Methodology for Generating Datasets with Characteristic Diagnostic Parameters of Rare Diseases Using the Example of Pompe Disease, Gaucher Disease and Smith-Lemli-Opitz Syndrome

Ann-Christin Liebers-Kyungay

Otto-von-Guericke-University Magdeburg: Otto von Guericke Universität Magdeburg

Klaus Mohnike (✉ klaus.mohnike@med.ovgu.de)

Otto von Guericke Universität Magdeburg <https://orcid.org/0000-0001-5113-8657>

Corine van Lingen

Azienda sanitaria universitaria Friuli Centrale

Anita Bressan

Azienda sanitaria universitaria Friuli Centrale

Katja Palm

Otto-von-Guericke-University Magdeburg: Otto von Guericke Universität Magdeburg

Athanasia Ziagaki

Charite University Hospital Berlin: Charite Universitätsmedizin Berlin

Research

Keywords: diagnosis of rare diseases, HPO-Terms, late-onset Pompe disease, Gaucher disease Type I, Smith-Lemli-Opitz syndrome

Posted Date: November 19th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1077021/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

Finding a diagnosis for rare diseases is a challenge for patients and those treating them. Establishing a uniform methodology for specifying the symptoms of a patient seems useful. This, as well as a database with clinical parameters reported in patients already diagnosed with the corresponding disease or that have led to the diagnosis, would facilitate the global data exchange between specialists and subsequently diagnosis. The aim of this work is to develop standardized data sets with the most frequent symptoms exemplarily for the three rare diseases late-onset Pompe disease, Gaucher disease Type I and Smith-Lemli-Opitz syndrome (SLOS).

Methods and results

A systematic literature review of characteristic symptoms and diagnostic criteria was performed for each of the three disorders. These parameters were converted into vocabulary standardized by *The Human Phenotype Ontology* (HPO), so-called HPO terms. Subsequently, a retrospective analysis of the patient files of 23 late-onset Pompe disease patients, 21 Gaucher disease Type I patients and 25 SLOS patients was carried out together with the University Children's Hospital Magdeburg and the Center of excellence for Rare Metabolic Diseases at the Charité Berlin. Features present in $\geq 40\%$ of the cohort and collected simultaneously in a certain minimum number of patients were filtered out. The analysis resulted in data sets with 20 diagnostic parameters for late-onset Pompe disease, 16 features for Gaucher disease Type I and 17 parameters for SLOS. After the statistical evaluation, the results were discussed comparatively with similar studies exemplarily for SLOS.

Conclusion

The developed datasets for the three diseases provide a good basis for expansion with further patient examples and for extending the methodology to other diseases with the aim of improving the diagnostic pathway and thus the health care of patients with rare diseases.

Background

The European Union defines a disease as rare if no more than 5 out of 10,000 people are affected by this disease. Of the approximately 30,000 known diseases, about 8,000 are counted as rare diseases. In Germany, about 4 million people live with a rare disease, Europe-wide about 30 million. Due to their rarity, the knowledge of these diseases among physicians in normal care is often rather low and the number of corresponding experts is low. Diagnosis is often delayed and care can also be difficult, so that affected patients often feel left alone.

Publications and case reports exist on most diseases. However, the varying availability of corresponding studies and the frequent lack of guidelines for the individual diseases pose a challenge for those treating patients who are confronted with these diseases. At the same time, variable and diverse descriptions of the symptomatology of diseases in publications make it difficult to search for a specific disease based on the symptoms of those affected with an unclear, possibly rare disease.

With the establishment of self-help groups, the promotion of drug development, the creation of information portals such as ZIPSE (Central Information Portal on Rare Diseases) or se-atlas (Care Atlas for People with Rare Diseases) or the founding of the National Action Alliance for People with Rare Diseases (NAMSE), a lot is already being done in Germany to make things easier for all those involved. At the European level, the European Reference Networks (ERNs)

are intended to improve prevention and diagnosis and to guarantee high-quality healthcare for patients with rare diseases throughout Europe. The Clinical Patient Management System (CPMS) designed for the ERNs is intended to facilitate cross border diagnostic and therapeutic consultations and the safe exchange of patient data through the cooperation of specialists across European borders in compliance with data protection regulations.

In order to facilitate this exchange, it seems useful to establish a uniform methodology for specifying the symptoms of a patient. Furthermore, a type of database with clinical parameters that are reported in patients already diagnosed with the corresponding disease or that have led to the diagnosis would facilitate the exchange of data and the subsequent diagnosis.

Within the scope of this work, data sets with the most frequent symptoms and diagnostic criteria for three exemplarily selected rare diseases are to be developed on the basis of a literature analysis as well as the examination of patients of two reference centers. For these data sets, a standardized word form is to be chosen that enables European or even worldwide exchange and can also be used for the CPMS.

The three genetically determined rare metabolic diseases selected as examples are late-onset Pompe disease, Gaucher disease Type I and Smith-Lemli-Opitz syndrome (SLOS).

Late-onset Pompe disease, a lysosomal storage disease, results from a genetic deficiency of the enzyme acid alpha-1,4-glucosidase, which hydrolyzes glycogen to glucose. The resulting accumulation of glycogen in the lysosomes due to the enzyme defect leads to damage of glycogen-storing organs, such as muscles. The prognosis can be improved by enzyme replacement therapy, also depending on the severity of the enzyme deficiency.

Gaucher disease Type I, another lysosomal metabolic disorder, is caused by a deficiency of the lysosomal enzyme beta-glucocerebrosidase, due to a defect in the GBA gene. Glucocerebroside, an unsplit metabolic product of glycosphingolipids, accumulates and forms Gaucher cells. In Type I patients hepatosplenomegaly, bone involvement and hematologic system changes may be present. Type I can be treated by enzyme replacement therapy.

SLOS is caused by a genetic defect in cholesterol biosynthesis, so that 7-dehydrocholesterol cannot be converted into cholesterol due to a lack of 7-dehydrocholesterol reductase. The disease is characterized by numerous craniofacial dysmorphic signs, 2-3 toe syndactyly, psychomotor retardation and organ malformations. Symptomatic treatment with cholesterol supplements and/or the HMG-CoA-reductase inhibitor simvastatin is performed.

For all three diseases, there are already published studies of the symptomatology and diagnostic parameters of differently sized patient cohorts. However, the literature research carried out shows how difficult it is to bring symptom descriptions from different studies to a common denominator, to summarize them or even to derive diagnoses for patients who have not yet been diagnosed, due to the different focal points and vocabulary used. Compressed and standardized data sets should contribute to facilitating the path to diagnosis and the exchange with specialists for those treating and affected by the disease.

Methods

In cooperation with the University Children's Hospital Magdeburg and the Center of excellence for Rare Metabolic Diseases at the Charité Berlin, three hereditary metabolic diseases were selected for the development of the data sets: Late-onset Pompe disease, Gaucher disease Type I and SLOS. For these diseases, both centers have their own experience and larger patient collectives.

First, a systematic literature analysis was performed with regard to characteristic symptoms and diagnostic criteria of the three diseases mentioned. For this purpose, the book Vademecum metabolicum [1], specialist portals such as Orphanet and OMIM (Online Mendelian Inheritance in Men) as well as numerous other publications were used. The result was a table with 67 diagnostic parameters for Pompe disease. In addition to a literature analysis for late-onset Pompe disease, the parameters for the infantile-onset form of the disease were also researched. For Gaucher's disease, a table of 63 symptoms and diagnostic criteria was obtained, again differentiating between Gaucher types I, II and III. The literature search for SLOS syndrome yielded 84 symptoms and diagnostic criteria. If a classification of the frequency of occurrence of the diagnostic parameter was found in the literature, it was adopted in the table created. There is a subdivision of frequency into *very frequent*, *frequent*, *occasional* and *rare*, whereby mainly *very frequent*, *frequent* and partly *occasional* occurring features were adopted in the table. However, since these frequency data are not available for all parameters, a complete ordering of the parameters according to their frequency as reported in the literature is not possible.

Subsequently, these terms, formulated freely or based on the sources, were converted into vocabulary standardized by *The Human Phenotype Ontology* (HPO), so-called HPO terms. HPO is an online database of standardized vocabulary of phenotypic abnormalities of human diseases [2]. This standardization facilitates the global exchange of disease data [2].

In the third step, medical records of patients with confirmed diagnosis of the respective diseases were analyzed according to these criteria. A retrospective analysis of the medical records of 23 patients with late-onset Pompe disease, 21 patients with Gaucher disease Type I and 25 patients with SLOS was performed. In the tables created for the literature analysis, it was marked which of the parameters occur in which patient. A subdivision was made into *characteristic present*, *characteristic not present* or *no specification*. In this context, the table was expanded to include characteristics that were not included in the table during the literature research but did occur in patients.

This resulted in percentage frequencies with corresponding confidence intervals for the occurrence of a characteristic within the patient group. There were two variants of the calculation: in the first variant, the percentages were calculated for the entire patient cohort of a disease for each characteristic, including the patients with missing information. In the second variant, the patients without information for this characteristic were excluded. Accordingly, the percentage frequencies of occurrence here are partly related to smaller numbers of patients. With regard to the latter variant, characteristics were filtered out that were detected and documented in $\geq 40\%$ of the patients assessed according to this characteristic and collected simultaneously in a certain minimum number of patients.

Results

The evaluation of the frequency of occurrence of the diagnostic parameters in the examined patients following the literature analysis was carried out separately for each of the three diseases. The results of the examinations of the patient cohorts for the individual diseases are presented below.

Pompe disease, late onset form

A total of 23 patients were included in the analysis, 10 female and 13 male. All patients included in the analysis were diagnosed with late-onset Pompe disease (LOPD), either due to an abnormal sequence analysis of the GGA gene (documented in 21/23 patients) and/or the measurement of reduced activity of lysosomal alpha-1,4-glucosidase (in 15/23 patients). Age at diagnosis ranged from 12 to 73 years, with a mean of 39.5 years. 7/23 (30%) patients had a

positive family history of LOPD. Table 1 presents the 20 studied characteristics that were present in ≥ 8 patients and reached percent frequencies $\geq 40\%$. Twelve of these features had frequencies of occurrence $\geq 60\%$.

Table 1
Presentation of the frequency of occurrence of diagnostic parameters in Pompe disease patients

Feature	HPO-Term	No. of patients	Percentage
Musculoskeletal system			
Reduced bone mineral density	HP:0004349	14/23	60.9
Hyperlordosis	HP:0003307	11/23	47.8
Back pain	HP:0003418	11/23	47.8
Skeletal muscle atrophy	HP:0003202	19/23	82.6
Hypotonia	HP:0001252	11/23	47.8
Progressive muscle weakness	HP:0003323	21/23	91.3
Pelvic girdle muscle weakness	HP:0003749	13/23	56.5
Scapular winging	HP:0003691	10/23	43.5
Abnormality of movement	HP:0100022	17/23	73.9
Respiratory system			
Respiratory insufficiency	HP:0002093	15/23	65.2
Sleep disturbance	HP:0002360	13/23	56.5
Nervous system			
Tetraparesis	HP:0002273	11/23	47.8
Reduced tendon reflexes	HP:0001315	11/23	47.8
Laboratory tests			
Glycogen accumulation in muscle fiber lysosomes	HP:0030231	11/13	84.6
Elevated serum creatine kinase concentration	HP:0003236	22/22	100
Elevated circulating alanine aminotransferase concentration	HP:0031964	23/23	100
Elevated circulating aspartate aminotransferase concentration	HP:0031956	23/23	100
Increased circulating lactate dehydrogenase concentration	HP:0025435	16/21	76.2
Confirmation of diagnosis			
Abnormal enzyme/coenzyme activity (acid α -Glucosidase)	HP:0012379	15/15	100
Abnormal gene sequence analysis of GAA		21/21	100

Gaucher disease Type I

The patient records of 21 patients diagnosed with Gaucher disease Type I were analyzed, 11 female and 10 male. The age at diagnosis ranged from 7 to 81 years, with a mean of 34.2 years. A positive family history of Gaucher disease was known in 3 patients. Sixteen features occurred with frequencies $\geq 40\%$ and were simultaneously present in ≥ 7 patients (Table 2). Ten of these features had percent frequencies of at least 60%.

Table 2
Frequency of occurrence of diagnostic parameters in Gaucher disease patients

Feature	HPO Term	No. of patients	Percentage
Musculoskeletal system			
Abnormality of the skeletal system	HP:0000924	14/17	82.4
Bone pain	HP:0002653	11/19	57.9
Avascular necrosis	HP:0010885	8/17	47.1
Herniation of intervertebral nuclei	HP:0008441	8/17	47.1
Abdominal organs			
Hepatomegaly	HP:0002240	18/21	85.7
Splenomegaly	HP:0001744	18/20	90.0
Blood			
Anemia	HP:0001903	12/19	63.2
Thrombocytopenia	HP:0001873	16/19	84.2
Cardiovascular system			
Hypertension	HP:0000822	7/17	41.2
Constitution			
Fatigue	HP:0012378	12/19	63.2
Laboratory tests			
Elevated serum acid phosphatase	HP:0003148	7/17	41.2
Increased circulating ferritin concentration	HP:0003281	10/19	52.6
Elevated CSF chitotriosidase 1 level	HP:0032652	14/19	73.7
Confirmation of diagnosis			
Abnormal enzyme/coenzyme activity (reduced activity of beta-glucosidase)	HP:0012379	13/13	100
Gauchercells in liver biopsy or bone marrow biopsy		10/10	100
Characteristic sequence analysis of genes		15/15	100

Smith-Lemli-Opitz Syndrome

Twenty-five cases of patients with SLOS were analyzed. Of these, 9 patients were female and 16 patients were male. The diagnosis was made on the basis of an elevated 7-dehydrocholesterol (7DHC) level in plasma or, in one case, amniotic fluid, or on the basis of genetic analysis. The mean age at diagnosis was 12.8 months of life with a range from the 9th month of gestation (equivalent in calculation to 0 months of life) to 74 months of life. Nine of the examined patients had a known positive family history. Table 3 shows that 17 features occur with a frequency of at least 40% and were simultaneously present in ≥ 9 patients. An exception is the feature *Cryptorchidism*, which, because it can only be applied to the 16 male patients, must only be present in at least 6 patients. Eight of the 17 characteristics occur in at least 60% of the cases.

Table 3
Frequency of occurrence of diagnostic parameters in patients with Smith-Lemli-Opitz syndrome

Feature	HPO-Term	No. of patients	Percentage
Facial asymmetry			
Micrognathia	HP:0000347	11/21	52.4
Wide nasal bridge	HP:0000431	11/21	52.4
Anteverted nares	HP:0000463	13/21	61.9
Low-set ears	HP:0000369	13/23	56.5
Eye			
Ptosis	HP:0000508	11/24	45.8
Head			
Microcephaly	HP:0000252	18/25	72.0
Musculoskeletal system			
2-3 toe syndactyly	HP:0004691	23/25	92.0
Hypotonia	HP:0001252	12/24	50.0
Feeding difficulties	HP:0011968	18/25	72.0
Genitourinary system			
Cryptorchidism	HP:0000028	8/16	50.0
Development			
Failure to thrive	HP:0001508	10/24	41.7
Global developmental delay	HP:0001263	16/24	66.7
Sleep			
Sleep disturbance	HP:0002360	12/25	48.0
Immune system			
Recurrent infections	HP:0002719	17/24	70.8
Laboratory tests			
Elevated 7-dehydrocholesterol	HP:0010569	24/25	96.0
Elevated 8-dehydrocholesterol	HP:0003462	24/25	96.0
Hypocholesterolemia	HP:0003146	13/24	54.2

Discussion

The standardized English-language vocabulary developed by the Human Phenotype Ontology includes more than 13,000 terms to describe signs, symptoms, or phenotypic manifestations that characterize specific diseases. Despite this comprehensive registry, converting the terms used by clinicians for documentation into HPO terms proved challenging in some cases. Different vocabulary was used for the same symptoms, or there were minimal differences or variations in the expression or presence of symptoms, making it difficult to assign them to a specific term. One possibility here is the categorization and hierarchical ordering of HPO terms from a generalized abnormality with various subgroups to increasingly detailed characterizations. Thus, in some cases, assignment to a somewhat more general HPO term was necessary. Furthermore, since HPO provides only phenotypic characteristics, this also explains why some very common traits were not standardized by an HPO term. In the future, due to the continuous expansion and updating of the HPO database, a more precise characterization of the symptomatology will be possible. However, this shows that the use of a standardized vocabulary is advantageous for the documentation and exchange of the symptomatology of certain diseases.

Another methodological challenge was the retrospective nature of the study. It is true that both reference centers involved care for their patients according to a detailed care scheme specific to each disease, so that certain diagnoses and examinations are always carried out. However, there are patients who came from other centers or from other countries, so that in some cases only the diagnosis and no or incomplete findings at diagnosis were available. For this reason, and also due to different types of documentation, no information was available for some parameters that were established in the literature analysis. In some cases, it was also not possible to determine whether the characteristic had actually not been investigated or whether it had been investigated but was not present and therefore not documented. Accordingly, the evaluation of some characteristics was only possible on the basis of a smaller group of patients than the original number, since missing data had to be subtracted.

If, as planned, the data sets are expanded by other physicians or within the framework of studies with further patient examples, this will increase their informative value. At the same time, biases that arise due to differences in the focus of different physicians in the study will be compensated for. It should also be considered to what extent an additional temporal classification of the occurrence of the characteristics can be documented. It would then be possible to distinguish between initial symptoms and late symptoms. This precision of the information could further facilitate the diagnostic pathway. The retrospective study carried out here proved to be unsuitable for this purpose because the documentation did not always show when and which symptoms appeared for the first time. Possibly, interviews with the patients in addition to the analysis of the files could provide more concrete information.

Furthermore, a comparative analysis of the results obtained with the existing literature is necessary. This will be done in this rather methodical work using the SLOS as an example.

Smith-Lemli-Opitz syndrome

Typical features of SLOS that also contribute to the diagnosis include craniofacial malformations, psychomotor retardation, malformations of organs and extremities, failure to thrive and developmental disorders. Both in the literature and in the current study, these features are described, but with different emphases and percentages. Table 4 presents the results of the analysis of SLOS patients and 6 other studies in a comparative manner.

Table 4

Comparative presentation of the frequencies of occurrence of diagnostic parameters in Smith-Lemli-Opitz patients

Feature	Present study	Kelley and Hennekam, 2000 [8]	Donoghue et al, 2018 [7]	Ryan et al, 1998 [6]	Qu��lin et al, 2011 [5]	Balogh et al, 2012 [4]	Smith et al, 1964 [11]
Number of patients	25	167	14	49	10 (Fetus)	13	3
Country	Germany, Magdeburg	USA, The Netherlands	Australia	UK	France	Hungary	USA
Study design	Retro	Survey	Retro	Survey	Retro	Survey	Case reports
Sex, F/M	9/16	167	7/7	14/35	5/5	7/6	0/3
Clinical features n (%)							
Micrognathia	11/21 (52,4)	N.A.	7 (50,0)	33 (67,4)	9 (90,0)	6 (46,2)	3 (100)
Wide nasal bridge	11/21 (52,4)	N.A.	7 (50,0)	34 (69,4)	N.A.	4 (30,8)	3 (100)
Anteverted nares	13/21 (61,9)	90%	5 (35,7)	34 (69,4)	9 (90,0)	3 (23,1)	3 (100)
Low-set ears	13/23 (56,5)	N.A.	6 (42,9)	23 (46,9)	N.A.	4 (30,8)	N.A.
Ptosis	11/24 (45,8)	69%	3 (21,4)	26/44 (59,1)	N.A.	N.A.	2 (66,7)
Microcephaly	18 (72,0)	80%	N.A.	32/40 (80,0)	9 (90,0)	6 (46,2)	3 (100)
2-3 toe syndactyly	23 (92,0)	85%	13 (92,9)	43/48 (89,6)	10 (100,0)	12 (92,3)	2 (66,7)
Cryptorchidism	8/16 (50,0)	N.A.	4/6 (66,7)	6/11 (54,6)	N.A.	2/6 (33,3)	2 (66,7)
Hypotonia	12/24 (50,0)	N.A.	N.A.	N.A.	N.A.	4 (30,8)	Hypertonicity (3/3)
Feeding difficulties	18 (72,0)	N.A.	N.A.	43 (87,76) (required nasogastric tube feeding)	N.A.	4 (30,8)	3 (100)
Failure to thrive	10/24 (41,7)	N.A.	N.A.	N.A.	N.A.	N.A.	3 (100)
Global developmental delay	16/24 (66,7)	N.A.	N.A.	N.A.	N.A.	N.A.	3 (100)
Recurrent infections	17/24 (70,8)	N.A.	2 (14,3)	10/49 (20,4)	N.A.	N.A.	2 (33,3)

Feature	Present study	Kelley and Hennekam, 2000 [8]	Donoghue et al, 2018 [7]	Ryan et al, 1998 [6]	Quélin et al, 2011 [5]	Balogh et al, 2012 [4]	Smith et al, 1964 [11]
Sleep disturbance	12 (48,0)	N.A.	N.A.	16/23 (69,6)	N.A.	N.A.	N.A.

Craniofacial features

Nineteen of the 21 craniofacial features found in the literature research were also documented in the case analysis of patients. Six of these met the stated criteria (positive in ≥ 9 patients and percentage frequency of $\geq 40\%$):

Microcephaly (18/25; 72.0%), *Micrognathia* (11/21; 52.4%), *Wide nasal bridge* (11/21; 52.4%), *Anteverted nares* (13/21; 61.9%), *Low-set ears* (13/23; 56.5%), and *Ptosis* (11/24; 45.8%). All these features are also described in other SLOS patient cohorts, mostly with similar percentages [3–8]. Due to their frequent occurrence, even in larger cohorts, these six characteristics mentioned can be counted among the important diagnosis-indicating symptoms. *Ptosis* cannot be used as a feature in fetuses because it is not yet observable. In addition, other craniofacial features are typical. These are described again and again, but occur in the cohort examined here and partly also in other studies with lower or very different frequencies: *Cleft palate* (present study: 28.6%, Donoghue et al. : 57.1% [7], Kelley and Hennekam: 51% [8]), *Bifid uvula* (present study: 15.0%; Donoghue et al.: 21.4% [7]), *Retrognathia* (present study: 25.0%; Quélin et al.: Microrethrogathia 90.0% [5]), *Short neck* (present study: 5.0%; Donoghue et al.: 14.3% [7], Quélin et al.: 100% [5]), *Epicanthus* (present study: 35.0%, Cherstvoy et al.: 62.5% [9]), *Hypertelorism* (present study: 10%; Quélin et al. : 70% [5]).

Other typical features, on the other hand, were hardly or not at all documented in the patients studied here. These include, for example, the occurrence of a *Cataract* in only one case (1/24; 4.2%) compared to other studies in which percentage frequencies of 28.5% [7] or 12% [6] noted. *Bitemporal narrowing* is reported in the cohorts of Donoghue et al. [7] and Quélin et al. [5] with frequencies of 14.3% and 50% [5]. In contrast, *Biparietal narrowing* is documented in only one patient in the cohort studied.

These very different frequencies could be explained on the one hand by the fact that an assessment of the facies was probably rarely carried out using a standardized list of features. Often, the features that catch the examiner's eye are documented, but no explicit search is made for specific dysmorphic features. On the other hand, however, there is a very broad spectrum of possible craniofacial abnormalities that vary greatly in severity. This is also reflected in the distinction made by some between Type I with milder manifestations and Type II with more severe manifestations. With regard to the creation of data sets to facilitate diagnosis, this observation argues in favor of concentrating on the frequently occurring features.

Organs and extremities

Various organ malformations can be detected in SLOS patients. In the cohort studied, *2-3 toe syndactyly* is one of the most frequent symptoms, occurring in 92.0%. The constancy of this symptom for SLOS is confirmed by other studies [3–8]. Other abnormalities of the extremities, such as *Postaxial polydactyly*, described by Quélin et al. in 40% of cases [5] and by Ryan et al. in 49% of cases [6], were documented in only one patient (1/24; 4.2%) in the currently studied cohort. Similarly, the feature *Proximal placement of thumb* was noted for only 33.3% (8/24) of patients in the cohort

studied, compared with 49% of patients in Ryan et al. [6] and 70% of patients in Quélin et al. [5]. Other abnormalities of the hands, such as different *Abnormalities of palmar creases* are described: *Extended distal palmar crease* (71.4%) [5], *Horizontal upper palmar creases* (100%) [3], *Single transverse palmar crease* (28.6%) [7] and in the current study also *Single transverse palmar crease* with a frequency of 25%.

In addition, genital anomalies are repeatedly described. In the current study, the feature *Cryptorchidism* is one of the most frequent features, being present in 50% of the patient cohort (in 8/16 male patients). Other parameters from the genital system do not occur in more than 40% of the patients studied: *Ambiguous genitalia* (8.3%), *Hypoplasia of penis* (6.25%), *Hypospadias* (37.5%). Ryan et al. (5) report the incidence of genital anomalies as 91% (32/35), Kelley and Hennekam report genital abnormalities in 74% of patients [8]. If the four subgroups of genital anomalies mentioned in the current study were combined, abnormalities would be found in 11/24 patients (45.8%), still significantly fewer than in Ryan et al. and Balogh et al. Donoghue et al. also describes higher frequencies: *Hypospadias* (3/6; 50%), *Micropenis* (2/6; 33.3%) and *Undescended testes* (4/6 patients; 66.7%) (6). Since genital anomalies are listed as very characteristic features in many descriptions of the syndrome, the rather low frequency of occurrence in this studied cohort is a remarkable observation, to be confirmed by standardized data collection.

Multiple organ malformations and functions are repeatedly described in SLOS patients. These include *heart defects*, *renal anomalies*, *lung anomalies*, *gastrointestinal abnormalities* and *brain structural anomalies* [8]. These were also detected in the cohort studied, but only in very small numbers of patients, far below $\leq 40\%$. The highest frequency was still shown by the structural brain anomalies in 6/24 patients.

In the area of the musculoskeletal system, the feature (*Muscular*) *hypotonia* was documented in the cohort studied here, with an occurrence in 50% of cases. Balogh et al. also described this feature for 30.8% of patients [4]. In contrast, the first describers of the syndrome Smith, Lemli and Opitz describe the occurrence of moderate muscular hypertonicity in all three patients [3] so that the features seem to be opposite. This may be explained by an observed transition of postnatal muscular hypotension into later hypertension [10].

Thriving and development

SLOS is further characterized by failure to thrive. Thus, the parameter *Feeding difficulties* (18/25; 72.0%) is one of the most frequent features of this study. It is also described in many other cohorts studied with frequencies of 30.8% [4], 87.8% [6] and 100% [6][3]. The features *Feeding difficulties* and *Failure to thrive* (10/24; 41.7%) are characteristic of the early childhood of SLOS-patients.

In few cohorts *Global developmental delay* was also studied. In the current study, this was noted in 66.7% of patients while in the description of the syndrome by Smith et al. [3] in all three patients. In other publications, however, *Mental retardation* is documented rather or partly additionally: 92% [6], 97% [8], 71.4% [7] and 100% [3]. In the current study, however, this parameter with a frequency of 33.3% is not one of the more frequent characteristics.

It should also be noted that the parameters *Sleep disturbance* and *Recurrent infections* are among the most frequently occurring parameters in the cohort studied with 45.8% and 70.8%, respectively. While the feature *Recurrent infections* is not reported in any of the other studies used for comparison with a percentage frequency of $\geq 40\%$, the feature *Abnormal sleep pattern* is only described by Ryan et al. with a frequency of 70% [6]. The infrequent consideration of the feature *Recurrent infections* as well as the divergent results of the parameter *Sleep disturbance* may be due to the fact that the presence of these symptoms must always be explicitly asked.

It can be stated that the parameters determined as the most frequent features for SLOS on the basis of the current study were almost all documented in the published studies used for comparison with a percentage frequency of occurrence of $\geq 40\%$. Exceptions are the characteristics *Recurrent infections* and *Hypotonia*.

Conclusion

The aim of the study was to create data sets for three exemplarily selected rare diseases with the most important symptoms of these diseases leading to diagnosis in a standardized vocabulary. For this purpose, a literature review was conducted for late onset Pompe disease, Gaucher disease Type I and SLOS, followed by a retrospective examination of patient cohorts with regard to symptoms and diagnostic parameters. Using this methodology, 20 parameters presented in the form of HPO terms were identified as the most frequent features for late-onset Pompe disease, 16 for Gaucher disease Type I and 17 for SLOS. It was found that a retrospective analysis based on patient records makes standardization difficult due to missing data and different documentation. It can therefore be concluded that documentation using a template would have facilitated this. In turn, the developed data sets can be used to create this template. It is true that the patient cohorts for rare diseases were relatively large (Pompe: 23 patients, Gaucher: 21 patients, SLOS: 25 patients). However, the discussion conducted for SLOS indicates that there are features that are under- or over-represented in the current study compared to other publications. Thus, in order to obtain as complete and accurate a picture as possible, it is recommended that this study be supplemented with additional patient cases or cohorts to further expand and improve the data sets for CPMS. This and an application of the methodology to other diseases can improve the diagnosis and treatment of patients with rare diseases.

Declarations

Ethics approval

The ethical committee of the Otto-von-Guericke university in Magdeburg approved the study.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

INEA/CEF/ICT/A2018/1816992

Authors' contributions

ACLK analyzed and interpreted the patient data and was a major contributor in writing the manuscript. AZ as well as KM conceptualized and designed the study, recruited patients and revised manuscript. CVL and AB revised the manuscript for intellectual content. KP contributed to patient recruitment and revised the manuscript. All authors read and approved the manuscript.

Acknowledgements

We would like to thank Anke Lux (Medical faculty - University Hospital Magdeburg, Institute of Biometry and Medical Informatics) for her assistance with statistical analysis. Moreover, we thank Laura Paneghetti, PhD for comments that greatly improved the manuscript.

References

1. Zschocke J, Hoffmann GF. Vademecum Metabolicum: Diagnose und Therapie erblicher Stoffwechselkrankheiten. 4., vollständig überarb. Edition. Stuttgart: Schattauer; 2012.
2. Human Phenotype Ontology n.d. <https://hpo.jax.org/app/> (accessed April 16, 2021).
3. Smith DW, Lemli L, Opitz JM. A newly recognized syndrome of multiple congenital anomalies. The Journal of Pediatrics 1964;64:210–7. [https://doi.org/10.1016/S0022-3476\(64\)80264-X](https://doi.org/10.1016/S0022-3476(64)80264-X).
4. Balogh I, Koczok K, Szabó GP, Török O, Hadzsiev K, Csábi G, et al. Mutational Spectrum of Smith-Lemli-Opitz Syndrome Patients in Hungary. MSY 2012;3:215–22. <https://doi.org/10.1159/000343923>.
5. Quélin C, Loget P, Verloes A, Bazin A, Bessièrès B, Laquerrière A, et al. Phenotypic spectrum of fetal Smith–Lemli–Opitz syndrome. European Journal of Medical Genetics 2012;55:81–90. <https://doi.org/10.1016/j.ejmg.2011.12.002>.
6. Ryan AK, Bartlett K, Clayton P, Eaton S, Mills L, Donnai D, et al. Smith-Lemli-Opitz syndrome: a variable clinical and biochemical phenotype. Journal of Medical Genetics 1998;35:558–65. <https://doi.org/10.1136/jmg.35.7.558>.
7. Donoghue SE, Pitt JJ, Boneh A, White SM. Smith-Lemli-Opitz syndrome: clinical and biochemical correlates. Journal of Pediatric Endocrinology and Metabolism 2018;31:451–9. <https://doi.org/10.1515/jpem-2017-0501>.
8. Kelley RI, Hennekam RCM. The Smith-Lemli-Opitz syndrome. Journal of Medical Genetics 2000;37:321–35. <https://doi.org/10.1136/jmg.37.5.321>.
9. Cherstvoy ED, Lazjuk GI, Ostrovskaya TI, Shved IA, Kravtsova GI, Lurie IW, et al. The smith-lemli-opitz syndrome: A detailed pathological study as a clue to a etiological heterogeneity. Vichows Archiv A Pathol Anat 1984;404:413–25. <https://doi.org/10.1007/BF00695225>.
10. Cunniff C, Kratz LE, Moser A, Natowicz MR, Kelley RI. Clinical and biochemical spectrum of patients with RSH/Smith-Lemli-Opitz syndrome and abnormal cholesterol metabolism. American Journal of Medical Genetics 1997;68:263–9. [https://doi.org/10.1002/\(SICI\)1096-8628\(19970131\)68:3<263::AID-AJMG4>3.0.CO;2-N](https://doi.org/10.1002/(SICI)1096-8628(19970131)68:3<263::AID-AJMG4>3.0.CO;2-N).
11. Weigel JFW, Kiess W, Schulz M, Liebermann S, Pfäffle RW, Bührdel P. Das Smith-Lemli-Opitz-Syndrom. Kinder- und Jugendmedizin 2005;5:178–82. <https://doi.org/10.1055/s-0037-1617867>.