A novel classification system for research reporting in rare and progressive genetic conditions

CLARA D M VAN KARNEBEEK[1,2,7]
DANIËL BEUMER[3]
COLLEEN PAWLIUK[4]
HELLY GOEZ[5]
SARA MOSTAFAVI[1]
GAIL ANDREWS[4]
ROSE STEELE[6]
HAROLD SIDEN[4,7]

1 Centre for Molecular Medicine and Therapeutics, BC Children's Hospital Research Institute, University of British Columbia, Vancouver, BC, Canada.
2 Departments of Pediatrics and Clinical Genetics, Amsterdam University Medical Centres, Amsterdam, The Netherlands;
3 Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands.
4 BC Children's Hospital Research Institute, Vancouver, BC, Canada;
5 Division of Pediatric Neurology, Department of Pediatrics, University of Alberta, Stollery Children's Hospital, Edmonton, AB;
6 School of Nursing, Faculty of Health, York University, Toronto, ON;
7 Department of Pediatrics, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada.

Correspondence to Clara D M van Karnebeek, BC Children's Hospital Research Institute, Centre for Molecular Medicine and Therapeutics, University of British Columbia, 950 W 28th Ave, Vancouver, BC V5Z 4H4, Canada. E-mail: cvankarnebeek@cw.bc.ca

ABBREVIATION

CTT Charting the Territory

ABSTACT

AIM To create a classification system for severe, rare, and progressive genetic conditions for use in research reporting.

METHOD A modified Delphi consensus technique was used to create and reach agreement on a new system of condition categories. Interrater reliability was tested via two rounds of an online survey whereby physicians classified a subset of conditions using our novel system. Overall percentage agreement and agreement above chance were calculated using Fleiss' kappa (κ).

RESULTS Eleven physicians completed the first Delphi, with an overall agreement of 76.4%, the κ value was 0.57 (95% confidence interval 0.51–0.63), indicating moderate agreement (0.41–0.60) above chance. Based on the first survey several categories were described in more detail. The second survey confirmed a classification system with 12 categories, with an overall percentage agreement among the participants of 82.6%. The overall mean κ value was 0.71 (95% confidence interval 0.65–0.77), indicating substantial agreement (0.61–0.80).

INTERPRETATION Our new system was useful in categorizing a broad range of rare childhood diseases and may be applicable to other rare disease studies; further validation in larger cohorts is required.

A prospective study entitled ‘Charting the Territory’ (CTT) focused on children and their families living with conditions that are progressive, without cure or life-prolonging therapy (Category 3 according to the Association for Children’s Palliative Care and the Royal College of Paediatrics and Child Health).[1–3] CTT was a 5-year longitudinal, descriptive, correlational study that aimed to determine and document the clinical progression of each child’s condition and the associated biopsychosocial spiritual experiences of the parents and behaviour of siblings. In CTT, 258 families with 275 children, either newly diagnosed or with established illness, were recruited across nine North American children’s hospitals. The children were followed over time for symptom trajectory, and physical and social function, while their parents and siblings reported on their own well-being. Every child had a medical condition that made them eligible for the study, and many had a specific disease that was identified using diagnostic criteria (although we use condition and disease interchangeably, the term diagnosis in this paper is used to indicate a child’s specific disease). By the end of enrolment there were 113 individual diagnoses included in the cohort, which led to a challenge in data reporting. Treating all of the conditions as a single construct would hide any meaningful differences that might exist between conditions. However, providing information about 113 individual diagnoses that demonstrate great phenotypic/clinical overlap would obscure common patterns. We realized that we needed a classification system to group the conditions into categories that would highlight the similarities and differences between these conditions, while providing enough definition to create meaningful categories for data reporting in the research setting.

Review of disease classification systems

We undertook an assessment of available classification schemes to find a system that would help us establish a concrete variable for data analysis with a reasonable number of categories (i.e. <113 individual diagnoses). A list of each child’s condition was extracted from data collected at baseline for the CTT study. Two of the authors, with expertise in biochemical/clinical genetics, and complex conditions/paediatric palliative care (CvK, HS), then applied several widely used classification systems to group each of the 113 individual diagnoses. Classification systems assessed included the International Classification of Diseases, 11th Edition (ICD-11),[4] Orphanet,[5] SNOMED CT,[6] the Association for Children’s Palliative

Care/Royal College of Paediatrics and Child Health categories,[1] and the widely referenced Complex Chronic Conditions categories.[7,8]

ICD-11 is a diagnostic classification system based on the phenotype of the conditions.[6,9] Although ICD has a reputation as the criterion standard for classification of diseases and procedure, especially for health services research, we found that it was inadequate for the very rare conditions seen in CTT. Although ICD-11 has addressed many of the issues we found in the previous versions, including adding thousands of rare diseases with unique identifiers, conditions were sometimes not classified, were in overly broad categories, were too specific, or were categorized in ways that were counterintuitive and, thus, would not be useful for grouping conditions for reporting our research study.[9,10]

Orphanet is a multi-partner European consortium providing a reference portal for information on rare diseases with a powerful and comprehensive list of diseases.[7] SNOMED CT is a medical terminology standard maintained and distributed by SNOMED International. Both these systems are multi-hierarchical whereby diseases are classified into multiple categories and often also found under multiple subclasses within those categories.[11] There is no guidance on the Orphanet or SNOMED classification systems as to which category is the primary or most clinically relevant of the multiple hierarchies and, thus, there is no way to select for one label that allows grouping for each condition. These multi-hierarchical systems are very powerful in describing the many dimensions of the condition phenotype and enabling cross-linkages among conditions that share similar specific features (e.g. cerebellar hypoplasia). However, while multi-hierarchical approaches are valuable, they do not fulfil the requirements of a knowledge translation utility that bridges research findings to clinical implementation. Thus, neither of these widely used systems would meet our needs for data reporting.

A fundamental step towards a more comprehensive disease classification was made in Scriver's Online Metabolic and Molecular Bases of Inherited Disease,[12] which divides diseases into two groups. First, diseases that involve only one functional system or affect only one organ or anatomical system. In this group, presenting symptoms are uniform and diagnosis is straightforward even when the basic biochemical lesion gives rise to systemic consequences. Second, diseases in which the basic biochemical lesion either affects a metabolic pathway common to a large number of cells or organs, or is restricted to one organ, but gives rise to humoral and systemic consequences. This latter category is further subdivided based on

pathophysiology and symptomatology.[13] Despite the value of this classification system, 30 of the 113 individual diagnoses on our list were not represented in Scriver's Online Metabolic and Molecular Bases of Inherited Disease and so we determined that it would not be a useful classification system for analysing data.

Previous attempts at developing basic classifications for epidemiology, planning, and needs assessment, and health services research purposes for rare, severe, chronic, or life-threatening conditions in children have been reported in the UK and the USA.[1,7,8,14] The UK approach, developed by the charity Association for Children's Palliative Care (renamed Together for Short Lives) and the Royal College of Paediatrics and Child Health describes the broad spectrum of life-threatening conditions encountered in children eligible for paediatric palliative care.[1] An extension of this approach by Hain et al. maps information from children admitted to children's hospices or seen by paediatric palliative specialist teams onto the Association for Children's Palliative Care/Royal College of Paediatrics and Child Health categories.[14] All of the conditions studied in the CTT project were classified under Category 3 of this framework: conditions without cure or pathophysiologically directed interventions, with treatment of symptoms only. Another widely quoted system, developed in the USA by Feudtner et al. describes complex chronic conditions as 'any medical condition that can be reasonably expected to last at least 12 months (unless death intervenes), and to involve either several different organ systems, or one organ system severely enough to require specialty pediatric care and probably some period of hospitalization in a tertiary care center'.[7,8] These categories were all found to be too broad and did not further delineate among gene-based neurological and metabolic conditions (the main disease physiology for the CTT cohort).[7,8] As a result, they would not show any meaningful differences among the conditions.

The systems assessed all failed to provide either a useful level of specificity or generalization for the conditions in question. None of these approaches specifically focused on rare, progressive childhood conditions due to genetic disorders, nor would they enable more detailed classification for our cohort. A system that describes individual conditions based on phenotype or gene defect is too specific for the generalist clinician, while collective terms such as 'rare genetic disease' are too broad to assist the subspecialist or researcher. Furthermore, while the development of a system based on a single taxonomical construct (e.g. morphology, physiology, gross/histological pathology, gene alteration, billing, or health services utilization)

would provide a logical classification system, it undermines the utility of shared understanding.[15]

Therefore, we needed to develop a new classification system that would support both analysis of the data using a discrete variable and communication of study findings in a broader, clinically relevant manner. The purpose of this paper is to present a utilitarian method of categorization designed to usefully describe and distinguish types of life-threatening progressive neurological, metabolic, chromosomal, and monogenic conditions at the cohort level. In addition, we report results from an interrater reliability analysis of our initial system that led to subsequent improvement of the classification system overall.

METHOD

A Delphi method is a way to explore and achieve consensus about a problem from a group of informed researchers or clinicians. It allows communication via face-to-face discussion, mail, telephone, and – more recently – online approaches that enable experts to weigh in with their opinion on a select problem. Feedback from the experts results in achieving consensus or agreement on the problem and is popularly used in medical and health services research as a way to identify priorities in policy or education.[16] We modified this technique slightly in order to develop a clear definition of the categories through face-to-face discussion and then expanded the field of experts via e-mail to include multiple respondents off-site, and solicit their input in order to achieve consensus via two separate online surveys. These three phases of the study are described below.

Phase 1

Four of the authors (CvK, HG, SM, HS) initially attempted to classify the 113 individual diagnoses according to aetiology, pathology, or clinical phenotype. However, given the varying levels of modern scientific knowledge on the one hand (e.g. many of the clinical trajectories of the diseases are not known or published in the literature, which is the topic of a symptom scoping review undertaken separately to this report)[17] and the significant overlaps between conditions for each of these constructs on the other hand, this approach did not provide sufficient information for a complete classification system. Often a majority of the diseases could be grouped by a clinical phenotype (symptomatology) or pathology (mechanism of the

disease function), but their heterogeneous aetiology made this classification category clinically irrelevant.

Next, we set out to develop a new classification system that combined different constructs, i.e. specific descriptors of a condition determined by different analyses: clinical (patient observation); anatomical (imaging or pathology); physiological (laboratory tests, in vitro experiments); and genetic defects or disturbances (chromosomal/molecular analysis). Using a modified Delphi technique, initial categories were developed in an iterative approach, testing definitions against the sample conditions' list in a non-blinded fashion. Our phase 1 initial classification system resulted in 12 condition categories.

Phase 2

In order to validate the initial categories, we conducted two Delphi survey rounds to validate the utility, reach consensus on the categories, and assess usefulness of the resulting classification system. Twenty physicians with expertise in the relevant subspecialty fields at various institutions across Canada were approached and 11 participated. These physicians worked in the fields of inherited metabolic diseases (n=5), complex paediatrics (n=3), neurology (n=2), and clinical genetics (n=1). In an online survey, they were provided with 28 individual diagnoses from the CTT study, with definitions for 10 of the 12 proposed categories. Two categories were not included in the survey, namely Severe Neurological Impairment – Not Yet Diagnosed and Other Conditions Not Otherwise Specified because these categories should only be applied in special circumstances (i.e. if the patient does not have a diagnosis or if the patient has a known diagnosis that clearly does not fit into any other category). We chose these 28 individual diagnoses from the CTT cohort for sampling based on relevance, difficulty in categorization, severity of the disease, and frequency of the disorder in our patient population. Participants were provided with a definition of each category and asked to assign each of the 28 individual diagnoses to one of the categories. At the end of the survey, participants were given the opportunity to provide written feedback to the investigators that was incorporated into the second survey round (phase 3) of the Delphi. The participants were blinded to the response of others. The surveys were created in REDCap, a software solution to support clinical and translational research.[18]

Phase 3

Based on the results of the first survey completed in phase 2, we formulated more specific definitions for each category. This revised classification system formed the basis for the second survey round issued to the same participants, and was again rated on agreement for interrater reliability analysis.

Statistical analysis

The Fleiss' kappa (κ) statistic was calculated using Microsoft Excel (Redmond, WA, USA) to assess the proportion rate of observed agreement adjusted for the proportion of agreement expected to occur by chance.[19–21] Both the per category κ and the overall κ were computed, noting that the overall κ is computed as a weighted average of the per-category values. Guidelines regarding range of agreement suggest the following: 0 to 0.20, slight agreement; 0.21 to 0.40, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, substantial agreement; 0.81 to 1, almost perfect agreement.[22,23] To determine if any participant was an outlier for rater agreement, the analyses were repeated by removing each participant systematically and then rerunning the analysis.

RESULTS

Overall percentage agreement among the participants in phase 2 was 76.4%. The overall mean value for interrater reliability was 0.57 (95% confidence interval 0.51–0.63), indicating moderate agreement. When the analysis was repeated by systematically removing each participant in turn (thereby leaving 10 of the 11 raters in each repeated analysis), mean κ values remained essentially unchanged, suggesting that ratings remained consistent across all raters without significant outliers. See Table I for the κ rating of each category. Based on the results of the first Delphi round definitions in several categories were described in more detail.

In phase 3, eight of the 11 participants responded, with an overall percentage agreement among the participants of 82.6%. For interrater reliability, the overall mean κ value was 0.71 (95% confidence interval 0.65–0.77), indicating that 'substantial' agreement was achieved among the raters. When the analysis was repeated by systematically removing each participant in turn (thereby leaving seven of the eight raters in each repeated analysis), mean κ values remained essentially unchanged, suggesting that ratings remained consistent across all raters

without significant outliers. See Table I for the results of each category. This round confirmed a classification system with 12 categories listed (Table SI, online supporting information). All 113 individual diagnoses were eventually classified by four of the authors (CvK, HG, SM, HS) into the 12 categories for use in statistical analysis and reporting, with 109 of these shown in Table SII (online supporting information) as cures have since been developed for some conditions.

DISCUSSION

The goal of this endeavour was to develop a classification system with broad clinical relevance that enables cohort-level data analysis of rare, progressive childhood diseases. Specifically, the objective was to create a high-level set of defined categories that is intuitively sensible, extends beyond a single researcher's or clinician's opinion, and is not limited to the work of a single discipline. The system was designed to be practical in assisting with reporting the analysis of a broad range of clinical and psychosocial data regarding rare childhood diseases whose phenotypes include metabolic derangements, neurological conditions, and impairment of other organ systems. Also, the system may enhance uniformity in clinical management. The system was built from the 'ground up' by a multidisciplinary group of clinicians with research expertise; our modified Delphi attempted to reach consensus among this group and sufficient interrater reliability was demonstrated for the system we present.

For every condition, the depth of knowledge derived from the different underlying constructs will differ. For most, the clinical symptomatology is known, whereas insights into pathophysiology at anatomical, organ, cellular, and genetic levels are increasingly limited. An inverse pyramid depicts this incremental change (Fig. 1). To illustrate using X-linked adrenoleukodystrophy,[24] the *clinical* presentation is a neurodegenerative disease with electrolyte and hormonal disturbances primarily affecting males. The *anatomical* change is the leukodystrophy seen on neuroimaging; the *pathophysiology* occurs in two organs (central nervous system and adrenal glands), specifically those cells most vulnerable to the peroxisomal (organelle) dysfunction; and the causal *gene* is *ABCD1* located on the X chromosome, in which a large number of mutations have been reported to cause adrenoleukodystrophy. While our categories are based upon a combination of each of the four constructs, the clinical symptomatology is over-represented as it is best known and is also often the unifying construct

across the many diseases. In our system adrenoleukodystrophy is placed with other genetic conditions in the category ‘Lysosomal/peroxisomal leukodystrophy’, on the basis of their shared clinical (neurodegeneration) and anatomical (leukodystrophy) features, while acknowledging the distinctions in pathophysiology (lysosomes vs peroxisomes) and genetic defects.

A limitation of this study is the potential for bias given the relatively small number and non-random selection of physicians participating in the study and the fact they were not blinded in Phase 1. However, we do consider our raters to be representative of the spectrum of child health clinicians in different, but overlapping, fields who work in the field of paediatric neurodegenerative diseases. A second limitation of our resulting system is that it is not based on a single convenient construct, such as genotype, protein, or physiology. While a uniform approach to developing categories based on one construct may be intellectually satisfying, our current state of knowledge of the complex interplay of genotype and phentoype is not yet ready for a unifying system that will enable disease categorization for analysis of symptoms, treatments, or outcomes that the frontline clinician and translational researcher need. Third, the classification system shows a variability in performance, depending on the category of disease. For example, epileptic encephalopathies performed better than neuromuscular diseases in both rounds; and one category, congenital disorders of glycosylation, performed worse in the second than the first survey round. Notwithstanding, we do believe that our categorization system represents a reasonable initial approach yet emphasizes the need for further studies in larger cohorts of randomly selected blinded clinicians, representing different specialties, for robust validation and optimization. Furthermore, we expect this classification system to be further adapted and refined according to the experience of researchers using the system to analyse clinical data in this field.

CONCLUSION

This system was useful in categorizing a broad range of rare childhood diseases, including inborn errors of metabolism, neurological conditions, and impairment of other organ systems; the physician survey showed sufficient interrater reliability. The resultant categories will help to support further research and programme development in paediatric palliative care, complex care, neurodegenerative, genetics and metabolic diseases, and health services research. This classification system supports research as it provides a way to meaningfully group similar

conditions in a way that the currently available classification systems assessed for this study did not. With the advent and continued development of ‘omics’ technologies (epigenomics, metabolomics, proteomics), the genetic basis of an increasing number of rare, severe progressive diseases will be determined and their pathophysiology, as well as anatomical and clinical spectra, further characterized.[25] The classification system itself will be adapted to reflect these novel insights. While our study focused on rare, progressive conditions, mostly in the palliative care setting, we anticipate that our classification system may well be applied in other studies and settings, notably in research where the grouping of rare, individual diagnoses continues to be necessary to show clinically relevant differences at the cohort level.

ACKNOWLEDGEMENTS

The primary study (CTT) was supported by the Canadian Institutes of Health Research Grant MOP-89984. The authors would like to thank all the children and families who participated in the CTT study. We also thank the research assistants in each study setting for their contribution to data collection, and the site lead collaborators who enabled access to families and acted as lead investigators in the various settings. We gratefully acknowledge Ms Brigid Wong for her assistance in gathering information on the conditions and classification systems, Dr Sylvia Stockler for input on the initial iteration of categorization, Dr B. Kuzeljevc for contributing his statistical expertise, and Ms C. Sowerbutt for edits. Finally, we would like to thank the physicians participating in the testing of the inter-rater reliability of our new classification system. The authors have stated that they had no interests that might be perceived as posing a conflict or bias.

REFERENCES

1. Association for Children with Life-Threatening or Terminal Conditions and their Families (ACT). A guide to the development of children's palliative care services (3rd edition) [Internet]. Bristol: ACT/RCPCH, 2009. Available at: https://edisciplinas.usp.br/pluginfile.php/4282446/mod_folder/content/0/ACT.%20A%20Guide%20to%20the%20Development%20%20of%20Childrens%20Palliative%20Care%20Services.pdf?forcedownload=1 (accessed 10 August 2017).

2. Siden H, Steele R, Brant R, et al. Designing and implementing a longitudinal study of children with neurological, genetic or metabolic conditions: charting the territory. *BMC Pediatr* 2010; **10**: 67.

3. Siden H, Steele R. Charting the territory: children and families living with progressive life-threatening conditions. *Paediatr Child Health* 2015; **20**: 139–44.

4. World Health Organization (WHO). The ICD-11 classification of mental and behavioral disorders: clinical and diagnostic guidelines [Internet]. Available at: https://icd.who.int/browse11/l-m/en. Geneva: WHO, 2018 (accessed 10 August 2017).

5. OrphaNet. Available at: http://www.orpha.net (accessed 10 August 2017).

6. International Health Terminology Standards Development Organisation. SNOMED CT. Available at: https://www.nlm.nih.gov/healthit/snomedct/ (accessed 30 November 2017).

7. Feudtner C, Christakis DA, Connell FA. Pediatric deaths attributable to complex chronic conditions: a population-based study of Washington State, 1980–1997. *Pediatrics* 2000; **106**(Suppl.): 205–9.

8. Feudtner C, Feinstein JA, Zhong W, Hall M, Dai D. Pediatric complex chronic conditions classification system version 2: updated for ICD-10 and complex medical technology dependence and transplantation. *BMC Pediatr* 2014; **14**: 199.

9. Kodra Y, Fantini B, Taruscio D. Classification and codification of rare diseases. *J Clin Epidemiol* 2012; **65**: 1026–7.

10. 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.

11. 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**: 803–8.

12. McGraw Hill Companies. Scriver's Online Metabolic and Molecular Bases of Inherited Diseases (OMMBID). Available at: http://ommbid.mhmedical.com (accessed 10 August 2017).

13. Saudubray J, Charpentier C. Clinical phenotypes: diagnosis/algorithms. In Valle D, Beaudet A, Vogelstein B, editors. Online molecular and metabolic bases of inherited disease [Internet]. New York: McGraw-Hill, 2014. Available at: http://ommbid.mhmedical.com/content.aspx?bookid=971&sectionid=62670964&jumpsectionID=62671211 (accessed 10 August 2017).

14. Hain R, Devins M, Hastings R, Noyes J. Paediatric palliative care: development and pilot study of a 'Directory' of life-limiting conditions. *BMC Palliat Care* 2013; **12**: 43.

15. Loscalzo J, Kohane I, Barabasi AL. Human disease classification in the postgenomic era: a complex systems approach to human pathobiology. *Mol Syst Biol* 2007; **3**: 124.

16. Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ* 1995; **311**: 376–80.

17. Brander G, Pawliuk C. Research-embedded health librarians as facilitators of a multidisciplinary scoping review. *J Can Health Libraries Assoc* 2017; **38**: 38–43.

18. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009; **42**: 377–81.

19. Zaiontz C. Real statistics using Excel [Internet]. Available at: http://www.real-statistics.com/reliability/intraclass-correlation/ (accessed 27 November 2018).

20. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960; **20**: 37–46.

21. Gwet KL, Handbook of inter-rater reliability: the definitive guide to measuring the extent of agreement among raters. 4th ed. Gaithersburg, MD: Advanced Analytics, 2014.

22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–74.

23. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med* 2005; **37**: 360–3.

24. Kemp S, Berger J, Aubourg P. X-linked adrenoleukodystrophy: clinical, metabolic, genetic and pathophysiological aspects. *Biochim Biophys Acta* 2012; **1822**: 1465–74.

25. Armstrong L, Biancheri R, Shyr C, et al. AIMP1 deficiency presents as a cortical neurodegenerative disease with infantile onset. *Neurogenetics* 2014; **15**: 157–9.

Table I: Kappa results for first and second survey rounds[a]

Category	First survey round mean kappa	Second survey round mean kappa
Congenital disorders of glycosylation	1.0	0.88
Epileptic encephalopathies	0.74	0.93
Lysosomal storage/peroxisomal diseases	0.39	0.76
Multiorgan congenital abnormalities	0.56	0.74
Mitochondrial encephalopathies/myopathies	0.72	0.94
Neurodegenerative diseases	0.41	0.53
Neuromuscular diseases	0.77	0.57
Other inborn errors of metabolism	0.21	0.40
Small-molecules diseases	0.34	0.72
Structural central nervous system abnormalities	0.76	0.79

[a]Two conditions are not included in the survey rounds: Severe Neurological Impairment – Not Yet Diagnosed and Other Conditions Not Otherwise Specified.

Figure 1: Inverse pyramid illustrating the four different constructs used for the classification systems using the example X-linked adrenoleukodystrophy.

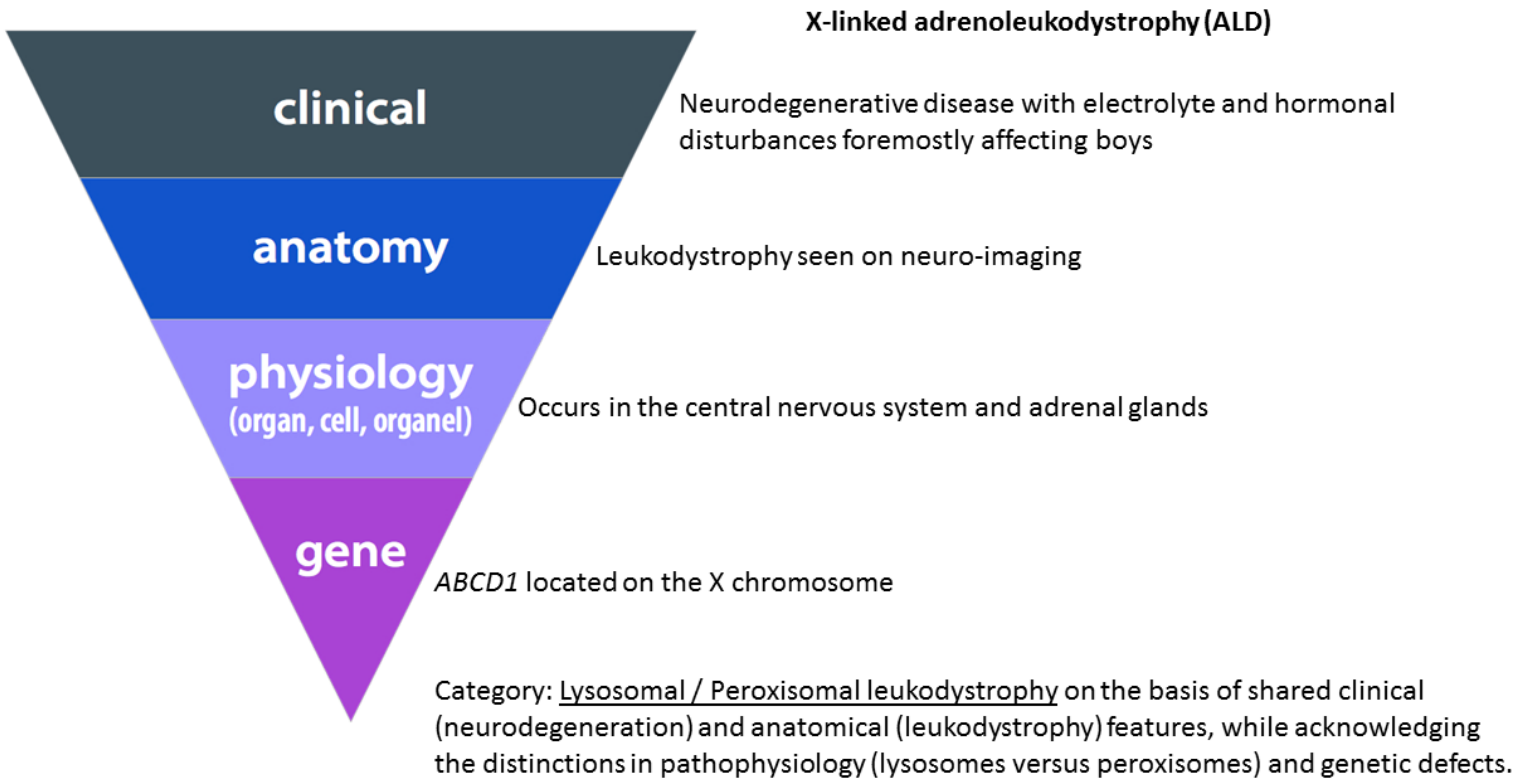

Table S1. Categories and definitions of rare, progressive childhood conditions

Category	Definition
1. Congenital Disorders of Glycosylation	Inborn errors of metabolism due to a gene defect encoding protein glycosylation, confirmed either by an abnormal pattern of transferrin isoelectric focusing and/or molecular analysis.
2. Epileptic Encephalopathies	Severe brain disorders of early age that manifest with multiform, intractable seizures causing progressive psychomotor delay and often episodes of reduced level of consciousness. The epileptic syndrome represents either a known gene disorder or a descriptive seizure disorder. These conditions involve deterioration and may therefore be confused with another category – *Neurodegenerative Diseases.* In the Epileptic Encephalopathies' group however, the seizure disorder is both a cause and an expression of disease.
3. Lysosomal Storage / Peroxisomal Diseases	Group of disorders caused by inborn errors of lysosomal or peroxisomal metabolism.
4. Multi-Organ Congenital Abnormalities	Congenital abnormalities in multiple organs, clinically recognizable as a known monogenic syndrome and/or due to a confirmed pathologic (sub-)microscopic numeric or structural chromosomal variant identified on karyotype / chromosome micro-array.

5. Mitochondrial Encephalo-/Myopathies	Clinically ascertained dysfunction of brain and/or muscle caused by mitochondrial disease (defined as abnormal respiratory chain complexes and/or nuclear or mitochondrial DNA mutations or rearrangements) and/or defects in pyruvate metabolism.
6. Neurodegenerative Diseases	These are progressive diseases of brain deterioration in brain cellular structure and function, inherent in neurons or supporting cells. The underlying gene or pathway is known and manifests in functional, histological and/or imaging changes. Seizures are often a complication of these diseases and therefore may resemble *Epileptic Encephalopathies*. In the Neurodegenerative Diseases' group however, the underlying pathology represents degeneration that is manifested in functional decline and often accompanying seizures.
7. Neuromuscular Diseases	These diseases are characterized primarily by muscle weakness and occur via mechanisms involving dysfunction of either muscle tissue or the peripheral nerves. In many cases, they can be considered "degenerative" as there is evidence of progressive tissue deterioration. They are distinct, however, from the *Neurodegenerative disease* category in that the primary clinical concern is weakness and not Central Nervous System involvement with cognitive and functional impairment and/or seizures.
8. Other Conditions Not Otherwise Specified	This category should be applied very infrequently and only in special circumstances. It will be used when a condition clearly does not fit into any other category. An example would be a phenotypically described condition that has many manifestations and different etiologies (e.g. due to different genes, toxins, etc.). The "end result" phenotype will be variable as well – unlike the situation for example with *Epileptic encephalopathy / Neurodegenerative disease conditions*, which also have variable etiologies, but a more homogenous phenotype.

9. Other Inborn Errors of Metabolism	These are other inborn errors of metabolism that do not belong to the above outlined categories. They are not, however, *Other Conditions Not Otherwise Specified (NOS)* in that a biochemical pathway is known to be involved. This diverse group is difficult to characterize. The broad range means that inclusion of a condition in this group may hinder the utility of this group for detailed analysis. Over time, as more discoveries are made around the etiologies and pathophysiologies of these conditions, some of them will be re-categorized in the future.
10. Severe Neurological Impairment – Not Yet Diagnosed (SNI-NYD)	Many children are affected by severe impairments in all functional domains due to abnormalities of the brain, without a known cause. Patients have profound limitations with intellectual disability, immobility, lack of communication, and a need for feeding support. In addition, they often have seizures, GERD, orthopedic problems and dystonia. These impairments are also routinely found in children without an etiologic diagnosis. We have chosen the term Severe Neurological Impairment – Not Yet Diagnosed (SNI-NYD).
11. Small Molecules Diseases	This group includes organic acidurias, urea cycle defects, amino-acidopathies, carbohydrate metabolism, and defects in purine /pyrimidine metabolism. It usually presents clinically as intoxication when toxic metabolites build up and/or delay in fuel provision.
12. Structural Central Nervous System Abnormalities	Congenital abnormalities of the central nervous system (ascertained on MRI / CT), that represent the patient's most prominent phenotypic feature, and for which the underlying defect has not yet been identified.

Table S2. List of 110 of the 113 Individual Diagnoses from “Charting the Territory”, organized by category

Category	Conditions from CTT Study (OMIM Number, if available)	
1. Congenital Disorders of Glycosylation	• Congenital Disorder of Glycosylation Type 1A (212065)	• Congenital Disorder of Glycosylation Type 1D (601110)
2. Epileptic Encephalopathies	• Sodium channel, voltage-gated, type 1, alpha subunit (SCN1A) (182389) • Dravet Syndrome (607208)	• Lennox-Gastaut Syndrome (615369, 616346, 617113, 618141) • West Syndrome (300672, 308350, 613477, 613722, 615006, 616139, 616341, 617065, 617929) • Early Myoclonic Encephalopathy (300868)

3. Lysosomal Storage / Peroxisomal Diseases	• GM1 Gangliosidosis (230500, 230600, 230650) • GM2 Gangliosidosis (Tay Sachs Disease) (272800) • Infantile Refsum Disease (601539) • Krabbe Disease (245200) • Leukodystrophy • Metachromatic Leukodystrophy (250100) • Mucolipidosis Type IV (252650) • Mucopolysaccharidosis Type I (Hurler Syndrome) (607014) • Mucopolysaccharidosis Type I (Hurler-Scheie Syndrome) (607015) • Mucopolysaccharidosis Type II (Hunter Syndrome) (309900) • Mucopolysaccharidosis Type III (Sanfilippo Syndrome A) (252900) • Mucopolysaccharidosis Type III (Sanfilippo Syndrome B) (252920) • Mucopolysaccharidosis Type III (Sanfilippo Syndrome), Type not specified • Mucopolysaccharidosis Type IV (Morquio Syndrome), Type not specified • Mucopolysaccharidosis Type IV A (Morquio Syndrome A) (253000) • Neuronal Ceroid Lipofuscinosis Type 3 (Batten Disease) (204200) • Neuronal Ceroid Lipofuscinosis Type 6 (601780) • Niemann-Pick Disease Type A (257200) • Niemann-Pick Disease Type C (257220, 607625) • X-linked Adrenoleukodystrophy (300100)

4. Multi-Organ Congenital Abnormalities	• Achondroplasia (100800) • Aicardi Syndrome (304050) • Bowen-Conradi Syndrome (211180) • Chromosome 11, 18 Partial Trisomy • Chromosome 11q Partial Trisomy • Chromosome 12p Mosaicism Tetrasomy (Pallister-Killian Syndrome) (601803) • Chromosome 13 Trisomy (Patau Syndrome) • Chromosome 13 Trisomy with Mosaicism • Chromosome 13 Trisomy, 10 Partial Monosomy • Chromosome 14 Deletion • Chromosome 18 into 5 deletion • Chromosome 18 Trisomy (Edwards Syndrome) (300484) • Chromosome 18q Deletion (601808) • Chromosome 1p36 Monosomy (607872) • Chromosome 21 Trisomy (with congenital heart and liver malformations) (190685) • Chromosome 22 Mosaic Trisomy • Chromosome 4p Deletion (Wolf–Hirschhorn Syndrome) (194190) • Chromosome 5 Trisomy • Chromosome 5p12.2 Deletion (Cri-du-chat) (123450) • Chromosome 5q14.3 Deletion Syndrome (612881) • Chromosome 6/4 Unbalanced Translocation • Chromosome 69 Triploidy, xxy • Chromosome 79, 23q, q24.3 Deletion 46xx • Chromosome Unbalanced Rearrangment (Partial Trisomy 7, Monosomy 6) • Klippel-Trenaunay Syndrome (149000) • Malignant Infantile Osteopetrosis (259700, 611490, 259720) • Marshall-Smith Syndrome (602535) • Mowat-Wilson Syndrome (235730) • Schinzel-Giedion Syndrome (269150) • Systemic Juvenile Xanthogranulomatosis • Tuberous Sclerosis (191100, 613254)

	• Chromosome 22q11.2 Deletion (DiGeorge Syndrome) (188400) • Chromosome: 3 Inversion • Chromosome: 3p25 Partial Monosomy, 5p15.3 Trisomy	
5. Mitochondrial Encephalo-/Myopathies	• Leigh's Syndrome (Complex I Deficiency) (25600) • Mitochondrial Disorder (LBSL, DARS2 Gene Defect) • Mitochondrial Cytopathy (53000) • Mitochondrial Complex I Deficiency (252010) • Mitochondrial Complex III Deficiency (124000, 615157, 615158, 615159, 615160, 615453, 615824, 615838, 616111) • Mitochondrial Complex IV Deficiency (220110)	• Mitochondrial Complex V Deficiency (604273, 614052, 614053, 615228) • MELAS Syndrome (540000) • Pyruvate Carboxylase Deficiency (266150) • Pyruvate Dehydrogenase Deficiency (313170)

6. Neurodegenerative Diseases	• Aicardi-Goutieres Syndrome (225750, 610181, 610181, 610333, 612952, 615010, 61846) • Atypical Rett Syndrome (312750) • Infantile Neuroaxonal Dystrophy (256600) • Juvenile Huntington's Disease (143100)	• Pantothenate Kinase-Associated Neurodegeneration (234200) • Pelizaeus-Merzbacher Disease (312080) • Rett Syndrome (312750)
7. Neuromuscular Diseases	• Centronuclear Myopathy (160150, 255200, 614408, 614807, 615959, 310400) •	• Desminopathy (601419)
8. Other Conditions Not Otherwise Specified	• Dilated Cardiomyopathy with Ataxia (610198)	• KIF1A Gene Defect
9. Other Inborn Errors of Metabolism	• Glucose-6-phosphate Dehydrogenase Deficiency (300908)	• Lesch-Nyhan Syndrome (300322) • Smith-Lemli-Opitz Syndrome (270400)

10. Severe Neurological Impairment – Not Yet Diagnosed (SNI-NYD)	• SNI-NYD: Dystonic Cerebral Palsy • SNI-NYD: Encephalopathy •	• SNI-NYD: Seizure Disorder • SNI-NYD: Microcephaly •
11. Small Molecules Diseases	• ArgininosuccinateLyase Deficiency (207900) • Glutaric Acidemia Type I (231670) • Glutaric Aciduria Type 2 (231680) • Hyperargininemia (207800)	• Non-ketotic Hyperglycinemia (605899) • Propionic Acidemia (606054) • Presumed 3-Hydroxyisobutyric Aciduria (not confirmed) (236795)

12. Structural Central Nervous System Abnormalities	• Agenesis of the Corpus Collosum • Alobar Holoprosencephaly (157170, 609637, 610829) • Cerebral Dysgenesis • Congenital Hydrocephalus (236600, 615219) • Congenital Microcephaly • Joubert Syndrome (213300, 610688, 612291, 614173, 614424, 614464, 614615, 614970, 615636 616490, 616654)	• Lissencephaly • Miller-Dieker Syndrome (Lissencephaly Type 1) (247200) • Pontocerebellar Hypoplasia Type II (277470, 612389, 612390, 613811, 617026) • Pontocerebellar Hypoplasia Type III (608027) • Schizencephaly (269160) • SemilobarHoloprosencephaly (157170, 609637, 610829) • Lobar Holoprosencephaly (157170, 609637, 610829)