HACKATHON 2021
SENECA COLLEGE

Obtaining insights on ALS and other rare diseases through Open data

—

By Leonardo Patricelli (Project Manager and data scientist)
Nikolas Trivanovic (UX/UI Designer)

Introduction

‘With research, possibilities are limitless’. The importance of this truth cannot be overstated, as continual research is necessary in order to unlock the mysteries often presented by rare diseases. In a field where concrete solutions are not immediate, it is essential to mine the different research we already have, to make connections where none existed previously; to complement that information with new findings; and to listen to the voices of rare disease patients, who often provide the foundations of what is known about these medical enigmas.

Till date there are thousands of rare diseases in the world. Six to seven thousand diseases are considered rare and new diseases are regularly described in medical literature. Diseases that are rare are serious, often chronic and progressive. According to various research, usually signs are observed at birth or in childhood.In the United States, a rare disease is defined as a condition that affects fewer than 200,000 people in the US. Just in Canada, a rare condition affects about 1 in 12 canadians.

Rare diseases are hard to study because it’s complicated to gather info about them. If we had more info about them, maybe doctors may be able to find a cure or to improve the lives of people affected by this kind of disease.

Fortunately, Orphadata provides to the scientific community a comprehensive, high-quality and freely-accessible dataset related to rare diseases and orphan drugs, in a reusable format. For this challenge, we analysed the Orphadata dataset to build an application able to visualize aggregated statistics related to rare diseases and to return important information about them, in order to help doctors and researchers in their job while diagnosing a rare disease.

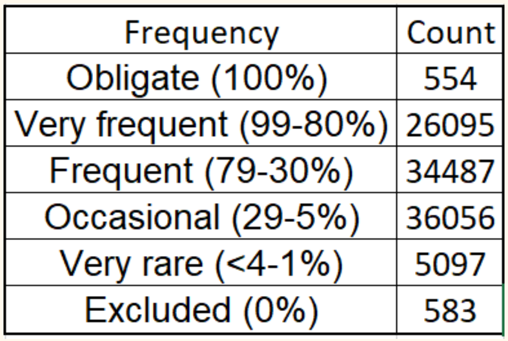
Data and technologies

For this challenge, we used the data available on the Orphadata website. The mission of Orphadata is to provide the scientific community with a comprehensive, high-quality and freely-accessible dataset related to rare diseases and orphan drugs, in a reusable format.

Once downloaded, the data were processed in a Python environment to get insights and to build an automatic way to explore the dataset.

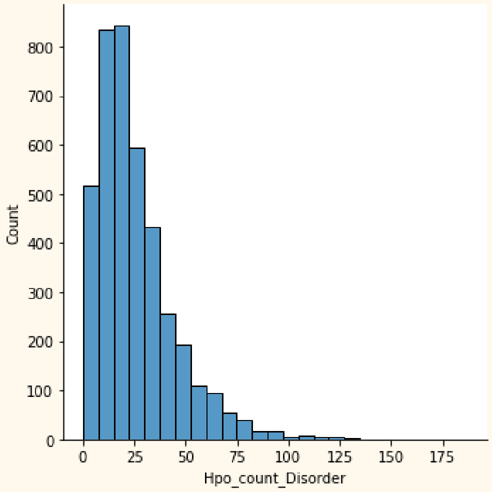
Human Phenotype Anthology (HPO)

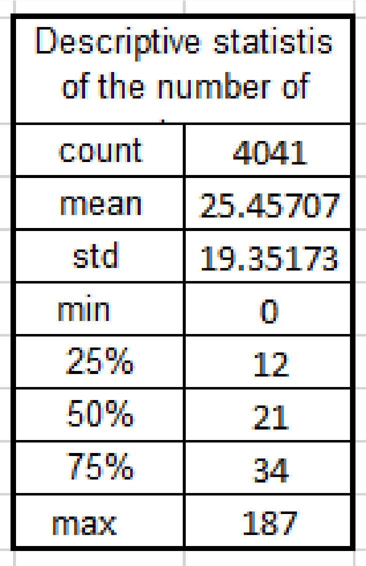
The Orphanet inventory of rare diseases is based on Human Phenotype Ontology (HPO) terms, a standardized and controlled terminology covering phenotypic abnormalities in human diseases. The annotation is characterized by frequency; let's have a look at the frequency categories and their distribution:



Frequency	Count
Obligate (100%)	554
Very frequent (99-80%)	26095
Frequent (79-30%)	34487
Occasional (29-5%)	36056
Very rare (<4-1%)	5097
Excluded (0%)	583

We can also look at the distribution of the number of HPOs associated with each disease.

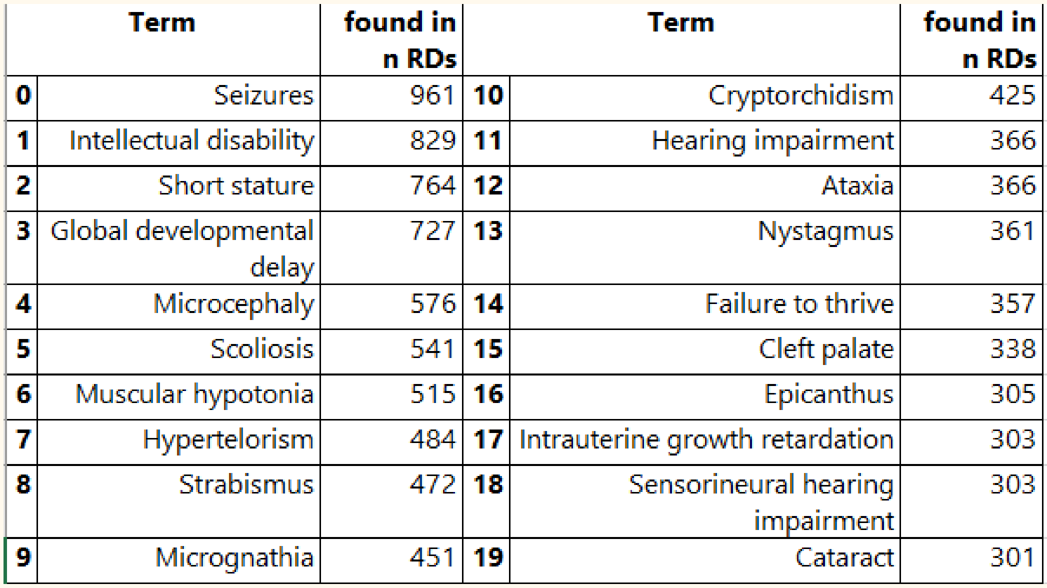




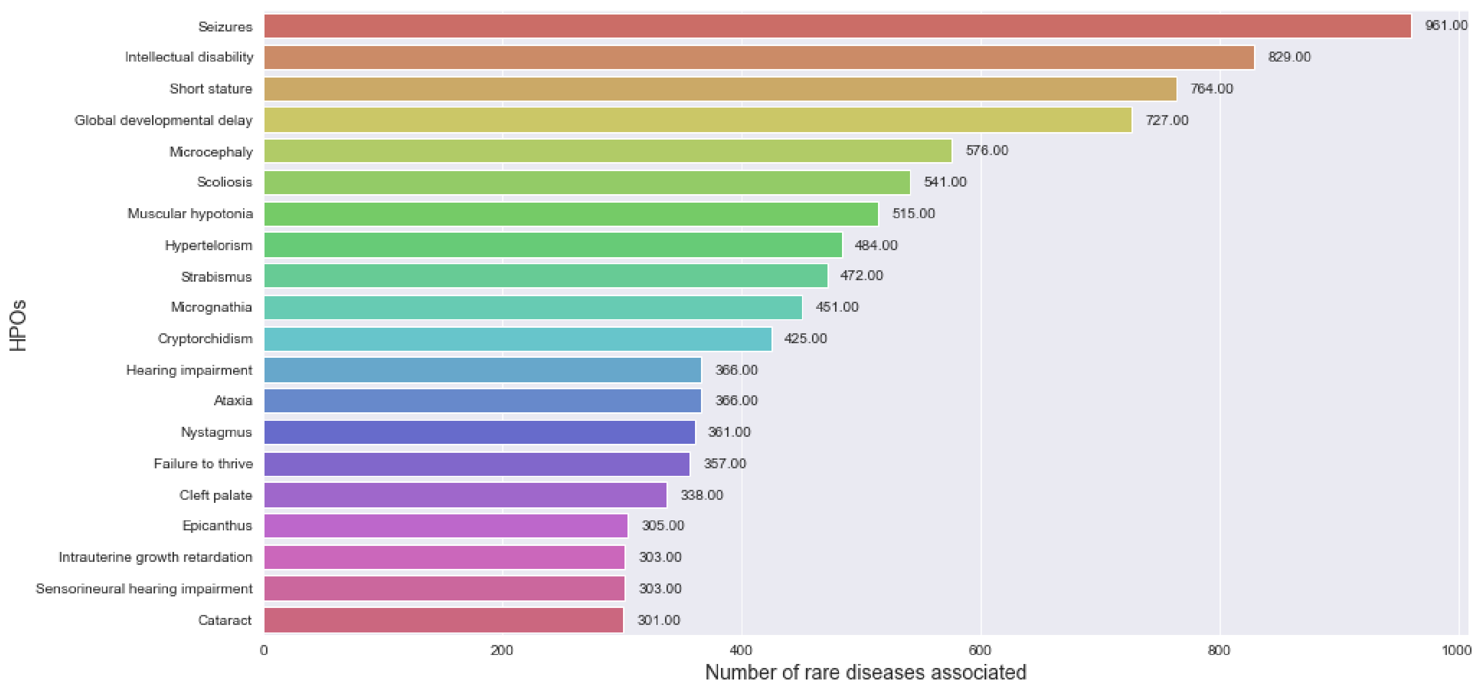
Descriptive statistis of the number of	
count	4041
mean	25.45707
std	19.35173
min	0
25%	12
50%	21
75%	34
max	187

It seems that each rare disease is associated on average with 25 HPOs, and 50% of them have 21 HPOs or less. Looking at the quantiles and at the histogram bins, we can have an idea of the distribution of the HPOs among the diseases.

It is also possible to look at the single phenotype to see in how many rare diseases it is present. In the table below we show the 20 most common Phenotypes with the count of the rare diseases involved.

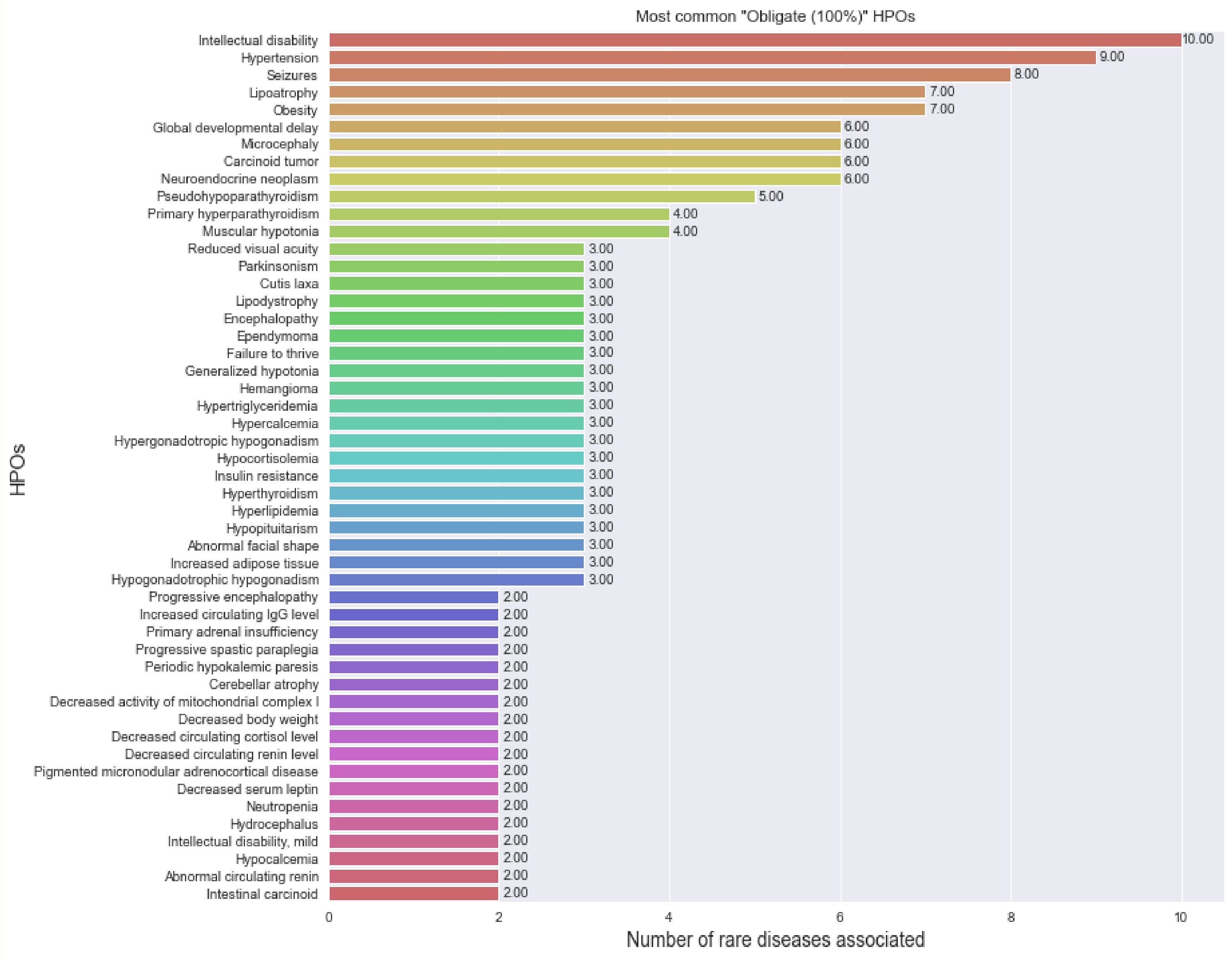


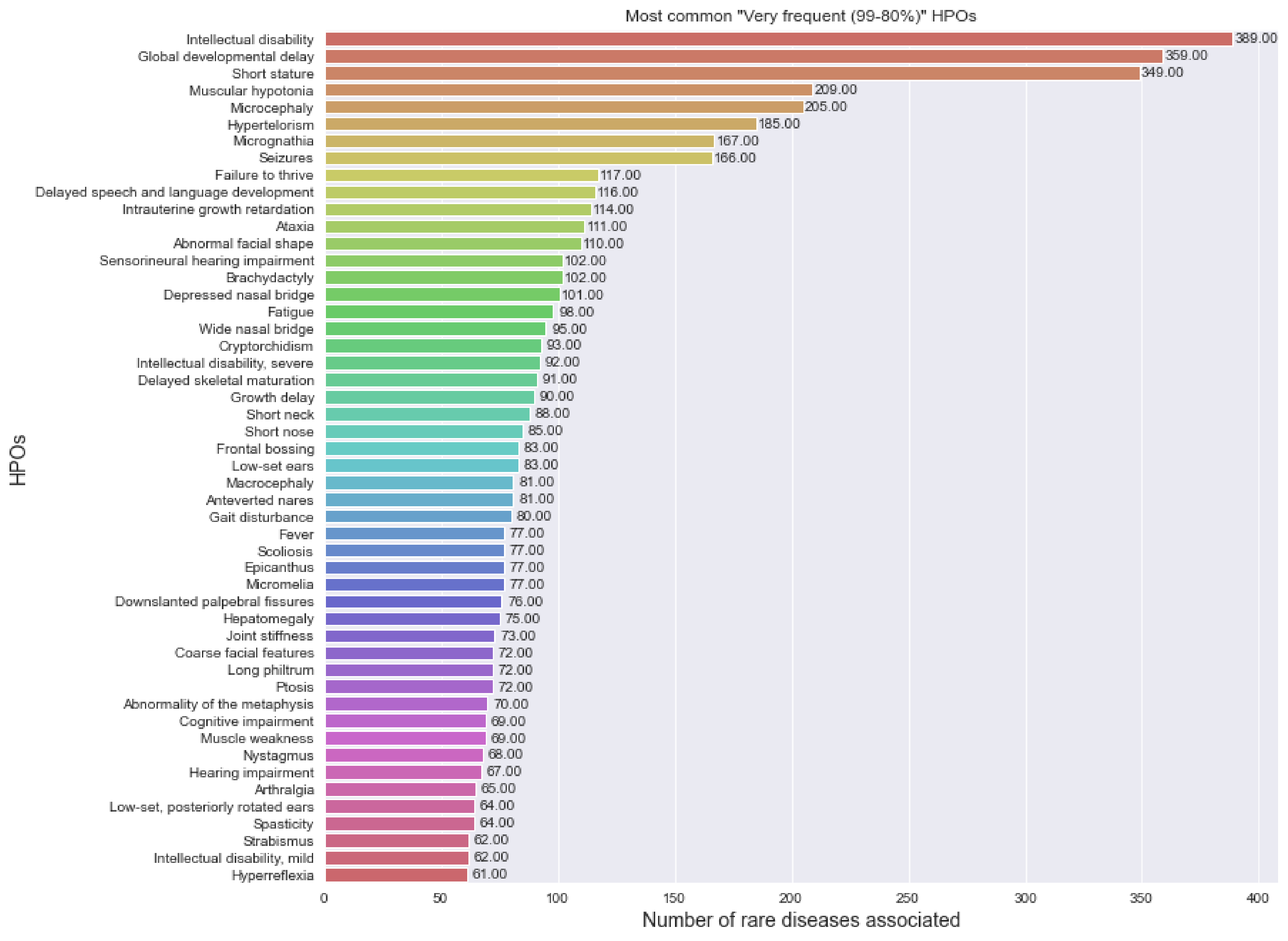
	Term	found in n RDs		Term	found in n RDs
0	Seizures	961	**10**	Cryptorchidism	425
1	Intellectual disability	829	**11**	Hearing impairment	366
2	Short stature	764	**12**	Ataxia	366
3	Global developmental delay	727	**13**	Nystagmus	361
4	Microcephaly	576	**14**	Failure to thrive	357
5	Scoliosis	541	**15**	Cleft palate	338
6	Muscular hypotonia	515	**16**	Epicanthus	305
7	Hypertelorism	484	**17**	Intrauterine growth retardation	303
8	Strabismus	472	**18**	Sensorineural hearing impairment	303
9	Micrognathia	451	**19**	Cataract	301

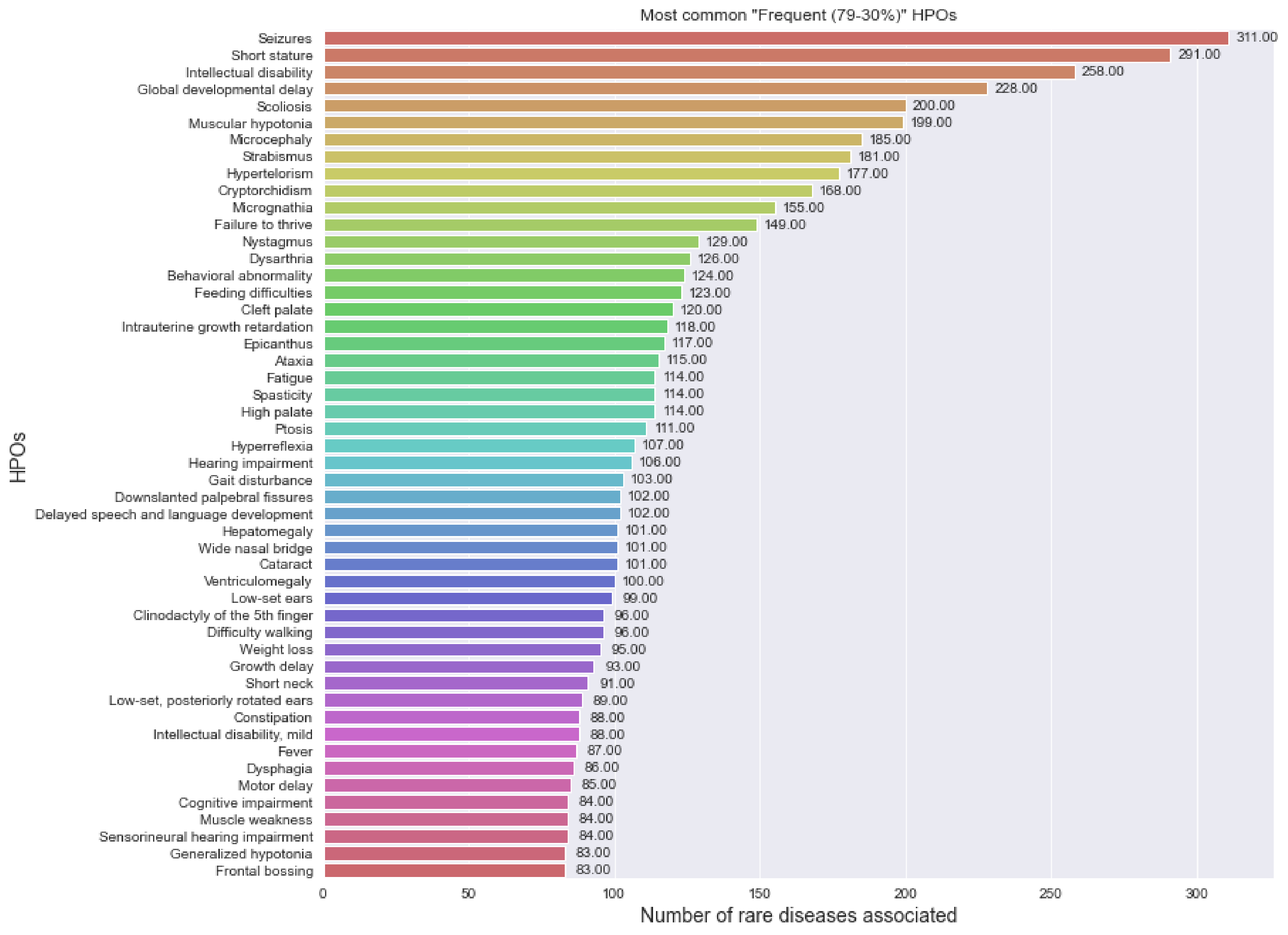


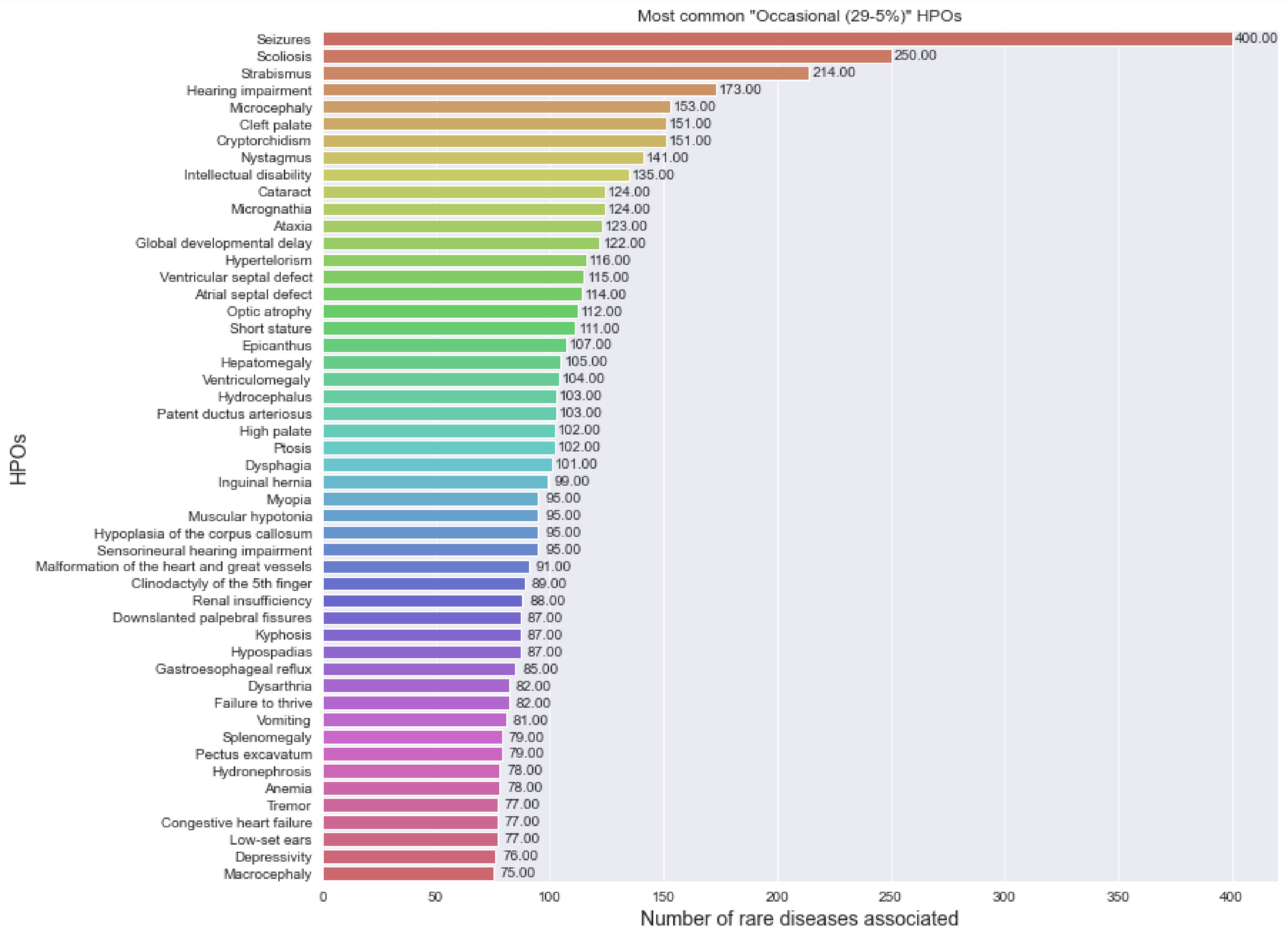
It has been identified that most cases are associated with **seizures** and **intellectual disability** followed by **short stature** and a **global development delay.**

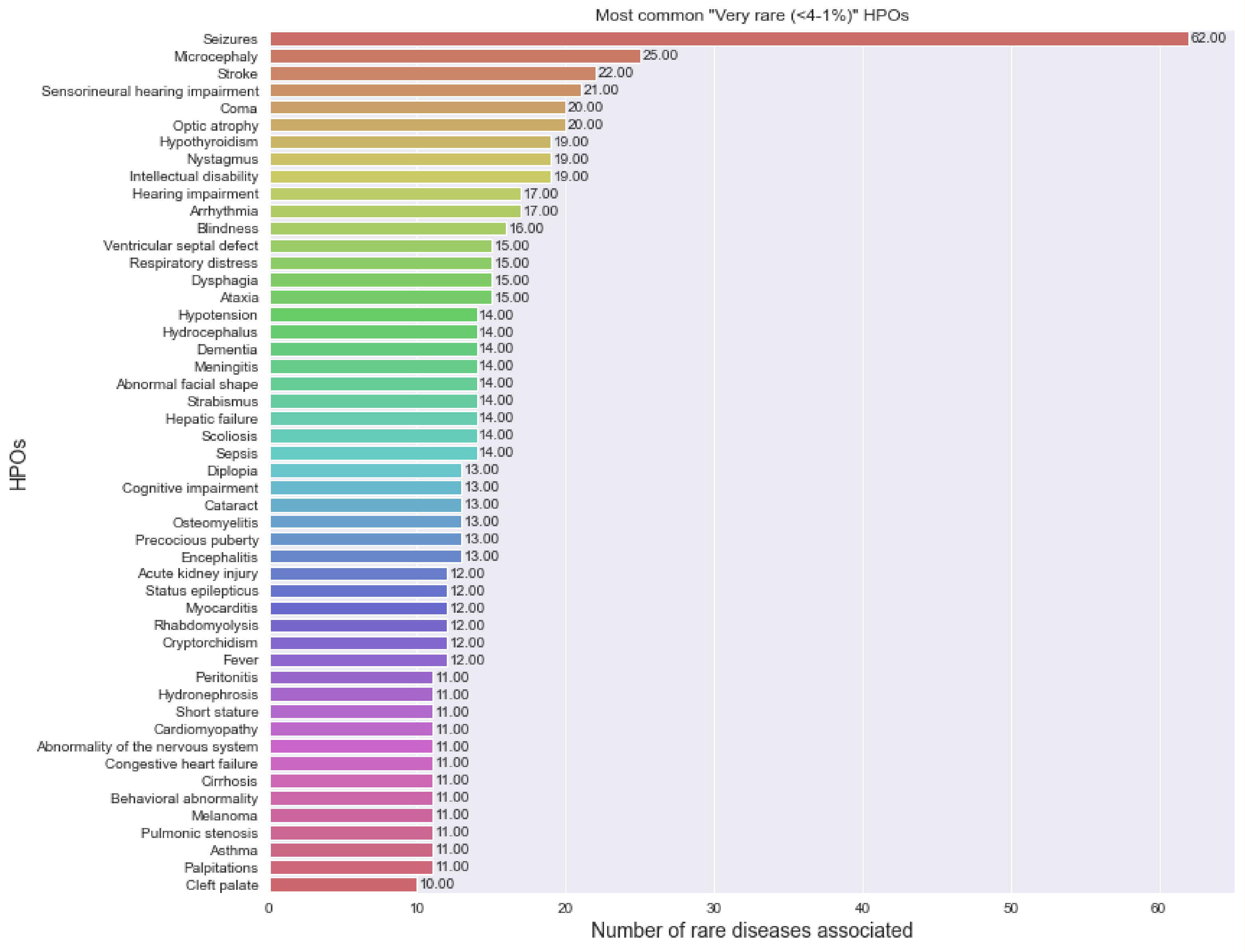
It is also possible to look at the HPOs per frequency class in order to understand what are the most frequent phenotypes in total (and per category).

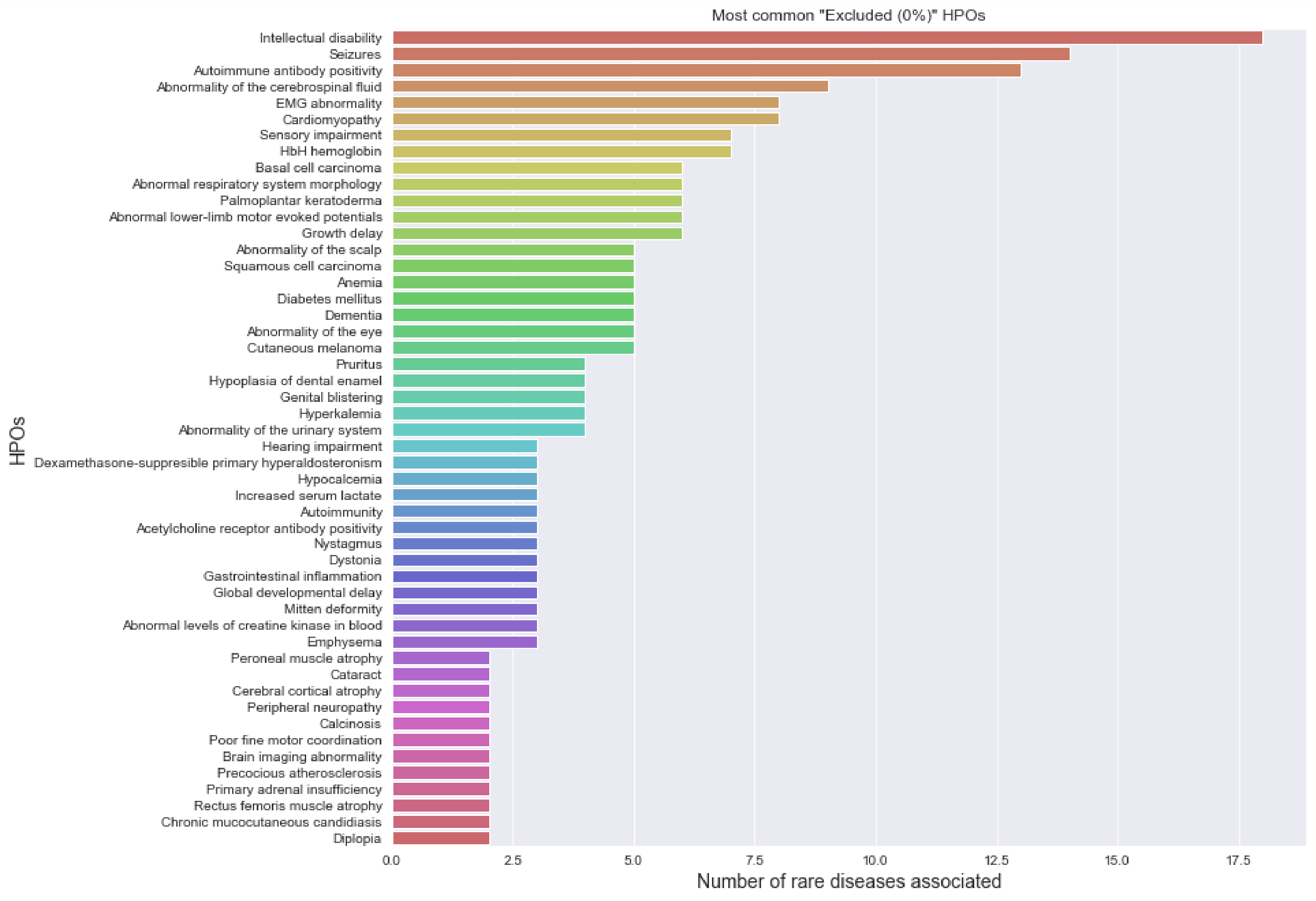

Most common "Obligate (100%)" HPOs
HPOs
Number of rare diseases associated
0
2
4
6
8
10
Intellectual disability
10.00
Hypertension
9.00
Seizures
8.00
Lipoatrophy
7.00
Obesity
7.00
Global developmental delay
6.00
Microcephaly
6.00
Carcinoid tumor
6.00
Neuroendocrine neoplasm
6.00
Pseudohypoparathyroidism
5.00
Primary hyperparathyroidism
4.00
Muscular hypotonia
4.00
Reduced visual acuity
3.00
Parkinsonism
3.00
Cutis laxa
3.00
Lipodystrophy
3.00
Encephalopathy
3.00
Ependymoma
3.00
Failure to thrive
3.00
Generalized hypotonia
3.00
Hemangioma
3.00
Hypertriglyceridemia
3.00
Hypercalcemia
3.00
Hypergonadotropic hypogonadism
3.00
Hypocortisolemia
3.00
Insulin resistance
3.00
Hyperthyroidism
3.00
Hyperlipidemia
3.00
Hypopituitarism
3.00
Abnormal facial shape
3.00
Increased adipose tissue
3.00
Hypogonadotrophic hypogonadism
3.00
Progressive encephalopathy
2.00
Increased circulating IgG level
2.00
Primary adrenal insufficiency
2.00
Progressive spastic paraplegia
2.00
Periodic hypokalemic paresis
2.00
Cerebellar atrophy
2.00
Decreased activity of mitochondrial complex I
2.00
Decreased body weight
2.00
Decreased circulating cortisol level
2.00
Decreased circulating renin level
2.00
Pigmented micronodular adrenocortical disease
2.00
Decreased serum leptin
2.00
Neutropenia
2.00
Hydrocephalus
2.00
Intellectual disability, mild
2.00
Hypocalcemia
2.00
Abnormal circulating renin
2.00
Intestinal carcinoid
2.00


Most common "Very frequent (99-80%)" HPOs
HPOs
Number of rare diseases associated
0
50
100
150
200
250
300
350
400
Intellectual disability 389.00
Global developmental delay 359.00
Short stature 349.00
Muscular hypotonia 209.00
Microcephaly 205.00
Hypertelorism 185.00
Micrognathia 167.00
Seizures 166.00
Failure to thrive 117.00
Delayed speech and language development 116.00
Intrauterine growth retardation 114.00
Ataxia 111.00
Abnormal facial shape 110.00
Sensorineural hearing impairment 102.00
Brachydactyly 102.00
Depressed nasal bridge 101.00
Fatigue 98.00
Wide nasal bridge 95.00
Cryptorchidism 93.00
Intellectual disability, severe 92.00
Delayed skeletal maturation 91.00
Growth delay 90.00
Short neck 88.00
Short nose 85.00
Frontal bossing 83.00
Low-set ears 83.00
Macrocephaly 81.00
Anteverted nares 81.00
Gait disturbance 80.00
Fever 77.00
Scoliosis 77.00
Epicanthus 77.00
Micromelia 77.00
Downslanted palpebral fissures 76.00
Hepatomegaly 75.00
Joint stiffness 73.00
Coarse facial features 72.00
Long philtrum 72.00
Ptosis 72.00
Abnormality of the metaphysis 70.00
Cognitive impairment 69.00
Muscle weakness 69.00
Nystagmus 68.00
Hearing impairment 67.00
Arthralgia 65.00
Low-set, posteriorly rotated ears 64.00
Spasticity 64.00
Strabismus 62.00
Intellectual disability, mild 62.00
Hyperreflexia 61.00


Most common "Frequent (79-30%)" HPOs
HPOs
Seizures
311.00
Short stature
291.00
Intellectual disability
258.00
Global developmental delay
228.00
Scoliosis
200.00
Muscular hypotonia
199.00
Microcephaly
185.00
Strabismus
181.00
Hypertelorism
177.00
Cryptorchidism
168.00
Micrognathia
155.00
Failure to thrive
149.00
Nystagmus
129.00
Dysarthria
126.00
Behavioral abnormality
124.00
Feeding difficulties
123.00
Cleft palate
120.00
Intrauterine growth retardation
118.00
Epicanthus
117.00
Ataxia
115.00
Fatigue
114.00
Spasticity
114.00
High palate
114.00
Ptosis
111.00
Hyperreflexia
107.00
Hearing impairment
106.00
Gait disturbance
103.00
Downslanted palpebral fissures
102.00
Delayed speech and language development
102.00
Hepatomegaly
101.00
Wide nasal bridge
101.00
Cataract
101.00
Ventriculomegaly
100.00
Low-set ears
99.00
Clinodactyly of the 5th finger
96.00
Difficulty walking
96.00
Weight loss
95.00
Growth delay
93.00
Short neck
91.00
Low-set, posteriorly rotated ears
89.00
Constipation
88.00
Intellectual disability, mild
88.00
Fever
87.00
Dysphagia
86.00
Motor delay
85.00
Cognitive impairment
84.00
Muscle weakness
84.00
Sensorineural hearing impairment
84.00
Generalized hypotonia
83.00
Frontal bossing
83.00
0
50
100
150
200
250
300
Number of rare diseases associated


Most common "Occasional (29-5%)" HPOs
HPOs
Seizures 400.00
Scoliosis 250.00
Strabismus 214.00
Hearing impairment 173.00
Microcephaly 153.00
Cleft palate 151.00
Cryptorchidism 151.00
Nystagmus 141.00
Intellectual disability 135.00
Cataract 124.00
Micrognathia 124.00
Ataxia 123.00
Global developmental delay 122.00
Hypertelorism 116.00
Ventricular septal defect 115.00
Atrial septal defect 114.00
Optic atrophy 112.00
Short stature 111.00
Epicanthus 107.00
Hepatomegaly 105.00
Ventriculomegaly 104.00
Hydrocephalus 103.00
Patent ductus arteriosus 103.00
High palate 102.00
Ptosis 102.00
Dysphagia 101.00
Inguinal hernia 99.00
Myopia 95.00
Muscular hypotonia 95.00
Hypoplasia of the corpus callosum 95.00
Sensorineural hearing impairment 95.00
Malformation of the heart and great vessels 91.00
Clinodactyly of the 5th finger 89.00
Renal insufficiency 88.00
Downslanted palpebral fissures 87.00
Kyphosis 87.00
Hypospadias 87.00
Gastroesophageal reflux 85.00
Dysarthria 82.00
Failure to thrive 82.00
Vomiting 81.00
Splenomegaly 79.00
Pectus excavatum 79.00
Hydronephrosis 78.00
Anemia 78.00
Tremor 77.00
Congestive heart failure 77.00
Low-set ears 77.00
Depressivity 76.00
Macrocephaly 75.00
0
50
100
150
200
250
300
350
400
Number of rare diseases associated


Most common "Very rare (<4-1%)" HPOs
HPOs
Seizures 62.00
Microcephaly 25.00
Stroke 22.00
Sensorineural hearing impairment 21.00
Coma 20.00
Optic atrophy 20.00
Hypothyroidism 19.00
Nystagmus 19.00
Intellectual disability 19.00
Hearing impairment 17.00
Arrhythmia 17.00
Blindness 16.00
Ventricular septal defect 15.00
Respiratory distress 15.00
Dysphagia 15.00
Ataxia 15.00
Hypotension 14.00
Hydrocephalus 14.00
Dementia 14.00
Meningitis 14.00
Abnormal facial shape 14.00
Strabismus 14.00
Hepatic failure 14.00
Scoliosis 14.00
Sepsis 14.00
Diplopia 13.00
Cognitive impairment 13.00
Cataract 13.00
Osteomyelitis 13.00
Precocious puberty 13.00
Encephalitis 13.00
Acute kidney injury 12.00
Status epilepticus 12.00
Myocarditis 12.00
Rhabdomyolysis 12.00
Cryptorchidism 12.00
Fever 12.00
Peritonitis 11.00
Hydronephrosis 11.00
Short stature 11.00
Cardiomyopathy 11.00
Abnormality of the nervous system 11.00
Congestive heart failure 11.00
Cirrhosis 11.00
Behavioral abnormality 11.00
Melanoma 11.00
Pulmonic stenosis 11.00
Asthma 11.00
Palpitations 11.00
Cleft palate 10.00
0
10
20
30
40
50
60
Number of rare diseases associated

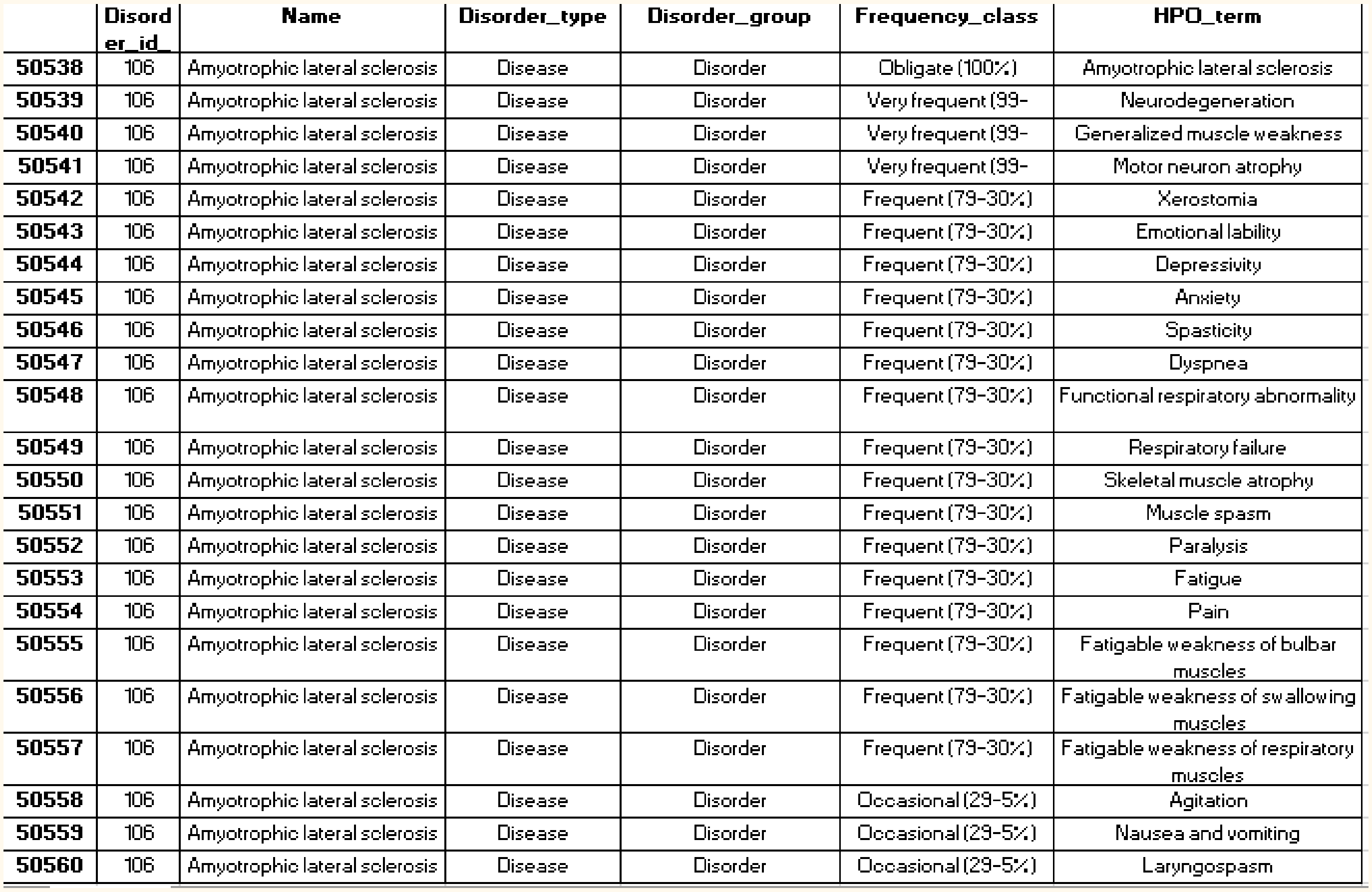


Looking at the single disease (ALS)

It is also possible to select a single disease to know what HPOs are related to and the frequency class associated.

In the case of ALS, for example, we see that **neurodegeneration**, **generalized muscle weakness** and **motor neuron atrophy** are very frequent. Below is displayed the full table with all the HPO related to the ALS and their frequency class.

The same output can be obtained for all the diseases.

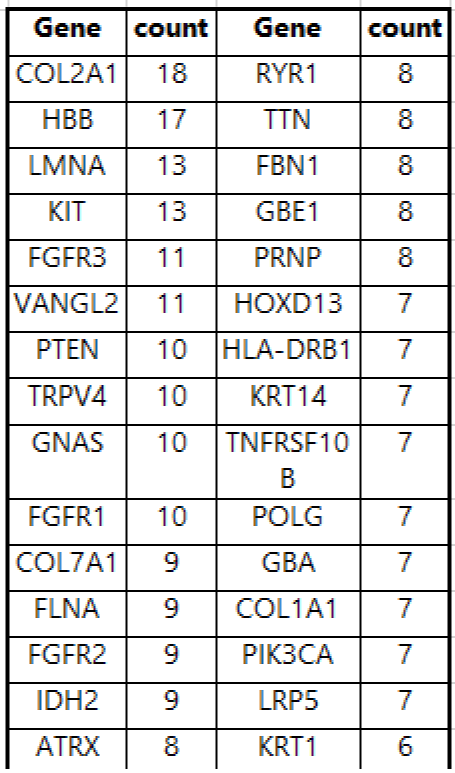


	Disorder_id_	Name	Disorder_type	Disorder_group	Frequency_class	HPO_term
50538	106	Amyotrophic lateral sclerosis	Disease	Disorder	Obligate (100%)	Amyotrophic lateral sclerosis
50539	106	Amyotrophic lateral sclerosis	Disease	Disorder	Very frequent (99-	Neurodegeneration
50540	106	Amyotrophic lateral sclerosis	Disease	Disorder	Very frequent (99-	Generalized muscle weakness
50541	106	Amyotrophic lateral sclerosis	Disease	Disorder	Very frequent (99-	Motor neuron atrophy
50542	106	Amyotrophic lateral sclerosis	Disease	Disorder	Frequent (79-30%)	Xerostomia
50543	106	Amyotrophic lateral sclerosis	Disease	Disorder	Frequent (79-30%)	Emotional lability
50544	106	Amyotrophic lateral sclerosis	Disease	Disorder	Frequent (79-30%)	Depressivity
50545	106	Amyotrophic lateral sclerosis	Disease	Disorder	Frequent (79-30%)	Anxiety
50546	106	Amyotrophic lateral sclerosis	Disease	Disorder	Frequent (79-30%)	Spasticity
50547	106	Amyotrophic lateral sclerosis	Disease	Disorder	Frequent (79-30%)	Dyspnea
50548	106	Amyotrophic lateral sclerosis	Disease	Disorder	Frequent (79-30%)	Functional respiratory abnormality
50549	106	Amyotrophic lateral sclerosis	Disease	Disorder	Frequent (79-30%)	Respiratory failure
50550	106	Amyotrophic lateral sclerosis	Disease	Disorder	Frequent (79-30%)	Skeletal muscle atrophy
50551	106	Amyotrophic lateral sclerosis	Disease	Disorder	Frequent (79-30%)	Muscle spasm
50552	106	Amyotrophic lateral sclerosis	Disease	Disorder	Frequent (79-30%)	Paralysis
50553	106	Amyotrophic lateral sclerosis	Disease	Disorder	Frequent (79-30%)	Fatigue
50554	106	Amyotrophic lateral sclerosis	Disease	Disorder	Frequent (79-30%)	Pain
50555	106	Amyotrophic lateral sclerosis	Disease	Disorder	Frequent (79-30%)	Fatigable weakness of bulbar muscles
50556	106	Amyotrophic lateral sclerosis	Disease	Disorder	Frequent (79-30%)	Fatigable weakness of swallowing muscles
50557	106	Amyotrophic lateral sclerosis	Disease	Disorder	Frequent (79-30%)	Fatigable weakness of respiratory muscles
50558	106	Amyotrophic lateral sclerosis	Disease	Disorder	Occasional (29-5%)	Agitation
50559	106	Amyotrophic lateral sclerosis	Disease	Disorder	Occasional (29-5%)	Nausea and vomiting
50560	106	Amyotrophic lateral sclerosis	Disease	Disorder	Occasional (29-5%)	Laryngospasm

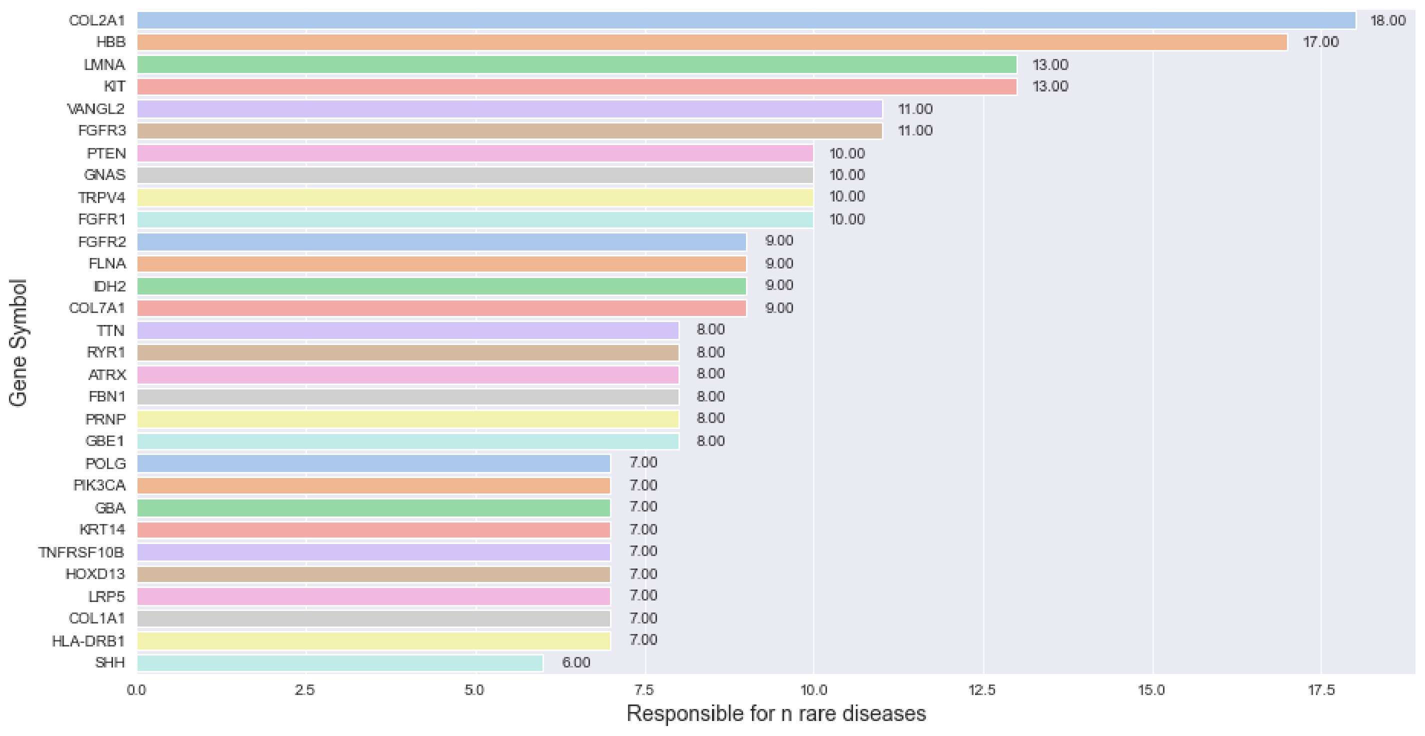
Genes and their loci

In order to better define rare disorders of genetic origin, Orphanet provides information on every gene related to a rare disorder. This information includes the genetic international nomenclature, the gene typology and the chromosomal location. Orphanet also defines the relationship between genes and their related rare disorders and provides evidence for establishing these gene-disorder relationships.

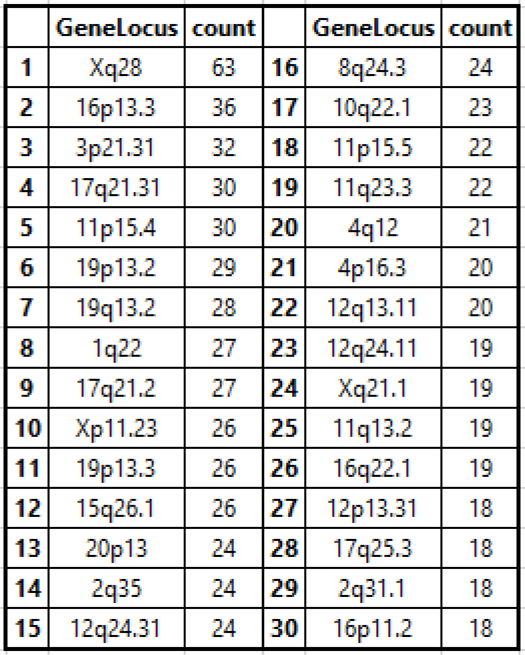
The table below displays the 30 genes (actually their official symbol) most responsible for rare diseases with the count of associated diseases.



Gene	count	Gene	count
COL2A1	18	RYR1	8
HBB	17	TTN	8
LMNA	13	FBN1	8
KIT	13	GBE1	8
FGFR3	11	PRNP	8
VANGL2	11	HOXD13	7
PTEN	10	HLA-DRB1	7
TRPV4	10	KRT14	7
GNAS	10	TNFRSF10B	7
FGFR1	10	POLG	7
COL7A1	9	GBA	7
FLNA	9	COL1A1	7
FGFR2	9	PIK3CA	7
IDH2	9	LRP5	7
ATRX	8	KRT1	6

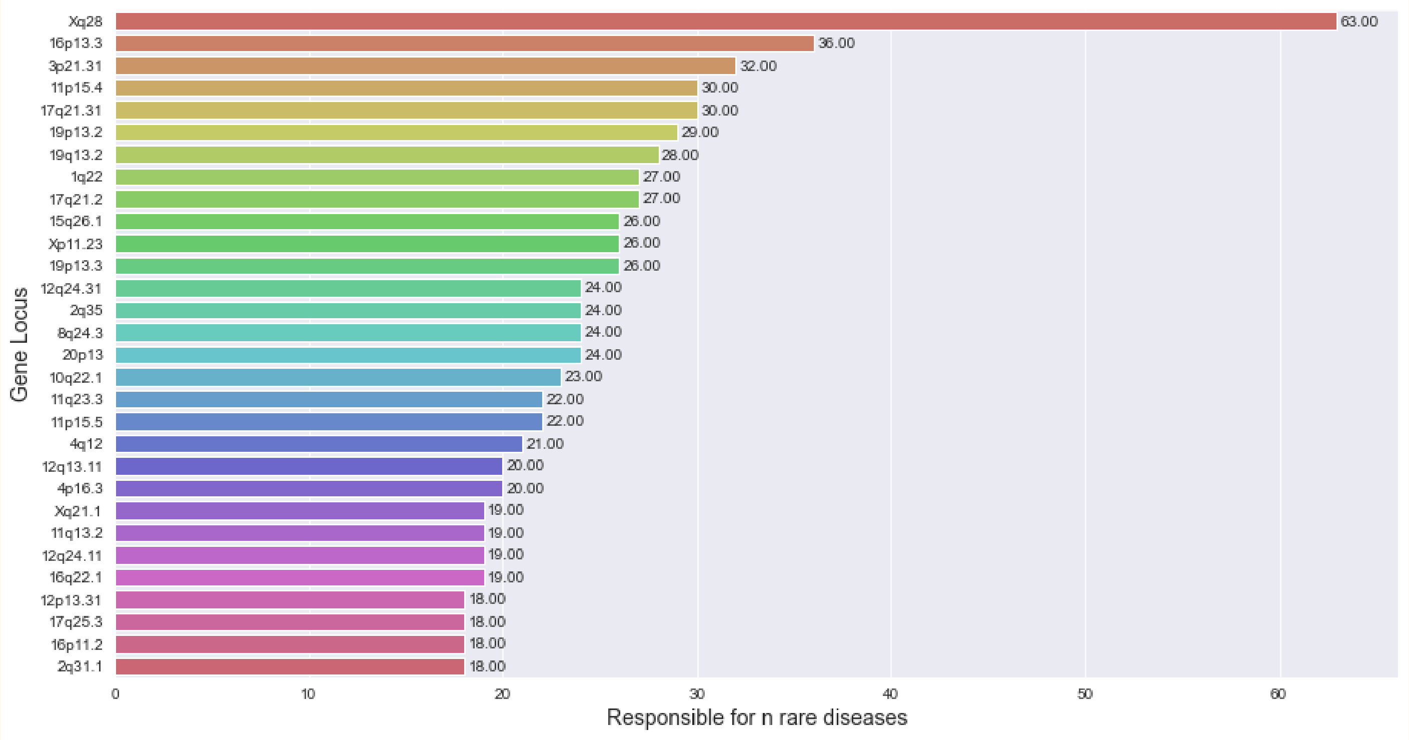


In the data, we can find also information related to the location of the Gene. We grouped the genes by their location in order to find the locations responsible for most of the rare diseases in the dataset.



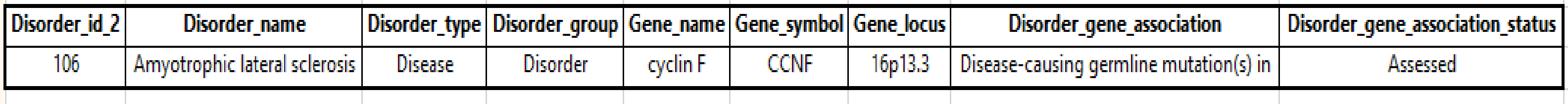
	GeneLocus	count		GeneLocus	count
1	Xq28	63	**16**	8q24.3	24
2	16p13.3	36	**17**	10q22.1	23
3	3p21.31	32	**18**	11p15.5	22
4	17q21.31	30	**19**	11q23.3	22
5	11p15.4	30	**20**	4q12	21
6	19p13.2	29	**21**	4p16.3	20
7	19q13.2	28	**22**	12q13.11	20
8	1q22	27	**23**	12q24.11	19
9	17q21.2	27	**24**	Xq21.1	19
10	Xp11.23	26	**25**	11q13.2	19
11	19p13.3	26	**26**	16q22.1	19
12	15q26.1	26	**27**	12p13.31	18
13	20p13	24	**28**	17q25.3	18
14	2q35	24	**29**	2q31.1	18
15	12q24.31	24	**30**	16p11.2	18

These tables can provide insights on what genes should be tested first during the diagnosis of a rare disease, or in what location to see for genetic anomalies. Knowing what are the most “problematic” genes and their location can speed up the process and lead to faster diagnoses.



Looking at the single disease (ALS)

It is also possible to retrieve the genetic information for every disease in the dataset. For the ALS, we can see that it is related to the Gene ***cyclin F (***wth related ***symbol)*** located in position ***16p13.3***, with the related association between gene and disorder (in this case ***Disease-causing germline mutations***)



Disorder_id_2	Disorder_name	Disorder_type	Disorder_group	Gene_name	Gene_symbol	Gene_locus	Disorder_gene_association	Disorder_gene_association_status
106	Amyotrophic lateral sclerosis	Disease	Disorder	cyclin F	CCNF	16p13.3	Disease-causing germline mutation(s) in	Assessed

Functional consequences (Disabilities)

The Orphanet inventory of rare diseases is annotated with activity limitation/participation restriction (functional consequences), using the Orphanet Functioning Thesaurus, derived and adapted from the International Classification of Functioning, Disability and Health – Children and Youth (ICF-CY, WHO 2007). The information provided is assessed taking into account the whole patient population affected by the disease, receiving standard care and management (specific and/or symptomatic management, prevention and prophylaxis, devices and aids, care and support).

Each functional consequence is annotated with the following:

- Frequency in the patient population:
  - o Very frequent: more than 80%
  - o Frequent: between 30% and 80%
  - o Occasional: fewer than 30%
- Temporality:

o Permanent limitation/restriction: the functional consequence is present throughout the life of the patient. It can be congenital, secondary to loss of a skill or participation. It can be a direct or indirect consequence of the disease or of its treatment.

o Transient limitation/restriction: the functional consequence occurs during acute episodes, periodic crises or relapses. It resolves or reduces spontaneously or by the action of treatment or care.

o Delayed acquisition: a skill or participation is performed later than by a healthy person.

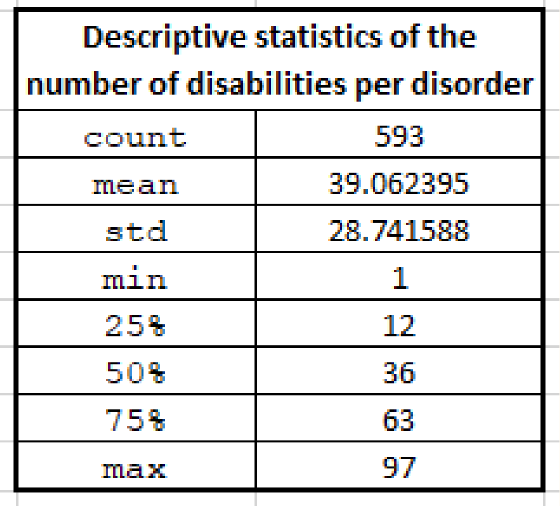
- Degree of severity:
  - o Low: activity or participation can be carried out with little difficulty by the patient alone.
  - o Moderate: activity or participation can be carried out with some technical and/or human assistance
  - o Severe: activity or participation cannot be carried out without substantial technical and/or human assistance.
  - o Complete: activity or participation cannot be carried out, even with technical and/or human assistance.
  - o Unspecified: limitation/restriction is difficult to quantify or highly variable between patients (ranging from 'Low' to 'Complete').
- Loss of ability when relevant, defined by the progressive and definitive loss of a skill or participation over the course of the disease.

A functional limitation is stated to be « undefined » when the current knowledge does not enable information about the extent of the consequences on daily life to be provided.

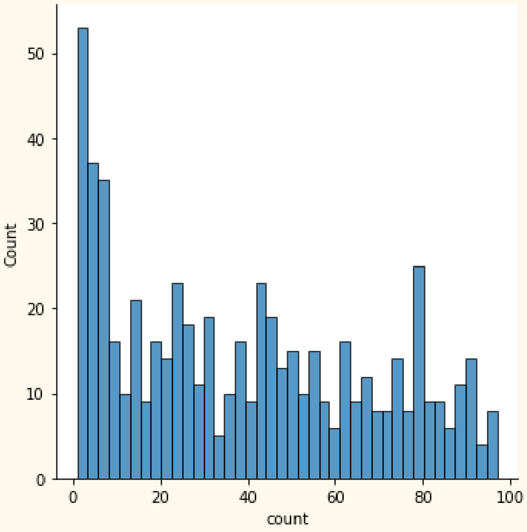
The unaffected activities and participation are not listed.

Environmental factors that may have an impact on the daily activities of the patients are also identified and listed when possible.

First of all, let’s have a look at the distribution of the number of disabilities per disorder

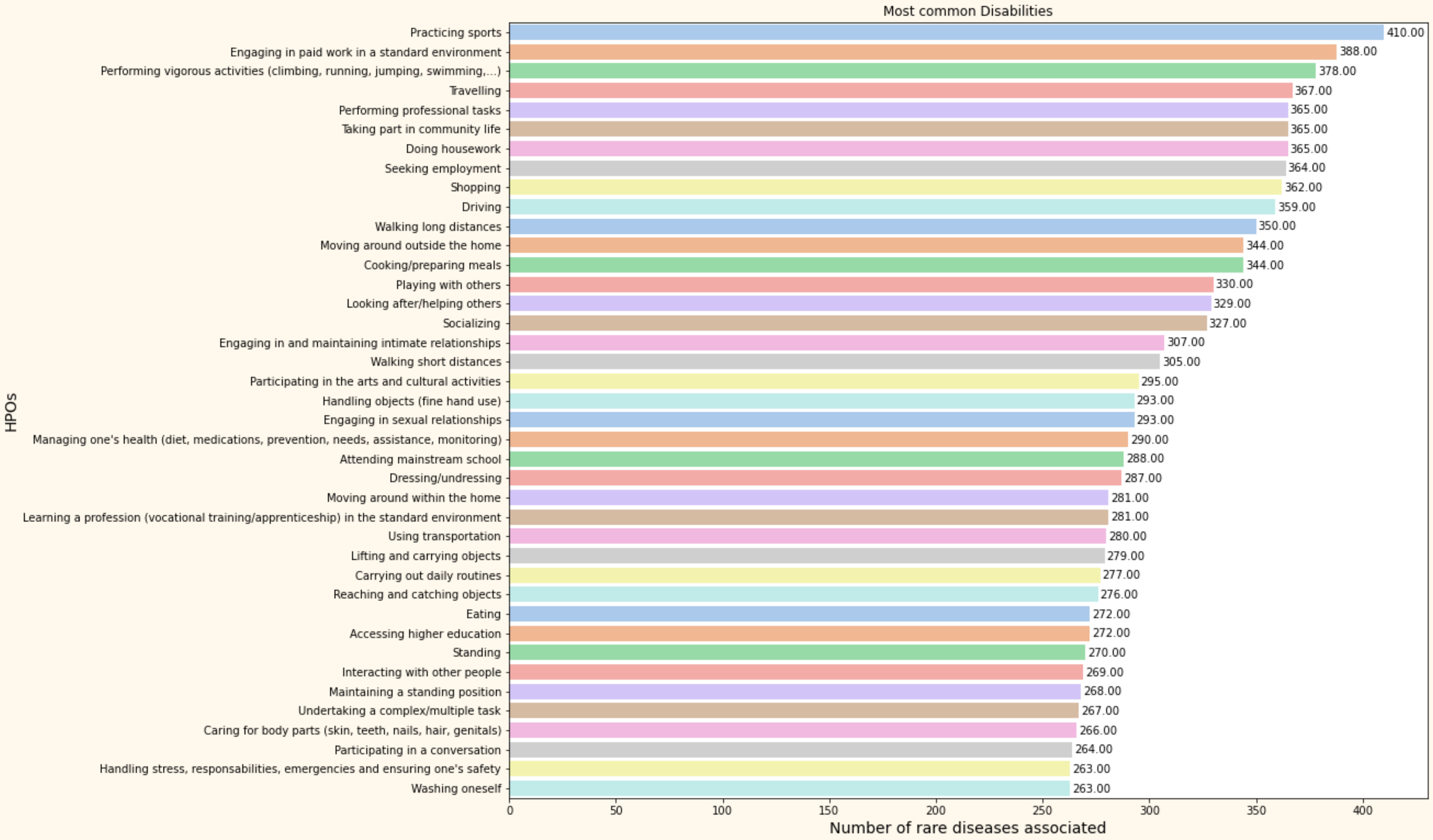


Descriptive statistics of the number of disabilities per disorder	
count	593
mean	39.062395
std	28.741588
min	1
25%	12
50%	36
75%	63
max	97



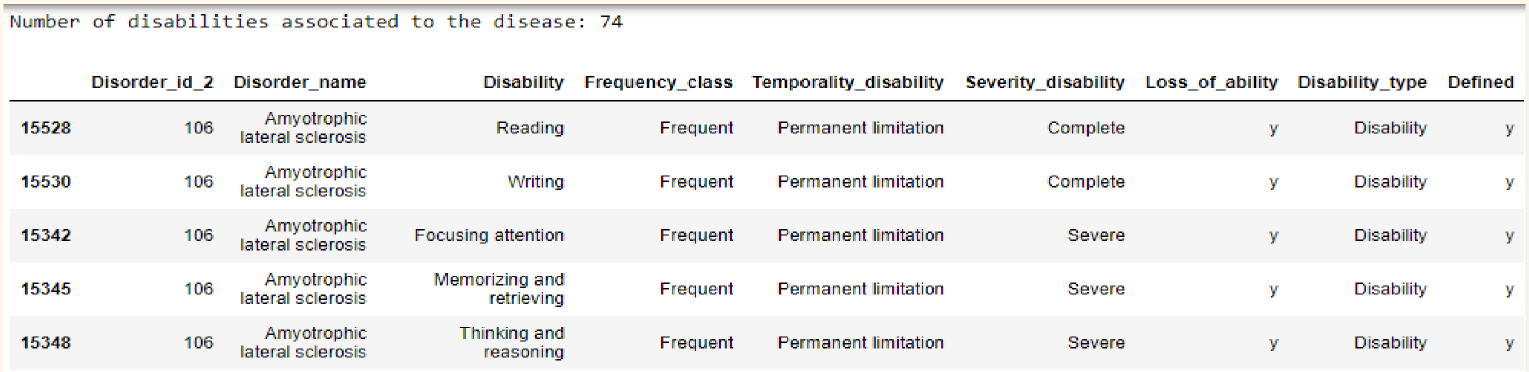
It seems that on average each disease has 39 disabilities related, and half of them has 36 diseases or less (the median). Looking at the barplot we can have a better idea of the distribution. It seems that it varies a lot (in fact this justifies the high standard deviation, with a value of 28).

Here we can have a look at the most common disabilities :



Looking at the single disease (ALS)

It's also possible to analyze what are the disabilities related to a single disease. Looking at the ALS we can notice that the disease has 74 associated disabilities. Our application sort the values by *Frequency class* and *severity*. We are going to pick just the top 5; of these 74 disabilities, **reading** and **writing** result as ***complete*** disabilities, while **Focusing attention**, **memorizing and retrieve** and **thinking and reasoning** are also frequent, but their severity is "***severe***".



```
Number of disabilities associated to the disease: 74
```

	Disorder_id_2	Disorder_name	Disability	Frequency_class	Temporality_disability	Severity_disability	Loss_of_ability	Disability_type	Defined
15528	106	Amyotrophic lateral sclerosis	Reading	Frequent	Permanent limitation	Complete	y	Disability	y
15530	106	Amyotrophic lateral sclerosis	Writing	Frequent	Permanent limitation	Complete	y	Disability	y
15342	106	Amyotrophic lateral sclerosis	Focusing attention	Frequent	Permanent limitation	Severe	y	Disability	y
15345	106	Amyotrophic lateral sclerosis	Memorizing and retrieving	Frequent	Permanent limitation	Severe	y	Disability	y
15348	106	Amyotrophic lateral sclerosis	Thinking and reasoning	Frequent	Permanent limitation	Severe	y	Disability	y

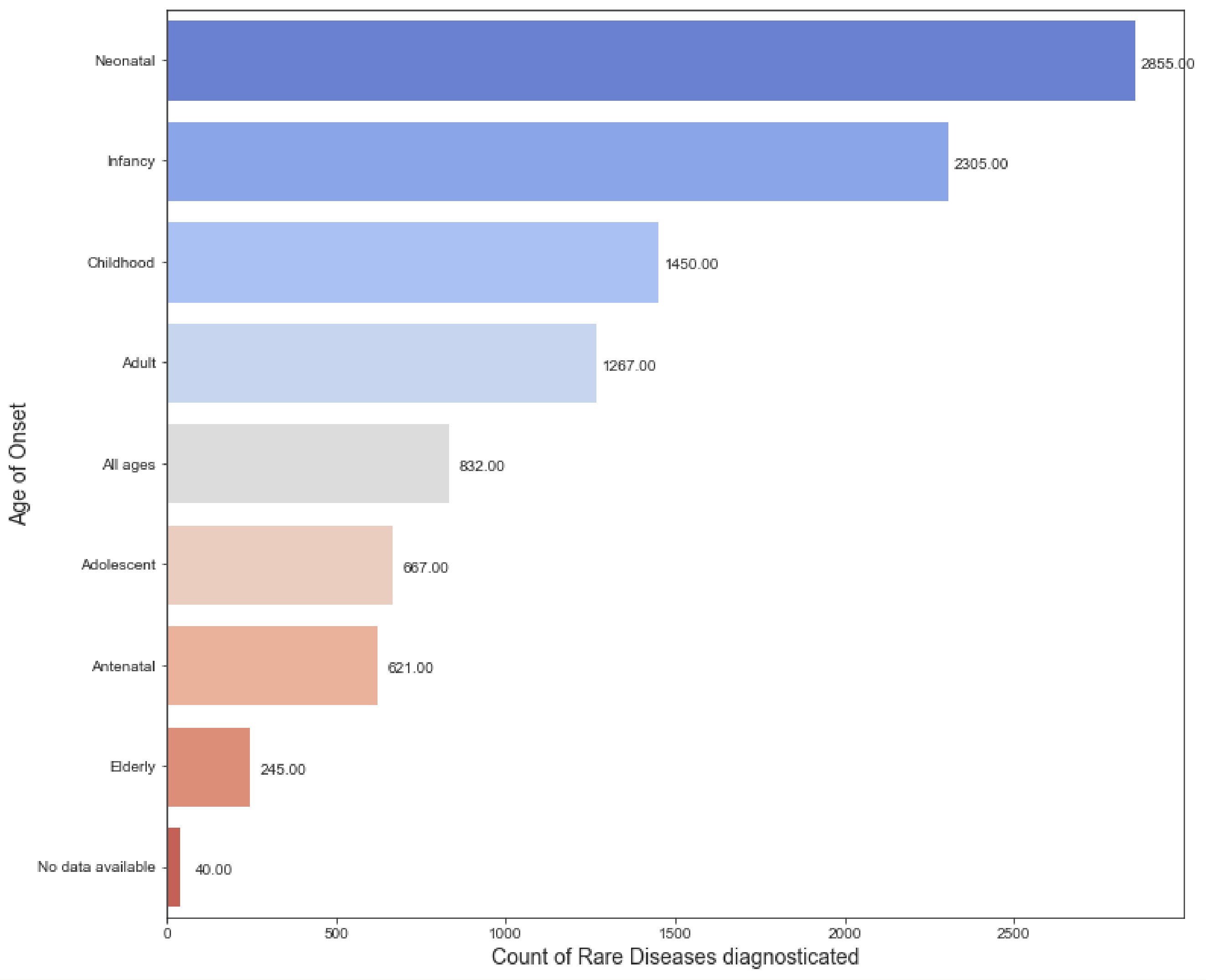
Average Age of onset and of death

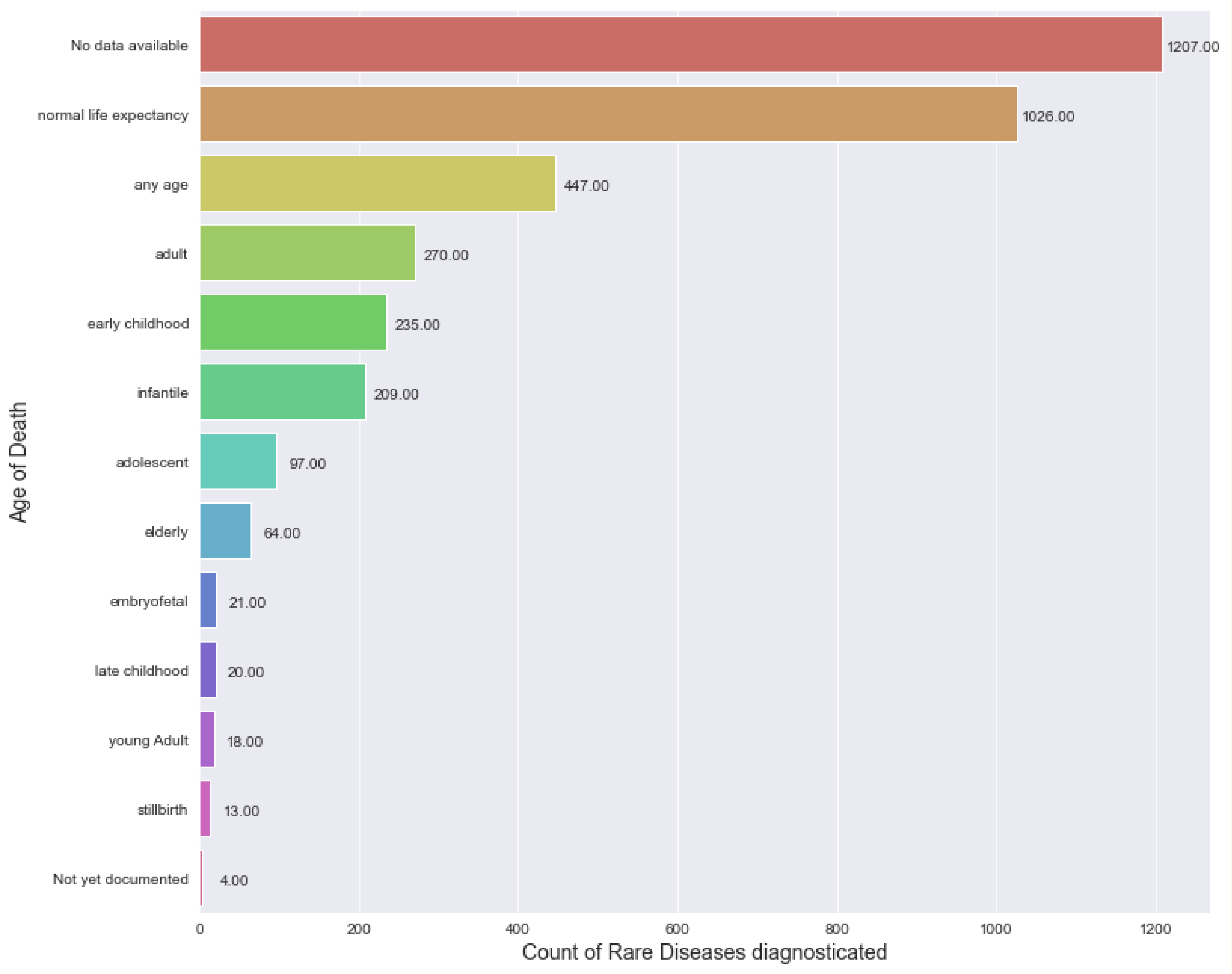
In this section, we are going to present the information related the average age of when the disease is discovered, and the average age of death. To estimate those values, Orphadata referred to published studies and their validity is taken for granted and not re-assessed.

Here an explanation of the variables:

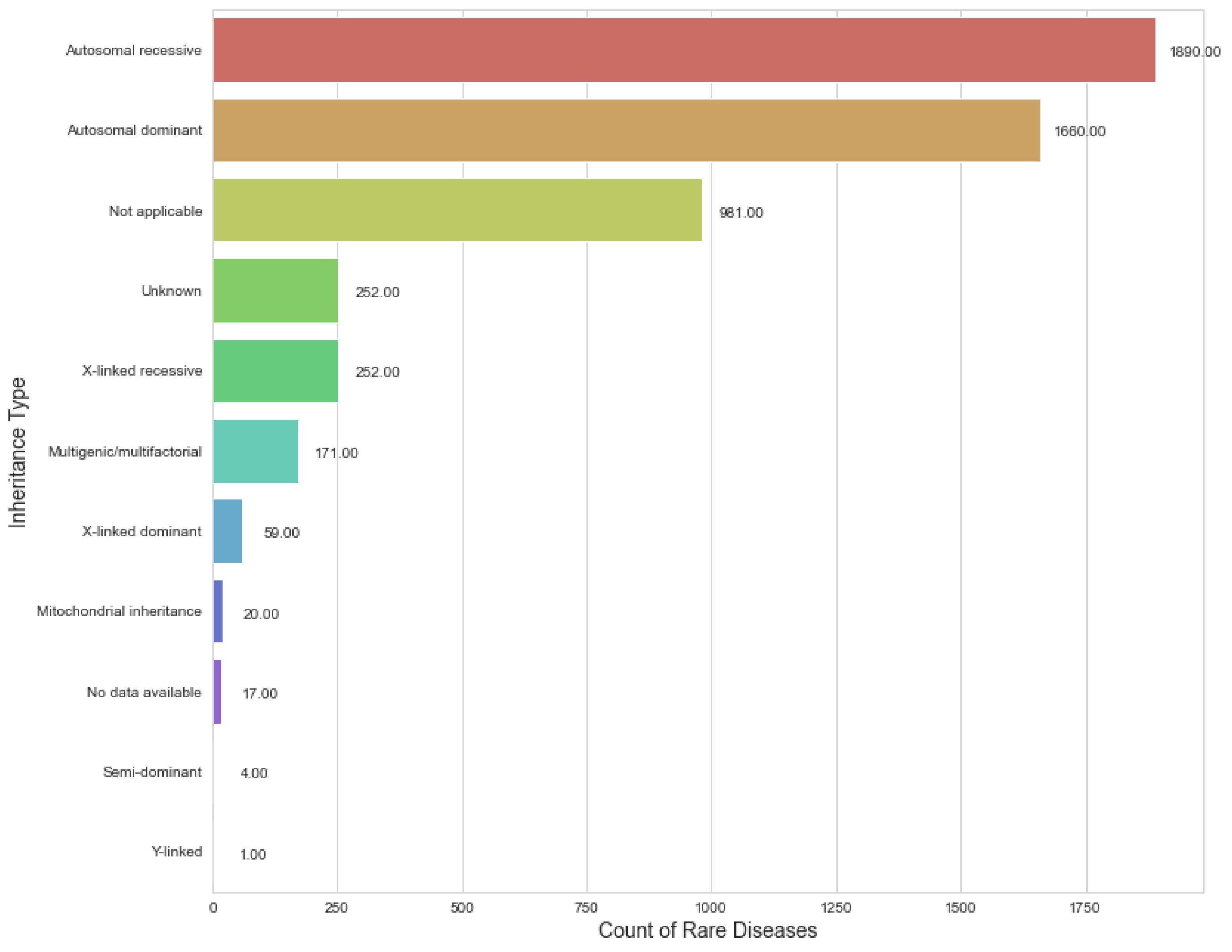
- **AverageAgeOfOnset**: classes based on the estimated average age of clinical entity onset.
- **AverageAgeOfDeath**: classes based on the estimated average age at death for a given clinical entity. There are twelve different population age groups.
- **TypeOfInheritance**: type(s) of inheritance associated with a given clinical entity. There are thirteen different types of inheritance.

Looking at the **AverageAgeOfOnset,** it seems that most of the rare disease is found in early age (infact *Neonatal*, *Infancy*, and *Childhood* are the most recurrent classes). For what concerns the **AverageAgeOfDeath**, we have missing data for about a third of the disease, while the second third of diseases allow the patient to have normal life expectancy, and the restant third is distributed mostly between early ages and adult age.


Age of Onset
Neonatal
Infancy
Childhood
Adult
All ages
Adolescent
Antenatal
Elderly
No data available
2855.00
2305.00
1450.00
1267.00
832.00
667.00
621.00
245.00
40.00
0
500
1000
1500
2000
2500
Count of Rare Diseases diagnosticated

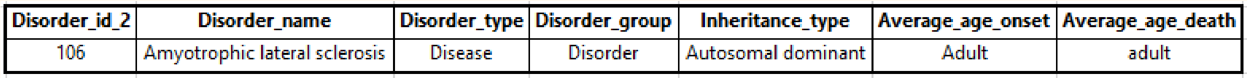


According to the Orphadataset, most of the rare diseases have an inheritance ***Autosomal recessive*** or ***Autosomal Dominant***. Unfortunately, there is a large part of the dataset in which this information is “*not applicable*”.



Looking at the single disease (ALS)

This information can be retrieved for every disease. Looking at the ALS disease, we can notice that it has an ***autosomal dominant*** inheritance, and on average it is diagnosed in ***adult age*** and led soon to death, since the average death age is also ***adult***.



Disorder_id_2	Disorder_name	Disorder_type	Disorder_group	Inheritance_type	Average_age_onset	Average_age_death
106	Amyotrophic lateral sclerosis	Disease	Disorder	Autosomal dominant	Adult	adult

Prevalence

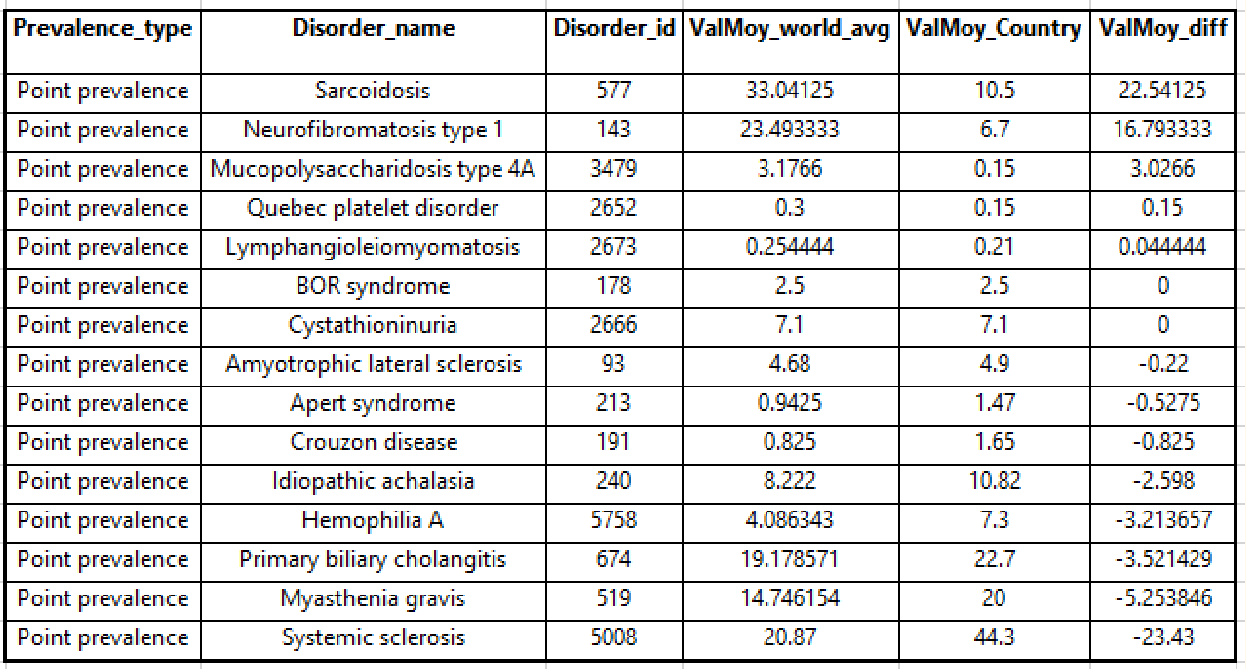
According to Wikipedia, the prevalence is the proportion of a particular population found to be affected by a medical condition (typically a disease or a risk factor such as smoking or seat-belt use) at a specific time. It is derived by comparing the number of people found to have the condition with the total number of people studied, and is usually expressed as a fraction, a percentage, or the number of cases (ex. per 10,000 or 100,000 people).

The prevalence is calculated per countries, and can assume 3 values:

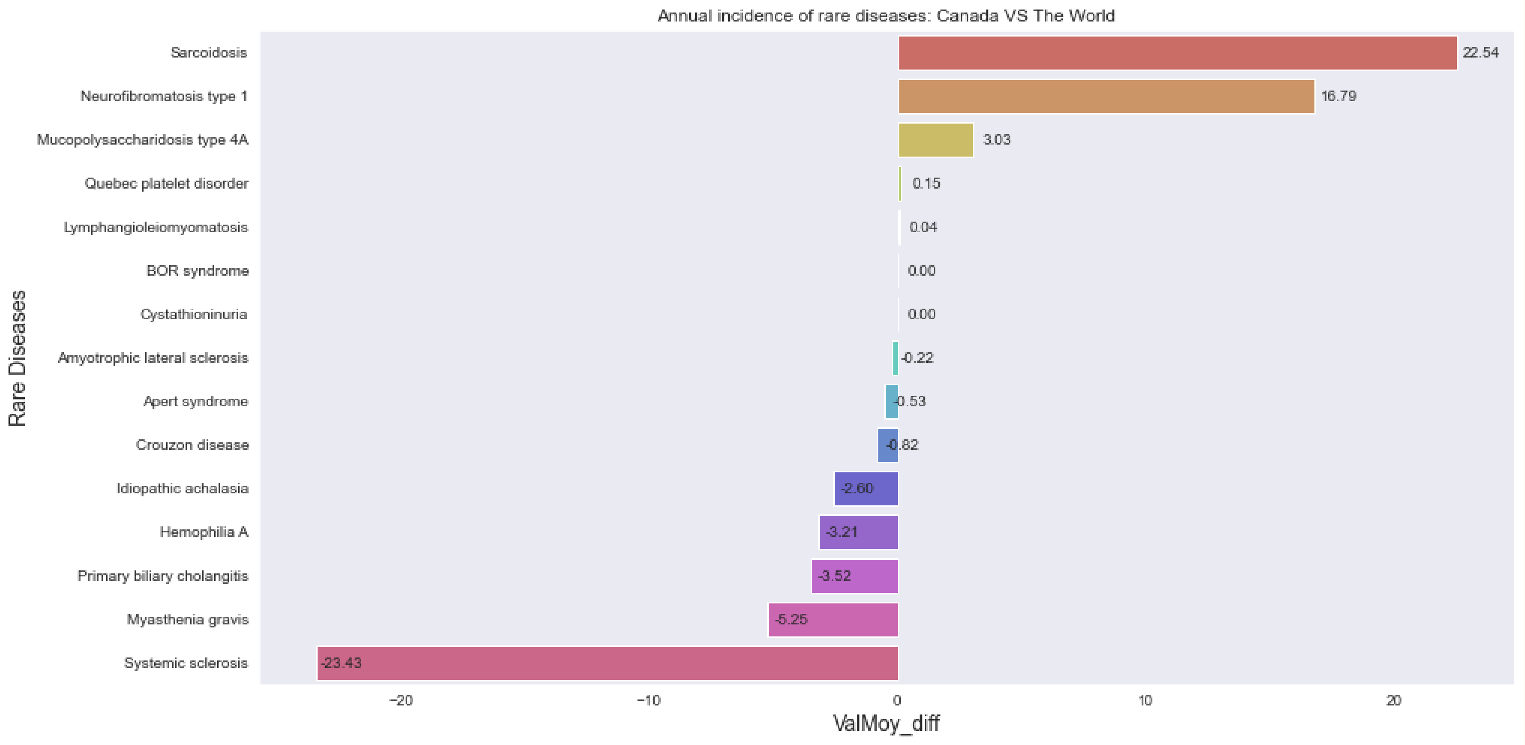
1. Point Prevalence: Number of cases scaled up to the general population at a given time;
2. Annual Incidence: is a measure of the probability of occurrence of a given medical condition in a population within a year;
3. Prevalence at birth: Number of cases observed at birth relative to the number of children born alive at a given moment;

With our implementation, Selecting a country and the type of prevalence, it is possible to explore the "spread" of rare disease compared to the rest of the world.

In this case, selecting Canada and "Point Prevalence", we can compare the Point prevalence of the Canadian rare disease with the rest of the world.
The Column "ValMoy_diff" contains the difference between the world average prevalence of a particular disease less the canadian value. Positive values mean that Canada has a higher prevalence, while negative values that Canada has less prevalence. If the difference value is close to 0, than the value of world average and country is very similar.

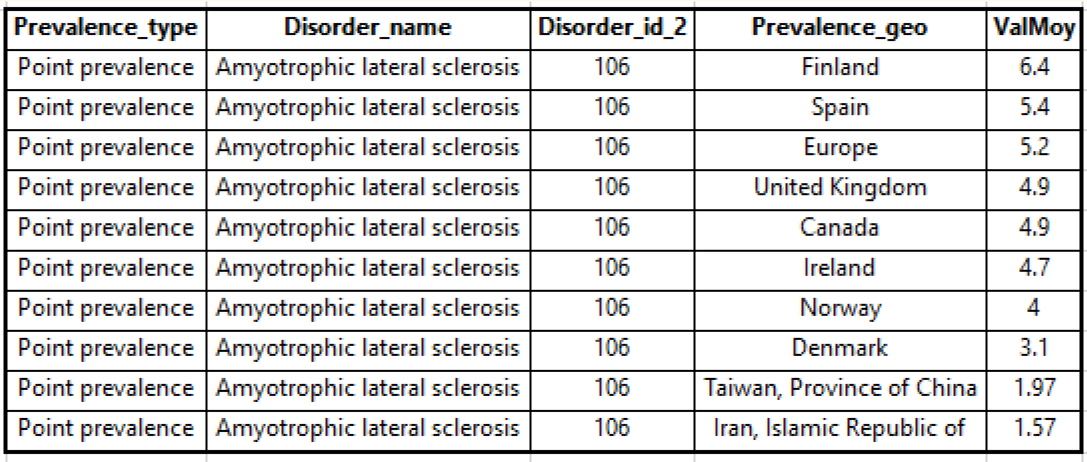


Prevalence_type	Disorder_name	Disorder_id	ValMoy_world_avg	ValMoy_Country	ValMoy_diff
Point prevalence	Sarcoidosis	577	33.04125	10.5	22.54125
Point prevalence	Neurofibromatosis type 1	143	23.493333	6.7	16.793333
Point prevalence	Mucopolysaccharidosis type 4A	3479	3.1766	0.15	3.0266
Point prevalence	Quebec platelet disorder	2652	0.3	0.15	0.15
Point prevalence	Lymphangioleiomyomatosis	2673	0.254444	0.21	0.044444
Point prevalence	BOR syndrome	178	2.5	2.5	0
Point prevalence	Cystathioninuria	2666	7.1	7.1	0
Point prevalence	Amyotrophic lateral sclerosis	93	4.68	4.9	-0.22
Point prevalence	Apert syndrome	213	0.9425	1.47	-0.5275
Point prevalence	Crouzon disease	191	0.825	1.65	-0.825
Point prevalence	Idiopathic achalasia	240	8.222	10.82	-2.598
Point prevalence	Hemophilia A	5758	4.086343	7.3	-3.213657
Point prevalence	Primary biliary cholangitis	674	19.178571	22.7	-3.521429
Point prevalence	Myasthenia gravis	519	14.746154	20	-5.253846
Point prevalence	Systemic sclerosis	5008	20.87	44.3	-23.43



Looking at the single disease (ALS)

Selecting a particular disease, we are able to see the spread among the countries.



Prevalence_type	Disorder_name	Disorder_id_2	Prevalence_geo	ValMoy
Point prevalence	Amyotrophic lateral sclerosis	106	Finland	6.4
Point prevalence	Amyotrophic lateral sclerosis	106	Spain	5.4
Point prevalence	Amyotrophic lateral sclerosis	106	Europe	5.2
Point prevalence	Amyotrophic lateral sclerosis	106	United Kingdom	4.9
Point prevalence	Amyotrophic lateral sclerosis	106	Canada	4.9
Point prevalence	Amyotrophic lateral sclerosis	106	Ireland	4.7
Point prevalence	Amyotrophic lateral sclerosis	106	Norway	4
Point prevalence	Amyotrophic lateral sclerosis	106	Denmark	3.1
Point prevalence	Amyotrophic lateral sclerosis	106	Taiwan, Province of China	1.97
Point prevalence	Amyotrophic lateral sclerosis	106	Iran, Islamic Republic of	1.57

In this case, we can notice that Finland and Spain and Europe have the highest value, so we can conclude that they have more cases.
Thanks to this, researchers can see where certain diseases are more spread and can give a “hint” for future researches.

Implementations for the Healthcare system

This data can help Healthcare systems in making faster diagnosys and researchers to retrieve useful information during their job.

For example, looking at the information related to the genes, it is possible to implement a procedure that starts from testing the genes and the genes loci associated with the highest number of rare diseases.
For example, starting with analysing the genes COL2A1, HBB, LMNA and KIT the researcher can check 61 rare diseases, and looking at the loci Xq28, 16p13.3, and 3p21.31 it is possible to check 131 rare diseases.

Or if a doctor suspects a patient may have a particular rare disease, he can search for the HPOs related to the disease and check if the patient has them or not.

Thanks to these tools, it is possible to speed up the diagnosis process.
Making faster diagnosis has two main advantages:

- Doctors can spend less time on the diagnosis and more on the treatment;
- Testing the genes starting from the ones involved in the higher number of rare diseases can help to save money, since each gene test has a cost of around 300 USD.

In addition, Researchers can use these Apps to retrieve geographical information with the prevalence app and check the spread of a rare disease around the world, and use it as basys to make hypotheses on why certain rare disease are less likely in some countries and more in others.