**Detail of the scripts**

**Number of groups of disorders, disorders and subtypes of disorders**

We open the xml called "product1\_en\_cross\_jdbor\_evo", and select only the active clinical entities. They include the Head of classification (DisorderFlag=128), the Historical entities (DisorderFlag=512), and the On-line (DisorderFlag=1).

For the groups of disorders, we look for the label ‘’Group of disorders” in the Name of DisorderGroup, into the active clinical entities. We put them in a list, and then we count the number of elements in the list, that corresponds to the number of groups of disorders.

Concerning the disorders and the subtypes of disorders, we do exactly the same, respectively with the labels “Disorder” and “Subtype of disorder”.

**Number of disorders uniclassified vs multiclassified (excluding genetic disorders)**

We open all the xml which the name begins by “product3”. We only delete the 'product3\_156\_en\_jdbor\_evo.xml' because it is about genetic informations and we don’t focus on it in this analysis. We get 34 xml, and each one corresponds to a class of disorder (Rare cardiac disease, rare cardiomyopathy, etc…).

From these xml, we select the node “Disorder” and we put all the informations about it into a frame. We get back a list with 34 classes, each class is represented by a frame with the OrphaCodes and the DisorderTypes corresponding to the disorders of the class. Some disorders and so OrphaCodes are duplicated into the same class, we remove them. Then, we only keep the informations about the disorders (excluding groups and subtypes of disorder), we filter the DisorderType and retain “Malformation syndrome”, “Disease”, “Morphological anomaly”, “Clinical syndrome”, “'Particular clinical situation in a disease or syndrome”, and “Biological anomaly”. Finally, we count how many times an OrphaCode is present in all the 34 classes. Now we just have to count how many disorders are present once in the 34 classes; they are uniclassified. And if an OrphaCode is present at least twice in the 34 classes, the disorder is multiclassified.

**Number of active genetic disorders**

We open the xml "product1\_en\_cross\_jdbor\_evo", and select the node “DisorderList” that contains all the disorders. Between these disorders, we get the active clinical entities (Head of classification, Historical entities and On-line) by filtering the DisorderFlag (respectively equal to 128, 512 and 1). Then, we only keep the informations about the disorders (excluding groups and subtypes of disorder), we filter the DisorderGroup and retain “Disorder”. We just have to count the number of disorders that we get. We will need this number later.

Now we open the xml 'product3\_156\_en\_jdbor\_evo.xml' that contains the genetic datas. We select the node “Disorder” that have all the disorders with their OrphaCodes and DisorderType. We only keep the informations about the disorders (excluding groups and subtypes of disorder), so we filter the DisorderType and retain “Malformation syndrome”, “Disease”, “Morphological anomaly”, “Clinical syndrome”, “'Particular clinical situation in a disease or syndrome”, and “Biological anomaly”. Then we just have to count the number of disorders that are present to obtain the number of genetic disorders. To calculate the number of non-genetic disorders, we do the number of disorders calculated previously minus the number of genetic disorders.

**Distribution by preferential parent**

We open the xml "product1\_en\_cross\_jdbor\_evo", and select the node “DisorderList” that contains all the disorders. Between these disorders, we get the active clinical entities (Head of classification, Historical entities and On-line) by filtering the DisorderFlag (respectively equal to 128, 512 and 1). Then, we only keep the informations about the disorders (excluding groups and subtypes of disorder), we filter the DisorderGroup and retain “Disorder”. We get the OrphaCodes of each active clinical entities which is a disorder. We will need it later.

We just have to count the number of disorders that we get. We will need this number later.

We open the xml ‘product7\_linear\_en\_jdbor\_evo.xml’ that contains the datas about preferential parent, and we select the node “DisorderList”, in which there are all the disorders and their corresponding informations. In these informations, if the element “TargetDisorder” is present, it means that there is a preferential parent. So we keep only the OrphaCode and Preferential Parent of the disorders by filtering with the presence of “TargetDisorder”.

Then, we merge by the OrphaCode the frames that we got with the product1 and product3 to obtain all the active clinical entities which are a disorder and have a preferential parent. And finally, we count the number of OrphaCodes (so of disorders) by preferential parent to have the distribution of preferential parent.

**Number of disorders aligned**

We open the xml "product1\_en\_cross\_jdbor\_evo", and select the node “DisorderList” that contains all the disorders. Between these disorders, we get the active clinical entities (Head of classification, Historical entities and On-line) by filtering the DisorderFlag (respectively equal to 128, 512 and 1). Then, we only keep the informations about the disorders (excluding groups and subtypes of disorder), we filter the DisorderGroup and retain “Disorder”. We just have to count the number of disorders that we get. We will need this number later.

To have all the disorders that are aligned at least once, we take a look at the node “ExternalReferenceList” and if the count “.attrs” is superior to zero, it means that there is at least an alignment. So we keep these disorders, they represent the disorders with an alignment.

To get the disorders with at least one OMIM or ICD-10, we still take a look at the node “ExternalReferenceList”, and if the “Source” is equal to “OMIM” or “ICD-10”, we keep the disorder (by keeping its OrphaCode). We need to remove duplicates because some disorders can have several alignments with the same source. Then we just have to count how many disorders we have. Moreover, to calculate the percentage of disorders aligned with at least one OMIM or ICD-10, we just divide the number of disorders with at least one OMIM or ICD-10 by the number of disorders (previously calculated), and multiply by 100.

We do the same for the disorders with at least one OMIM, and those with at least one ICD-10.

To have the disorders aligned with at least one OMIM by an exact relationship, we still take a look at the node “ExternalReferenceList” for all the disorders, and two conditions must be filled: the “Source” is equal to “OMIM”; and the Name of the DisorderMappingRelation is equal to “E (Exact mapping: the two concepts are equivalent)”. Then we keep the disorder (by keeping its OrphaCode). We need to remove duplicates because some disorders can have several exact alignments with the same source. Then we just have to count how many disorders we have. Moreover, to calculate the percentage of disorders aligned with at least one OMIM by an exact relationship, we just divide the number of disorders with at least one OMIM by an exact relationship by the number of disorders (previously calculated), and multiply by 100.

We do exactly the same for the number and the percentage of disorders aligned with at least one ICD-10 by an exact relationship.

**Distribution by medical speciality**

We open all the 35 xml which are called “product3”. Each product3 corresponds to a medical speciality. From these xml, we select the node “Disorder” and we put all the informations about it into a frame. For all the disorders, we paste the name of the xml where the disorder is from (corresponding to the medical speciality). Every disorder is mainly represented in a frame by its Name, OrphaCode and Medical Speciality. Some disorders and so OrphaCodes are duplicated into the same class, we remove them. Then, we only keep the informations about the disorders (excluding groups and subtypes of disorder), we filter the DisorderType and retain “Malformation syndrome”, “Disease”, “Morphological anomaly”, “Clinical syndrome”, “'Particular clinical situation in a disease or syndrome”, and “Biological anomaly”. Finally, we count how many disorders (so OrphaCodes) are present in each medical speciality.

**Number of definitions**

* In English

We open the xml "product1\_en\_cross\_jdbor\_evo", and select the node “DisorderList” that contains all the disorders. Between these disorders, we get the active clinical entities (Head of classification, Historical entities and On-line) by filtering the DisorderFlag (respectively equal to 128, 512 and 1). Then, we only keep the informations about the disorders (excluding groups and subtypes of disorder), we filter the DisorderGroup and retain “Disorder”. Then, we look for the disorders with a definition. If there a Name in TextSection of the SummaryInformation, it means that there is a definition. We just have to count the number of disorders that fill this condition. To have the percentage of disorders with a definition, we divide the number of disorders with a definition by the number of disorders and we multiply by 100.

* In an other language

We open the xml “product1” corresponding to the language that we want. Then it is exactly the same process than in English, except for the step excluding the groups and subtypes of disorders. Here we look if the active clinical entities in the language of our choice are present in the disorders in English. If yes, we select them as disorders, and pursue the same steps than for the definitions in English.