare maxillo-facial surgical disease
Rare infertility
Rare abdominal surgical disease
Rare surgical cardiac disease
Rare surgical thoracic disease
Rare allergic disease
90
82
77
57
46
44
25
56
25
21
15
8
7
6
4
1
74
1
11
0
Count by classification, multiclassed entities
Count by preferential parent
0
500
1000
1500
2000
2500
Disease count

Figure 4