UNIVERSITY OF MODENA AND REGGIO EMILIA

PhD Program of Clinical and Experimental Medicine (CEM)

XXXV Doctorate Cycle

Director of the PhD Program: Prof. Marco Vinceti

Novel strategies for diagnosing and understanding the natural history of developmental and epileptic encephalopathies: models from Dravet Syndrome

Supervisor: Prof. Francesca Darra

Co-supervisor: Prof. Giuseppe Biagini

Candidate: Tommaso Lo Barco

Summary

ABSTRACT (ENGLISH) **4**

ABSTRACT (ITALIANO) **5**

ABBREVIATIONS **6**

INTRODUCTION **8**

DEVELOPMENTAL AND EPILEPTIC ENCEPHALOPATHIES (DEES) 8

Epilepsy: definition and classification *8*

Developmental and Epileptic Encephalopathies: definition and approaches to care *9*

DRAVET SYNDROME (DS) 10

Definition and main clinical features *10*

Epidemiology *11*

DS as monogenic encephalopathy *12*

Epilepsy in DS *12*

"NEXT GENERATION PHENOTYPING" 13

Electronic health records (EHR) *13*

Clinical Data Warehouses (CDW) and Medical Data Mining in health care *14*

Data mining and Machine Learning in rare diseases *14*

Natural Language Processing (NLP) and Unified Medical Language System (UMLS) *15*

REFERENCES 16

AIMS AND THESIS OVERVIEW **22**

CHAPTER 1 ": "A REGISTRY FOR DRAVET SYNDROME: THE ITALIAN EXPERIENCE" **24**

ABSTRACT 24

INTRODUCTION 26

METHODS 27

RESULTS 30

DISCUSSION 42

REFERENCES 45

CHAPTER 2: "IMPROVING EARLY DIAGNOSIS OF RARE DISEASES USING NATURAL LANGUAGE PROCESSING IN UNSTRUCTURED MEDICAL RECORDS: AN ILLUSTRATION FROM DRAVET SYNDROME" **48**

ABSTRACT 48

OBJECTIVES 49

MATERIALS AND METHODS 50

RESULTS 52

DISCUSSION 60

STUDY LIMITATIONS 62

Conclusion 63
References 63

CHAPTER 3: "OUTLINING THE NATURAL HISTORY OF RARE DISEASES USING NATURAL LANGUAGE PROCESSING OF NARRATIVE UNSTRUCTURED RECORDS: THE EXAMPLE OF DRAVET SYNDROME" 67

Abstract 67
Introduction 68
Methods 69
Results 71
Discussion 74
Conclusion 76
References 76

CHAPTER 4: "A EUROPEAN PILOT STUDY IN DRAVET SYNDROME TO DELINEATE WHAT REALLY MATTERS FOR THE PATIENTS AND FAMILIES" 79

Abstract 79
Introduction 80
Methods 80
Results 81
Discussion 84
References 86

CHAPTER 5 "ADAPTIVE BEHAVIOUR IN ADOLESCENTS AND ADULTS WITH DRAVET SYNDROME" 90

Abstract 90
Introduction 91
Method 92
Results 94
Discussion 101
Study Limitations 103
Conclusion 103
References 104
Appendix 106

CONCLUSION AND FUTURE PERSPECTIVES 109

References 114

Abstract (English)

Developmental and epileptic encephalopathies (DEEs) are a group of diseases characterized by developmental impairment and phases of plateauing and regression induced by epileptic activity. Treatment aims to reduce seizures and establish appropriate rehabilitation interventions.
With the rising emerging potential for precision medicine, early diagnosis and acknowledgment of the natural history of DEEs are crucial to define targeted treatments and promptly detect/prevent specific comorbidities.

Dravet syndrome (DS) is a well-known type of DEE. In this research project, we explored novel strategies for early diagnosing and understanding the natural history of DS.
We described Residras, the international registry of Dravet syndrome, as a disease-registry model which aims to improve and enlarge the longitudinal phenotypization through a multicenter structured data collection. Alongside, we investigated the use of an unstructured data mining procedure—natural language processing (NLP)—to optimize time of diagnosis and outline the natural history. With the aim of expanding exclusive clinician perspectives, we surveyed the caregivers' point-of-view around meaningful outcomes in clinical studies regarding DS. Finally, we tested the feasibility of using a common tool to monitor adolescents and adults with DS, stressing limitations and highlighting potential new findings on natural history by using disease-specific instruments.

These studies further delineate the natural history of DS and can help tailor patient-centered outcome measures in future clinical trials. Moreover, these approaches could be transferred to other lesser-known DEEs to decrease time for diagnosis and outline the natural history.

Abstract (Italiano)

Le encefalopatie dello sviluppo ed epilettiche (ESE) costituiscono un gruppo di malattie caratterizzate da un'alterazione dello sviluppo psicomotorio ed epilessia, con fasi di peggioramento del quadro cognitivo/comportamentale causate da un'accentuazione dall'attività epilettica. Il trattamento ha lo scopo di ridurre le crisi epilettiche e mettere in atto opportuni interventi di riabilitazione mirati.
Con il crescente potenziale della medicina di precisione, la diagnosi precoce ed il riconoscimento della storia naturale delle diverse ESE rappresentano punti cruciali per definire trattamenti specifici e rilevare prontamente, o prevenire, le comorbidità associate.

La Sindrome di Dravet (SD) è probabilmente la più nota e studiata ESE. In questo progetto di ricerca, abbiamo esplorato nuove strategie per la diagnosi precoce e la comprensione della storia naturale della SD.
Abbiamo descritto "Residras", il registro internazionale della SD, come un modello di registro di malattia che mira a migliorare e ampliare la caratterizzazione fenotipica su un piano longitudinale, attraverso una raccolta multicentrica, periodica, di dati strutturati. Parallelamente, abbiamo studiato l'uso di una procedura di "data mining" che utilizza dati non strutturati – il Natural Language Processing (NLP) – per ottimizzare i tempi di diagnosi e delineare la storia naturale. Con l'obiettivo di ampliare le prospettive da un punto di vista diverso da quello dei clinici, abbiamo interrogato i caregivers riguardo a possibili "outcome secondari" da considerare nell'elaborazione di uno studio clinico sulla SD. Infine, abbiamo testato la fattibilità dell'utilizzare uno strumento comunemente utilizzato nella pratica clinica per studiare il profilo adattivo, su adolescenti e adulti con SD, sottolineandone i limiti e mettendo in evidenza potenziali nuovi risultati sulla storia naturale utilizzando strumenti specifici per la malattia.

Questi studi contribuiscono a definire ulteriormente la storia naturale della SD e possono aiutare ad individuare "outcome secondari" per futuri studi clinici. Inoltre, questi approcci potrebbero essere trasferiti ad altre ESE, la cui storia naturale non è ancora nota, con l'obiettivo di diminuire il tempo per la diagnosi e delinearne la storia naturale.

Abbreviations

AE	Adverse Event
ABP	Adaptive Behavior Profile
ADHD	Attention Deficit Hyperactivity Disorder
ASM	Antiseizure medication
CDW	Clinical Data Warehouse
CT	Computerized Tomography
CUI	Concept Unique Identifier
DEE	Developmental and Epileptic Encephalopathies
DS	Dravet Syndrome
EMA	European Medicines Agency
ER	Emergency Room
FDA	Food and Drug Administration
FS	Febrile Seizures
EEG	Electroencephalogram
HER	Electronic Health Record
GEFS+	Genetic Epilepsy with Febrile Seizures Plus
GLUT1	Glucose Transporter 1
HPO	Human Phenotype Ontology
ICU	Intensive Care Unit
ILAE	International League Against Epilepsy
MRI	Magnetic Resonance Imaging
NGS	Next Generation Sequencing
NLP	Natural Language Processing
ORDO	Orphanet Rare Disease Ontology
PheWAS	Phenome-Wide Association Study
PROs	Patient Reported Outcomes
POMs	Primary Outcome Measures
RCT	Randomized Controlled Trial
SUDEP	Sudden Unexplained Death in Epilepsy
TSC	Tuberous Sclerosis Complex
UMLS	Unified Medical Language System
VABS	Vineland Adaptive Behavioural Scales
WISC	Wechsler Intelligence Scale for Children

WPPSI	Wechsler Preschool and Primary Intelligence test

Introduction

Developmental and Epileptic Encephalopathies (DEEs)

Epilepsy: definition and classification

An epileptic seizure is a transient occurrence of signs and/or symptoms due to abnormal excessive or synchronous neuronal activity in the brain. [1]
The definition of epilepsy has been remodeled during years. Epilepsy previously has been defined as at least two unprovoked seizures >24 h apart. [1] The current definition implies that epilepsy also can be considered to be present after one unprovoked seizure in individuals who have other factors that are associated with a high likelihood of a persistently lowered seizure threshold and therefore a high recurrence risk. [2]

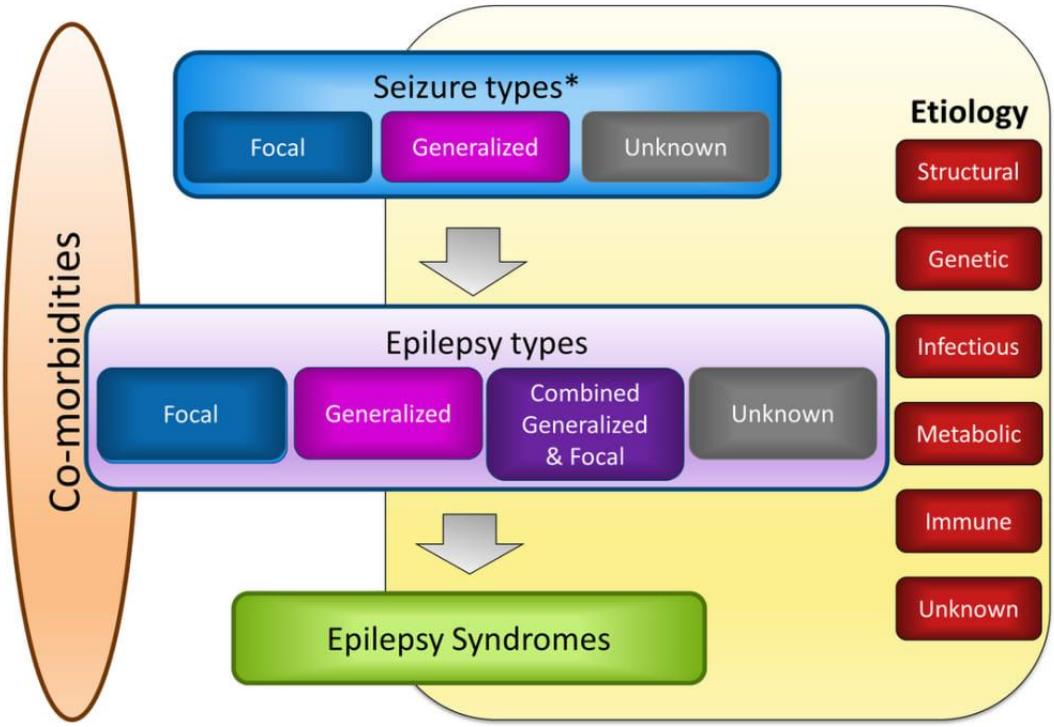
The appearance of epilepsy in an individual may have different implications related to the direct consequences of seizures and/or epileptic activity, and its underlying cause. The clinical picture of patients with epilepsy can therefore be very variable, in both inter-individual and intra-individual manner.

Epilepsy classification is the key clinical tool in evaluating an individual who is presenting with seizures. Classification serves many purposes: providing a framework for understanding the type of seizures that the patient has, the other seizure types that are more likely to occur in that individual, the potential triggers for their seizures, and often their prognosis. [3]
Classification also informs the risks of comorbidities including learning difficulties, intellectual disability, psychiatric features such as autism spectrum disorder, and mortality risk such as sudden unexpected death in epilepsy (SUDEP). It is notable that classification often guides the selection of antiepileptic therapies. [3]

The classification of epilepsies is organized into three different levels (Figure 1). The first level includes the classification of seizures' type (i.e., focal onset, generalized onset and unknown onset). The second level is devoted to the diagnosis of epilepsy type (i.e., focal epilepsy, generalized epilepsy, combined generalized and focal epilepsy, and unknown epilepsy group). The third level concerns the diagnosis of epilepsy syndrome, where a complete syndromic diagnosis can be made. [3] Classifications of both seizure type and epilepsy type consider the results of investigations conducted by using electroencephalography (EEG) and neuroimaging studies, together with other studies that explore the underlying etiology of the epilepsy. [3]

An epilepsy syndrome is a characteristic cluster of clinical and EEG features, often supported by specific etiological findings. The diagnosis of a syndrome in an individual with epilepsy frequently carries prognostic and treatment implications. Syndromes can be subdivided into those with generalized onset seizures, focal onset seizures, generalized and focal onset seizures, and developmental and/or epileptic encephalopathy or progressive neurological deterioration. [4]

Figure 1. ILAE Classification of epilepsies (from Scheffer et al., 2017)



Developmental and Epileptic Encephalopathies: definition and approaches to care

The term “epileptic encephalopathy” is defined as a condition where the epileptic activity itself contributes to severe cognitive and behavioral, impairments above and beyond what might be expected from the underlying pathology alone. [5]

In the 2017 ILAE Classification of Epilepsies, the new term of "developmental and epileptic encephalopathy" was promoted, referring to when cognitive functions are influenced by both the neurobiological process behind the epilepsy and seizure and interictal epileptiform activity. [3]

Usually, DEEs appear during early infantile or childhood periods, although any age group inclusive of adults can be affected. Most common causes of DEEs are genetic; other etiologies include structural (e.g. tuberous sclerosis complex (TSC), hypothalamic hamartomas, hemimegalencephaly), metabolic (e.g. pyridoxine or biotinidase deficiency, GLUT1 deficiency) and immune disorders (e.g. Rasmussen syndrome). [6]

In the clinical setting, the detection of a DEE entails a series of implications.

In fact, differently from the epileptic encephalopathies in which the clinical picture is normal at onset and usually returns to normality once resolved, DEEs are accompanied by comorbidities that tend to be chronic and possibly fluctuating in intensity in function of the epileptic activity.
Comorbidities are various and are often related to the underlying etiology, affecting cognitive and neurovegetative functions, behavior, language, movement, muscular tone, skeletal muscle apparatus, sleep, alimentation.
Nevertheless, epilepsy in DEE is often drug resistant, leading, in some cases, to frequent and prolonged hospitalizations due to long-lasting (convulsive) seizures and their complications (e.g. aspiration pneumonia, injuries, etc.).
All this, results in a severe impact on quality of life of affected patients and their caregivers. [7]

Dravet Syndrome (DS)

Definition and main clinical features

Dravet Syndrome is a genetic developmental and epileptic encephalopathy. [3] According to the ILAE classification (1989), typical DS is characterised by *febrile and afebrile generalised and unilateral, clonic or tonic-clonic, seizures, that occur in the first year of life in an otherwise normal infant and are later associated with myoclonus, atypical absences, and partial seizures. All seizure types are resistant to antiepileptic drugs. Developmental delay becomes apparent within the second year of life and is followed by definite cognitive impairment and personality disorders*. [8]
In variable percentages (25% to 71%), there is a family history of epilepsy or febrile seizures (FS). [9] [10] [11] [12]
Onset is during the first year of life, but sporadically patients who presented a first seizure after one year were reported. [13] [14]
Psychomotor development is normal during the first year, then a delay becomes evident. An unsteady gait develops after children start walking at a normal age. [15]
Language also starts at a normal age but progresses very slowly and many subjects cannot construct elementary sentences. [16] [17]
Fine motor abilities do not develop well, and patients are disturbed by myoclonus and lack of good hand-eye coordination, which prevents them from acquiring the ability to carry out daily activities and play. [18] [19] Autistic traits can be observed, but only a few children are actually autistic. [15]
Neurological signs also occur. Hypotonia is the most precocious sign, detectable around one year of age. [20] Ataxia appears when patients start walking and was reported in about 80%, persisting in more than one third of adults. [15] [20] [21] Pyramidal signs are found in 50%, and they tend to increase in periods with repeated convulsive seizures and decrease when interictal intervals become

longer. [20] [21] Interictal myoclonus is very frequent and appears at the same time as the myoclonic seizures, between the age of one and five, being also in many patients who do not experience them. [20] [22] [23]

Autonomic symptoms (as cold extremities, red or bluish colour of the extremities, alterations in sweating, large pupils, slow emptying of the stomach, episodes of tachycardia, facial and chest flushing) are also reported. [24]

In adults, extrapyramidal signs, including rigidity, dystonic postures and dystonic gait are reported; these patients have reduced motor abilities due to the combination of neurological signs and skeletal abnormalities, which can lead to a peculiar posture developing, "crouched gait".

Kyphosis and kyphoscoliosis, claw, and flat feet contribute to walking difficulties too. [15]

In addition to conditions described above, DS phenotype includes a series of other comorbidities.

Behavioural issues and ADHD are present in consistent percentages, setting in by the end of the first year. [18]

More than 70% of patients with DS have sleep problems, especially in sleep-wake transition disorders, and difficulty in initiating and maintaining sleep. [25] Sleep issues are often already present in the first 2 years of life. [26]

A major susceptibility to infection is also associated with DS, especially infections of the upper respiratory tract and pneumonia, that can be severe. However immunological studies did not show abnormalities that could explain it. [15] [27] [28]

No gross abnormalities were detected on early brain CT or MRI studies, although signs of slight or moderate, diffuse, cerebral or cerebellar atrophy, increased white matter signal have been occasionally reported. [20]

Epidemiology

DS is a rare disorder (classified as ORPHA 33069 in the Orphanet classification). Worldwide incidence has been reported to be between 1/40.000 and 1/15.700. [29] [30] [31] [32]

To date, it is likely that DS is underdiagnosed in adults. [33] [34] Given that most frequent mutations leading to DS arise *de novo*, it is expected that incidence of DS in developing countries reflects that of international studies; however, a limited number of studies describe incidence in these countries, where genetic epilepsies are among probably underdiagnosed and under-investigated, due to the fact that frequent seizures are often result of endemic parasitic disease, central nervous system infections, traumatic brain injury or perinatal insults. [35]

In some studies, males were more often affected than females in a ratio 2:1; in others, an equal percentage was seen between both genders. [29] [30] [36]

DS as monogenic encephalopathy

A clinical diagnosis of Dravet Syndrome is supported by the finding of SCN1A gene mutations in 85% of patients; 40% are missense mutations, 40% are truncating mutations, and the remaining are splice site changes. [12] [15] Most mutations are *de novo*, but familial mutations occur in 5-10% of cases and are usually missense; in these cases, family members with the same mutation have milder phenotype, consistent with GEFS+ spectrum. [23] Actually, since the same mutations have been associated with mild or severe types, a genotype-phenotype correlation is not consistently attempted. Other genes that have been reported to cause DS-like phenotypes include SCN2A, SCN8A, SCN9A, SCN1B, PCDH19, GABRA1, GABRG2, STXBP1, HCN1, CHD2 and KCNA2. [37]

Epilepsy in DS

DS course can be divided into three stages: a "febrile phase" in the first year, a "worsening phase" of variable duration, and a "stabilisation" phase; the age limit between the second and third phase are unclear. [38] [15]

In the "febrile phase", epilepsy starts as febrile convulsions in most patients. Body temperature variation is the main seizure trigger at onset; seizures often appear below the temperature level associated with fever and sometimes temperature increases after the end of the seizure. External temperatures can also represent an aggravating factor and some patients have more seizures in summer and in hot regions. Temperature sensitivity is also apparent when the child takes a hot bath. [15] In this period, seizures are clonic, generalized, focal or unilateral, prolonged in one third of the cases. 80% of patients have hemi-clonic and/or focal seizures, lasting over 15 min. [39] Convulsive status epilepticus is frequent.

From the second year, in the "worsening phase", motor seizures become more frequent but shorter, although their severity is still linked to mild hyperthermia. Then, myoclonic seizures and atypical absences appear and "obtundation" (non-convulsive) status may occur. [39]

In the "stabilization phase" seizures are less frequent, often occurring in sleep as tonic seizures in some [40] and in milder types they become rare, but rarely disappear completely. Atypical absences and myoclonic seizures often disappear, and cognitive troubles represent the main issue in this period.

At onset, EEG background activity is usually normal; in some cases, a rhythmic theta activity of 4-5 Hz is noted over the Rolandic and vertex areas. [20] [22] Background activity remains normal or slightly abnormal in 50% of cases; in the remaining cases it becomes slow and poorly organised, especially in period of multiple seizures. [20] When present, epileptic discharges are focal, multifocal

or generalised, and there is no relationship between the site of interictal anomaly and the site of seizure origin. [15] Sleep activity is usually well structured.

Photosensitivity has been reported as one of the main features of DS, occurring especially in patients with massive myoclonus; [20] [21] [39] it is often difficult to analyse it because it does not remain constant during the course of the disease. [15] In Dravet series of 31 patients, photosensitivity had onset between 3 months and 5 years and six months; in 7 patients duration was estimated from 3 months to 11 years and 6 months. [41] Photosensitivity persists in some cases and was observed in 12% of patients older than 18 years. [42]

"Next generation phenotyping"

Electronic health records (EHR)

An electronic health record (EHR) is the systematized collection of patient and population electronically-stored health information in a digital format. [43] EHRs may include a range of data, including demographics, medical history, medication, allergies, immunization status, laboratory test results, radiology images, vital signs, personal statistics like age and weight, and billing information. Comparing with paper-based records, EHRs offer a wide range of advantages.

Electronic records are more easily accessible by healthcare professionals. Indeed, EHRs are widely used for synchronizing patient records provided at different facilities, with many benefits for healthcare delivery. [44] [45] [46]

Remote medical services have rapidly expanded in the last decade; the use of telemedicine is making it easier for more people to access medical care (i.e., patients with restricted mobility) and can cut healthcare costs for patients and providers alike. EHR programs are essential in telemedicine to quickly access to patient information. [47]

Also teleradiology is a growth technology which employs EHRs: patient's radiological images (x-rays, CTs and MRIs) are transmitted from one location to another, for the purpose of sharing studies with other radiologists and physicians. [48]

Furthermore, an easy accessibility to personal medical data represents an advantage for patients; using smartphones or tablets for example, an instant access to their own health records at any time and any place can be possible. [49]

Nevertheless, digitization of forms facilitates the storage in clinical warehouses and the collection of data for several uses, as economic, epidemiologic or clinical studies.

Clinical Data Warehouses (CDW) and Medical Data Mining in health care

A CDW is a real time database that combines data from multiple clinical sources. It allows physicians to recover data for a single patient or to identify a cohort of subjects with common features. Typical data types which are often found within a CDW include: clinical laboratory test results, patient demographics, pharmacy information, radiology reports and images, pathology reports, hospital admission, discharge and transfer dates, ICD codes, discharge summaries and progress notes. [50]
CDW are used in the hospital setting for several purposes, for example to track prescribing trends, monitoring infectious diseases, monitoring the prescription of antibiotics, and so on. [51] The use of CDW could provide a wealth of knowledge about patients, their medical condition and outcome.
The database serves as a way to study the relationship and potential patterns between disease progression and management. The term "Medical Data Mining" has been coined for this method of research, and its use could serve as a valuable resource helping the future of healthcare in all facets of medicine. [52]
Data mining is the process of pattern discovery and extraction where huge amount of data is involved. The exponential growth of digital clinical data and clinical warehouse matches with the employ of data mining as new type of clinical research.
There is a vast potential for data mining applications in healthcare. [53] For example, they can serve to compare healthcare groups of parameters, such as resource utilization and lengths of stay, and to individuate areas for quality improvements. [54] Possible applications are in costs management, customer relationship management (for example: waiting times), and in detection of fraud and abuse. [55]
Among the most interesting applications, there is the chance to evaluate different aspects of a treatment: by comparing and contrasting causes, symptoms, and course of treatments, data mining can deliver an analysis of which course of action proves an effectiveness or bring out unknown side effects. [55] Particularly intriguing, given also the chance to track chronic disease, is the possibility of determining the most effective drug compounds for treating small populations that respond differently from the mainstream to certain drugs, and determining proactive steps that can reduce the risk of affliction. [55]
Nevertheless, by collating common symptoms data mining can also aid to elaborate diagnosis. [55]

Data mining and Machine Learning in rare diseases

Rare diseases occur in small percentage of population, and those people are often spread around the world. It is therefore clear that application of data mining is particularly providential for research in the field of rare diseases.

Cumulating data from all over the world could help to reach diagnosis, to know patients and track them, to highlight clinical features that are still scarcely known regarding symptoms, treatment and outcome. This point is particularly intriguing considering that the most common and important applications in data mining involve predictive modelling. [53]

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use in order to perform a specific task effectively without using explicit instructions, relying on patterns and inference instead of basing on sample data. [56] While data mining focuses on the discovery of previously unknown properties in the data, machine learning focuses on prediction, based on known properties learned from the training data.

In future, machine learning will dramatically improve the ability of health professionals to establish a prognosis. [57] In the clinical setting, prognosis is obviously elaborated on the basis of current knowledge, that in the rare disease context is often scarce and inadequate; by handling enormous numbers of predictors and combining them in a nonlinear and highly interactive ways, machine learning could help to make predictions where human knowledge has not yet arrived. [57] However, the limit to the application of machine learning in the field of rare diseases to date is still represented by the need to have millions of observations to reach acceptable performance levels; [58] hence, the need to create more and more rare diseases data sharing networks.

Furthermore, "*letting the data speak for themselves can be problematic: algorithms might overfit predictions to spurious correlations in the data, or multiple collinear, correlated predictors could produce unstable estimates. Either possibility can lead to overly optimistic estimates of the accuracy of a model and exaggerated claim about real-world performance. These concerns are serious and must be addressed by testing models on truly independent validation data sets, from different populations or periods that played no role in model development*". [57]

Natural Language Processing (NLP) and Unified Medical Language System (UMLS)

In EHRs, clinical data can be represented in structured or unstructured form. Structured data are created through constrained choices in the form of data entry devices (drop-down menus, check boxes and pre-filled templates). This type of data offer the advantage of easy searchability, reportability and analysing. Unstructured clinical data exist in the form of free text narratives; free text entries into the patient's health record allow providers to note comments or notions that are not provided by the constrained choice. [59]

Natural language processing (NLP) is a technology that allows the use of unstructured data by extracting data from free text. In the clinical setting NLP converts providers' notes and narratives into structured, standardized formats, for example exploiting UMLS. [60] [59]

The UMLS, or Unified Medical Language System, is a set of files and software that brings together many health and biomedical vocabularies and standards to enable interoperability between computer

systems. [61] English is mostly considered as the lingua franca of medical vocabulary, and it is assembled by integrating 153 medical vocabularies, including specialized terminologies (e.g. the Human Phenotype Ontology – HPO). [62] The UMLS consists of three components: Metathesaurus, a large multi-lingual vocabulary that contains millions of biomedical and health related concepts, their synonymous names, and their relationships; Semantic Network, composed of a set of broad subject categories, or *semantic types*, that provide a consistent categorization of all concepts represented in the Metathesaurus, and a set of useful and important relationships, or *semantic relations*, that exist between semantic types; the specialist Lexicon, developed to provide the lexical information needed for the NLP. [61] UMLS Metathesaurus contains about 3.2 million concepts identified by their unique identifier (the Concept Unique Identifier: CUI). A concept is a *cluster of synonymous terms* coming from various source vocabularies, also from different languages.
Using NLP and UMLS technologies, standardized clinical data can then be extracted from a medical text, and be used for different purposes.

References

[1] R. Fisher, W. van Emde Boas, W. Blume and e. al., " Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE)," *Epilepsia,* vol. 46, p. 470–472, 2005.

[2] R. Fisher, C. Acevedo, A. Arzimanoglou, A. Bogacz, J. Cross, C. Elger, J. J. Engel, L. Forsgren, J. French, M. Glynn, D. Hesdorffer, B. Lee, G. Mathern, S. Moshé, E. Perucca, I. Scheffer, T. Tomson, M. Watanabe and W. S., "ILAE official report: a practical clinical definition of epilepsy," *Epilepsia,* vol. 55, no. 4, pp. 475-482, 2014.

[3] I. Scheffer, S. Berkovic, G. Capovilla, M. Connolly, J. French, L. Guilhoto, E. Hirsch, S. Jain, G. Mathern, S. Moshé, D. Nordli, E. Perucca, T. Tomson, S. Wiebe, Y. Zhang and S. Zuberi, "ILAE classification of the epilepsies: Position paper of the ILAE Commission for Classification and Terminology," *Epilepsia,* vol. 58, no. 4, pp. 512-521. 2017.

[4] E. Wirrell, R. Nabbout, I. Scheffer, T. Alsaadi, A. Bogacz, J. French, E. Hirsch, S. Jain, S. Kaneko, K. Riney, P. Samia, O. Snead, E. Somerville, N. Specchio, E. Trinka, S. Zuberi, S. Balestrini, S. Wiebe, H. Cross, E. Perucca, L. Moshém and T. P., "Methodology for clasification and definition of epilepsy syndromes with list of syndromes: Report of the ILAE Task Force on Nosology and Definitions," *Epilepsia,* vol. 63, no. 6, pp. 1333-1348, 2022.

[5] A. Berg, S. Berkovic, M. Brodie and e. al., “Revised terminology and concepts for organizaion of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005-2009,” *Epilepsia,* vol. 51. pp. 676-685, 2010.

[6] S. Raga, N. Specchio, S. Rheims and W. JM, “Developmental and epileptic encephalopathies: recognition and approaches to care.,” *Epileptic Disorders,* vol. 23, no. 1. pp. 40-52, 2021.

[7] K. Gallop, A. Lloyd, J. Olt and M. J, “Impact of developmental and epileptic encephalopathies on caregivers: a literature review,” *Epilepsy Behav.,* vol. 124, p. 108324, 2021.

[8] ILAE, “Proposal for revised classification of epilepsies and epileptic syndromes. Commission on Classification and Terminology of the ILAE,” *Epilepsia,* vol. 30. pp. 389-99, 1989.

[9] T. Ohki, K. Watanabe, K. Negoro and e. al., “Severe myoclonic epilepsy in infancy: Evolution of seizures,” *Seizures,* vol. 6, pp. 219-24, 1997.

[10] F. Ragona, D. Brazzo, I. De Giorgi and e. al, “Dravet syndrome: Early clinical manifestations and cognitive outcome in 37 Italian patients,” *Brain Development,* vol. 32, pp. 71-77, 2010.

[11] M. Mancardi, P. Striano, E. Gennaro and e. al, “Familial occurrence of febrile seizures and epilepsy in severe myoclonic epilepsy of infancy patients with SCN1A mutations,” *Epilepsia,* vol. 47, pp. 1629-1635, 2006.

[12] C. Marini, D. Mei, T. Temudo and e. al, “Idiopathic epilepies with seizures precipitated by fever and SCN1A abnormalities,” *Epilepsia,* vol. 48, pp. 1678-1696, 2007.

[13] D. Hurst, “Severe myoclonic epilepsy in infancy,” *Pediatric Neurology,* vol. 3, pp. 269-272, 1987.

[14] J. Kearney, A. Wiste, U. Stephani and e. al, “Recurrent de novo mutations of SCN1A in Severe myoclonic epilepsy of infancy,” *Pediatric Neurology,* vol. 34, no. 2, pp. 116-120. 2006.

[15] C. Dravet and R. Guerrini, Dravet Syndrome, John Libbey Eurotext, 2011.

[16] P. Wang, P. Fan, W. Lee, C. Young, C. Huang and Y. Shen, “Severe myoclonic epilepsy in infancy: evolution of electroencephalographic and clinical features,” *Acta Paed Sin,* vol. 37, pp. 428-432, 1996.

[17] R. Caraballo and N. Fejerman, “Dravet Syndrome: A study of 53 patients,” *Epilepsy Res,* vol. 70. no. Suppl. 1. pp. S231-238, 2006.

[18] C. Cassé-Perrot, M. Wolff and C. Dravet, “Neuropsychological aspects of severe myoclonic epilepsy in infancy,” in *The Neuropsychology of Childhood Epilepsy*, New York, Plenum Press/Kluwer Academic, 2001. pp. 131-140.

[19] M. Wolff, C. Cassé-Perrot and C. Dravet, “Severe myoclonic epilepsy of infants (Dravet syndrome): Natural history and neuropsychological findings,” *Epilepsia,* vol. 47, no. Suppl.2, pp. 45-48, 2006.

[20] B. Dalla Bernardina, G. Capovilla, M. Gattoni, V. Colamaria, S. Bondavalli and M. Bureau, “Epilepsie myoclonique grave de la première année,” *Rev EEG Neurophysiol,* vol. 12, pp. 21-25, 1982.

[21] C. Dravet, M. Bureau and J. Roger, “Severe myoclonic epilepsy in infancy,” in *Epileptic Syndromes in Infancy, Childhood and Adolescence*, London, John Libbey, 1985, pp. 58-67.

[22] P. Giovanardi-Rossi, M. Santucci, G. Gobbi, A. Parmeggiani, A. Pini and G. Ambrosetto, “Long-term follow-up of severe myoclonic epilepsy in infancy,” in *Modern perspectives of child neurology*, Tokyo, Asahi Daily News, 1991. pp. 205-213.

[23] R. Nabbout, E. Gennaro and B. e. a. Dalla Bernardina, “Spectrum of SCN1A mutations in severe myoclonic epilepsy of infancy,” *Neurology,* vol. 60. pp. 1961-1967, 2003.

[24] B. Wical, D. Leighty, H. Wendorf, M. Tervo, R. Tervo and J. Maytum, “Signs of dysautonomia in children with Dravet syndrome,” *Epilepsia,* vol. 50. no. 3, p. 164, 2009.

[25] S. Licheni, J. Mcmahon, A. Schneider, M. Davey and I. Scheffer, “Sleep problems in Dravet syndrome: a modifiable comorbidity,” *Developmental medicine and child neurology,* 2017.

[26] A. Schoonjans, S. De Keersmaecker, M. Van Bouwel and B. Ceulemans, “More daytime sleepiness and worse quality of sleep in patients with Dravet Syndrome compared to other epilepsy patients,” *Official Journal of the European Paediatric Neurology Society,* vol. 23, pp. 61-69, 2018.

[27] M. Nieto, S. Roldan, B. Sanchez and e. al, “Estudio immunologico en pacientes con epilepsia mioclonica severa en la infancia,” *Rev Neurol (Spanish),* vol. 30. pp. 1-15, 2000.

[28] H. Oguni, K. Hayashi, Y. Awaya, Y. Fukuyama and M. Osawa, “Severe myoclonic epilepsy in infants - a reviewed based on the Tokyo Womes's medical university series of 84 cases,” *Brain Development,* vol. 23, pp. 736-48, 2001.

[29] D. Hurst, “Epidemiology of severe myoclonic epilepsy of infancy,” *Epilepsia,* vol. 31. pp. 397-400. 1990.

[30] L. Yakoub, O. Dulac, I. Jambaqué and P. Plouin, “Early diagnosis of severe myoclonic epilepsy in infancy,” *Brain Development,* vol. 14, pp. 299-303, 1992.

[31] A. Bayat, H. Hjalgrim and R. Moller, “The incidence of SCN1A-related Dravet syndrome in Denmark is 1:22,000: a population-based study from 2004 to 2009,” *Epilepsia,* vol. 56, no. 4, pp. e36-39, 2015.

[32] Y. Wu, J. Sullivan, S. McDaniel, M. Meisler, E. Walsh, S. Li and M. Kuzniewicz, “Incidence of Dravet Syndrome in a US Population,” *Pediatrics,* vol. 136, no. 5, pp. e1310-1315, 2015.

[33] F. Jansen, L. Sadlei, L. Harkins, L. Vadlamudi, S. Berkovic and e. al, “Severe myoclonic epilepsy of infancy (Dravet syndrome): Recognition and diagnosis in adults,” *Neurology,* vol. 67, pp. 2224-2226, 2006.

[34] M. Connolly, “Dravet Syndrome and Long-Term Course,” *Can J Neurol Sci,* vol. 43, no. Suppl 3, pp. S3-8, 2016.

[35] A. Esterhuizen, H. Mefford, R. Ramesar, S. Wang, G. Carvill and J. Wilmshurst, “Dravet Syndrome in South Africa infants: Tools for an early diagnosis,” *Seizure: European Journal of Epilepsy,* vol. 62, pp. 99-105, 2018.

[36] J. Skluzacek, K. Watts, O. Parsy and e. al, “Dravet syndrome and parent associations: The IDEA League experience with comorbid conditions, mortality, management, adaptation, and grief,” *Epilepsia,* vol. 52, no. Suppl. 2, pp. 95-101. 2011.

[37] D. Steel, J. Sysmonds, S. Zuberi and A. Brunklaus, “Dravet syndrome and its mimics: Beyond SCN1A,” *Epilepsia,* vol. 58, no. 11. pp. 1807-1816, 2017.

[38] I. Lambarri San Martin, C. Garaizar Axpe, E. Zuazo Zamalloa and e. al, “Epilepsia polimorfa de la infancia: revision de 12 casos,” *Anales Espanoles de Pediatria,* vol. 46, pp. 571-575, 1997.

[39] S. Gataullina and O. Dulac, “From genotype to phenotype in Dravet disease,” *Seizure,* vol. 44, pp. 58-64, 2017.

[40] E. Losito, M. Kuchenbuch, N. Chemaly, J. Laschet, C. Chiron, A. Kaminska and R. Nabbout, “Age-related "Sleep/nocturnal" tonic and tonic clonic seizure clusters are underdiagnosed in patients with Dravet Syndrome,” *Epilepsy Behav,* vol. 74, pp. 33-40. 2017.

[41] C. Dravet, M. Bureau, R. Guerrini, N. Giraud and J. Roger, “Severe myoclonic epilepsy in infants,” in *Epileptic Syndrome in Infancy, Childhood and Adolescence*, 75-88, Ed., London, John Libbey, 1992.

[42] M. Bureau and B. Dalla Bernardina, “Electroencephalographic characteristics of Dravet Syndrome,” *Epilepsia* , vol. 52, no. Suppl. 2, pp. 13-23, 2011.

[43] T. N. Gunter TD, “The emergence of national electronic health record architectures in the United States and Australia: models, costs, and questions,” *Journal of Medical Internet Research,* vol. 7, no. 1. p. e3, 2005.

[44] V. Papadouka, P. Schaeffer, A. Metroka, A. Borthwick, P. Tehranifar, J. Leighton, A. Aponte, R. Liao, A. Ternier, S. Friedman and N. Arzt, “Integrating the New York citywide

immunization registry and the childhood blood lead registry," *Journal of Public Health Management and Practice,* no. Suppl, pp. S72-80. 2004.

[45] P. Gioia, "Quality improvement in pediatric well care with an electronic record," in *Proceedings. AMIA Symposium*, 2001.

[46] S. Williams and W. Hollinshead, "Perspectives on integrated child health infromation systems: parents, providers, and public health," *Journal of Public Health Management and Practice,* no. Suppl, pp. S57-60. 2004.

[47] K. McCarthy, "The role of her in telemedicine," 2016. [Online].

[48] www.teleradiology-center.eu, "Teleradiology Center," 2015. [Online].

[49] M. Torrieri, "EHRs Go Mobile," 2012. [Online].

[50] S. Mac Kensie, M. Wyatt, R. Schuff, J. Tenenbaum and N. Anderson, "Practices and perspectives on building integrated data repositories: Results from 2010 CTSA survey," *Journal of the American Medical Informatics Association,* vol. 19, no. e1. pp. e119-e124, 2012.

[51] M. Samore, D. Lichtenberg, L. Saubermann, C. Kawachi and Y. Carmeli, "A Clinical Data Repository enhances hospital infection control," *Harvard Medical School,* pp. 56-60. 1997.

[52] J. Prather, D. Lobach, L. Goodwin, J. Hales, M. Hage and W. Hammond, "Medical Data Mining: Knowledge Discovery in a Clinical Data Warehouse," *Duke University Medical Center,* pp. 101-105, 1997.

[53] H. Koh and G. Tan, "Data mining applications in healthcare," *Journal of Healthcare Information Management,* vol. 19, no. 2, pp. 64-72, 2005.

[54] B. Schuerenberg, "An information excavation," *Health Data Management,* vol. 11. no. 6, pp. 80-82, 2003.

[55] A. Milley, "Healthcare and data mining," *Health Management Technology,* vol. 21. no. 8, pp. 44-47, 2000.

[56] C. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[57] Z. Obermeyer and E. Emanuel, "Predicting the Future - Big Data, Machine Learning, and Clinical Medicine," *The New England Journal of Medicine,* vol. 375, no. 13, pp. 1216-1219, 2016.

[58] A. Halevy, P. Norvig and F. Pereira, "The unreasonable effectiveness of data," *IEEE Intell Syst,* vol. 24, no. 2, pp. 8-12, 2009.

[59] H. Townsend, "Natural Language Processing and Clinical Outcomes: The Promise and Progress of NLP for Improved Care," *Journal of AHIMA,* vol. 84, no. 2, pp. 44-45, 2013.

[60] T. Rindflesch and M. Fiszman, “The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text,” *Journal of Biomedical Informatics,* vol. 36, no. 6, pp. 462-477, 2003.

[61] O. Bodenreider, “The Unified Medical Language System (UMLS): integrating biomedical terminology.,” *Nucleic Acids Res,* 2004.

[62] N. Garcelon, A. Neuraz, R. Salomon, N. Bahi-Buisson, J. Amiel, C. Picard, N. Mahlaoui, V. Benoit, A. Burgun and B. Rance, “Next generation phenotyping using narrative reports in a rare disease clinical data warehouse,” *Orphanet Journal of Rare Diseases,* vol. 13, no. 85, 2018.

Aims and thesis overview

Developmental and epileptic encephalopathies (DEEs) show considerable variability in terms of initial presentation; type and course of epilepsy; kind, appearance, and evolution of comorbidities; and life expectancy. Because of their rarity, diagnosis is often delayed, and the long-term outcome is unknown in most cases.

In the spirit of precision medicine, due to the rapid evolution of diagnostic techniques (molecular, biochemical, instrumental) in recent years, there has been a significant increase in the etiological diagnosis of DEEs. This development allows a new approach in terms of treatment and outcome definition, combining the syndromic framing of the epilepsy with the precise etiological characterization. To date, diagnostic criteria and phenotype description of rare diseases are mainly based on clinical data collections from multicenter studies, which are formed in the light of current knowledge. There is therefore a need to find new strategies to define new conditions to keep pace with the rapid progresses in the field of etiological definition for early development of targeted treatments.
In this thesis, Dravet syndrome (DS), an archetype of DEE, was used as a model to explore the feasibility of novel strategies for diagnosing and outlining the natural history with the purpose of extending the strategies to other lesser-known DEEs.

Disease registries are gaining increasing importance and are significantly contributing to enlarging the knowledge base around some rare diseases. Considering the rarity of DEEs, sharing data is crucial. We described the Residras registry (**chapter 1**), an international register dedicated to DS and other phenotypes related to SCN1A mutations, as a paradigm of registry for DEEs. This registry platform represents an invaluable asset for delineating natural history and for easing access to cohorts of patients that may benefit from clinical trials.

However, the clinical characterization of subjects affected by rare conditions can be only partial if acquired through structured data. Indeed, structured data only allow the collection of limited information and could result in the missing of rare and novel symptoms and signs. In **chapters 2 and 3**, we explored the use of natural language processing (NLP) in unstructured electronic medical records of patients with DS with the aim to individuate early diagnosis markers and retrace the natural history. This methodology would suggest an innovative, inexpensive, transposable, and sustainable approach to reduce time of diagnosis and outline the natural history for individuals with other DEEs.

Patients' needs vary over time in relation to the disease course and are part of the natural history itself. The US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) have encouraged patient reported outcomes as self-assessment of affected individual's health status. With this recommendation and acknowledging the centrality of the caregivers in the care management of individuals with DEEs, we explored caregivers' opinions on the outcome measures that matter in clinical trials in individuals with DS (**chapter 4**). In addition to seizures, the domains that a therapy in development for DS should target are highlighted in this study. Furthermore, domains are replicable for other DEEs.

The study of the adaptive functioning beyond the cognitive level and the impact of the disease in daily abilities are consistent in parts of the definition of the long-term outcome. In **chapter 5**, we explored the feasibility of using an adaptive behavior profile (ABP) assessment generated from a well-known measure as an instrument for outcome measures in adolescents and adults with DS. Results underline the need for different approaches in the assessment of different functions in comparison to those used for the general population and enhance the need for specific tools for individuals with a DEE.

Chapter 1 ": "A registry for Dravet syndrome: the Italian experience"[1]

Abstract

Objectives

[1] Simona Balestrini[a,b], Viola Doccini[a], Sabrina Giometto[c], Ersilia Lucenteforte[c], Salvatore De Masi[d], Elisa Giarola[e,f], Isabella Brambilla[e,f], Federica Pieroni[g], Marco Perulli[h], Domenica Battaglia[h], Nicola Specchio[i], Francesca Ragona[j], Tiziana Granata[j], Simona Pellacani[a], Annarita Ferrari[k], Carla Marini[l], Sara Matricardi[l], Elisabetta Cesaroni[l], Lucio Giordano[m], Patrizia Accorsi[m], Vittorio Sciruicchio[n], Paolo Tinuper[o,p], Tullio Messana[o], Angelo Russo[p], Dario Pruna[q], Margherita Nosadini[r], Valentina De Giorgis[s], Davide Caputo[t], Residras collaboration group, Lucio Parmeggiani[u], Tommaso Lo Barco[v,w], Francesca Darra[v], Bernardo Dalla Bernardina[e,v], and Renzo Guerrini[a]

[a] Neuroscience Department, Meyer Children's Hospital-University of Florence, Full Member of European Reference Network EpiCARE, Florence, Italy;

[b] Department of Clinical and Experimental Epilepsy, UCL Queen Square Institute of Neurology, London, and Chalfont Centre for Epilepsy, Buckinghamshire, UK;

[c] Unit of Medical Statistics, Department of Clinical and Experimental Medicine, University of Pisa, 56126 Pisa, Italy;

[d] Clinical Trial Center, University Hospital Careggi, Florence, Italy;

[e] Research Center for Pediatric Epilepsies Verona, IT, Verona, Italy;

[f] Dravet Italia Onlus, Italy; Rare Epilepsies Alliance, Italy; ERN EpiCare (Epag), Italy;

[g] Fondazione Toscana 'G. Monasterio', Pisa, Italy;

[h] Pediatric Neuropsychiatry, Dipartimento di Scienze della Salute della Donna, del Bambino e di Sanità Pubblica, Fondazione Policlinico Universitario Gemelli, IRCCS – Università Cattolica del Sacro Cuore, Rome, Italy;

[i] Department of Neuroscience, Bambino Gesù Children's Hospital, IRCCS, Full Member of European Reference Network EpiCARE, Rome, Italy;

[j] Department of Pediatric Neuroscience, Fondazione IRCCS Istituto Neurologico Carlo Besta, Full Member of European Reference Network EpiCARE, Milan, Italy;

[k] Department of Developmental Neuroscience, IRCCS Fondazione Stella Maris, Calambrone, 56128 Pisa, Italy;

[l] Department of Child Neuropsychiatry, Children's Hospital, Ancona, Italy;

[m] Paediatric Neurology and Psychiatry Unit, Spedali Civili Children's Hospital, University of Brescia, Brescia, Italy;

[n] Children Epilepsy and EEG Center, P.O. San Paolo, Bari, Italy;

[o] Department of Biomedical and Neuromotor Sciences, University of Bologna, Full Member of European Reference Network EpiCARE, Bologna, Italy;

[p] IRCCS Istituto delle Scienze Neurologiche di Bologna, UOC Neuropsichiatria dell'età pediatrica, Bologna, Italy;

[q] Child Neurology and Epileptology Unit, Paediatric Department, ARNAS Brotzu, Cagliari, Italy;

[r] Paediatric Neurology and Neurophysiology Unit, Department of Women's and Children's Health, University Hospital of Padova, Padova, Italy;

[s] Department of Child Neurology and Psychiatry, IRCSS Mondino Foundation, Pavia, Italy;

We describe the Residras registry, dedicated to Dravet syndrome (DS) and to other phenotypes related to SCN1A mutations, as a paradigm of registry for rare and complex epilepsies. Our primary objectives are to present the tools and framework of the integrative platform, the main characteristics emerging from the patient cohort included in the registry, with emphasis on demographic, clinical outcome, and mortality.

Methods

Standardized data of enrolled pediatric and adult patients were collected in 24 Italian expert centers, and regularly updated at least on a yearly basis. Patients were prospectively enrolled, at registry starting, but historical retrospective data were also included.

Results

At present, 281 individuals with DS and a confirmed SCN1A mutation are included. Most patients have data available on epilepsy (n=263) and their overall neurological condition (n=255), based on at least one follow-up update. Median age at first clinical assessment was 2 years (IQR 0-9) whilst at last follow-up was 11 years (IQR 5-18.5). During the seven year-activity of the registry, five patients died resulting in a mortality rate of 1.84 per 1000-person-years. When analyzing clinical changes over the first 5-year follow-up, we observed a significant difference in cognitive function ($P< 0.001$), an increased prevalence of behavioral disorders including attention deficit ($P< 0.001$) and autism symptoms ($P=0.04$), a significant worsening of language ($P= 0.001$) and intellectual disability ($P< 0.001$).

Significance

The Residras registry represents a large collection of standardized national data for the DS population. The registry platform relies on a shareable and interoperable framework which promotes multicenter high-quality data collection. Such integrated platform represents an invaluable asset for easing access to cohorts of patients that may benefit from clinical trials with emerging novel therapies, for drug

[t] Epilepsy Center, ASST Santi Paolo Carlo, Health Sciences Department University of Milan, Italy;

[u] Department of Pediatric Neurology, Bolzano Hospital, Bolzano, Italy;

[v] Child Neuropsychiatry, Department of Surgical Sciences, Dentistry, Gynecology and Pediatrics, University of Verona, Verona, Italy;

[w] PhD Program in Clinical and Experimental Medicine, University of Modena and Reggio Emilia, Modena, Italy.

*** Article in review**

safety monitoring, and for delineating natural history. Its framework makes it easily adaptable to other rare and complex epilepsy syndromes.

Introduction

Dravet syndrome (DS) is one of the most common developmental and epileptic encephalopathies with an incidence of 6.5 per 100 000 live births (95% CI 3.2–10.00). [1] Its hallmark clinical presentation is with prolonged, febrile and afebrile, generalized clonic or hemiclonic seizures with onset within the first year of life. Other seizure types with later onset include myoclonic, atypical absence and focal seizures. Episodes of seizure clusters and status epilepticus are frequent. In addition to seizures, most children, often during late infancy or early childhood, manifest cognitive, motor and behavioural impairment. [2] DS is caused by loss-of-function pathogenic variants in the gene coding for the α1 subunit of the sodium channel (SCN1A) which is important for action potential initiation in inhibitory GABAergic interneurons. [3] There is a wide spectrum of clinical entities caused by *SCN1A* mutations ranging from genetic epilepsy with febrile seizure plus (GEFS+) to DS, with variable disease course even within the core DS phenotype. [4]

Epilepsy is treatment-resistant although the seizure burden is highly variable. As in all severe developmental and epileptic encephalopathies, the most acceptable clinical poise should be individually established based on a balance between reduction in seizure severity and frequency and minimization of treatment-related adverse effects. [5] Controlled trials have demonstrated antiseizure efficacy of stiripentol in association with clobazam, and of add on cannabidiol and fenfluramine. [6] [7] [8] [9] However, there are no current treatments that address the overall disease, in addition to seizures. Precision medicine approaches such as antisense oligonucleotides (ASO) and adeno-associated virus (AAV)-delivered gene modulation are potential treatment options for DS which are currently being investigated on a research basis, with some now transitioning to clinical trials. [10]

Disease course is variable and not fully characterized, and the extent to which prompt treatment with the most effective medications can alter prognosis is unclear. The few available studies of outcomes in adulthood show that epilepsy severity progressively decreases from childhood to adolescence and throughout adulthood and reduced frequency of convulsive status epilepticus is associated with better seizure outcome. [11] On the other hand, seizure persistence in adolescents and adults correlates with cognitive and neurologic impairment, and there is ongoing cognitive dysfunction in adulthood, independent of seizure control. Early onset of seizures, especially myoclonic seizures, correlates with the severity of intellectual disability and language impairment. [12] [13] Most adults with DS require a considerable amount of support and are unable to live independently. [14] [15] [16] Also, the incidence of premature mortality including sudden unexpected death in epilepsy (SUDEP) is elevated

in childhood but data on this ominous outcome are lacking in adulthood. [17] [18]

Although DS is likely the most studied genetic developmental and epileptic encephalopathy and is the target of regulatory trials for orphan drug designation, there is still limited knowledge on some clinical aspects including natural history, risk of comorbidities, mortality, treatment response and safety. Such knowledge becomes crucial for the development of novel treatment strategies especially in view of identifying the potential for disease modifying effects of therapeutic interventions in the precision medicine framework.
Disease registries represent an ideal tool to move forward in research and improve knowledge in the field of rare diseases. [19]
The Italian Registry of DS and other Syndromes correlated with gene *SCN1A* and *PCDH19* mutation (Residras) was established to provide an integrative infrastructure for collection of standardized molecular and clinical data of patients with DS from national centres [20] where they were diagnosed and regularly followed-up.

The primary purposes of this study are to present: (1) the tools and framework of the integrative platform; (2) the main characteristics of the patient cohort i.e. demographic, clinical outcome, and mortality; (3) the concept of the Residras registry which we progressively adapted to the emerging diagnostic and therapeutic challenges of an epilepsy syndrome under intense clinical investigation; (4) to set the basis for a wider international reach of the registry, which is now being adopted in additional EU countries.

Methods

History of the Registry

The Dravet Registry "RESIDRAS" project started in 2010 under the auspices of the Scientific Committee of Dravet Italia Onlus, [21] an association of patients and physicians and focusing on research on DS and related syndromes. The Scientific Committee identified a minimum set of mandatory items to be included in the dataset. Members of the Scientific Committee participated with their centers in a pilot phase. Meyer Children's Hospital, in Florence, was selected as the coordinating center. Fondazione Monasterio, a public institution for healthcare research [22] with a particular experience in rare diseases registries, was identified as a partner for the platform creation, data storage, confidentiality and management of privacy. By June 2013, the dataset for the pilot phase was ready and the first patients were enrolled to test the process of data inclusion in the platform. Once the process was improved in terms of consistency, reliability, and ease, pilot centers started to enroll

all their patients, either new or in follow-up, and in April 2015 Residras was opened to all Italian centers upon their specific request and ethics committee approval.

Current format of the registry-coverage, interface for clinicians, data use and research applications

Registry coordination is entrusted by a Scientific Advisory Committee (ten expert clinicians on DS), an external Steering Committee for management ('Fondazione Toscana G. Monasterio'), an external Advisor on Rare Disease (National Centre on Rare Diseases), a patient association ('Dravet Italia Onlus') and three patients' representatives. [20]

Twenty-four centres are currently participating to data input (Figure 1). Written informed consent or assent was provided by adult patients with capacity to consent, or by families or legal guardians for children and adults lacking capacity.

The Residras registry contains clinical and epidemiological data that are compiled on a standardized online template during regular clinical visits. The information is schematically divided into the following sections: demographic features (date of birth, gender, parental details, place of birth, place of residence, willingness to be contacted on a regular basis to update the registry with longitudinal data and participate in a future clinical study), natural history (including description of pregnancy, delivery, neonatal period, neurodevelopment, neurological examination), clinical and genetic diagnosis, family history, age at seizure onset, mortality. Entries for longitudinal follow-up include: seizure assessment (i.e., type, frequency, episodes of status epilepticus, hospital admissions), neurological assessment (i.e., examination, cognition, behavior, formal neuropsychological testing), treatment (i.e., type, max dosage, duration, response, adverse events), EEG (i.e., date, background activity, epileptiform abnormalities, photosensitivity, recorded seizures), other investigations (e.g., neuroimaging). Formal cognitive assessment was performed by using the Brunet–Lezine or the Griffiths mental development scale under the age of two years, whilst for older ages the Wechsler Preschool and Primary Intelligence test (WPPSI), or the Wechsler Intelligence Scale for Children (WISC) and the Vineland Adaptive Behavioural Scales (VABS) were used. When formal testing was not available, cognition and behavior were evaluated based on descriptive assessment. Unavoidably, data were not systematically collected at the same time points for all patients, given the intrinsic design of the Registry, the different ages at first inclusion in the registry and not the always direct availability of neuropsychological expertise in the various centers.

As a quality control step, a set of mandatory fields need to be filled to save information provided in each section. Additional quality control elements include a centralized system to check for duplicate case enrolment. Residras applies the principles of Findability, Accessibility, Interoperability and

Reusability (FAIR) for humans and computers [23] thus enabling efficient analysis of data across multiple sources and making data 'as open as possible and as closed as necessary. [24] The platform has been customised for data collection specific to DS, and includes 14 out of 16 Common Data Elements (CDEs) for Rare Diseases Registration released by the Joint Research Centre of the European Commission. [25] The two elements not included are 6.3 (undiagnosed cases are not included) and 7 (there is no collection of biological samples). The ontology codes used include Unified Medical Language System (UMLS), [26] Human Phenotype Ontology (HPO),[27] Orphanet Rare Disease Ontology (ORDO). [28]

To access the system, each user is assigned a personal username and password. The online input and access to the data are restricted to health-care practitioners from each center. The access codes are generated by administrators once the user has signed a written agreement. Health-care practitioners have an online and secure access to patients' data. Only fully anonymized data are available to researchers and for analysis – all enrolled participants have a registry identification code which is automatically generated.

Investigators from expert centers who wish to access data for clinical research purposes are required to submit a research protocol to the internal Scientific Committee. When unanimous approval is obtained the proposing team can access a subset of anonymized data depending on the study requirements.

The interface layout has been designed ad hoc to facilitate navigation and allows the use of various tools integrated into the system. The Residras home page describes *SCN1A*-related conditions, the Residras platform (including aims, participating centers, statistical data, instruction to accredit a new center), and news in the field of *SCN1A*-related disease. [20]

Health-care practitioners can access with their personal login and complete or create a new patient follow-up. Patient health is summarized in a dashboard that helps physicians to display a detailed overview of the collected longitudinal data and to edit synopsizes or medical reports.

A diagnosis of DS is made based on the clinical definition and is distinct by other *SCN1A*-related epilepsies. [2] Although inclusion of patients in Residras is based on clinical criteria, for the purpose of this study we only considered DS individuals with a proven *SCN1A* mutation.

Statistical analysis

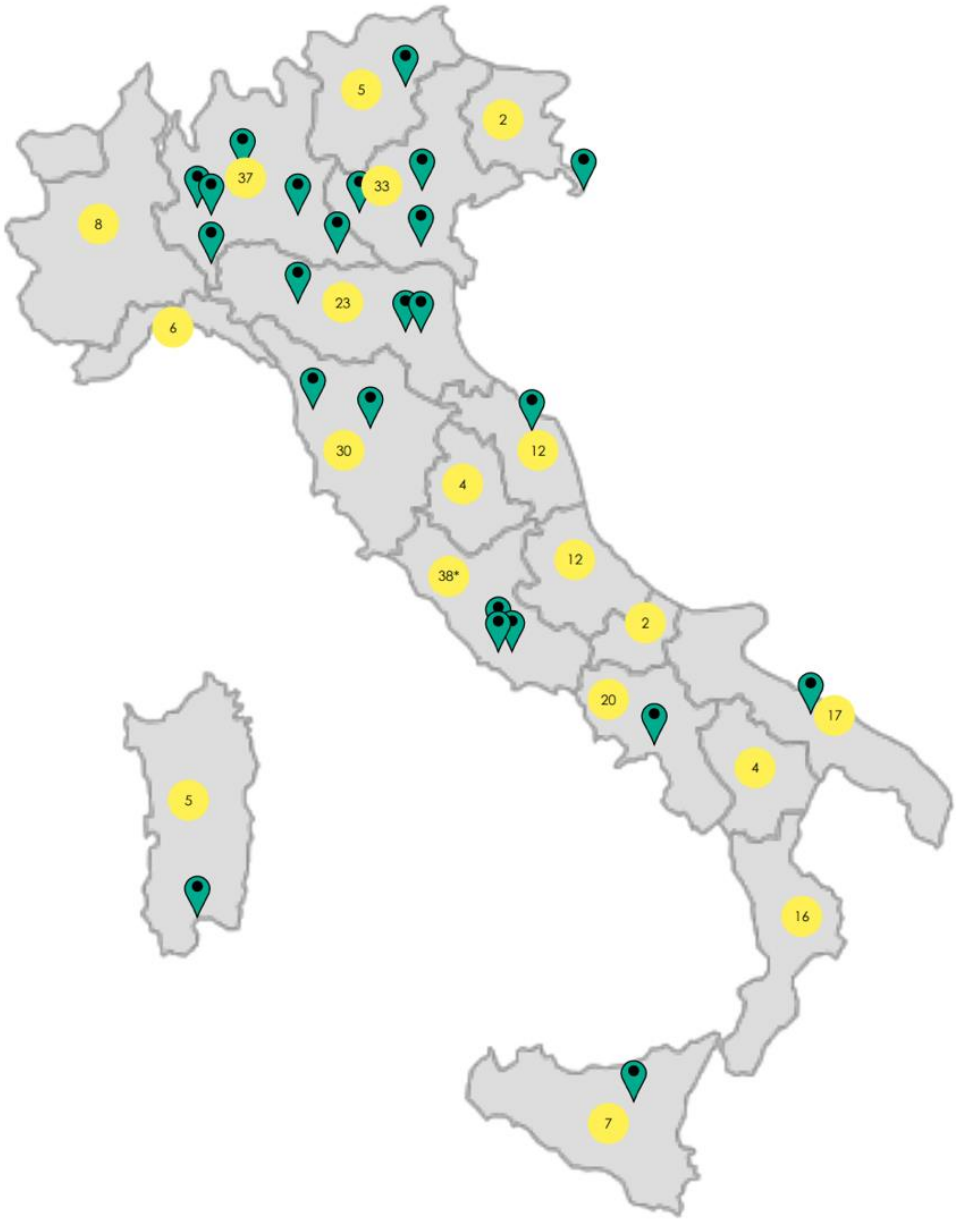
We present data as absolute number and percentage for categorical variable and median and interquartile range for continuous variables. We assessed clinical differences between first visit and after 5-year follow-up (Tables 1 and 2), and between first and last visit, using the chi-square test for independent data. When sensitivity analyses were performed using McNemar test for paired data, for subjects reporting both first and last visit data, results did not change. The mortality rate was

calculated as number of observed events (death) divided by the person time at risk. Alluvial diagrams were used to illustrate the flows of variation in clinical conditions, from the first visit (time 0), every six months until the fifth year of follow-up (time 60).

Results

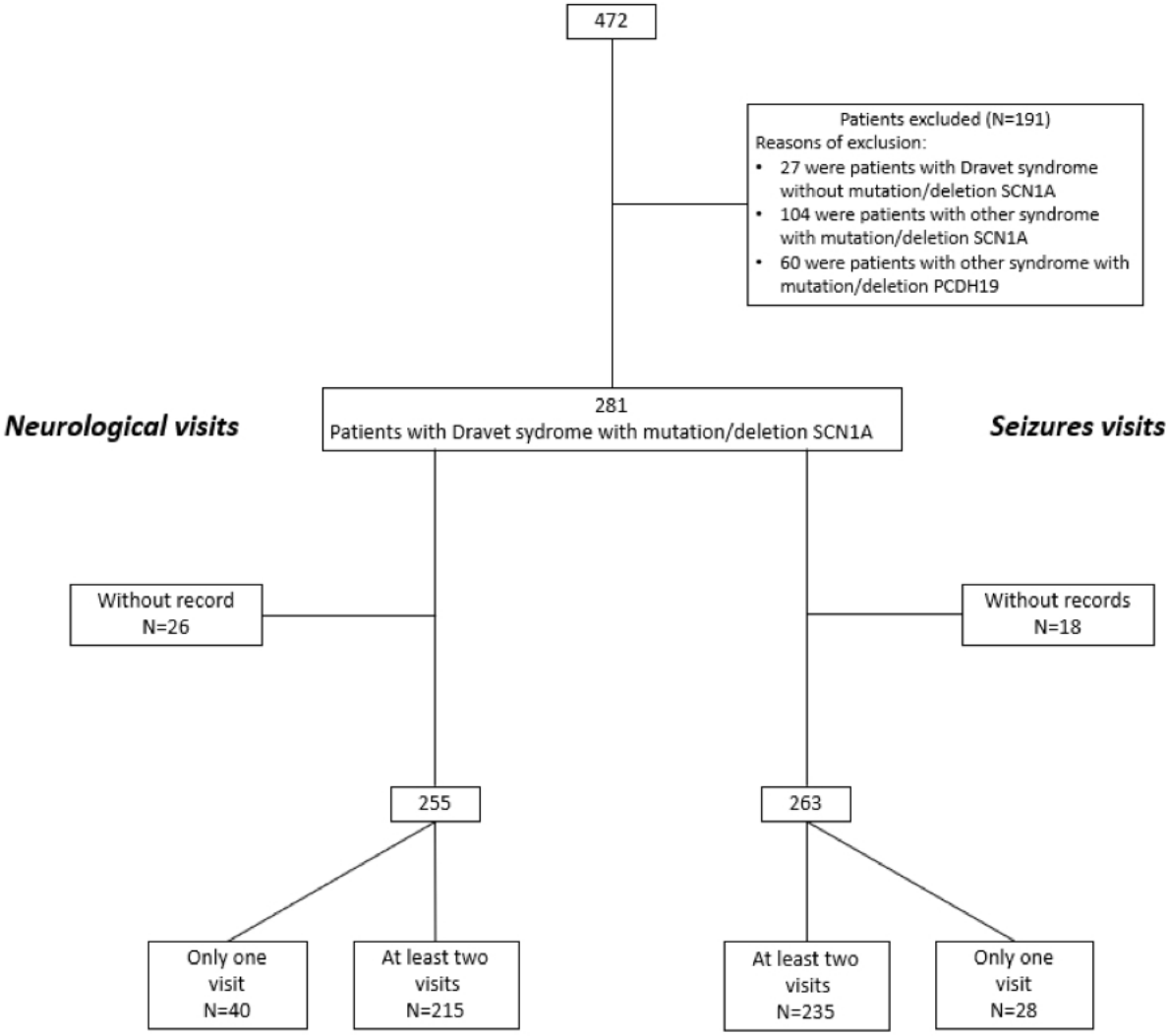
Although Residras has a nationwide coverage (Figure 1), patients' geographic distribution is non-homogenous with a lower proportional representation from the South of Italy. This is mainly related to the habit of families of seriously ill patients residing in the South to reach hospitals of the northern regions, seeking for a second opinion, and does not have any geographic epidemiological implication. To date, 281 individuals with DS with a confirmed *SCN1A* mutation have been enrolled, with regular updates at least on a yearly basis. There are available data on epilepsy (n=263) and on the overall neurological status (n=255), with at least one follow-up for most patients (Figure 2).

Figure 1. Coverage of Residras with illustration of the participating centre distribution (green pins) and number of enrolled patients who are resident in each italian region (yellow circles).



Historical retrospective data were also included, when enrolment date was during a follow-up visit and the patient had not been previously included in the registry or when the DS diagnosis was made at a later time after the first clinical evaluation. The time at the first visit was defined as the time at the inclusion in the registry for all patients, including the ones with enrolment date subsequent to first clinical assessment.

Figure 2. Flow chart illustrating current numbers of patients enrolled in Residras



Median follow-up was 5.5. years (2.8-11.3). Median age at the inclusion in the Registry was 2 years (IQR 0-9) whilst the median age at last follow-up was 11 years (IQR 5.5-17). During the seven year-activity of the registry, five patients had died with a mortality rate of 1.84 (95% CI 0.77-4.42) per 1000-person-years (median age at death 6.3 years, range 2.5-23.4).

The causes of death were diverse and included status epilepticus, cerebral haemorrhage, SUDEP, acute encephalopathy of unknown cause (with onset 40 days before death), and brain tumour.

There was a significant difference in cognitive function over the first five years of follow-up ($P< 0.001$) with evidence of a lower cognitive level after five years in the majority (median time to first deterioration in cognitive level 1.5 years, IQR 0.9-2.4).

There was a significant change in the educational progress over time; in particular, we observed increased prevalence of poor educational autonomy (from 17.6% at first visit to 45.5% after five years) ($P< 0.001$).

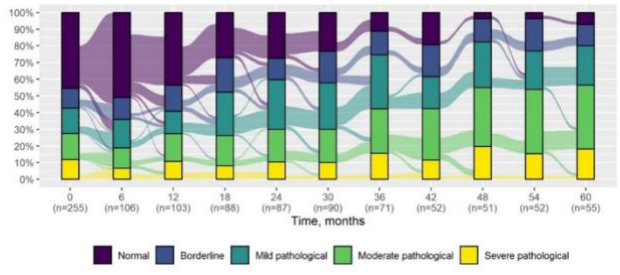
There was an increased prevalence of behavioral disorders over time including attention deficit (*P*< 0.001), and autism spectrum symptoms (*P*=0.04). Prevalence of neurological examination abnormalities increased at last visit (*P*< 0.001) in most patients (median time to first deterioration 1.6 years, IQR 0.9-2.4), including ataxia (*P*< 0.001) and motor instability (*P*=0.01).

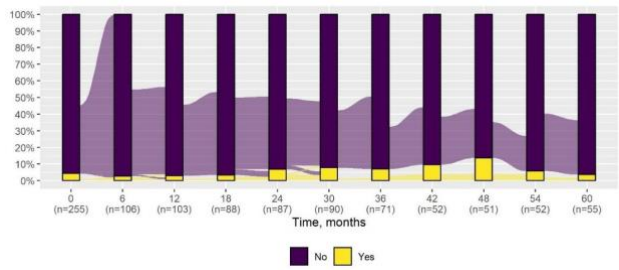
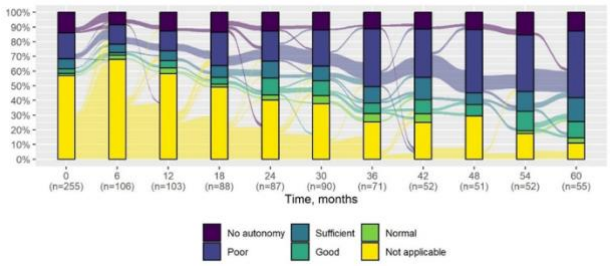
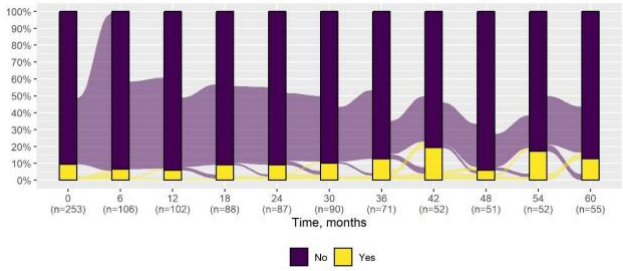
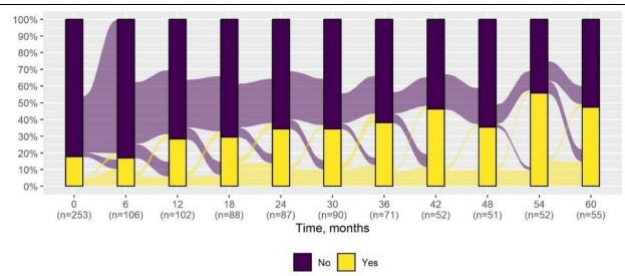
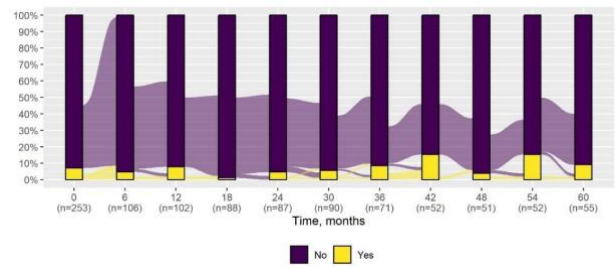
There was also a significant worsening of language (*P*= 0.001) (median time to first decreased level 1.4 years, IQR 0.8-2.3) and intellectual disability (*P*< 0.001) (median time to first decreased level 1.5 years, IQR 1.0-2.3), where these were formally assessable.

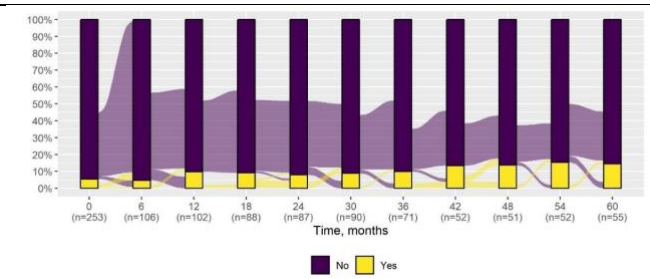
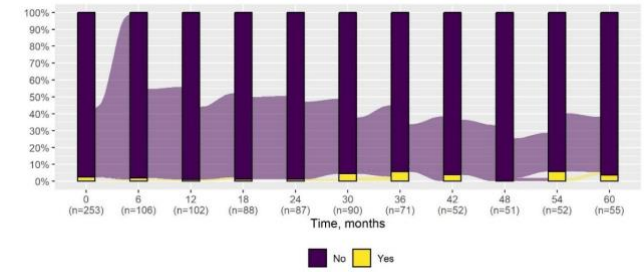
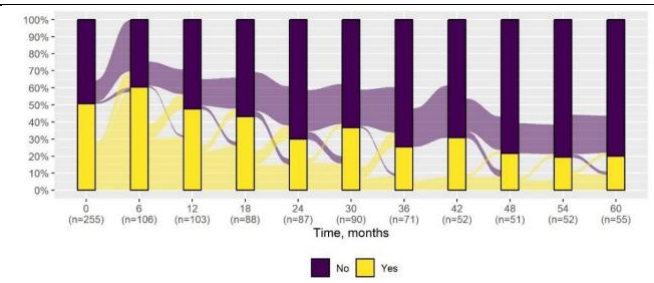
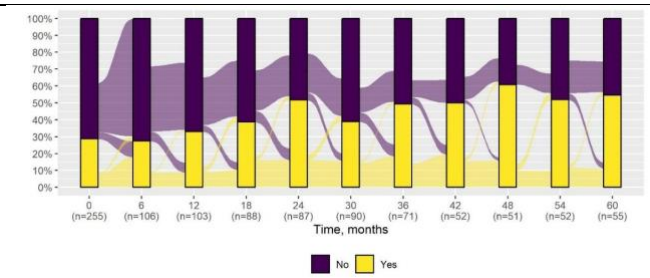
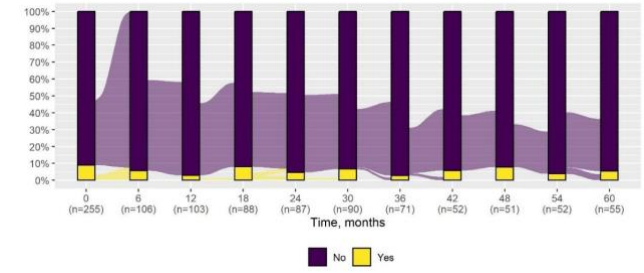
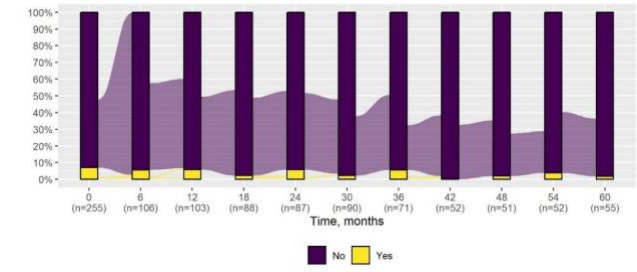
Significant differences related to the epilepsy features over the first 5-year follow-up included an increased prevalence of generalized seizures (*P*=0.03) and reduction of unilateral seizures (*P*= 0.001).

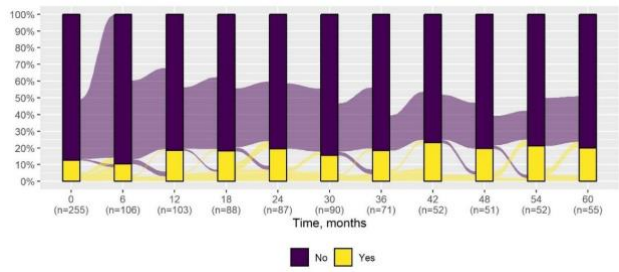
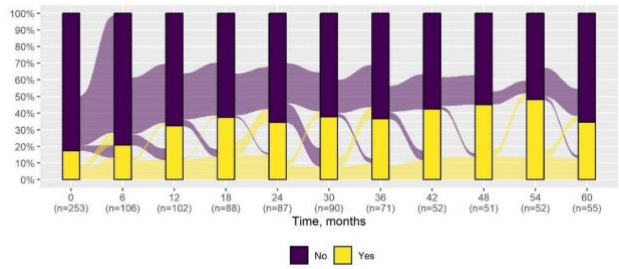
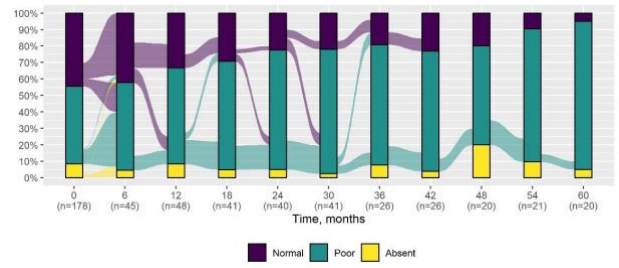
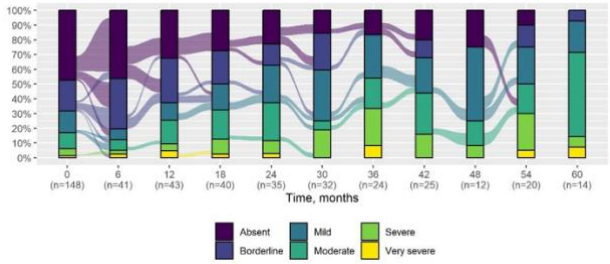
The main clinical variables included in the analysis, and their variation over the first 5-year follow-up are illustrated in Tables 1 and 2. The pattern of improvement, stability or deterioration related to each clinical variable is summarized in Tables 3 and 4.

Table 1. Neurological condition at first and last visit at 5th year with alluvial diagrams showing variation over time

	First visit	Last visit at 5th year	
N	255	55	
Gender Females	132 (51.8)	27 (49.1)	
Age Median (IQR)	2.0 (0-9.0)	6.0 (5.0-9.5)	
Cognitive function			
Frame* Normal Borderline Mild Pathological Moderate Pathological Severe Pathological	116 (45.5) 30 (11.8) 39 (15.3) 40 (15.7) 30 (11.8)	4 (7.3) 7 (12.7) 13 (23.6) 21 (38.2) 10 (18.2)	

Autism spectrum	11 (4.3)	2 (3.6)	
Educational Autonomy*			
Normal	4 (1.6)	2 (3.6)	
Good	8 (3.1)	6 (10.9)	
Sufficient	17 (6.7)	9 (16.4)	
Poor	45 (17.6)	25 (45.5)	
No Autonomy	36 (14.1)	7 (12.7)	
Not Applicable	145 (56.9)	6 (10.9)	
*Behavior**			
Oppositional defiant disorder	24 (9.5)	7 (12.7)	
Attention deficit*	45 (17.8)	26 (47.3)	
Impulse control disorder	18 (7.1)	5 (9.1)	

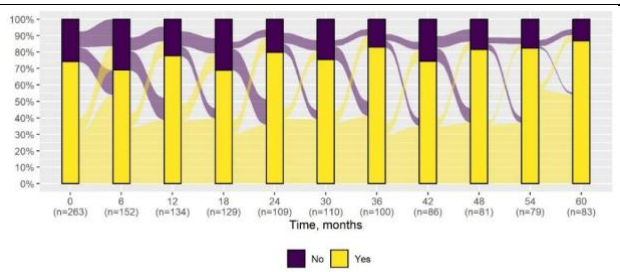
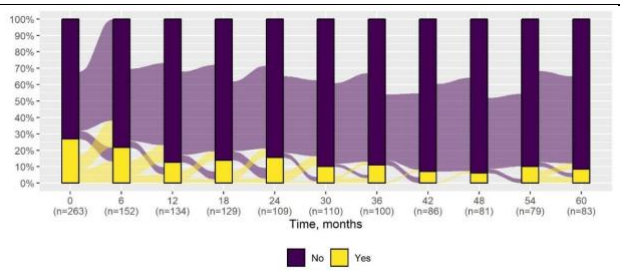
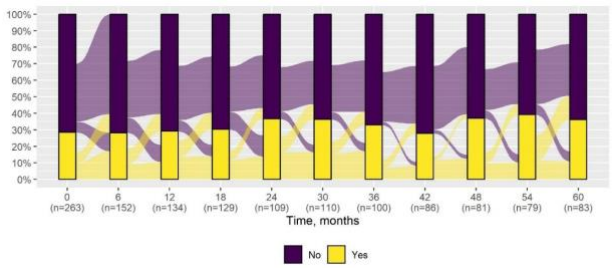
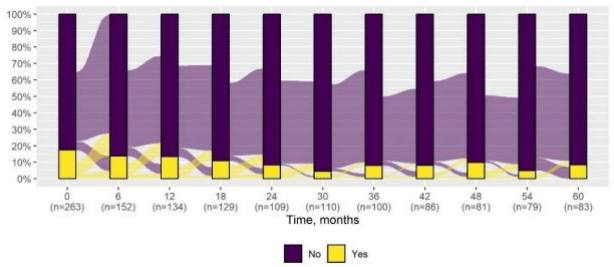
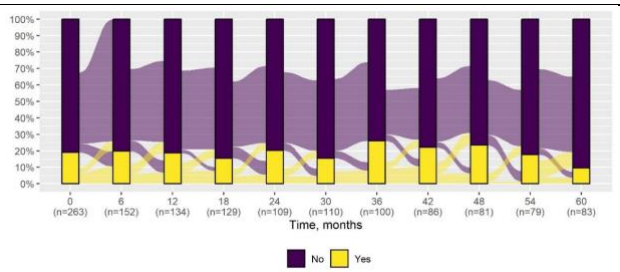
Autism spectrum symptoms*	14 (5.5)	8 (14.5)	
Obsessive compulsive disorder	6 (2.4)	2 (3.6)	
Neurological examination			
Normal*	129 (50.6)	11 (20)	
Ataxia*	73 (28.6)	30 (54.5)	
Pyramidal signs	23 (9.0)	3 (5.5)	
Extrapyramidal signs	18 (7.1)	1 (1.8)	

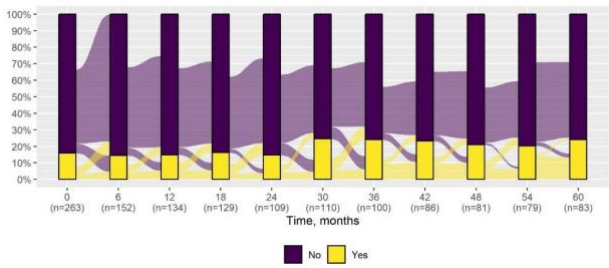
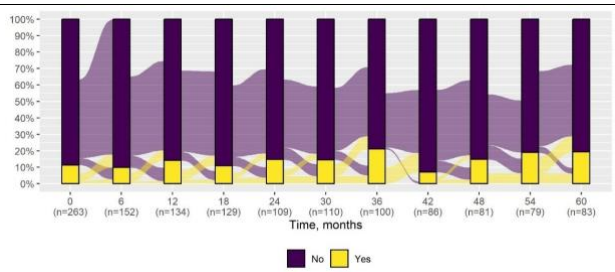
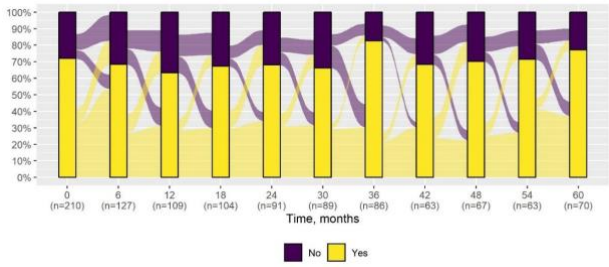
Action myoclonus	32 (12.5)	11 (20)	
Motor Instability*	44 (17.4)	19 (34.5)	
Neuropsychological assessment			
Language* Normal Poor Absent Not assessable	 79 (44.4) 84 (47.2) 15 (8.4) 32	 1 (5.0) 18 (90.0) 1 (5.0) 2	
Intellectual disability* Absent Borderline Mild Moderate Severe Very Severe Not assessable	 70 (47.3) 31 (20.9) 22 (14.9) 16 (10.8) 7 (4.7) 2 (1.4) 32	 - 1 (7.1) 3 (21.4) 8 (57.1) 1 (7.1) 1 (7.1) 2	

* statistically significant difference between first and last visit (p-value < 0.05)

Table 2. Epilepsy features at first visit and after 5-year follow-up with alluvial diagrams showing variation over time

	First visit	Last visit at 5th year	
N	**263**	**83**	

Age, years Median (IQR)	 2.0 (0-7.8)	 11.0 (5.0-18.0)	
Generalized seizures*	195 (74.1)	72 (86.7)	
Unilateral seizures*	71 (27.0)	7 (8.4)	
Focal-onset seizures	75 (28.5)	30 (36.1)	
Status epilepticus	46 (17.5)	7 (8.4)	
Massive myoclonias	50 (19.0)	8 (9.6)	
Frequency Daily Sporadic	 22 (45.8) 26 (54.2)	 3 (37.5) 5 (62.5)	

Absence seizures	42 (16.0)	20 (24.1)	
Frequency Daily Sporadic	 18 (43.9) 23 (56.1)	 7 (35.0) 13 (65.0)	
Reflex seizures	20 (7.6)	5 (6.0)	
Intermittent light stimulation	4 (1.5)	1 (1.2)	
Pattern	4 (1.5)	1 (1.2)	
Eyes closure	1 (0.4)	0 (0.0)	
Self-stimulation	4 (1.5)	1 (1.2)	
Seizure clusters	30 (11.4)	16 (19.3)	
Febrile seizures	151 (71.9)	6 (75.0)	

* statistically significant difference between first and last visit (p-value < 0.05)

Table 3. Pattern of changes in neurological condition at 5th year.

	N (%)
Cognitive function	
Deterioration by One class Two classes Three classes	 10 (18.2) 18 (32.7) 11 (20.0)

Four classes	1 (1.8)
Improvement	
One class	1 (1.8)
Stability	14 (25.5)
Autism spectrum disorder	
Deterioration	1 (1.8)
Improvement	1 (1.8)
Stability	196 (96.4)
Behavior	
Oppositional defiant disorder	
Deterioration	7 (12.7)
Improvement	2 (3.6)
Stability	46 (83.6)
Attention deficit	
Deterioration	16 (29.1)
Improvement	1 (1.8)
Stability	38 (69.1)
Impulse control disorder	
Deterioration	4 (7.3)
Improvement	1 (1.8)
Stability	50 (90.9)
Autism spectrum symptoms	
Deterioration	6 (10.9)
Improvement	0 (0.0)
Stability	149 (89.1)
Obsessive compulsive disorder	1 (1.8)
Deterioration	1 (1.8)
Improvement	53 (96.4)
Stability	
Neurological examination	
Normal	
Deterioration	19 (34.5)
Improvement	1 (1.8)
Stability	35 (63.6)

Ataxia	
Deterioration	16 (29.1)
Improvement	2 (3.6)
Stability	37 (67.3)
Pyramidal signs	
Deterioration	2 (3.6)
Improvement	1 (1.8)
Stability	52 (94.5)
Extrapyramidal signs	
Deterioration	1 (1.8)
Improvement	1 (1.8)
Stability	53 (96.4)
Action myoclonus	
Deterioration	7 (12.7)
Improvement	2 (3.6)
Stability	46 (83.6)
Motor instability	
Deterioration	12 (21.8)
Improvement	2 (3.6)
Stability	41 (74.5)
Neuropsychological assessment	
Language	
Deterioration	7 (31.8)
Improvement	2 (9.1)
Stability	10 (45.5)
Not assessable	3 (13.6)
Intellectual disability	
Deterioration by	
One class	1 (6.7)
Two classes	4 (26.7)
Three classes	2 (13.3)
Four classes	1 (6.7)
Five classes	1 (6.7)
Improvement	

One class	1 (6.7)
Stability	2 (16.7)
Not assessable	3 (20.0)

Table 4. Pattern of changes in epilepsy features at 5th year.

	N (%)
Generalized seizures	
Deterioration	14 (16.9)
Improvement	8 (9.6)
Stability	61 (73.5)
Unilateral seizures	
Deterioration	6 (7.2)
Improvement	21 (25.3)
Stability	56 (67.5)
Focal onset seizures	
Deterioration	6 (7.2)
Improvement	21 (25.3)
Stability	56 (67.5)
Status epilepticus	
Deterioration	6 (7.2)
Improvement	18 (21.7)
Stability	59 (71.1)
Massive myoclonus	
Deterioration	5 (6.0)
Improvement	18 (21.7)
Stability	60 (72.3)
Absence seizures	
Deterioration	12 (14.5)
Improvement	6 (7.2)
Stability	65 (78.3)
Seizure clusters	
Deterioration	14 (16.9)
Improvement	10 (12.0)
Stability	59 (71.1)

Febrile seizures	
Deterioration	6 (7.2)
Improvement	10 (12.0)
Stability	45 (54.2)
Missing	22 (26.5)

Discussion

Residras is based on a user-friendly platform that facilitates data collection and analysis of patients with *SCN1A*-related epilepsies, and on a network of 24 expert centers across Italy with specific expertise in rare epilepsy syndromes, with a scope to include further national and international centers and promote a longitudinal standardized data collection. The Residras registry might represent a paradigmatic example to homogenize data collection and facilitate research in rare and complex epilepsy syndromes. During the process, several control steps were put in place to ensure high data quality such as centralized control for duplicates, set of mandatory fields, use of standardized ontology codes, and a robust security infrastructure. All these processes are time demanding, which may in part explain why there are no established registries for rare and complex epilepsy syndromes are in place. However, Residras has ample potential for expansion at both national and international level so to constitute a unified source of longitudinal phenotypic data for DS and other *SCN1A*-related epilepsies.

We adopted two different approaches to analyze clinical variation over time: we focused on the first 5-year follow-up to increase specificity of the disease evolution from onset, although clinical assessment at the 5-year time point was not available for most patients (Table 1 and 2); we then analyzed variation from first to last visit but noting that follow-up length was variable and therefore results are less specific to interpret disease course.

Our data confirm worsening of cognitive ability over time in DS, as already reported in cross-sectional and retrospective studies. [12] [29] A minority of patients exhibited normal cognitive skills at last follow-up. Although we excluded febrile convulsions and GEFS+ phenotypes from the analysis, there were some borderline DS phenotypes that were included and explain this finding. The spectrum of the syndrome is not sharply delimited and less severe forms may still be compatible with a relatively good cognitive outcome. This is not predictable early after onset and excluding patients with less severe outcomes from the syndrome spectrum in the aftermath as they do not fit the core definition would generate artefacts. We highlight how the items evaluating cognitive and behavioral skills contribute to capture the granularity of the complex neurodevelopmental phenotype and may

serve as a robust basis to assess the impact of existing and novel treatment strategies. Although patient inclusion in the registry does not require the mandatory use of specific neuropsychological testing, the more systematic use of formal assessment by the appropriate scales should be encouraged.

We note that the number of individuals with autistic symptoms is low although an increased prevalence is observed during the first 5-year follow-up; although formal assessment for a diagnosis of autism spectrum disorder was not regularly available, the increase over time is likely consequent to difficulties in diagnosing autistic features under the age of three years.

Premature mortality is a recognized unfavorable outcome in DS although most available data are from children [17] [18] and there are no survival analyses in the long term. In a cohort of 100 consecutively recruited DS individuals, mortality rate was 15.84 (98% CI 9.01-27.85) per 1000-person-years whilst the rate of SUDEP was 9.32/1000-person-years. Living individuals had a median follow-up of 17 years, whilst the median age at death was 7 years. [18] We observed a mortality rate of 1.84 (95% CI 0.77-4.42) per 1000-person-years, with a median age at death of 6 years. Of the five deaths reported, one was due to SUDEP and one was epilepsy-related (i.e., status).
Multiple factors may concur in generating the far lower mortality rates we observed with respect to previous reports. [18] Firstly, mortality data emerging from a registry need to be interpreted with caution as death may occur before a diagnosis is made or before a patient is included, or after the latest follow-up data entry. In addition, differences in rates of premature mortality may be partly explained by diverse healthcare provision and treatment strategies.

The Cooper et al. [18] DS cohort was more heterogeneous in terms of geographic origin, and was established almost twenty years earlier, which implies a longer follow-up and little or no access to more recently introduced drugs with a proven efficacy in DS, particularly stiripentol, cannabidiol and fenfluramine. [6] [7] [8] [9] Finally, improved management strategies have been applied over the years, including earlier diagnosis in larger number if patients and avoidance of inappropriate drugs such as sodium channel blockers or exceedingly sedative drug regimens. All these factors may contribute to reduced mortality rates in subsequent generations of patients.

Additional highly relevant assets provided by the registry include its unique reliability in delineating the natural history of DS, the longitudinal comparison of medical practice between DS expert centres, the availability of a wide clinical and genetic spectrum of patients with SCN1A-related epilepsies whose characteristics, homogeneously recorded, are available for genotype-phenotype correlations and comprehensive characterization of comorbidities occurring at different ages.

Accuracy in managing a registry for DS carries has limitations, as it may be expected for a severe disease starting early in life, associating chronic disability with periods of acute exacerbations and comorbidities that require multiple medical interventions and treatment adjustments. Unavoidably, some data are missing, inclusion of follow-up information might not respect the set deadlines, the various centers may apply different levels of accuracy in reporting relevant among completeness in data inclusion, application of clinical diagnostic criteria for milder DS forms or other *SCN1A*-related epilepsies is not necessarily uniform.

Registry curation is time-consuming and might be particularly challenging for those centers with high number of patients but limited personnel resources. The use of rescue medications might be limited if data are not timely included in the registry. Our registry was partly supported by limited funding raised by a no-profit patient association ('Dravet Italia Onlus'), and specific allocation of funding to support personnel dedicated to systematic and timely registry curation in each center would have undoubtedly enhanced accuracy in data collection.

A complication in registry curation for a chronic disease that certainly exists across all health systems is related to transition to adult care. Unless there is continuity of care, for example, in a specialized institution, patients may transition to adult neurology centers with limited expertise in DS and propensity to adhere to a registry conceived for an infantile onset disorder.
To mitigate this limitation, we aim to increase the registry coverage and improve the transition system for rare and complex epilepsy syndromes. [30]
We are planning to promote the registry activities and recruit additional centers so improve national coverage and limit the number who may be lost to follow-up.
Multiple associations for specific genetic disorders are being founded with often the aim to establish dedicated registries, but it may become challenging for the treating physicians to fill different registries with variable formats and become familiar with them. Applying the Residras model to other rare and complex genetic epilepsies, would be of great help in gathering up for each of them a critical mass of homogeneously stored information on epidemiology, disease course, and easing access to cohorts of patients that may benefit from clinical trials and drug safety monitoring. Therapy development for rare diseases faces several specific challenges, including small populations for clinical studies, difficulty in determining relevant outcome measures and endpoints, and poorly understood natural history. Residras is easily adaptable to other genetic epilepsies and could already be adopted, 'as is', to multiple rare and complex epilepsy disorders, including for example the ones associated with SCN2A, SCN8A, KCNQ2, KCNT1- mutations, and many others.

References

[1] Symonds JD, Elliott KS, Shetty J, Armstrong M, Brunklaus A, Cutcutache I, et al. Early childhood epilepsies: epidemiology, classification, aetiology, and socio-economic determinants. Brain. 2021;144:2879–91.

[2] Mei D, Cetica V, Marini C, Guerrini R. Dravet syndrome as part of the clinical and genetic spectrum of sodium channel epilepsies and encephalopathies. Epilepsia. 2019;60 Suppl 3:S2–7.

[3] Yu FH, Mantegazza M, Westenbroek RE, Robbins CA, Kalume F, Burton KA, et al. Reduced sodium current in GABAergic interneurons in a mouse model of severe myoclonic epilepsy in infancy. Nat Neurosci.;9:1142–9.

[4] Li W, Schneider AL, Scheffer IE. Defining Dravet syndrome: An essential pre-requisite for precision medicine trials. Epilepsia;62:2205–17.

[5] Wirrell EC, Laux L, Donner E, Jette N, Knupp K, Meskis MA, et al. Optimizing the Diagnosis and Management of Dravet Syndrome: Recommendations From a North American Consensus Panel. Pediatr Neurol.;68:18-34.e3.

[6] Chiron C, Marchand MC, Tran A, Rey E, d'Athis P, Vincent J, et al. Stiripentol in severe myoclonic epilepsy in infancy: a randomised placebo-controlled syndrome dedicated trial. STICLO study group. Lancet. 2000;356:1638–42.

[7] Devinsky O, Cross JH, Laux L, Marsh E, Miller I, Nabbout R, et al. Trial of Cannabidiol for Drug-Resistant Seizures in the Dravet Syndrome. N Engl J Med. 2017;376:2011–20.

[8] Lagae L, Sullivan J, Knupp K, Laux L, Polster T, Nikanorova M, et al. Fenfluramine hydrochloride for the treatment of seizures in Dravet syndrome: a randomised, doubleblind, placebo-controlled trial. Lancet. 2019;394:2243–54.

[9] Nabbout R, Mistry A, Zuberi S, Villeneuve N, Gil-Nagel A, Sanchez-Carpintero R, et al. Fenfluramine for Treatment-Resistant Seizures in Patients With Dravet Syndrome Receiving Stiripentol-Inclusive Regimens: A Randomized Clinical Trial. JAMA Neurol. 2020;77:300–8.

[10] Isom LL, Knupp KG. Dravet Syndrome: Novel Approaches for the Most Common Genetic Epilepsy. Neurotherapeutics. 2021;18:1524–34.

[11] Akiyama M, Kobayashi K, Yoshinaga H, Ohtsuka Y. A long-term follow-up study of Dravet syndrome up to adulthood. Epilepsia. 2010;51:1043–52.

[12] Ragona F, Granata T, Dalla Bernardina B, Offredi F, Darra F, Battaglia D, et al. Cognitive development in Dravet syndrome: a retrospective, multicenter study of 26 patients. Epilepsia. 2011;52:386–92.

[13] Darra F, Battaglia D, Dravet C, Patrini M, Offredi F, Chieffo D, et al. Dravet syndrome: Early electroclinical findings and long-term outcome in adolescents and adults. Epilepsia. 2019;60 Suppl 3:S49–58.
[14] Catarino CB, Liu JYW, Liagkouras I, Gibbons VS, Labrum RW, Ellis R, et al. Dravet syndrome as epileptic encephalopathy: evidence from long-term course and neuropathology. Brain. 2011;134:2982–3010.
[15] Genton P, Velizarova R, Dravet C. Dravet syndrome: the long-term outcome. Epilepsia. 2011;52 Suppl 2:44–9. Epilepsia Open.
[16] Takayama R, Fujiwara T, Shigematsu H, Imai K, Takahashi Y, Yamakawa K, et al. Long-term course of Dravet syndrome: a study from an epilepsy center in Japan. Epilepsia. 2014;55:528–38.
[17] Sakauchi M, Oguni H, Kato I, Osawa M, Hirose S, Kaneko S, et al. Mortality in Dravet syndrome: search for risk factors in Japanese patients. Epilepsia. 2011;52 Suppl 2:50–4.
[18] Cooper MS, Mcintosh A, Crompton DE, McMahon JM, Schneider A, Farrell K, et al. Mortality in Dravet syndrome. Epilepsy Res. 2016;128:43–7.
[19] Dawkins HJS, Draghia-Akli R, Lasko P, Lau LPL, Jonker AH, Cutillo CM, et al. Progress in Rare Diseases Research 2010-2016: An IRDiRC Perspective. Clin Transl Sci. 2018;11:11–20.
[20] Residras. Dravet registry [Internet]. [cited 2022 Jan 3]. Available from: https://www.dravet-registry.com/registry
[21] Dravet Italia Onlus [Internet]. [cited 2022 Jan 3]. Available from: https://www.dravet.it
[22] Fondazione Monasterio [Internet]. [cited 2022 Jan 3]. Available from: https://www.monasterio.it/
[23] Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3:160018.
[24] Mons, Barend, Neylon, Camerond, Velterop, Jane, Dumontier, Michelf, da Silva Santos, Luiz Olavo Bonino, Wilkinson, Mark. Cloudy, increasingly FAIR; Revisiting the FAIR Data guiding principles for the European Open Science Cloud. Information Services & Use. 37(1):49–56.
[25] European Commission. Set of Common Data Elements [Internet]. [cited 2022 Jan 3]. Available from: https://eu-rd-platform.jrc.ec.europa.eu/set-of-common-dataelements_en
[26] Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res. 2004;32(Database issue):D267-270.
[27] Human Phenotype Ontology (HPO) [Internet]. [cited 2022 Jan 3]. Available from: https://hpo.jax.org/app/
[28] Orphanet Rare Disease Ontology (ORDO) [Internet]. [cited 2022 Jan 3]. Available from: https://bioportal.bioontology.org/ontologies/ORDO

[29] Wolff M, Cassé-Perrot C, Dravet C. Severe myoclonic epilepsy of infants (Dravet syndrome): natural history and neuropsychological findings. Epilepsia. 2006;47 Suppl 2:45–8.
[30] Balestrini S, Guerrini R, Sisodiya SM. Rare and Complex Epilepsies from Childhood to Adulthood: Requirements for Separate Management or Scope for a Lifespan Holistic Approach? Curr Neurol Neurosci Rep. 2021;21:65

Chapter 2: "Improving early diagnosis of rare diseases using Natural Language Processing in unstructured medical records: an illustration from Dravet syndrome" [2]

Abstract

Background

The growing use of Electronic Health Records (EHRs) is promoting the application of data mining in health-care. A promising use of big data in this field is to develop models to support early diagnosis and to establish natural history.

Dravet Syndrome (DS) is a rare Developmental and Epileptic Encephalopathy that commonly initiates in the first year of life with febrile seizures (FS). Age at diagnosis is often delayed after two years, as it is difficult to differentiate DS at onset from FS.

We aimed to explore if some clinical terms (concepts) are significantly more used in the electronic narrative medical reports of individuals with DS before the age of two years compared to those of individuals with FS. These concepts would allow an earlier detection of patients with DS resulting in an earlier orientation toward expert centers that can provide early diagnosis and care.

Methods

Data were collected from the Necker Enfants Malades Hospital using a document-based data warehouse, Dr Warehouse, which employs Natural Language Processing, a computer technology

[2] Tommaso Lo Barco[a,b], Mathieu Kuchenbuch[a,c], Nicolas Garcelon[c], Antoine Neuraz[d], Rima Nabbout[a,c]

[a] Department of Pediatric Neurology, Necker-Enfants Malades Hospital, Centre de Référence Épilepsies Rares, member of ERN EPICARE, Université de Paris, Paris, France.

[b] Child Neuropsychiatry, Department of Surgical Sciences, Dentistry, Gynecology and Pediatrics, University of Verona, Italy.

[c] Université de Paris, Imagine Institute, Data Science Platform, INSERM UMR 1163, F-75015, Paris, France

[d] University of Paris, Paris, France; INSERM UMR1138, Centre de Recherche des Cordeliers, Paris, France; Department of Medical Informatics, University Hospital Necker-Enfants Malades, APHP, Paris, France.

Published on Orphanet Journal of Rare Disease, 2021 Jul 13;16(1):309. doi: 10.1186/s13023-021-01936-9. PMID: 34256808; PMCID: PMC8278630.

consisting in processing written information. Using Unified Medical Language System Meta-thesaurus, phenotype concepts can be recognized in medical reports.
We selected individuals with DS (DS Cohort) and individuals with FS (FS Cohort) with confirmed diagnosis after the age of four years. A phenome-wide analysis was performed evaluating the statistical associations between the phenotypes of DS and FS, based on concepts found in the reports produced before two years and using a series of logistic regressions.

Results

We found significative higher representation of concepts related to seizures' phenotype distinguishing DS from FS in the first phases, namely the major recurrence of complex febrile convulsions (long-lasting and/or with focal signs) and other seizure-types. Some typical early onset non-seizure concepts also emerged, in relation to neurodevelopment and gait disorders.

Conclusion

Narrative medical reports of individuals younger than two years with FS contain specific concepts linked to DS diagnosis, which can be automatically detected by software exploiting NLP. This approach could represent an innovative and sustainable methodology to decrease time of diagnosis of DS and could be transposed to other rare diseases.

Objectives

Electronic health records (EHRs) contain healthcare data of individuals and population electronically-stored in a digital format. [1] In the last decade, the use of EHRs has become part of routine care across the majority of developed countries. [2]
Through data mining techniques, this growing use of EHRs is allowing the development of predictive models aimed to individuate high risk patients and support prevention initiatives. [3][4] As well, models to support diagnosis and treatment of rare diseases are emerging. [5][6]
EHRs consist of structured and unstructured data. Structured data are produced through constrained choices (drop-down menus, check boxes and pre-filled templates as in registries), whereas unstructured clinical data exist in the form of free text narratives and are often used in clinical care for medical reports. [7] Combining Natural Language Processing (NLP) technology and UMLS (Unified Medical Language System), providers' notes and narratives can be converted into structured, standardized formats, usable for data mining. [8][9][10]
Dravet Syndrome (DS) is a rare disorder, with a worldwide incidence between 1/40.000 and 1/15.700. [11] DS is a genetic developmental and epileptic encephalopathy with onset in first year of life,

characterized at onset by febrile seizures and convulsive status epilepticus in otherwise healthy infants.[12] Starting by the second year, individuals present multiple seizure types (clonic, tonic-clonic, motor and non-motor onset focal seizures, myoclonic, atypical absences), that are often drug resistant, with developmental slowing leading to definite cognitive impairment. [13] Diagnosis is easier after the age of two as more pathognomonic seizure types and other symptoms are present from this age. Genetic testing shows a pathogenic variant in SCN1A in over 85% of cases reinforcing the diagnosis suspicion, but this testing might take months and is not available for all individuals with suspected DS.[14] However there is a need for early diagnosis in order to avoid worsening therapies and to establish best therapy protocol as seizure control might be partly related to cognitive improvement and a better quality of life.[15]

Early diagnosis of individuals with DS is often delayed as it is difficult to differentiate at onset from Febrile Seizures (FS).[16] These two conditions present substantial clinical differences, leading to exclude one on other diagnosis but might be overlapping at onset. Even if physician awareness of Dravet syndrome has markedly improved in last decades,[17] time to diagnosis is still over two years,[18] and it remains underdiagnosed in adult population and in developing countries. [19][20]

Using data mining, we analysed clinical reports produced before the age of two years for individuals with confirmed DS and FS with the aim of identifying specific terms (concepts) allowing early DS suspicion and reducing diagnosis delay. We then explored the differences between the concepts in the reports of two subgroups of individuals with DS: patients with suspected diagnosis before the age of 2 years and patients for whom diagnosis was suspected after the age of two.

Materials and Methods

Data were collected from Necker Enfants Malades Hospital, a paediatric University hospital belonging to the Assistance Publique Hopitaux de Paris group (400 paediatric beds, 200 adult beds), which is a national and European reference center for rare and undiagnosed diseases, including the reference a centre for rare epilepsies.

DrWarehouse [21]is a document-based open-source data warehouse oriented toward narrative clinical reports from the Electronic health records (EHRs). It contains more than 4.5 million clinical free-text documents produced at Necker Hospital from 2009, for more than 465,000 individuals and more than 20 departments. DrWarehouse uses UMLS Metathesaurus to recognize phenotype concepts inside narrative medical reports. In this manuscript, the word “concepts” will refer to phenotypes extracted automatically from hospital reports, without a priori, by using a UMLS subset of 20.000 phenotypic words or expressions.

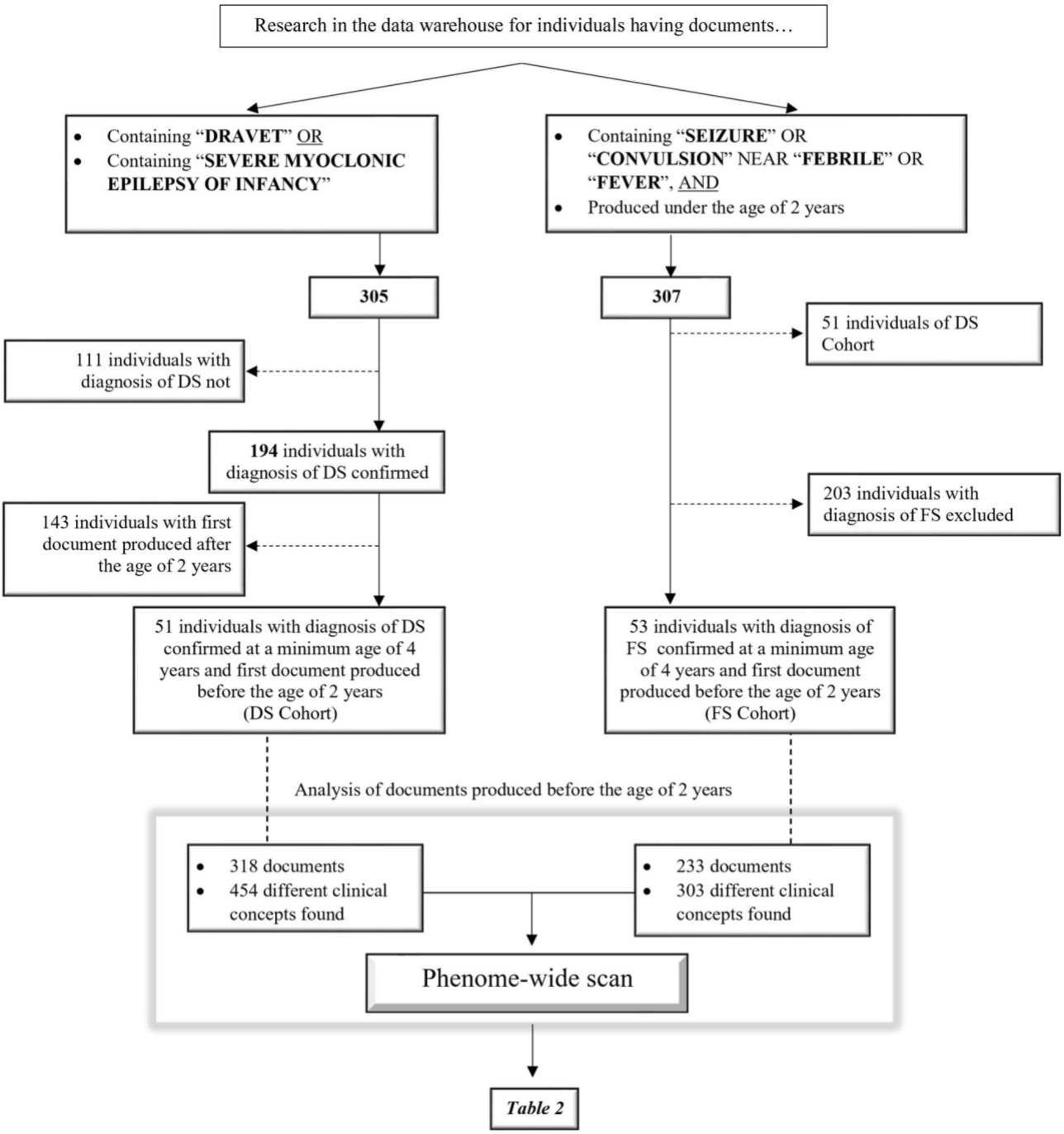
By using the appropriate research field in DrWarehouse, we searched all individuals who presented in their medical reports the word “Dravet” or “Severe Myoclonic Epilepsy of Infancy” at least in one clinical document. We then selected from this group all individuals that had a definite diagnosis of DS based on clinical and genetic criteria, and evaluated after the age of four where the full blown syndrome can be confirmed. We finally included from this group individuals with at least one clinical report before the age of two years and this final selection constituted the “Dravet Syndrome Cohort” (DS Cohort).

Subsequently, we searched in the data of DrWarehouse all individuals whose medical reports produced before the age of two presented the words “seizure” / “seizures” or “convulsion”/ “convulsions” in proximity (max 5 words away) to “fever” or “febrile”. From this group, we excluded the individuals of the DS Cohort and individuals in which febrile seizures was a symptom of a more complex condition (infections involving the central nervous system, other encephalopathies, structural brain injury, detected genetic or metabolic pathologies, or epilepsies). The “Febrile Seizures’cohort” (FS cohort) included the individuals from this group aged over year where we confirmed the diagnosis of febrile seizures based on EHRs or by telephone interviewing of the family (FS Cohort) (Figure 1).

The phenome-wide scan consists in comparing the distribution of phenotypes between two groups (cases and controls) and estimates the association between the phenotypes and the groups. These associations are assessed sequentially. [22] [23] We evaluated the statistical associations between the phenotypes and the cohorts DS and FS, using a series of multivariate logistic regressions adjusted on gender and age. For the analysis, we used concepts found in clinical reports with a minimum number of occurrences of three individuals, excluding negations and those associated to family members. The p-values were corrected for multiple testing using a false discovery rate (FDR) methodology.

We also compared the phenotype differences in the DS Cohort between the subgroup where diagnosis of DS was confirmed or suspected before the age of two years, and the subgroup where DS diagnosis was not reported.

Figure 1. Flowchart of the selection procedures and constitution of the cohorts



Results

"Dravet Syndrome Cohort" (DS Cohort)

The term "Dravet" and/or "severe myoclonic epilepsy of infancy" appeared in 305 individuals present in the warehouse: 194 of them had a final diagnosis of DS in the last document on the database, 51 had at least one document produced under the age of two years. All had a clinical and genetic diagnosis of DS. These individuals constituted the DS Cohort.

DS cohort included 28 males and 23 females. The mean age at first seizure was 5.5 months (min 2 – max 12). The average age of the first produced document was 1.05 years, median is 1.15 (min 0.25 -

max 1.98). The average length of the follow-up of these individuals was 5.68 years, median 4.98 (min 3,75- max 13,42).

In order to compare early characteristics of this population with a population with FS at the same age, documents produced exclusively before two years were selected, for a total of 318 documents (mean: 6.24; median: 3 for each individual). 3484 concepts were extracted from the abovementioned documents (mean: 10.9 per document), 454 of which were unique concepts. Concepts present in almost 10% of the population are listed in a decreasing order in the Table 1. The most prevalent concepts were "Seizures" (found in 48 individuals – 94%), "Fever" (43 individuals – 84%), "Epilepsy" (42 individuals – 82%), "Dravet Syndrome" (37 individuals – 73%), "Convulsions" (31 individuals – 61%).

"Febrile Seizure Cohort" (FS Cohort)

The research of the words "seizure" or "convulsion" in individuals' reports close to the words "febrile" or "fever", limited to documents produced by the first two years of life and excluding individuals of DS Cohort, led to 256 subjects. After exclusion of other aetiologies, we included all 53 subjects with a diagnosis of febrile seizures. Diagnosis was confirmed after age four by reviewing child's medical history, neurological and developmental outcome in the available medical files in addition to a telephone interview with the family.

This cohort was constituted of 17 females and 36 males. The mean age of the first document produced was 1.18 years, while median was 1.3 (min 0.30 - max 1.96). The mean duration of follow-up was 4.20 years, median 4.02 (min 3,70 - max 5.57). The mean age at first seizure was 12.4 months (min 4 – max 21) with 1 individual having an onset before six months and 23 before 12 months.

In order to compare phenotypes of FS Cohort with DS Cohort at the same age (before the age of two years), documents produced exclusively before two years were selected, for a total of 233 documents (mean 4.4; median 3 for each individual). From these, 2053 concepts have been extrapolated (mean 8.8 concepts per document), 303 of which were unique concepts.

The concepts present in more than 10% of individuals are shown in Table 1. The most prevalent concepts were "Fever" (found in 48 individuals – 91%), "Seizures" (44 individuals – 83%), "Convulsions" (40 individuals – 75%), "Febrile Seizures" (37 individuals – 70%), "Epilepsy" (35 individuals – 66%).

Table 1. Comparison between concepts found in more than 10% of individuals of DS Cohort (left) and FS Cohort (right). CUI = Concept Unique Identifiers.

UMLS CUI Code	Concept	DS cohort	FS cohort

		Found in Individuals	Frequency (>10%)	Found in Individuals	Frequency (>10%)
C0036572	Seizures	48	94%	44	83%
C0015967	Fever	43	84%	48	91%
C0014544	Epilepsy	42	82%	35	66%
C0751122	Dravet Syndrome	37	73%		
C0234972	Convulsions	31	61%	40	75%
C1419856	SCN1A	28	55%		
C0009952	Febrile Seizures	21	41%	37	70%
C0026827	Hypotonia	21	41%	19	36%
C2825055	Recurrence	21	41%	30	57%
C0234535	Clonic	17	33%	9	17%
C1705236	Deviation	17	33%		
C3809175	Prolonged Seizure	16	31%	8	15%
C2830004	Drowsiness	16	31%	17	32%
C0038220	Status Epilepticus	15	29%	8	15%
C0259972	Ketogenic Diet	13	25%		
C0235195	Sedation	13	25%		
C3887612	Psychomotor Agitation	12	24%	7	13%
C0009450	Infection	12	24%	9	17%
C0027066	Myoclonia	11	22%		
C1698630	Virosis	11	22%	11	21%
C0030193	Pain	10	20%	6	11%
C0013144	Falling Asleep	10	20%	12	23%
C0026837	Hypertonus	10	20%	8	15%
C0029877	Otitis	10	20%	13	25%
C0522336	Rolling of eyes	10	20%	18	34%
C0010200	Cough	10	20%	9	17%
C0004134	Ataxia	9	18%		
C0006271	Bronchiolitis	9	18%	11	21%
C0009443	Rhinitis	9	18%	14	26%
C0085639	Fall	8	16%		
C0424927	Education	8	16%		
C0024115	Pneumonia	8	16%		
C0684320	Regression	8	16%		
C0020517	Allergies	7	14%		
C0751495	Focal Seizures	7	14%		
C0015672	Fatigue	7	14%	7	13%

C0026205	Myosis	7	14%		
C0036973	Shiverings	7	14%		
C0035561	Side	6	12%		
C0031350	Pharingitis	6	12%	8	15%
C0424230	Psychomotor Delay	6	12%		
C0027441	Nasopharyngitis	6	12%		
C0549209	Startle	6	12%		
C0003123	Anorexia	5	10%		
C0034642	Rales	5	10%		
C0270844	Tonic Seizures	5	10%	6	11%
C0010520	Cyanosis	5	10%	9	17%
C0011991	Diarrhea	5	10%		
C0017160	Gastroenteritis	5	10%		
C0019209	Hepatomegaly	5	10%		
C0013384	Movement Disorder	5	10%		
C0013604	Edema	5	10%		
C0232483	Reflux	5	10%		
C0035203	Respiration	5	10%		
C0037763	Spasms	5	10%		
C1504405	Pyramidal Syndrome	5	10%		
C0008049	Chickenpox	5	10%	7	13%
C0596002	Reflex			19	36%
C0271429	Acute Otitis Media			9	17%
C0855324	Normal Pulse Pressure			8	15%
C0042963	Vomiting			8	15%
C0034150	Purpura			7	13%
C0494475	Tonic-clonic seizures			6	11%

Comparison of DS and FS cohorts

DS cohort was constituted of 54% of males and 46% of females while in FS cohort, gender comparison showed significant difference with 68% of males and 32% of females (p=0.009)
The different length of follow-up at our centre among the two cohorts shows the higher medical needs for individuals with DS (mean 3.99 years, median 3.11) compared to individuals with FS (mean 1.82 years, median 1.37 years). Indeed, the follow-up at our centre often stops when the diagnosis of FS is confirmed, and children are usually referred back to their paediatrician or general practitioner.

The mean number of documents per individual produced during the same period (0-2 years), was higher in the population with DS (6.2 vs 4.4), as well as the mean number of concepts extrapolated per document (10.9 vs 8.8).

The phenome-wide comparison of both cohorts showed a different representation of a series of concepts (Table 2). Some of these concepts were related to seizures. Concept "Deviation" ($p<0.01$), which is found within sentences describing focal seizures, point out to a significant higher occurrence of focal seizures in DS cohort compared to FS cohort. The frequency of "prolonged seizures" concept was also significantly higher in DS cohort (31% compared to 15% in FS cohort, $p=0.05$. Another concept, "sedation", which was used in the medical reports with reference to the post-ictal phase or to the need of rescue medication showed a significant difference (25% in the DS Cohort, 0% in the FS Cohort; $p=0.02$). The concept "myoclonia" was not found in the FS Cohort, while was reported in 22% of individuals of DS Cohort ($p=0.02$), and the concept "clonic" was reported two folds in the DS Cohort compared to the FS one (33% versus 17%, $p=0.05$). The concept "febrile seizures" was significantly higher in the FS Cohort and was found in 70% of individuals compared to 41% of individuals of DS Cohort ($p=0.01$). Other non seizures concepts were found only in the DS Cohort, namely "ataxia" (18%; $p=0.02$), "regression" (16%; $p=0.03$) and "pneumonia" (16%; $p=0.03$).

In addition, a series of concepts were consistently more represented in the DS Cohort than in FS Cohort, without reaching a statistical significance as "status epilepticus" (29% versus 15%; $p=0.07$, OR=2.4), "startle" (12 versus 0%; $p=0.07$, OR=7.2), "psychomotor delay" (12 versus 0%; $p=0.07$, OR=7.2), "pyramidal syndrome" (10 versus 0%; $p=0.18$, OR=4.6) "hemiparesis" (8 versus 2%; $p=0.18$, OR=4.6) and "photosensitivity" (8 versus 0%; $p=0.18$, OR=4.6).

Table 2. Phenome-wide comparison of DS Cohort and FS Cohort. *CUI = Concept Unique Identifier*

UMLS CUI Code	Concept	Found in DS individuals (%)	Found in FS individuals (%)	Not found in DS individuals (%)	Not found in FS individuals (%)	OR	*p value*
C0751122	Dravet Syndrome	36 (70.6)	0 (0)	15 (29.4)	53 (100)	129.60	0..00
C1419856	SCN1A	28 (54.9)	0 (0)	23 (45.1)	53 (100)	65.74	0.00
C1705236	Deviation	17 (33..3)	4 (7.5)	34 (66.7)	49 (92.5)	6.37	0.00
C0259972	Ketogenic Diet	13 (25.5)	0 (0)	38 (74.5)	53 (100)	18.47	0.01
C0009952	Febrile Seizures	21 (41.2)	37 (69.8)	30 (58.8)	16 (30.2)	0.34	0.01
C0235195	Sedation	13 (25.5)	4 (7.5)	38 (74.5)	49 (92.5)	4.36	0.02
C0027066	Myoclonia	11 (21.6)	3 (5.7)	40 (78.4)	50 (94.3)	4.77	0.02
C0004134	Ataxia	9 (17.6)	1 (1.9)	42 (82.4)	52 (98.1)	11.57	0.02
C0024115	Pneumonia	8 (15.7)	0 (0)	43 (84.3)	53 (100)	10.05	0.03
C0684320	Regression	8 (15.7)	0 (0)	43 (84.3)	53 (100)	10.05	0.03
C0234535	Clonic	17 (33..3)	9 (17)	34 (66.7)	44 (83)	2.56	0.05
C0085639	Fall	8 (15.7)	2 (3.8)	43 (84.3)	51 (96.2)	4.93	0.05

C3809175	Prolonged Seizure	16 (31..4)	8 (15.1)	35 (29.6)	45 (84.9)	2.87	0.05
C0014544	Epilepsy	41 (80.4)	35 (66)	10 (19.6)	18 (34)	2.34	0.06
C0038220	Status Epilepticus	15 (29.4)	8 (15.1)	36 (70.6)	45 (84.9)	2.45	0.07
C0424230	Psychomotor Delay	6 (11.8)	0 (0)	45 (88.2)	53 (100)	7.20	0.07
C0549209	Startle	6 (11.8)	0 (0)	45 (88.2)	53 (100)	7.20	0.07
C0855324	Normal Pulse Pressure	2 (3.9)	8 (15.1)	49 (96.1)	52 (98.1)	0.24	0.08
C0026205	Myosis	7 (13.7)	2 (3.8)	44 (86.3)	51 (96.2)	4.22	0.08
C0036572	Seizures	47 (92..9)	44 (83)	4 (7.8)	9 (17)	2.94	0.08
C0271429	Acute Otitis Media	3 (5.9)	9 (17)	48 (94.1)	44 (83)	0.32	0.10
C0003123	Anorexia	5 (9.8)	0 (0)	46 (90.2)	53 (100)	5.87	0.11
C0013604	Edema	5 (9.8)	1 (1.9)	46 (90.2)	52 (98.1)	5.87	0.11
C0035203	Respiration	5 (9.8)	1 (1.9)	46 (90.2)	52 (98.1)	5.87	0.11
C0037763	Spasms	5 (9.8)	1 (1..9)	46 (90.2)	52 (98.1)	5.87	0.11
C0034150	Purpura	2 (3.9)	7 (13.2)	49 (96.1)	46 (86.8)	0.28	0.12
C0522336	Rolling of eyes	10 (19.6)	18 (34)	41 (80.4)	35 (66)	0.50	0.13
C0231218	Malaise	1 (2)	5 (9.4)	50 (98)	48 (90.6)	0.20	0.15
C0427008	Stiffness	1 (2)	5 (9.4)	50 (98)	48 (90.6)	0.20	0.15
C1336751	Flat	1 (2)	5 (9.4)	50 (98)	48 (90.6)	0.20	0.15
C3887612	Pyschomotor Agitation	12 (23.5)	7 (13.2)	39 (76.5)	46 (86.8)	2.11	0.15
C0042963	Vomiting	3 (5.9)	8 (15.1)	48 (94.1)	45 (84.9)	0.37	0.16
C0036973	Shiverings	7 (13.7)	3 (5.7)	44 (86.3)	50 (94.3)	2.76	0.16
C0751495	Focal Seizures	7 (13.7)	3 (5.7)	44 (86.3)	50 (94.3)	2.76	0.16
C2825055	Recurrence	21 (41.2)	30 (56.6)	30 (58.8)	23 (43.4)	0.58	0.17
C0013473	Eating Disorders	4 (7.8)	0 (0)	47 (92.9)	53 (100)	4.60	0.18
C0018989	Hemiparesis	4 (7.8)	1 (1.9)	47 (92.9)	52 (98.1)	4.60	0.18
C0020649	Hypotension	4 (7.8)	0 (0)	47 (92.9)	53 (100)	4.60	0.18
C0032290	Aspiration Pneumonia	4 (7.8)	0 (0)	47 (92.9)	53 (100)	4.60	0.18
C0037036	Hypersalivation	4 (7.8)	0 (0)	47 (92.9)	53 (100)	4.60	0.18
C0038450	Stridor	4 (7.8)	0 (0)	47 (92.9)	53 (100)	4.60	0.18
C0205721	Nosocomial Infection	4 (7.8)	1 (1.9)	47 (92.9)	52 (98.1)	4.60	0.18
C0333641	Atrophy	4 (7.8)	0 (0)	47 (92.9)	53 (100)	4.60	0.18
C0349506	Photosensitivity	4 (7.8)	0 (0)	47 (92.9)	53 (100)	4.60	0.18
C0428167	FiO2	4 (7.8)	1 (1..9)	47 (92.9)	52 (98.1)	4.60	0.18
C0865850	Acute Respiratory Insufficiency	4 (7.8)	0 (0)	47 (92.9)	53 (100)	4.60	0.18
C1504405	Pyramidal Syndrome	4 (7.8)	0 (0)	47 (92.9)	53 (100)	4.60	0.18

Analysis of the DS cohort in regard to the early diagnosis

In the DS cohort, we compared the subgroup of individuals who had DS diagnosis confirmed or suspected before the age of 2 years of age (n=36) versus the subgroup where the diagnosis of DS was not suspected (n=15). In the first, the term (concept) Dravet syndrome was reported in the clinical reports before the age of two years while none of the individuals of the second group had any use of this term suggesting that DS diagnosis was not suspected before the age of 2 years. The mean age at first seizure was 5.3 months (min 2 – max 12) in the subpopulation that received a diagnosis or a suspected diagnosis before age 2 and 6.1 months (min 2 – max 9) in the group without an early diagnosis (p=0.2). Individuals who received diagnosis within 2 years showed a higher rate of concepts as "seizures" (p<0.01), "fever" (p<0.01), "epilepsy" (p<0.01), "prolonged seizures" (p<0.01), "convulsions" (p=0.01), "myoclonia" (p=0.02) and "ataxia" (p=0.04) compared to the second group (table 3).

Table 3. Comparison between concepts found in more than 10% individuals of DS Cohort who received the diagnosis/suspicion of DS before (left) and after (right) the age of 2 years

UMLS CUI Code	Concept	DS cohort diag < 2 years		DS cohort diag > 2 years	
		Found in Individuals	Frequency	Found in Individuals	Frequency
C0751122	Dravet Syndrome	36	100%		
C0036572	Seizures	36	100%	11	79%
C0015967	Fever	34	94%	8	57%
C0014544	Epilepsy	33	92%	8	57%
C0234972	Convulsions	26	72%	5	36%
C1419856	SCN1A	22	61%	6	43%
C0026827	Hypotonia	17	47%	4	29%
C2825055	Recurrence	17	47%	4	29%
C0009952	Febrile Seizures	15	42%	5	36%
C3809175	Prolonged Seizure	15	42%		
C0234535	Clonic	14	39%	3	21%
C0038220	Status Epilepticus	13	36%	2	14%
C0596002	Reflex	13	36%	2	14%
C1705236	Deviation	12	33%	5	36%
C2830004	Drowsiness	12	33%	4	29%
C0259972	Ketogenic Diet	11	31%	2	14%
C0235195	Sedation	11	31%	2	14%
C1698630	Virosis	11	31%		
C3887612	Psychomotor Agitation	10	28%	2	14%

C0009450	Infection	10	28%		
C0027066	Myoclonia	10	28%		
C0004134	Ataxia	9	25%		
C0030193	Pain	9	25%		
C0013144	Falling Asleep	9	25%		
C0029877	Otitis	9	25%		
C0009443	Rhinitis	9	25%		
C0010200	Cough	9	25%		
C0424927	Education	8	22%		
C0522336	Rolling of eyes	8	22%	2	14%
C0020517	Allergies	7	19%		
C0006271	Bronchiolitis	7	19%	2	14%
C0015672	Fatigue	7	19%		
C0026837	Hypertonus	7	19%	3	21%
C0026205	Myosis	7	19%		
C0024115	Pneumonia	7	19%		
C0085639	Fall	6	17%		
C0494475	Tonic-clonic seizures	6	17%		
C0751495	Focal Seizures	6	17%		
C0031350	Pharingitis	6	17%		
C0424230	Psychomotor Delay	6	17%		
C0684320	Regression	6	17%	2	14%
C0549209	Startle	6	17%		
C0036973	Shiverings	6	17%		
C0003123	Anorexia	5	14%		
C0035561	Side	5	14%		
C0034642	Rales	5	14%		
C0013604	Edema	5	14%		
C0035203	Respiration	5	14%		
C0424230	Psychomotor Delay	5	14%		
C0008049	Chickenpox	5	14%		
C0004093	Asthenia	4	11%		
C0270844	Tonic Seizures	4	11%		
C0010520	Cyanosis	4	11%		
C0011991	Diarrhea	4	11%		
C0221725	Bronchial Obstruction	4	11%		
C0017160	Gastroenteritis	4	11%		
C0021400	Flu	4	11%		

C0037036	Hypersalivation	4	11%		
C0042769	Viral Infection	4	11%		
C0349506	Photosensitivity	4	11%		
C0032290	Aspiration Pneumonia	4	11%		
C1272641	Arterial Blood Pressure	4	11%		
C0027441	Nasopharyngitis	4	11%	2	14%
C0037763	Spasms	4	11%		
C0038450	Stridor	4	11%		
C1504405	Pyramidal Syndrome	4	11%		
C0018916	Angiome			2	14%
C0085584	Encephalopathy			2	14%
C0018989	Hemiparesis			2	14%
C0018991	Hemiplegia			2	14%
C0019209	Hepatomegaly			2	14%
C0232483	Reflux			2	14%

Discussion

This study shows that narrative medical reports produced before two years include several clinical concepts which are significantly associated with individuals with DS compared to FS, this latter condition representing the main differential diagnosis at the onset. These concepts are consistent with the main clinical findings constituting the criteria for differentiating DS from FS in first two years of life.

FS are usually reported after the first year with some cases initiating before 12 months. They are usually brief and generalized. [24] In our study, concepts referred to prolonged ("status epilepticus", "prolonged seizures", "sedation") and focal seizures ("deviation") are prominent in the DS cohort, emphasizing the higher tendency of individuals with DS to present at onset long lasting and focal febrile seizures compared to individuals with FS. [16][25][26] Importantly, individuals with DS develop different types of seizures as myoclonic or atypical absences in addition to the first seizures mimicking FS. We observed in our DS cohort concepts referring to seizures other than febrile convulsions, including "Myoclonia" and "startle", which is mostly used in narrative reports to depict myoclonic seizure semiology. [16][27][28] The concept "hemiparesis" was more frequent in the DS Cohort compared to FS one. This is consistent with the higher occurrence of transitory hemiplegia after long-lasting hemiclonic seizures, a type of seizure being quite suggestive of DS. [16][27][29]

Some important non-seizure concepts also emerged, differentiating the two cohorts. Subjects with DS and FS show a normal neurodevelopment at the seizure onset, but then psychomotor trajectories deviate. [26][30] In accordance, concepts related to psychomotor delay were found only in the DS Cohort ("Regression", "Psychomotor delay"). In addition, "Ataxia" was significantly more reported DS Cohort, reflecting the peculiar gait disorder commonly observed in individuals with DS, and representing an early motor-marker of this condition. [28][31]
Interestingly, the concept "febrile seizures" was found with significant higher frequency in the FS Cohort probably because it was used for a "diagnostic" purpose in the clinical reports.

The study was carried out in a tertiary epilepsy center, so it is plausible that some words have been chosen as a consequence of the clinical suspicion of Dravet Syndrome by highly experienced specialist in epileptology (e.g. "myoclonia", "ataxia"). However, many of the medical reports were done by physicians without a specific expertise in epilepsy or DS (e.g. emergency care or intensive care physicians), emphasizing the uniformity of expressions used for reporting disease and individuals description, and suggesting that most of key-concepts may have also been found into non-specialists medical reports (e.g. "deviation", "prolonged seizures", "startle").

Several studies show a substantial worldwide issue of diagnostic delay of DS, with a mean age at diagnosis that is usually over two years, resulting in "unnecessary, costly, and, at times, invasive testing, and use of ineffective therapies, which can exacerbate seizures, increase the risk of status epilepticus, and worsen cognitive outcome". [17][32][33][34] Moreover, DS is certainly less recognized in adult population and in developing countries. [19][20]

Computer-based models using EHRs able to suggest diagnosis/avoid misdiagnosis are gaining ground. [3][35] These models are mostly based on structured data, as image-based or laboratory data. [36][37] Recently, more complex models of artificial intelligence are emerging, which are able to elaborate diagnosis by extracting clinically relevant information from unstructured data in EHRs. [38][39]

On the basis of our findings, further extensive studies might focus on elaborating a specific computer algorithm which combines significant concepts and their age of appearance within narrative specialists and non-specialists reports, in order to automatically produce an alert signal suggesting possible diagnosis of DS.

Some results of our analysis set out some additional insights. For example, the major incidence of concept “pneumonia” in DS Cohort compared to FS Cohort appears to be relevant, since it can represent both a facilitator and a complication of inhalation during a long lasting convulsive seizure or a status epilepticus.[40] In addition, a number of concepts related to peri-ictal nosocomial and respiratory complications were found with higher frequency in reports of individuals with DS (“nosocomial infections”, “acute respiratory insufficiency”, “aspiration pneumonia”, “FiO2”, “stridor”) underlying that convulsive status epilepticus might be a life-threatening condition in this population. [40][41]

Furthermore, in this study the concept “Dravet Syndrome” was found in 72% of individuals of DS Cohort before the age of two years. This is concordant with the literature showing the early recognition of DS in France. [34]
Some clinical concepts were found with higher frequency in the reports of individuals who received the diagnosis/suspicion of DS before the age of two years: the “long-lasting seizure” concepts (“Status epilepticus”, “Prolonged seizures”, “Sedation”), the “myoclonic” concepts (“Myoclonia”, “Startles”), the “drug resistance” concepts (“Ketogenic diet”), as well as “Ataxia”, and “Photosensitivity”. Although statistical significance was not reached for all these concepts as sample was small, these findings may support that these clinical concepts are the most DS diagnosis orienting. We can hypothesise that individuals belonging to the sub-group who did not receive a diagnosis within two years presented a less “typical” phenotype. The diagnosis was made later than 2 years of age when the full blown syndrome is often complete with pharmacoresistant seizures and developmental plateauing. However, in this subgroup without early diagnosis with individuals presenting “intermediate” features between only FS and the “complete” DS clinical picture, the median age at first seizure was significantly lower than in FS cohort (6.1 months vs 12.4 months). This finding confirms that age at first seizure might be the strongest predictor of DS in infants who experience febrile seizures.[25]

Study Limitations

Word sense disambiguation poses a challenge in extracting meaningful data from unstructured text. Clinical notes often contain terms or phrases that have more than one meaning, [8] or that need for a contextualisation to understand the real clinical meaning. For example, concept “deviation” apparently do not link to a specific clinical feature, but in the narrative reports of individuals of both cohorts it was mostly used within the description of the seizure semeiology, thus referring to a focal seizure.

The presence of a clinical concept in a medical report does not necessary implies that the individual presents this clinical feature. For instance, the concept “spasms” that we found in five individuals of the DS Cohort, was used within the clinical description of paroxysmal motor events that could suggest epileptic spasms, but was not confirmed in any of them. Similarly, concept “Dravet Syndrome” could be found in reports of subjects who received the diagnosis, or in which a suspicion was made (i.e: “We see today patient X for the suspicion of Dravet Syndrome”). The method used by Dr Warehouse automatically classifies concepts according to polarity (negation/affirmation) and the experiencer (patient/family). But there may still be errors in the classification. In addition, the classification does not take into account the notion of hypothesis.

In this study, the FS population presents some “atypical” features; for instance, the frequency of the concept “status epilepticus” in these subjects is higher than expected in terms of incidence in individuals with febrile seizures. [42][43] This might be due to a preferential referral to university hospital of individuals with febrile long lasting seizures or febrile status epilepticus, as they might need further admission to ICU.

Conclusion

Narrative medical reports of individuals younger than two years with febrile seizures, contain different words depending if they have or will develop clinical phenotype of DS, or not. The elaboration of algorithm exploiting NLP on the basis of our work, could be useful to early individualize these individuals, in order to establish early diagnosis and adequate therapy that in some instances need to address them to expert epilepsy centres.

This methodology would represent an innovative, “cheap”, transposable and sustainable methodology to reduce time of diagnosis for individuals with Dravet Syndrome and other rare conditions.

Some “key early symptoms” often identified by the patients/care givers and the non-expert physicians are merely linked to a given known disease causing diagnosis delay. Using these symptoms and signs as alerts and warning signs can help to address patients earlier to expert centres for a definite diagnosis. The future step is to validate the impact of the implementing of these “warnings” in the electronic health records on shortening the patient’s odyssey to diagnosis and therapies.

References

[1] Gunter TD, Terry NP. The emergence of national electronic health record architectures in the United States and Australia: models, costs, and questions. J Med Internet Res. 2005 Mar 14;7(1):e3.

[2] Landi I, Glicksberg BS, Lee HC, Cherng S, Landi G, Danieletto M, et al. Deep representation learning of electronic health records to unlock patient stratification at scale. npj Digit Med. 2020;3:1–11.

[3] Olivera P, Danese S, Jay N, Natoli G, Peyrin-Biroulet L. Big data in IBD: a look into the future. Nat Rev Gastroenterol Hepatol. 2019 May 1;16(5):312–21.

[4] Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. Health Aff. 2014;33(7):1123–31.

[5] Shen F, Liu S, Wang Y, Wen A, Wang L, Liu H. Utilization of electronic medical records and biomedical literature to support the diagnosis of rare diseases using data fusion and collaborative filtering approaches. J Med Internet Res. 2018;20(10).

[6] Southall NT, Natarajan M, Lau LPL, Jonker AH, Deprez B, Guilliams T, et al. The use or generation of biomedical data and existing medicines to discover and establish new treatments for patients with rare diseases-recommendations of the IRDiRC Data Mining and Repurposing Task Force. Orphanet J Rare Dis. 2019;14(1):225.

[7] Garcelon N, Neuraz A, Salomon R, Bahi-Buisson N, Amiel J, Picard C, et al. Next generation phenotyping using narrative reports in a rare disease clinical data warehouse. Orphanet J Rare Dis. 2018;13:85.

[8] Townsend H. Natural language processing and clinical outcomes: the promise and progress of NLP for improved care. J AHIMA. 2013;84:44–5.

[9] Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res. 2004;32:D267-70.

[10] Rindflesch TC, Fiszman M. The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. J Biomed Inform. 2003;36(6):462–77.

[11] Wu YW, Sullivan J, McDaniel SS, Meisler MH, Walsh EM, Li SX, et al. Incidence of dravet syndrome in a US population. Pediatrics. 2015;136(5):e1310–5.

[12] Scheffer IE, Berkovic S, Capovilla G, Connolly MB, Guilhoto L, Hirsch E, et al. ILAE Classification of the Epilepsies Position Paper of the ILAE: Commission for Classification and Terminology. Epilepsia. 2017;58:512–21.

[13] Dravet C. The core Dravet syndrome phenotype. Epilepsia. 2011;52(SUPPL. 2):3–9.

[14] Hirose S, Scheffer IE, Marini C, De Jonghe P, Andermann E, Goldman AM, et al. SCN1A testing for epilepsy: Application in clinical practice. Epilepsia. 2013;54:946–52.

[15] Catarino CB, Liu JYW, Liagkouras I, Gibbons VS, Labrum RW, Ellis R, et al. Dravet syndrome as epileptic encephalopathy: Evidence from long-term course and neuropathology. Brain. 2011;134(10):2982–3010.

[16] Hattori J, Ouchida M, Ono J, Miyake S, Maniwa S, Mimaki N, et al. A Screening test for the prediction of Dravet syndrome before one year of age. Epilepsia. 2008;49(4):626–33.
[17] Lagae L, Brambilla I, Mingorance A, Gibson E, Battersby A. Quality of life and comorbidities associated with Dravet syndrome severity: a multinational cohort survey. Dev Med Child Neurol. 2018;60(1):63–72.
[18] Bremer A, Lossius MI, Nakken KO. Dravet syndrome - considerable delay in making the diagnosis. Acta Neurol Scand. 2012;125(5):359–62.
[19] Jansen FE, Sadleir LG, Harkin LA, Vadlamudi L, McMahon JM, Mulley JC, et al. Severe myoclonic epilepsy of infancy (Dravet syndrome): Recognition and diagnosis in adults. Neurology. 2006;67(12):2224–6.
[20] Connolly MB. Dravet Syndrome: Diagnosis and Long-Term Course. Can J Neurol Sci. 2016;43:S3–8.
[21] Garcelon N, Neuraz A, Salomon R, Faour H, Benoit V, Delapalme A, et al. A clinician friendly data warehouse oriented toward narrative reports: Dr. Warehouse. J Biomed Inform. 2018;80:52–63.
[22] Neuraz A, Chouchana L, Malamut G, Le Beller C, Roche D, Beaune P, et al. Phenome-wide association studies on a quantitative trait: application to TPMT enzyme activity and thiopurine therapy in pharmacogenomics. PLoS Comput Biol. 2013;9(12):e1003405.
[23] Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, et al. PheWAS: Demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics. 2010;26(9):1205–10.
[24] Baumann RJ. Technical report: treatment of the child with simple febrile seizures. Pediatrics [Internet]. 1999 [cited 2021 Jun 8];103(6 I):1278–9. Available from: https://pubmed.ncbi.nlm.nih.gov/10353983/
[25] Cetica V, Chiari S, Mei D, Parrini E, Grisotto L, Marini C, et al. Clinical and genetic factors predicting Dravet syndrome in infants with SCN1A mutations. Neurology. 2017;88(11):1037–44.
[26] Dravet C, Guerrini R. Dravet Syndrome. John Libbey Eurotext; 2011.
[27] Ohki T, Watanabe K, Negoro T, Aso K, Haga Y, Kasai K, et al. Severe myoclonic epilepsy in infancy: evolution of seizures. Seizure. 1997;6(3):219–24.
[28]Gataullina S, Dulac O. From genotype to phenotype in Dravet disease. Seizure. 2017;44:58–64.
[29] Ragona F, Brazzo D, De Giorgi I, Morbi M, Freri E, Teutonico F, et al. Dravet syndrome: Early clinical manifestations and cognitive outcome in 37 Italian patients. Brain Dev. 2010;32:71–7.
[30] Nabbout R, Chemaly N, Chipaux M, Barcia G, Bouis C, Dubouch C, et al. Encephalopathy in children with Dravet syndrome is not a pure consequence of epilepsy. Orphanet J Rare Dis. 2013;8(1):1–8.

[31] Verheyen K. Motor Development in children with Dravet syndrome. Dev Med Child Neurol. 2019;61:950–6.
[32] Wirrell EC, Laux L, Donner E, Jette N, Knupp K, Meskis MA, et al. Optimizing the Diagnosis and Management of Dravet Syndrome: Recommendations From a North American Consensus Panel. Pediatr Neurol. 2017;68:18-34.e3.
[33] Wirrell EC, Laux L, Franz DN, Sullivan J, Saneto RP, Morse RP, et al. Stiripentol in Dravet syndrome: results of a retrospective U.S. study. Epilepsia. 2013;54(9):1595–604.
[34] Nabbout R, Auvin S, Chiron C, Thiele E, Cross H, Scheffer IE, et al. Perception of impact of Dravet syndrome on children and caregivers in multiple countries: looking beyond seizures. Dev Med Child Neurol. 2019;61:1229–36.
[35] Shilo S, Rossman H, Segal E. Axes of a revolution: challenges and promises of big data in healthcare. Nat Med. 2020 Jan 1;26(1):29–38.
[36] Castaneda C, Nalley K, Mannion C, Bhattacharyya P, Blake P, Pecora A, et al. Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. J Clin Bioinforma. 2015;5(1):4.
[37] Fitipaldi H, McCarthy MI, Florez JC, Franks PW. A global overview of precision medicine in type 2 diabetes. Diabetes. 2018;67:1911–22.
[38] Liang H, Tsui BY, Ni H, Valentim CCS, Baxter SL, Liu G, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med. 2019;25(3):433–8.
[39] Hully M, Lo Barco T, Kaminska A, Barcia G, Cances C, Mignot C, et al. Deep phenotyping unstructured data mining in an extensive pediatric database to unravel a common KCNA2 variant in neurodevelopmental syndromes. Genet Med. 2021;
[40] Shmuely S, Sisodiya SM, Gunning WB, Sander JW, Thijs RD. Mortality in Dravet syndrome: A review. Epilepsy Behav. 2016;64:69–74.
[41] Kim Y, Bravo E, Thirnbeck CK, Smith-Mellecker LA, Kim SH, Gehlbach BK, et al. Severe peri-ictal respiratory dysfunction is common in Dravet syndrome. J Clin Invest. 2018;128(3):1141–53.
[42] Hesdorffer DC, Shinnar S, Lewis D V., Moshé SL, Nordli DR, Pellock JM, et al. Design and phenomenology of the FEBSTAT study. Epilepsia. 2012 Sep;53(9):1471–80.
[43] Vitaliti G, Castagno E, Ricceri F, Urbino A, Di Pianella AV, Lubrano R, et al. Epidemiology and diagnostic and therapeutic management of febrile seizures in the Italian pediatric emergency departments: A prospective observational study. Epilepsy Res. 2017;129:79–85.

Chapter 3: "Outlining the Natural History of Rare Diseases using Natural Language Processing of narrative unstructured records: The example of Dravet syndrome"

Abstract

Background

The growing use electronic health records is promoting the application of big data mining to diagnose and phenotype rare diseases. In this study we aimed to explore the utilization of electronic narrative clinical reports of a cohort of patients with Dravet Syndrome (DS) longitudinally followed at our center, in order to identify the capacity of this methodology to retrace the known natural history of this disease.

Methods

Data were collected at Necker Hospital in Paris, which integrates a document-based clinical data warehouse, employing Natural Language Processing (NLP) and UMLS (Unified Medical Language System) Metathesaurus to recognize phenotype concepts inside medical reports.
We selected all patients with confirmed diagnosis of DS and having a first document produced before the age of two years (DS Cohort). For each individual included in the DS Cohort, we asked to Dr Warehouse to randomly select 5 to 6 individuals followed at the Necker hospital, with: i) same gender, ii) same age at last report, iii) same number of reports ("general control population").
We then selected a second control population including individuals with same criteria, plus: iv) followed in the Pediatric Neurology Unit, v) including the concept "seizure" or "convulsion" in a report before the age of 2 years All clinical concepts were extracted from narrative reports ("neurological control population"). A phenome-wide association study was run, comparing DS Cohort and the control populations, to exclude concepts non-specific for DS. We then performed a qualitative analysis of the survived concepts basing on their median age of first appearance.

Results

Concepts related to the typical epilepsy features of DS emerged in the first years of life. A series of non-epilepsy related concepts emerged lately, referring to the plateauing in the development trajectory, gait and movement disorder, language impairment, behavior disorders, socialization issues.

Interpretation

The extraction of clinical terms by narrative reports of children with DS, allows to outline the natural history of this rare disease in early childhood. This model of "longitudinal phenotyping" could be applied for other rare conditions whose natural history is still poorly known.

Introduction

The growing use of Electronic Health Records (EHRs) is promoting the application of big data mining in the health-care field and encouraging the shift to precision medicine approaches. [1][2] Most innovative uses include the development of predictive models to individuate high risk patients and support prevention initiatives. [1][3][4] In addition, big data are allowing a better knowledge of some diseases with a finer delineation of diagnosis, clinical course and outcome. This constitutes a pioneering advance in the field of the rare diseases, in which knowledge is limited by definition, in the face of a growing identification of new genetic conditions, due to the diffusion of diagnostic techniques. [5][6][7][8]

We recently reported how the use EHRs could help to address the diagnosis of Dravet Syndrome (DS), a rare genetic Developmental and Epileptic Encephalopathy, through mining for main clinical concepts in the narrative medical reports by exploiting the Natural Language Processing (NLP). [9]

Natural history of DS is well described till early adulthood. [10] [11] The first stage is the "febrile stage", starting in first year of life in healthy children, and is characterized by presentation of convulsive seizures, mainly related to fever. [11] Seizures are often tonic-clonic, hemiclonic, or myoclonic, and can last more than 15 minutes, sometimes evolving to status epilepticus and necessitating high benzodiazepines and sedation in intensive care unit. [11] [12] [13] Second phase, between 1 and 4 years, is the "worsening and polymorphous seizures stage", characterized by the appearance of different seizure types: myoclonic seizures, atypical absences and focal seizures. [11] The psychomotor impairment becomes clear, including language disorders and delayed fine and gross motor skills. Attention deficit, hyperactivity, and sometimes autistic features also appear. [11] After this stage myoclonic and absence seizures tend to disappear, convulsive seizures decrease, and cognitive impairment persists together with social and behavioral issues, and gait deterioration. [11] [14]

In this study, we explored the utilization of data mining through NLP from narrative clinical reports to retrace the known natural history of DS, in order to validate this methodology in other rare and very rare diseases in childhood.

Methods

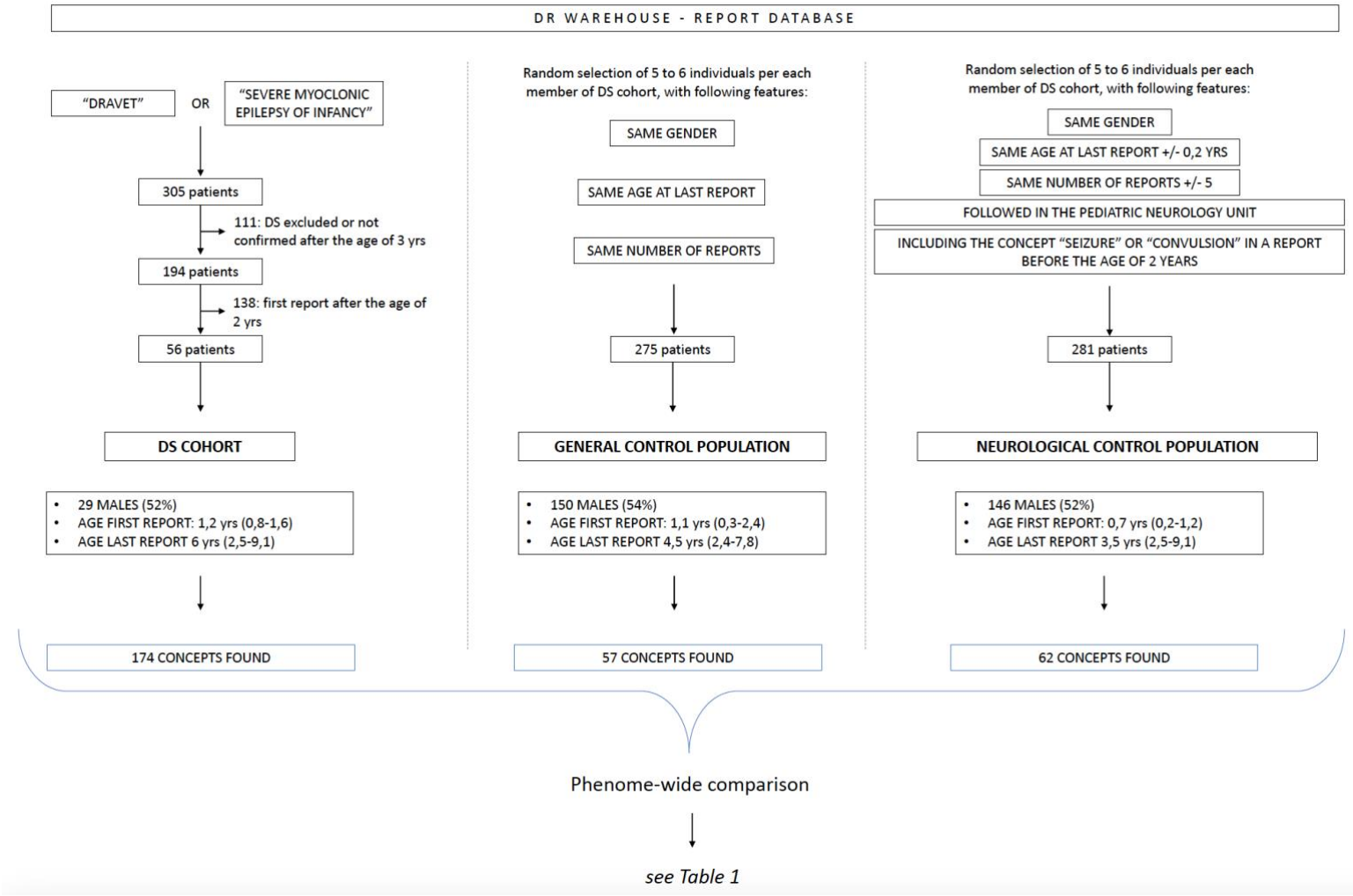
Data were collected at Necker Children Hospital in Paris. From 2015 Necker Hospital integrates Dr Warehouse®, [15] a document-based clinical data warehouse which employs NLP and UMLS (Unified Medical Language System) Metathesaurus [16] to recognize phenotype concepts inside medical reports. Dr Warehouse includes medical records of over 9 million for almost 840 000 patients.

By using the research field in Dr Warehouse®, we selected patients presenting the term “Dravet” or “Severe myoclonic epilepsy of infancy”, the previous denomination of Dravet Syndrome, at least in one document. We included patients with final diagnosis of DS confirmed after the age of three years and having a first document produced before the age of two years, in order to retrace the course of the disease. These patients constituted the “Dravet Syndrome Cohort” (DS Cohort). (Figure 1)

For each individual included in the DS Cohort, we asked to Dr Warehouse to randomly select 5 to 6 individuals followed at the Necker hospital, with: i) same gender, ii) same age at last report, iii) same number of reports. These subjects constituted the “general control population”. (Figure 1)

We then selected a second control population including individuals with same criteria, plus: iv) followed in the Pediatric Neurology Unit, v) including the concept “seizure” or “convulsion” in a report before the age of 2 years. To reach the same size of the “general control population”, individuals with +/- 0.2 years age at last report, and number of reports +/- 5 were also included. These subjects constituted the “neurological control population”. (Figure 1)

Figure 1. Flowchart of the selection procedures and constitution of the cohorts



All concepts were extracted from clinical reports of all cohorts, excluding concepts associated to negations and related to familial members. [15] In order to exclude clinical concepts non-specific for DS, we performed a phenome-wide association study using a COX survival model, including all concepts with a least 5 occurrences in DS cohort.

To find concepts specifically associated with DS, we performed a phenome-wide association study (PheWAS) using Cox regression.[17] A PheWAS is a series of exposed-non exposed association studies nested in a cohort, allowing to detect associations between the group of interest and all the available phenotypes. [18] [19] Traditionally PheWAS use logistic regression to assess the association. In this work, we preferred a Cox regression model which enables us to consider the age of first appearance of a phenotypic concept and not only the presence of the concept. Using a survival model also deals with the duration of follow-up and allows to increase the statistical power of the PheWAS.[17] As a PheWAS requires to perform a large series of statistical tests, a correction for multiple testing. We applied a false discovery rate correction using q=0.05. [20]

We then performed a qualitative analysis of the survived concepts basing on their median age of first appearance.

Results

Study cohorts

Concepts "Dravet" and/or "Severe myoclonic epilepsy of infancy" appeared in 305 subjects in the warehouse: 194 of them had a final diagnosis of DS in the last document, 56 had at least one document under the age of two years and constituted the DS cohort. This was constituted of 29 males and 27 females. Median age was 1.19 years (min 0.81 - max 1.57) at first report and 6 years (min 2.5 – max 9.1) at last report (median length of follow-up: 5.3 years). The median age at first seizure was 5.5 months (min 2 – max 12). Individual in this cohort had 983 documents (mean 19.3 per patient, min 3 – max 92) and a total of 9860 concepts were extracted from these reports. Distinct clinical concepts recognized were 174.

The general control cohort consisted of 275 individuals (150 males). Median age was 1.09 years (min 0.27 - max 2.42) at first report and 4.5 years (min 2.4 – max 7.8) at last report (median length of follow-up: 2.9 years). 57 distinct concepts were found.

The neurological control population is constituted of 281(146 males). Median age was 0.66 years (min 0.18 - max 1.24) at first report and 3.5 years (min 2.4 – max 6.8) at last report (median length of follow-up: 3 years). 62 distinct concepts were found.

Ages at first ($p<0.001$) and last report ($p=0.004$) are significantly different in the 3 populations.

Phenome-wide analysis and of concept appearance in DS Cohort

The phenome-wide comparison of three cohorts showed a different representation of several concepts. Clinical concepts which were found significantly more represented in the DS Cohort ($p<0.05$) are listed in Table 1, ordered by the median age of first appearance in a clinical report.

Concepts appearing within 2 years age are mostly related with the typical epilepsy features at the onset of DS, namely the occurrence of convulsive and prolonged seizures triggered by fever, often needing high doses of benzodiazepines and hospitalizations in ICU due to respiratory complications of status epilepticus: "Febrile seizures" (1.27 years), "Sedation" (1.4 years), "Status epilepticus"(1.43 years), "Convulsive seizures" (1.45 years), "Prolonged seizures" (1.57 years), "Aspiration pneumonia" (1.67 years). A series of concepts related to infectious diseases also emerged at this age, probably having acted as seizure triggers: "Pneumonia" (1.57 years), "Virose" (1.58 years), "Infection" (1.78 years), "Pharyngitis" (1.83 years), "Rhinitis" (1.85 years). Concepts as "Deviation"

(1.57 years), which is found in sentences describing seizure semiology, represent the higher tendency to present convulsions with focal signs comparing with patients who only show febrile seizures. The finding in the DS Cohort of concepts “Hemiclonic seizures” (1.28 years) and “Hemiparesis” (1.49 years) is coherent with the known recurrence of prolonged hemiclonic seizures, which are highly evocative of DS at this age, particularly if evolving to focal motor status epilepticus and leading to a transitory hemiplegia.

Subsequently, concepts related to new types of seizures and the establishment of a drug-resistant epilepsy appeared, coherently with natural history: “Myoclonic Seizures” (2.15 years), “Ketogenic Diet” (2.35 years), “Absence Seizures” (2.81 years), “Tonic seizures” (3.56 years), “Drug-resistant epilepsy” (4.05 years), “Seizure cluster” (4.37 years), “Focal Seizures” (4.38 years). Concept “Photosensitivity” was also found in the third year (2.63 years).

A series of non-epilepsy related concepts emerged after 2 years age, referring to the neurological features and comorbidities classically associated to DS. They mostly include terms related to the gait and movement disorder:[21] “Ataxia” (2.41 years), “Cerebellar syndrome” (2.88 years), “Fall” (3.1 years), “Impaired balance” (3.88 years), “Gait disorder” (4.46 years). Concepts related to the plateauing/slowing in the development trajectory appeared [“Psychomotor Retardation” (2.93 years), “Cognitive impairment” (3.23 years)], followed by several terms associated to behavior disorders: “Hyperactivity” (3.24 years), “Aggressiveness” (3.94 years), “Distractibility” (3.96 years), “Irritability” (4.02 years), “Behavior disorder” (4.1 years). In addition, in this period emerged concepts related to social interaction impairment and autism spectrum disorder: “Disruptive Developmental Disorder” (3.12 years), “Autism” (4.73 years), “Autism spectrum disorder” (6.95 years).

Concept	n/total(%)	P-value (FDR)	Hazard Ratio [95%CI]	Crosspoint (survival curve)	Median Age at first appearance in a report [min-max]
Myosis	7/19 (0.37)	4.45 × 10−4	5.33 [2.09;13.56]	1.000	1.06 [0.92-1.215]
Purpura	6/16 (0.38)	2.40 × 10−3	4.82 [1.75;13.31]	1.980	1.1 [0.92-1.6]
Seizures	54/319 (0.17)	2.36 × 10−12	2.91 [2.16;3.93]	1.000	1.195 [0.835-1.6675]
Fever	52/314 (0.17)	7.48 × 10−11	2.74 [2.02;3.72]	-	1.215 [0.8525-1.91]
Epilepsy	53/191 (0.28)	2.41 × 10−27	6.18 [4.45;8.59]	1.100	1.24 [0.94-1.87]
Rash	15/61 (0.25)	5.77 × 10−4	2.79 [1.56;5.00]	-	1.25 [0.895-3.15]
Febrile Seizures	26/64 (0.41)	4.81 × 10−16	8.03 [4.86;13.29]	0.790	1.27 [0.81-2.555]
Hemiclonic Seizures	5/7 (0.71)	2.23 × 10−4	22.01 [4.26;113.66]	0.350	1.28 [1.12-2.04]
Sedation	18/111 (0.16)	3.18 × 10−2	1.74 [1.05;2.89]	3.880	1.4 [1.09-2.525]
Status Epilepticus	25/63 (0.40)	5.55 × 10−14	6.95 [4.19;11.53]	1.010	1.435 [1.01-2.9175]
Convulsive seizures	43/159 (0.27)	3.07 × 10−19	5.11 [3.58;7.30]	1.020	1.45 [0.92-2.09]

Hemiparesis	12/57 (0.21)	9.04 × 10−3	2.34 [1.24;4.43]	0.870	1.49 [0.99-2.46]
Prolonged seizure	25/44 (0.57)	1.40 × 10−19	15.89 [8.73;28.92]	0.860	1.57 [1.21-2.79]
Deviation	25/117 (0.21)	1.33 × 10−5	2.68 [1.72;4.18]	0.820	1.575 [1.1825-3.07]
Pneumonia	12/62 (0.19)	2.32 × 10−2	2.08 [1.11;3.91]	-	1.575 [1.24-2.815]
Virose	20/69 (0.29)	4.76 × 10−7	3.81 [2.26;6.41]	4.080	1.58 [1.26-5.06]
Aspiration Pneumonia	7/19 (0.37)	1.08 × 10−3	4.81 [1.88;12.35]	1.000	1.67 [1.17-2.73]
Hepatic cytolisis	6/23 (0.26)	2.99 × 10−2	2.83 [1.11;7.23]	0.690	1.67 [1.41-2.81]
Infection	26/157 (0.17)	7.68 × 10−3	1.77 [1.16;2.70]	0.350	1.78 [1.085-3.42]
Pharyngitis	11/46 (0.24)	1.98 × 10−3	2.91 [1.48;5.74]	0.570	1.83 [0.81-2.74]
Tremors	11/34 (0.32)	4.08 × 10−4	3.82 [1.82;8.05]	1.520	1.83 [1.34-5.04]
Rhinits	19/111 (0.17)	1.82 × 10−2	1.81 [1.11;2.97]	5.390	1.855 [0.8625-3.0425]
Staring	7/32 (0.22)	3.26 × 10−2	2.49 [1.08;5.77]	1.810	1.88 [1.405-2.55]
Drooling	8/23 (0.35)	4.06 × 10−4	4.71 [2.00;11.13]	7.780	1.92 [0.64-3.98]
Somnolence	27/142 (0.19)	9.65 × 10−5	2.31 [1.52;3.51]	0.360	1.975 [1.1725-3.005]
Clonic movements	36/106 (0.34)	2.43 × 10−17	5.78 [3.85;8.68]	1.170	2.04 [1.21-3.095]
Anorexia	10/26 (0.38)	7.89 × 10−5	5.02 [2.25;11.19]	0.580	2.09 [1.15-5.57]
Myoclonic seizures	27/46 (0.59)	5.29 × 10−19	14.46 [8.03;26.03]	1.120	2.15 [1.23-3.995]
Eyes rolling	17/64 (0.27)	1.36 × 10−5	3.43 [1.97;5.97]	2.550	2.15 [1.37-2.845]
GEFS	6/9 (0.67)	8.44 × 10−5	16.37 [4.06;65.98]	2.520	2.21 [1.86-3.905]
Pyelonephritis	7/26 (0.27)	7.47 × 10−3	3.27 [1.37;7.78]	1.400	2.27 [0.82-2.89]
Ketogenic Diet	30/47 (0.64)	1.72 × 10−20	16.99 [9.34;30.91]	1.660	2.35 [1.385-4.765]
Psychomotor Agitation	33/134 (0.25)	3.73 × 10−9	3.27 [2.20;4.84]	0.530	2.38 [1.45-5.5]
Ataxia	24/54 (0.44)	1.28 × 10−14	8.29 [4.84;14.20]	-	2.41 [1.5-4.745]
Eating disorders	9/29 (0.31)	1.18 × 10−3	3.69 [1.68;8.11]	7.430	2.42 [1.32-4.82]
Rhinopharingitis	13/65 (0.20)	7.85 × 10−3	2.28 [1.24;4.19]	0.920	2.44 [1.26-3.01]
Photosensitivity	11/17 (0.65)	5.65 × 10−8	15.81 [5.84;42.82]	-	2.63 [2.03-5.095]
Atonia	7/9 (0.78)	2.55 × 10−5	29.32 [6.08;141.30]	0.580	2.64 [2.025-4.515]
Startle	15/43 (0.35)	4.71 × 10−6	4.48 [2.36;8.53]	1.470	2.7 [1.38-5.57]
Loss of consciousness	11/34 (0.32)	2.28 × 10−4	3.88 [1.89;7.97]	0.570	2.7 [2.08-4.665]
Mydriasis	5/18 (0.28)	2.18 × 10−2	3.35 [1.19;9.41]	-	2.72 [2.09-2.8]
Drowsiness	27/103 (0.26)	3.47 × 10−8	3.45 [2.22;5.35]	0.790	2.76 [1.3375-4.235]
Fatigue	29/119 (0.24)	3.88 × 10−9	3.54 [2.33;5.40]	0.870	2.76 [2.12-4.105]
Shout	5/13 (0.38)	4.59 × 10−3	5.06 [1.65;15.53]	-	2.8 [1.57-4.82]
Absence seizures	19/32 (0.59)	6.29 × 10−12	12.23 [5.99;24.98]	0.820	2.81 [1.98-5.51]
Cerebellar Syndrome	5/13 (0.38)	4.44 × 10−3	5.08 [1.66;15.55]	0.660	2.88 [2.19-5.57]
Psychomotor Retardation	20/105 (0.19)	5.40 × 10−3	2.00 [1.23;3.25]	2.050	2.93 [1.66-4.24]
Gait	20/42 (0.48)	4.23 × 10−12	8.56 [4.66;15.70]	-	2.955 [2.0925-5.26]
Fall	25/146 (0.17)	1.27 × 10−2	1.73 [1.12;2.66]	1.030	3.1 [2.1875-5.5175]
Disruptive Developmental Disorder	5/12 (0.42)	2.37 × 10−3	5.93 [1.88;18.71]	0.580	3.12 [2.84-3.96]
Ocular Deviation	5/10 (0.50)	7.80 × 10−4	8.39 [2.43;29.03]	0.720	3.16 [1.88-3.64]

Cognitive impairment	5/9 (0.56)	2.77 × 10−3	7.92 [2.04;30.75]	1.850	3.23 [1.425-5.105]
Insomnia	5/13 (0.38)	5.68 × 10−3	4.87 [1.59;14.96]	0.830	3.24 [2.35-4.45]
Hyperactivity	11/31 (0.35)	1.46 × 10−4	4.23 [2.01;8.90]	2.550	3.24 [2.745-6.66]
Sleep disorders	11/42 (0.26)	2.99 × 10−3	2.84 [1.43;5.67]	1.030	3.31 [2.01-6.16]
Viral infection	11/41 (0.27)	1.46 × 10−3	3.08 [1.54;6.17]	0.840	3.315 [1.1975-7.8975]
Tonic seizures	11/30 (0.37)	8.72 × 10−5	4.45 [2.11;9.37]	0.600	3.56 [1.845-8.075]
Impaired balance	7/16 (0.44)	2.27 × 10−4	6.42 [2.39;17.27]	1.040	3.88 [2.28-4.56]
Aggressiveness	7/19 (0.37)	4.93 × 10−3	3.86 [1.51;9.92]	-	3.94 [2.515-4.26]
Distractibility	5/9 (0.56)	8.18 × 10−4	9.50 [2.54;35.51]	1.600	3.96 [2.79-5.78]
Irritability	11/39 (0.28)	8.80 × 10−4	3.28 [1.63;6.60]	1.070	4.02 [1.52-6.21]
Drug-resistant epilepsy	6/18 (0.33)	7.18 × 10−3	3.85 [1.44;10.29]	7.340	4.055 [2.2775-5.345]
Behavior disorder	11/51 (0.22)	2.90 × 10−2	2.11 [1.08;4.11]	1.575	4.1 [2.5475-6.28]
Ketosis	6/12 (0.50)	4.05 × 10−4	7.77 [2.49;24.20]	1.380	4.3 [2.6525-5.4375]
Seizure cluster	14/27 (0.52)	2.05 × 10−8	8.86 [4.13;18.99]	1.150	4.375 [2.3325-7.605]
Focal seizures	17/55 (0.31)	3.11 × 10−6	3.91 [2.20;6.93]	0.940	4.38 [1.475-7.265]
Inappetence	7/17 (0.41)	7.29 × 10−4	5.39 [2.03;14.32]	0.350	4.45 [1.98-8.25]
Gait disorder	7/12 (0.58)	1.93 × 10−5	12.22 [3.88;38.53]	3.070	4.46 [2.49-5.06]
Social interaction	15/24 (0.62)	8.65 × 10−10	13.49 [5.87;30.97]	0.490	4.59 [3.92-5.885]
Autism	8/12 (0.67)	4.09 × 10−6	16.87 [5.07;56.14]	0.440	4.73 [2.41-7.7]
Myoclonic Epilepsy	10/15 (0.67)	9.04 × 10−7	15.11 [5.11;44.66]	0.940	5.61 [2.9475-6.76]
Nephrolithiase	7/10 (0.70)	1.58 × 10−4	14.04 [3.57;55.26]	1.250	5.82 [4.56-6.56]
Flat feet	7/22 (0.32)	4.25 × 10−2	2.57 [1.03;6.42]	1.100	5.95 [4.38-9.015]
Autism spectrum disorder	5/10 (0.50)	7.13 × 10−3	5.69 [1.60;20.20]	2.220	6.955 [4.5425-9.8025]

Discussion

This study shows that concepts extrapolated through NLP from narrative clinical reports of children of this DS cohort and their appearance during the clinical course, follows the same outline as what is known in DS.

In particular, as confirmed in our results, the first phase (1-2 years) of the DS is characterized by the appearance of convulsive/hemiconvulsive prolonged seizures in healthy children, often triggered by fever and infections, sometimes needing for hospitalization in ICUs to stop status epilepticus. [10][11][12][13][22][23][24] Photosensitivity can be early reported in DS and its recognition can add to the criteria of a positive diagnosis. [22][25] Then (after 2-3 years), coherently with our findings, new seizure types appear and epilepsy become drug resistant. [10][12][24][26] Concomitantly, typical neurological signs emerge, mainly involving gait and cognitive impairment,[10][24][26][27] and subsequently (by 3-4 years), a series of comorbidities are outlined, mainly involving behavior and social domains. [28][29][30]

Interestingly, also some “minor” features emerged in the analysis and completed the phenotypic spectrum. Disorders of appetite and eating difficulties are reported among all ages in patients with DS;[31] in our cohort several concepts related to issues of alimentation emerged by the second year of life (“Eating disorder”, “Inappetence”, “Anorexia”). Sleep disorders have also been reported in DS, appearing early in childhood;[32] in our population, several terms related to sleep issues emerged by the third year of life (“Sleep Disorder”, “Insomnia”).

Many rare epilepsies are reported increasingly with the advances of medical genetics, particularly with the advent of Next Generation Sequencing (NGS). [33]

The growing diffusion of these diagnosis techniques is allowing the rapid identification of new genetic conditions for which long-term course is mostly unknown. However, as the description of these entities is recent, natural history and outcome are lacking in most of them. In addition, if data acquisition is limited to structured and prospective natural history registry data, this might take many years from the time of report of the new epilepsy or disease. In these cases, unstructured clinical reports through data mining can be a plausible alternative as we showed in this study for DS.

Free text report express nuanced, unexpected, and unexplained signs or symptoms regarding the patient case, [5] and therefore are particularly useful in describing rare disease phenotypes. Interestingly, these reports were done in some instances by non-specialist in epileptology but by emergency care or intensive care physicians, emphasizing the uniformity of expression of the disease and patients’ description.

The main limitation of this study is the short follow-up. Life-long follow-up was not possible as our center, acting as a national center for rare epilepsies, often help patients to get back to proximity care when the diagnosis is confirmed, and the care program is established. Finally, we restricted our study to a small number of patients as we kept on patients who were followed since the onset of their disease of the first year of onset in order to avoid dating a clinical feature erroneously by the time of the clinic although it appeared before. Further studies involving adolescents and adults longitudinally followed from the onset are needed to confirm the validity of this methodology on the long-term course of the disease.

Other limitations are due to describing a phenotype by using NLP to free text reports; for instances, clinical notes often contain terms or phrases that have more than one meaning, and some concepts can be used referring to a condition that is not necessary present, but may be only previewed, suspected, or mentioned for other reasons.

Conclusion

We showed that the extraction of clinical terms by narrative reports of children with DS, having regard to the known phenotypic variability, allows to outline the natural history of this rare disease in early childhood. Therefore, we think that this model of "longitudinal phenotyping" could be applied for other rare conditions where natural history is still poorly known.

References

[1] Olivera P, Danese S, Jay N, et al. Big data in IBD: a look into the future. Nat Rev Gastroenterol Hepatol. 2019; 16(5):312–21.

[2] Shilo S, Rossman H, Segal E. Axes of a revolution: challenges and promises of big data in healthcare. Nat Med. 2020; 26(1):29–38.

[3] Bates DW, Saria S, Ohno-Machado L, et al. Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. Health Aff. 2014; 33(7):1123–31.

[4] Ahmed Z. Precision medicine with multi-omics strategies, deep phenotyping, and predictive analysis. 1st ed. Precision Medicine. Elsevier Inc.; 2022. 1–25 p.

[5] Garcelon N, Neuraz A, Salomon R, et al. Next generation phenotyping using narrative reports in a rare disease clinical data warehouse. Orphanet J Rare Dis. 2018; 13(1):85.

[6] Hully M, Lo Barco T, Kaminska A, et al. Deep phenotyping unstructured data mining in an extensive pediatric database to unravel a common KCNA2 variant in neurodevelopmental syndromes. Genet Med. 2021; .

[7] Hallowell N, Parker M, Nellåker C. Big data phenotyping in rare diseases: some ethical issues. Genet Med. 2019; 21(2):272–4.

[8] Brasil S, Pascoal C, Francisco R, et al. Artificial Intelligence (AI) in Rare Diseases: Is the Future Brighter? Genes (Basel). 2019; 10(12).

[9] Lo Barco T, Kuchenbuch M, Garcelon N, et al. Improving early diagnosis of rare diseases using Natural Language Processing in unstructured medical records: an illustration from Dravet syndrome. Orphanet J Rare Dis. 2021; 16(1).

[10] Dravet C. The core Dravet syndrome phenotype. Epilepsia. 2011; 52(SUPPL. 2):3–9.

[11] Dravet C, Oguni H. Dravet syndrome (severe myoclonic epilepsy in infancy). In: Handbook of Clinical Neurology. Elsevier B.V.; 2013. p. 627–33.

[12] Cross JH, Caraballo RH, Nabbout R, et al. Dravet syndrome: Treatment options and management of prolonged seizures. Epilepsia. 2019; 60(S3):S39–48.

[13] Kim Y, Bravo E, Thirnbeck CK, et al. Severe peri-ictal respiratory dysfunction is common in Dravet syndrome. J Clin Invest. 2018; 128(3):1141–53.
[14] Darra F, Battaglia D, Dravet C, et al. Dravet syndrome: Early electroclinical findings and long-term outcome in adolescents and adults. Epilepsia. 2019; 60(S3):S49–58.
[15] Garcelon N, Neuraz A, Salomon R, et al. A clinician friendly data warehouse oriented toward narrative reports: Dr. Warehouse. J Biomed Inform. 2018; 80:52–63.
[16] Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res. 2004; 32:D267-70.
[17] Hughey JJ, Rhoades SD, Fu DY, et al. Cox regression increases power to detect genotype-phenotype associations in genomic studies using the electronic health record. BMC Genomics. 2019; 20(1).
[18] Neuraz A, Chouchana L, Malamut G, et al. Phenome-wide association studies on a quantitative trait: application to TPMT enzyme activity and thiopurine therapy in pharmacogenomics. PLoS Comput Biol. 2013; 9(12).
[19] Denny JC, Ritchie MD, Basford MA, et al. PheWAS: Demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics. 2010; 26(9):1205–10.
[20] Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B. 1995; 57(1):289–300.
[21] Verheyen K. Motor Development in children with Dravet syndrome. Dev Med Child Neurol. 2019; 61:950–6.
[22] Ragona F, Brazzo D, De Giorgi I, et al. Dravet syndrome: Early clinical manifestations and cognitive outcome in 37 Italian patients. Brain Dev. 2010; 32:71–7.
[23] Ohki T, Watanabe K, Negoro T, et al. Severe myoclonic epilepsy in infancy: evolution of seizures. Seizure. 1997; 6(3):219–24.
[24] Gataullina S, Dulac O. From genotype to phenotype in Dravet disease. Seizure. 2017; 44:58–64.
[25] Dalla Bernardina B, Capovilla G, Gattoni MB, et al. Epilepsie myoclonique grave de la premiere annee. Rev Electroencephalogr Neurophysiol Clin. 1982; 12(1):21–5.
[26] Nabbout R, Chemaly N, Chipaux M, et al. Encephalopathy in children with Dravet syndrome is not a pure consequence of epilepsy. Orphanet J Rare Dis. 2013; 8(1):1–8.
[27] Chieffo D, Battaglia D, Lucibello S, et al. Disorders of early language development in Dravet syndrome. Epilepsy Behav. 2016; 54:30–3.
[28] Lagae L, Brambilla I, Mingorance A, et al. Quality of life and comorbidities associated with Dravet syndrome severity: a multinational cohort survey. Dev Med Child Neurol. 2018; 60(1):63–72.

[29] Knupp KG, Scarbro S, Wilkening G, et al. Parental Perception of Comorbidities in Children With Dravet Syndrome. Pediatr Neurol [Internet]. 2017; 76:60–5. Available from: http://dx.doi.org/10.1016/j.pediatrneurol.2017.06.008

[30] Li BM, Liu XR, Yi YH, et al. Autism in Dravet syndrome: Prevalence, features, and relationship to the clinical characteristics of epilepsy and mental retardation. Epilepsy Behav. 2011; 21(3):291–5.

[31] Knupp KG, Scarbro S, Wilkening G, et al. Parental Perception of Comorbidities in Children With Dravet Syndrome. Pediatr Neurol. 2017; 76:60–5.

[32] Schoonjans AS, De Keersmaecker S, Van Bouwel M, et al. More daytime sleepiness and worse quality of sleep in patients with Dravet Syndrome compared to other epilepsy patients. Eur J Paediatr Neurol. 2019; 23(1):61–9.

[33] Symonds JD, McTague A. Epilepsy and developmental disorders: Next generation sequencing in the clinic. Vol. 24, European Journal of Paediatric Neurology. W.B. Saunders Ltd; 2020. p. 15–23.

Chapter 4: "A European pilot study in Dravet Syndrome to delineate what really matters for the patients and families"[3]

Abstract

We aimed to identify caregivers' opinions on the outcome measures that matter in clinical trials in individuals with Dravet syndrome (DS). We conducted a prospective European multicentre study based on a 11-closed questions' survey developed by the French reference centre for rare epilepsies and DS patients' advocacy groups. Items included questions on seizures and daily life outcomes that a clinical trial on a therapy for individuals with DS should target. Statistical analyses were performed to evaluate the impact of the country of residence and of the patients' age.

The survey was answered by 153 caregivers (68%: France, 28%: Germany and 24%: Italy) for individuals with DS. Individuals with DS included 86 males (mean age of 11.4 [interquartile:7-20.4] years). Families ranked as important almost all the items proposed. However, items related to daily life had the highest rank in all 3 countries compared to items about seizures (p=0.02). Increase of individuals' age was associated with a higher age at diagnosis (ρ =0.26, p=0.02) and a less important impact of seizure duration (ρ =-0.25, p=0.005) and of the need of hospital referral (ρ =-0.26, p=0.005). These data can help tailor patient-centered outcome measures in future clinical and real-life trials for DS.

[3] Nicole Chemaly[a,b,c], Mathieu Kuchenbuch[a,b], Théo Teng[a], Elodie Marie[d], Gianluca D'Onofrio[a,e], Tommaso Lo Barco[a,f], Isabella Brambilla[g], Silke Flege[h], Anne-Sophie Hallet[d], Rima Nabbout[a,b,c]

[a] Department of Pediatric Neurology, Reference Centre for Rare Epilepsies, Hôpital Necker-Enfants Malades, APHP, Member of ERN EpiCARE, Paris, France

[b] Laboratory of Translational Research for Neurological Disorders, INSERM MR1163, Imagine Institute, Paris, France

[c] Université de Paris, Paris, France

[d] Alliance Syndrome Dravet, 45330 Malesherbes, France

[e] School of Pediatrics, Department of Women and Child Health, University of Padua

[f] Child Neuropsychiatry, Department of Surgical Sciences, Dentistry, Gynecology and Pediatrics, University of Verona, Verona, Italy

[g] Dravet Italia Onlus, 37010 Affi, Italy

[h] Dravet-Syndrom e.V., Turmblick 1. 04416 Markkleeberg, Germany

***Published on Epilepsia Open. 2021 Nov 7. doi: 10.1002/epi4.12557. Epub ahead of print. PMID: 34747137.**

Introduction

Determining what really matters, i.e., the meaningful outcomes for the health of individuals, is a key point in clinical trials, especially for rare diseases. Indeed, rare diseases generally involve several organs with high inter-individual variability leading to great disparity in terms of clinical presentations and consequences in the daily life.

The use of primary outcome measures (POMs) elaborated by practitioners might be simplistic and misrepresents the multiple facets of the impact of a disease.[1] In epilepsy, evaluation of clinical trials is mainly based on POMs targeting efficacy, like responder rate (defined as the number of affected individuals with at least 50% reduction in total seizure frequency) or the proportion of subjects who achieved seizure-free status,[2][3][4][5] and safety, using incidence of adverse events (AEs) and withdrawals rate due to AEs.[6][7] However, it is important to question the meaning of these endpoints, particularly in the context of developmental and epileptic encephalopathy (DEE) characterized by major drug resistance and associated comorbidities beyond seizures.

Dravet Syndrome (DS), one of the archetype of DEEs, is commonly related to pathogenic variant of SCN1A leading to a loss-of-function of voltage dependent sodium channel.[8][9][10][11][12] This DEE is associated with various degrees of intellectual disabilities, autism spectrum disorder in almost one third and behavioral disorders which incidence increases with age.[8][9][13] Several additional features are frequently reported: eating and sleep disorders, gait deterioration, dysautonomia, and a higher predisposition to infections.[14][15][16]

Research into the impact of DS beyond seizures increased in the last years and allowed to describe DS whole phenotype based on families and practitioners' reports.[14][15][17][18][19][20] In addition, the burden-of illness of individuals with Dravet syndrome also have a high impact on the caregivers. Compared to caregivers of individuals with difficult-to-treat epilepsy, caregivers of DS had higher depression scores and were more likely to change their employment status, including leaving their job.[21]

The aim of this study is to explore the domains that really matters for individuals with DS and their families emphasizing that a therapy targeting these domains would have a positive meaningful impact on their outcome.

Methods

Survey development

We developed, with three national patients' advocacy groups of DS in Europe ("Alliance Syndrome Dravet" in France, "Dravet-Syndrom e.V." in Germany and "Dravet Italia Onlus"), a 11-closed

questions survey based on our preliminary surveys[17][22] and other literature reports.[14][15][18] This survey explored different items related to current seizures (frequency, duration, seizures requiring rescue therapy, seizures needing of referral to emergency room (ER) or intensive care unit (ICU)) and to daily life aspects (sleep, eating disorders, language, motor skills, daily activity, behavior, communication, and interaction) using a Likert's scale from 1 to 5 (not important at all = 1, not important = 2, neutral = 3, important = 4 and highly important = 5).

Participants

This study was a prospective cohort study with convenience sampling. The survey was filled during annual associations meeting (France and Germany) or shared online with families for a period of 6 weeks. For every individual with DS, a unique caregiver completed the survey. Written informed consent to participate in this study was provided by the participants. This study was approved by the ethics committee of our institution (Necker Hospital, APHP).

Statistics

Results are expressed as the average ± standard deviation in case of Likert's scale data and as median [25th-75th percentile] otherwise. To study the impact of countries on the responses, we used one-way ANOVA in case of homogeneity of variance (Levene test). Otherwise, we used a more robust test called Brown Forsythe test with the same factors.[23] Bonferroni post-hoc tests after ANOVAs or Tamhane posthoc tests after Brown Forsythe tests were then applied in case of significance. For qualitative data, khi2 tests were used to study the presence or the absence of significant difference between the different countries. We correlated different quantitative answers to affected individuals' age using Spearman's rank correlation coefficient (Rho) coefficients.
To illustrate the possible correlation with age, we presented the data in relation to three age groups: <6 years, 6-12 years and >12 years.

Results

A total of 153 surveys were filled by parents with 96.5% of response rate (missing data: 81/2295). Table 1 summarized the characteristic of the population. There was no significant difference between the three countries regarding demographic characteristics.

Table 1. Demographic characteristics

	Total	France	Italy	Germany	*p*
N (%)	153	73 (47.7)	37 (24.2)	43 (28.1)	*ns*

Sex (m/f)	86/67	41/32	19/18	26/17	*ns*
Current age (y)	11.4 [7-20]	11 [5-20.5]	10.8 [6.6-17-7]	12.8 [9-23.8]	*ns*
Age at seizure onset (m)	5 [3.5-6.5]	5 [3.5-7]	5 [3.5-6]	5 [3-6]	*ns*
Age at diagnosis (y)	18 [12-33.6]	13.2 [9.6-27.6]	18 [12-72]	21.6 [13.2-33.6]	*ns*

Concerning the age at diagnosis, a significant correlation with patients' age was identified (Table 2, ρ =0.26, p=0.002). A lower age of diagnosis was associated with the younger age group.

Families ranked as important almost all the items proposed (Table 2). They rated similarly communication, sleep, behavior, daily activity, and motor skills (4.05+/-1.34) compared to lower scores for items related to seizures (3.96+/-1.22,. p=0.02) (Figure 1A).

For items regarding seizures, the highest score was achieved for seizure frequency (4.25+/-0.78) followed by seizure duration (4.15+/-1.11).

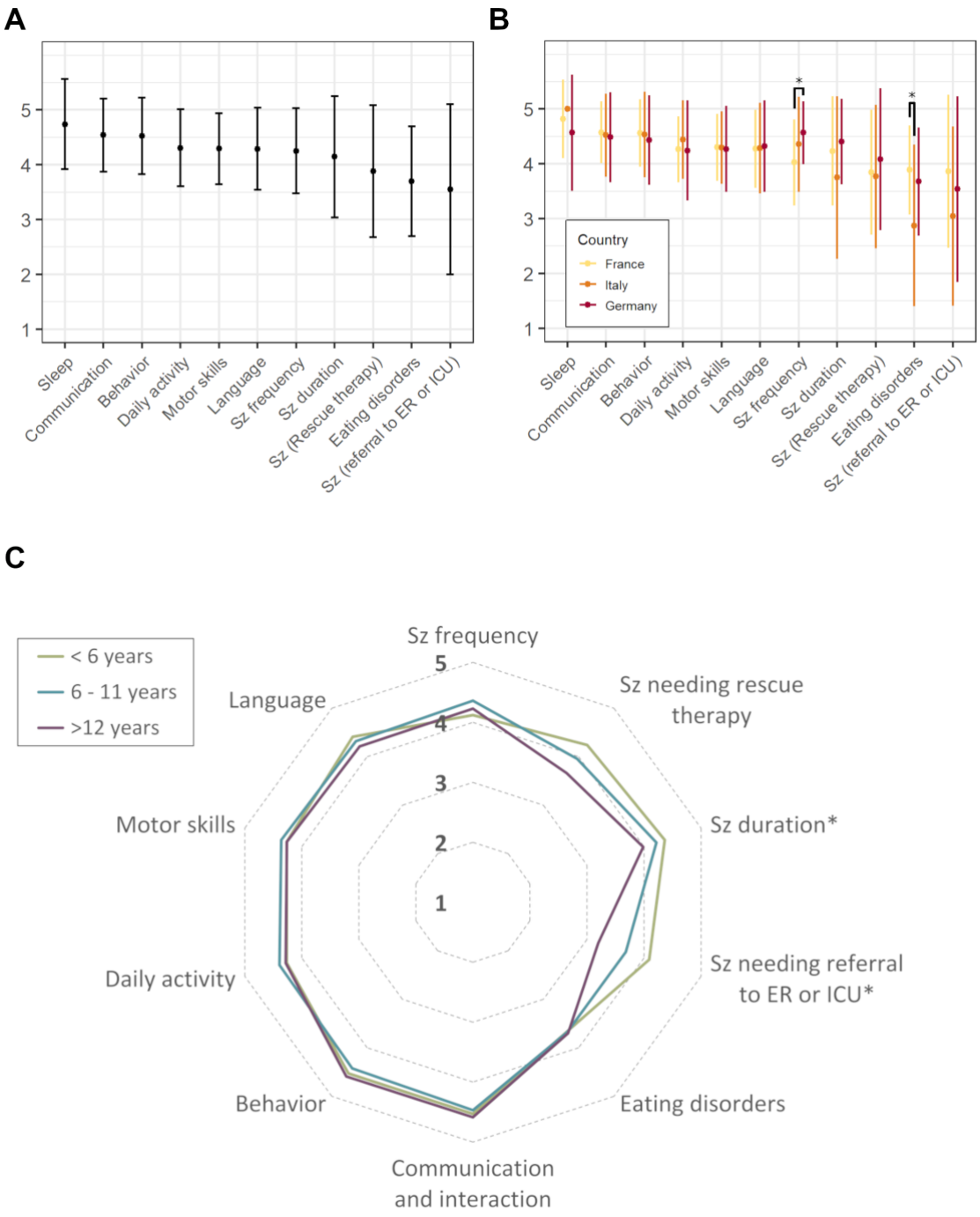
Caregivers in the three countries agreed on the importance with a descending order for sleep, communication, behavior, daily activities, motor skills, language, seizure duration, seizures requiring rescue therapy, and seizures necessitating referral to ER or ICU. In relation to age, only seizure duration and the need of referral to ER or ICU were negatively correlated with individuals' age, i.e., the highest scores were reported in the youngest individuals (ρ =-0.25, p=0.005 for seizure duration and ρ =-0.26, p=0.005 for seizure with referral to ER or ICU) (Figure 1C). There were few significant differences in the evaluation of the different items according to countries (Figure 1B). These differences were mainly about seizure frequency, which had higher score in Italy compared to France (4.57+/-0.55 in Italy and 4.03+/-0.77 in France, p= 0.0005, Germany: 4.36+/-0.72, p=ns) and those of eating disorders, which was higher in France compared to Germany (3.89+/- 0.8 in France and 2.88+/-1.45 in Germany, p=0.004, Italy: 3.68+/-0.95, p=ns).

Table 2: Impact of age on the different domains. The statistical impact of age was identified using Spearman's rank test and illustrated using three groups of affected individuals, namely <6 years, between 6 and 12 years and > 12 years.

	Total	< 6 years	6-12 years	>12 years	Spearman's p	*p*
Countries (France/Italy/Germany)	73/37/43	22/3/9	19/14/14	32/20/20	-	-
Current age (years)	11.4 [7-20.4]	4 [3.3 - 4.5]	8.8 [8 - 10.7]	21 [16.3 - 26.5]	-	-
Age at seizure onset (months)	5 [3.5 - 6]	5.5 [4 - 6.8]	5 [3 - 6.7]	4 [3.5 - 6]	ns	ns

Age at diagnosis (months)	18 [12 - 33.6]	13.2 [9.6 - 21.6]	18 [12 - 27.6]	20.4 [12 – 60.5]	0.26	0.002
Behavior	4.52 +/- 0.7	4.55 +/- 0.64	4.43 +/- 0.85	4.58 +/- 0.56	ns	ns
Communication and interaction	4.54 +/- 0.67	4.55 +/- 0.71	4.47 +/- 0.72	4.59 +/- 0.51	ns	ns
Daily activity	4.31 +/- 0.7	4.24 +/- 0.75	4.39 +/- 0.61	4.28 +/- 0.71	ns	ns
Motor skills	4.29 +/- 0.65	4.24 +/- 0.7	4.36 +/- 0.64	4.26 +/- 0.57	ns	ns
Language	4.29 +/- 0.75	4.39 +/- 0.81	4.32 +/- 0.75	4.21 +/- 0.61	ns	ns
Sleep	4.27+/-0.83	4.52+/-0.85	4.2+/-0.96	4.21+/-0.51	ns	ns
Seizure frequency	4.25 +/- 0.78	4.12 +/- 0.87	4.36 +/- 0.74	4.24 +/- 0.6	ns	ns
Seizure duration	4.15 +/- 1.11	4.38 +/- 1.16	4.22 +/- 1.09	4 +/- 0.99	-0.25	0.005
Seizure (Rescue Therapy)	3.88 +/- 1.21	4.25 +/- 1.31	3.96 +/- 1.23	3.67 +/- 0.79	ns	
Eating disorders	3.7 +/- 1	3.65 +/- 1.17	3.68 +/- 0.88	3.72 +/- 0.73	ns	ns
Seizure (referral to ER or ICU)	3.55 +/- 1.55	4.09 +/- 1.65	3.68 +/- 1.49	3.22 +/- 1.26	-0.26	0.005

Figure 1. Caregivers' opinions (Likert's scales from 1: not at all to 5: very important) about the domains that a therapy should improve for their children with Dravet syndrome (A), same results according to the 3 countries (B) and to the different age groups (C).



Discussion

Caregivers of individuals with DS across 3 EU countries expressed their needs for therapies that improve behavior, communication, sleep, daily activities, motor skills and language beyond their efficacy on seizures. These are the first direct results from families across 3 countries in Europe supporting smaller studies hypothesis.[17][22]

In order to improve the evidence of efficacy in clinical trials, Food and Drug Agency in 2009 and the European Medicines Agency in 2010.[6][24] have encouraged the Patient Reported Outcomes (PROs) as self-assessment of affected individuals' health status and validated it as a possible

secondary endpoint to complement the evaluation of clinical trials. The use of PROs in clinical trials, defined as “any report coming directly from patients, without interpretation by physicians or others, about how they function or feel in relation to a health condition and its therapy”,[25] has increased significantly since 2005.[26][27][28]

The development of specific PROs for DS may seem anecdotic because there are generic PROs, such as the health-related quality of life. However, generic PROs are not accurate enough to assess the quality of life of individuals with rare diseases and intellectual disability.[29] This is why, given the lack of standardized PROs dedicated to individuals with rare diseases, the International Rare Diseases Consortium has decided to set up the Patient-centered outcome measures (PCOMs) initiatives.[30] Determining the domains that are important to individuals with DS and their families is the first step of PRO development.[31][32]

Our study showed that different needs can emerge in individuals with DS. In addition, the major needs can vary with age.[12] These results are correlated to the 3-phases of natural history of individuals with DS.[9] In the first two phases, seizures are at the forefront. During the first 15 to 18 months, affected individuals present seizures triggered by fever often prolonged evolving to status epilepticus. Till around 5-6 years, individuals show different types of seizures as atypical absences, focal and tonic seizures with frequency drug resistant epilepsy in addition to the emergence of developmental slowing and behavioral disorders. Finally, in the third phase, seizures often decrease in term of frequency [9][18] and might become nocturnal and brief.[33] Intellectual disability and behavior problems move to the front scene with the families struggling for the education and rehabilitation special needs19. In this survey, families rated a decreasing need with age for a therapy targeting seizures’ reduction and referral to ER or ICU.

This data can be interesting in designing age related outcomes as ER and ICU needs are significantly more frequent in infants and pre-school children with DS compared to adolescents and adults.[18] However, the need of treatment to reduce seizure frequency and of the need to rescue treatment remains stable with age highlighting the persistence of high drug resistance throughout life.[34] Refining the age-related outcomes in DEEs might be the first step toward a precision design of CTs in such rare diseases.

Another key finding of this study is the age at diagnosis of affected individuals (18 [12- 33.6] months), showing a significant decrease in the age of diagnosis in the youngest individuals. These data confirm the improvement in the early diagnosis of DS over the last years18,35,36. Importantly, this earlier diagnosis age might question a younger age of inclusion in RCTs where the median age of inclusion in recent trials was between 7.6 and 9.3 years.[2][4][5][37][38] An earlier therapy can be a clue for a better neurodevelopmental outcome.[39]

Some limitations must be highlighted. This is a cross-sectional study to assess the impact of age on caregivers' expectations regarding what should treatment target. A longitudinal study will be probably more efficient to identify the evolution of caregivers' perspectives. However, to date, there is no study with this design probably due to its complexity and the rarity of this pathology. The convenient sample of this study might have led to a selection bias. Indeed, the identification of affected individuals through national families' associations might encourage the recruitment of families with specific profile and individuals with possibly more severe phenotypes. This survey is not accompanied by a qualitative study of the patients' opinions using for example Delphi methodology,[40] as we previously reported in DS.[17][22]

However, the design of this study is complex, time consuming, requires the definition of experts and does not allow us to have as large a population as in this study.[41]

In conclusion, this study highlights the domains that a therapy in development for DS should target in addition to seizures. target non-epileptic features. The next step would be to develop measurable and reproducible scales adding these items to seizures frequency as outcome measures for coming trials. For more accuracy and precision, an age related approach might refine these measures. This shift in our thinking in developing outcomes measures with more participatory approaches is urgent to establish in the era of gene therapy. Indeed, these therapies based on correction of the underlying genetic defect aim to rescue the genetic defect and to change the present path of affected individuals achieving disease modifying therapies beyond seizures decrease.[39]

References

[1] Morel T, Cano SJ. Measuring what matters to rare disease patients - Reflections on the work by the IRDiRC taskforce on patient-centered outcome measures. Orphanet J Rare Dis. 2017; 12:1–13.

[2] Devinsky O, Cross JH, Laux L, Marsh E, Miller I, Nabbout R, et al. Trial of cannabidiol for drug-resistant seizures in the dravet syndrome. N Engl J Med. 2017; 376:2011–20.

[3] Devinsky O, Patel AD, Cross JH, Villanueva V, Wirrell EC, Privitera M, et al. Effect of cannabidiol on drop seizures in the lennox–gastaut syndrome. N Engl J Med. 2018; 378:1888–97.

[4] Miller I, Scheffer IE, Gunning B, Sanchez-Carpintero R, Gil-Nagel A, Perry MS, et al. Dose-Ranging Effect of Adjunctive Oral Cannabidiol vs Placebo on Convulsive Seizure Frequency in Dravet Syndrome: A Randomized Clinical Trial. JAMA Neurol. 2020; 77:613–21.

[5] Nabbout R, Mistry A, Zuberi S, Villeneuve N, Gil-Nagel A, Sanchez-Carpintero R, et al. Fenfluramine for Treatment-Resistant Seizures in Patients with Dravet Syndrome Receiving Stiripentol-Inclusive Regimens: A Randomized Clinical Trial. JAMA Neurol. 2020; 77:300–8.

[6] European Medicine Agency-Committee for medicinal products for human use. Guideline on clinical investigation of medicinal products in the treatment of epileptic disorders. [Internet]. 2010 [cited 2020]. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinicalinvestigation- medicinal-products-treatment-epileptic-disorders-revision-2_en.pdf

[7] Ben-Menachem E, Sander JW, Privitera M, Gilliam F. Measuring outcomes of treatment with antiepileptic drugs in clinical trials. Vol. 18, Epilepsy and Behavior. 2010. p. 24–30.

[8] Wolff M, Cassé-Perrot C, Dravet C. Severe myoclonic epilepsy of infants (Dravet syndrome): Natural history and neuropsychological findings. Epilepsia. 2006; 47:45–8.

[9] Dravet C. The core Dravet syndrome phenotype. Epilepsia. 2011; 52:3–9.

[10] Claes L, Del-Favero J, Ceulemans B, Lagae L, Van Broeckhoven C, De Jonghe P. De novo mutations in the sodium-channel gene SCN1A cause severe myoclonic epilepsy of infancy. Am J Hum Genet. 2001; 68:1327–32.

[11] Wu YW, Sullivan J, McDaniel SS, Meisler MH, Walsh EM, Li SX, et al. Incidence of dravet syndrome in a US population. Pediatrics. 2015; 136:e1310–5.

[12] Scheffer IE, Nabbout R. SCN1A-related phenotypes: Epilepsy and beyond. Epilepsia. 2019; 60:S17–24.

[13] Ragona F, Granata T, Bernardina BD, Offredi F, Darra F, Battaglia D, et al. Cognitive development in Dravet syndrome: A retrospective, multicenter study of 26 patients. Epilepsia. 2011; 52:386–92.

[14] Knupp KG, Scarbro S, Wilkening G, Juarez-Colunga E, Kempe A, Dempsey A. Parental Perception of Comorbidities in Children With Dravet Syndrome. Pediatr Neurol. 2017; 76:60–5.

[15] Villas N, Meskis MA, Goodliffe S. Dravet syndrome: Characteristics, comorbidities, and caregiver concerns. Epilepsy Behav. 2017; 74:81–6.

[16] Rodda JM, Scheffer IE, McMahon JM, Berkovic SF, Graham HK. Progressive gait deterioration in adolescents with Dravet syndrome. Arch Neurol. 2012; 69:873–8.

[17] Nabbout R, Auvin S, Chiron C, Irwin J, Mistry A, Bonner N, et al. Development and content validation of a preliminary core set of patient- and caregiver-relevant outcomes for inclusion in a potential composite endpoint for Dravet Syndrome. Epilepsy Behav. 2018; 78:232–42.

[18] Lagae L, Brambilla I, Mingorance A, Gibson E, Battersby A. Quality of life and comorbidities associated with Dravet syndrome severity: a multinational cohort survey. Dev Med Child Neurol. 2018; 60:63–72.

[19] Sinoo C, de Lange IML, Westers P, Gunning WB, Jongmans MJ, Brilstra EH. Behavior problems and health-related quality of life in Dravet syndrome. Epilepsy Behav. 2019; 90:217–27.

[20] Skluzacek J V., Watts KP, Parsy O, Wical B, Camfield P. Dravet syndrome and parent associations: The IDEA League experience with comorbid conditions, mortality, management, adaptation, and grief. Epilepsia. 2011; 52:95–101.
[21] Strzelczyk A, Schubert-Bast S, Bast T, Bettendorf U, Fiedler B, Hamer HM, et al. A multicenter, matched case-control analysis comparing burden-of-illness in Dravet syndrome to refractory epilepsy and seizure remission in patients and caregivers in Germany. Epilepsia. 2019; 60:1697–710.
[22] Nabbout R, Auvin S, Chiron C, Thiele E, Cross H, Scheffer IE, et al. Perception of impact of Dravet syndrome on children and caregivers in multiple countries: looking beyond seizures. Dev Med Child Neurol. 2019; 61:1229–36.
[23] Brown MB, Forsythe AB. The Anova and Multiple Comparisons for Data with Heterogeneous Variances. Biometrics. 1974; 30:719.
[24] FDA, HHS. Guidance for Industry Use in Medical Product Development to Support Labeling Claims Guidance for Industry. Clin Fed Regist. 2009; :1–39.
[25] Patrick DL, Burke LB, Powers JH, Scott JA, Rock EP, Dawisha S, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. In: Value in Health.
[26] Vodicka E, Kim K, Devine EB, Gnanasakthy A, Scoggins JF, Patrick DL. Inclusion of patient-reported outcome measures in registered clinical trials: Evidence from ClinicalTrials.gov (2007-2013). Contemp Clin Trials. 2015; 43:1–9.
[27] Mercieca-Bebber R, Williams D, Tait MA, Roydhouse J, Busija L, Sundaram CS, et al. Trials with patient-reported outcomes registered on the Australian New Zealand Clinical Trials Registry (ANZCTR). Qual Life Res. 2018; 27:2581–91.
[28] Mercieca-Bebber R, King MT, Calvert MJ, Stockler MR, Friedlander M. The importance of patient-reported outcomes in clinical trials and strategies for future optimization. Patient Relat Outcome Meas. 2018; Volume 9:353–67.
[29] Arends M, Hollak CEM, Biegstraaten M. Quality of life in patients with Fabry disease: A systematic review of the literature. Vol. 10. Orphanet Journal of Rare Diseases. 2015.
[30] International rare diseases research Consortium. Patient-Centered Outcome Measures in the Field of Rare Diseases [Internet]. 2016. p. 31. Available from: https://irdirc.org/wpcontent/uploads/2017/12/PCOM_Post-Workshop_Report_Final.pdf
[31] Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD, Sloan JA. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? Value Heal. 2007; 10:S94–105.
[32] Snyder CF, Watson ME, Jackson JD, Cella D, Halyard MY, Sloan JA. Patient-reported outcome instrument selection: Designing a measurement strategy. Value Heal. 2007; 10:S76–85.

[33] Losito E, Kuchenbuch M, Chemaly N, Laschet J, Chiron C, Kaminska A, et al. Age-related "Sleep/nocturnal" tonic and tonic clonic seizure clusters are underdiagnosed in patients with Dravet Syndrome. Epilepsy Behav. 2017; 74:33–40.
[34] De Liso P, Chemaly N, Laschet J, Barnerias C, Hully M, Leunen D, et al. Patients with dravet syndrome in the era of stiripentol: A French cohort cross-sectional study. Epilepsy Res. 2016; 125:42–6.
[35] Strzelczyk A, Kalski M, Bast T, Wiemer-Kruel A, Bettendorf U, Kay L, et al. Burden-ofillness and cost-driving factors in Dravet syndrome patients and carers: A prospective, multicenter study from Germany. Eur J Paediatr Neurol. 2019; 23:392–403.
[36] de Lange IM, Gunning B, Sonsma ACM, van Gemert L, van Kempen M, Verbeek NE, et al. Influence of contraindicated medication use on cognitive outcome in Dravet syndrome and age at first afebrile seizure as a clinical predictor in SCN1A-related seizure phenotypes. Epilepsia. 2018; 59:1154–65.
[37] Devinsky O, Patel AD, Thiele EA, Wong MH, Appleton R, Harden CL, et al. Randomized, dose-ranging safety trial of cannabidiol in Dravet syndrome. Neurology. 2018; 90:e1204–11.
[38] Devinsky O, Nabbout R, Miller I, Laux L, Zolnowska M, Wright S, et al. Long-term cannabidiol treatment in patients with Dravet syndrome: An open-label extension trial. Epilepsia. 2019; 60:294 302.
[39] Nabbout R, Kuchenbuch M. Impact of predictive, preventive and precision medicine strategies in epilepsy. Nat Rev Neurol. 2020; 16:674–88.
[40] Jones J, Hunter D. Qualitative Research: Consensus methods for medical and health services research. Bmj. 1995; 311:376.
[41] Goodman CM. The Delphi technique: a critique. J Adv Nurs. 1987; 12:729–34.

Chapter 5 "Adaptive behaviour in adolescents and adults with Dravet Syndrome"[4]

Abstract

Aim

To explore the feasibility of using an adaptive behaviour profile (ABP) assessment generated from a well-known measure—the Vineland Adaptive Behaviour Scales, Second Edition (VABS-II)—as an instrument for outcome measures in adolescents and adults with Dravet Syndrome (DS).

Methods

We administered the VABS-II to 35 adolescents and adults with DS and collected epilepsy history and neurological features at the time of assessment. We conducted a cross-sectional analysis of VABS-II raw-scores and performed cluster analysis to identify different subgroups. We then explored possible relationships between clinical and epilepsy features, ABPs, and age.

Results

Most participants obtained the minimum standard scores in the various VABS-II subdomains, while the raw-score analysis outlined inter- and intra-individual differences among skills. We found two subpopulations: one with a "lower" ABP and one with a "higher" ABP, corresponding respectively to individuals in whom myoclonic seizures or generalized spike-and-wave activity are present ("complete phenotype") or absent ("incomplete phenotype").

[4] Lo Barco Tommaso[a,b,c], Offredi Francesca[a], Castino Eva[a], Proietti Jacopo[a,c], Cossu Alberto[a], Fiorini Elena[a], Fontana Elena[a], Cantalupo Gaetano[a,c], Dalla Bernardina Bernardo[a,c], Darra Francesca[a,c]

[a] Child Neuropsychiatry Unit, Department of Surgical Sciences, Dentistry, Gynecology and Pediatrics, University of Verona, Full Member of European Reference Network EpiCARE, Verona, Italy

[b] PhD Program in Clinical and Experimental Medicine, University of Modena and Reggio Emilia, Modena, Italy

[c] Research Centre for Pediatric Epilepsies Verona, IT, Verona, Italy

*** Published on Developmental Medicine and Child Neurology. 2022 Oct 31. doi: 10.1111/dmcn.15448. Epub ahead of print. PMID: 36316303.**

Interpretation

This study further delineates the natural history of DS. The assessment of an ABP through the VABS-II raw-score analysis provides a means by which to illustrate profiles of adaptive behaviour in adolescent and adults with DS, but shows limitations related to poor sensitivity in measuring fine clinical details. There is a need for new and more specific tools to monitor patients with developmental and epileptic encephalopathies.

Introduction

Developmental and epileptic encephalopathies (DEEs) are a group of diseases characterized by developmental impairment and phases of plateauing or regression induced by the epileptic activity that contribute to cognitive outcomes.[1] It is a complex and heterogeneous group, with wide variability of impaired functioning, primarily due to genetic alteration.[2]

Dravet Syndrome (DS) is a well-known type of DEE caused by congenital mutations in the *SCN1A* gene.[1][3] Symptom onset is in the first year of life with convulsive seizures and status epilepticus in otherwise healthy infants. During childhood, a drug-resistant epilepsy occurs together with developmental slowing, leading to cognitive impairment. Language disturbances, motor disorders, and social and behavioural issues complete the clinical picture.

In DS, and in DEEs in general, targeted treatment aims to not only reduce the seizure burden, but to also establish appropriate rehabilitation interventions aimed at preventing and minimizing comorbidities, which largely contribute to exacerbation of the clinical picture. Reduction of comorbidities is also a need frequently expressed by caregivers and families of affected individuals.[4][5]

Rehabilitation planning is challenging for subjects with DEEs. Interventions are tailored to the individual according to both personal abilities and vulnerabilities, which can be long-lasting; treatment goals, therefore, need to be modified according to the subject's age.

The assessment of cognitive ability alone does not provide enough useful information to highlight individual skills and competences, as well as plan and modify a treatment program over time. This is especially true for adults with DEEs, for whom IQ scores typically show a "floor effect." Conversely, the evaluation of daily-life abilities, communication, socialization, and motor skills is useful for identifying treatment targets and potentially monitoring response to an intervention.[6]

In this study, we sought to explore the utility of an adaptive behaviour profile (ABP) assessment for outcome measures in adolescents and adults with Dravet Syndrome (DS) through the use of the Vineland Adaptive Behaviour Scales, Second Edition (VABS-II).[7] While these features have been reported in the paediatric literature for DS, there is limited understanding of their presentation in adolescence and adulthood.[8][9] The current study may, therefore, contribute to the literature by further delineating the natural history of DS and provide suggestions regarding intervention targets and concrete goals through forward-looking rehabilitation programs.

Method

The current investigation is a monocentric study conducted at the Child Neuropsychiatry Unit of the University-Hospital of Verona, Italy.

The subject sample was comprised of 35 individuals with a clinical diagnosis of DS, longitudinally followed since 1978 and examined at our centre from 2017 to 2021.

Epilepsy history was collected for each subject, including age and presence of fever at first seizure, age at first afebrile seizure, recurrence of convulsive status epilepticus, seizure type according to 2017 ILAE classification,[10] and presence of reflex seizures.

The Vineland Adaptive Behaviour Scales, Second Edition (VABS-II)[7] was administered as an interview by a trained neuropsychologist to the caregivers of 10 adolescents (12-18 years) and 25 adults (>18 years) with DS between 2017 and 2021. The VABS-II is largely used in clinical practice with individuals with intellectual disability, including adults and subjects with DEEs.[8][11][12] This questionnaire explores the adaptive behaviour profile from birth to 90+ years of age through four domains: Communication (receptive, expressive, and written communication skills), Daily Living Skills (personal, domestic and community interaction skills), Socialization (interpersonal relationships, play and leisure time, and coping skills) and Motor Skills (gross and fine). A composite score is also provided, summarizing the individual's skills in all four domains. Higher scores suggest higher adaptive functioning, while lower scores suggest lower adaptive functioning.

Standard scores (SS, mean =100. standard deviation=15) were reported for all individuals. Of note, standard scores are not calculated for the Motor Skills domain on the VABS-II due to a ceiling effect over the age of 6 years. Adaptive levels were derived from standard scores, differentiating the following groups: low (SS=20-70), moderately low (SS=71-85); adequate (SS=86-114), moderately high (SS=115-129), and high (SS=130-160).

To explore the adaptive behaviour profile more deeply, we introduced an Adjusted Raw Score (ARS) obtained by dividing each participant's raw score by their age-expected raw score in all domains to obtain a percentage. We then conducted a cross-sectional analysis of VABS-II scores for the whole sample, considering both standard scores and adjusted raw scores.

Seizure frequency at the time of VABS-II administration was assessed from a seizure diary maintained by caregivers. Concomitantly, the presence/absence of ataxic gait, cortical myoclonus, pyramidal signs, and parkinsonism/bradykinesia was obtained through a complete neurological examination. Language ability was also tested during the medical visit and classified into four categories: "absent", "single words", "short sentences", "simple conversations/normal". The existence of autistic traits was assessed during the course of clinical observation and through specific questions to caregivers.

After generating a descriptive analysis of the entire cohort's VABS-II data and electroclinical variables, we used a data-driven approach to explore potential differences in the population: the two-step cluster analysis. The two-step cluster analysis is an exploratory tool implemented in IBM SPSS Statistics (version 25.0) designed to reveal natural groupings (or clusters) within a dataset.[13] It uses a distance measure to separate groups and then a probabilistic approach to choose the optimal subgroup model.[14] This analysis was performed separately using VABS-II adjusted raw scores and electroclinical variables to identify:

- subgroups with different ABPs (independently from electroclinical variables)
- subgroups with different epileptic phenotypes (independently from VABS-II data).

The number of clusters was determined automatically.

We then compared groups based on cluster results to explore possible relationships between epilepsy features and ABPs. We compared variables not used in the respective cluster analysis, including outcome variables, using cluster membership as a grouping variable, by means of Analysis of Variance (ANOVA), Mann-Whitney U-test, or Fisher exact test, respectively for continuous or categorical variables, as appropriate.

We further investigated the possible influence of age at test administration on VABS-II results by means of Spearman's correlation analysis.

Finally, a descriptive analysis of the ABPs of the adult population was performed, reporting the most common well-mastered skills as well as vulnerabilities.

Ethical approval was not required by the institution's research ethics committee.

Participants' caregivers have given informed consent for publication of the following results.

Results

Demographic and clinical features

The sample was composed of 35 individuals (15 males) born between 1972 and 2008. *SCN1A* genetic testing revealed pathogenic variants in 31 out of 32 tested subjects.

Median age at first seizure was 5 months (range 2-11 months). First seizure occurred during fever in 17 individuals, (49%, median: 6 months, range: 3-11 months) and without fever in 18 subjects (51%, median: 4 months, range: 2-9 months) (Figure 3). Prolonged convulsive seizures were experienced by a majority of subjects (83%). Common seizure types were focal onset non-motor seizures (88%) and hemiclonic seizures (80%). Twenty-four subjects had absence seizures (68%), and 19 subjects experienced absence status epilepticus (54%). Myoclonic seizures were exhibited by 22 subjects (63%). Reflex seizures occurred in 19 subjects (54%); 16 subjects (46%) experienced reflex seizures triggered by flashing lights and 12 subjects (34%) experienced self-triggered reflex seizures.

At adaptive behaviour profile assessment, the age of subjects ranged between 12 and 46 years (median: 20 years).
Seizures occurred daily/almost daily in 3 subjects (8.6%, median: 26 years, range: 15-34 years), weekly in 9 subjects (25.7%, median: 31 years, range: 18-38 years), monthly in 12 subjects (34.3%, median: 18 years, range: 12-46 years), were sporadic in 4 subjects (11.4%, median: 20 years, range: 16-27 years), and were absent in 7 subjects (20%, median: 20 years, range: 13-32 years).
Twentysix individuals (74%) showed tonic/tonic-clonic/tonic vibratory seizures; 4 had focal seizures, 1 (15 years) still showed myoclonic and absence seizures.
At neurological examination:
Ataxic gait was seen in 21 subjects (60%, median: 21 years, range: 13-46 years) and was absent in 14 subjects (40%, median: 20 years, range: 12-38 years).
Cortical myoclonus was found in 21 subjects (60%, median: 26 years, range: 14-46 years) and was not seen in 14 subjects (40%, median: 19 years, range: 12-38 years).
Parkinsonism/bradykinesia was observed in 8 subjects (22.9%, median: 19 years, range: 13-36 years) and was absent in 27 subjects (77.1%, median: 21 years, range: 12-46 years).
Pyramidal signs were seen in 7 subjects (20%, median: 34 years, range: 26-46 years) and were absent in 28 subjects (80%, median: 19 years, range: 12-37 years).
Language was absent in 5 individuals (14.3%, median: 26 years, range: 13-34). Of the remaining subjects, one individual produced isolated words (26 years), 16 subjects used only short sentences

(45.7%, median: 24 years, range: 12-46), and 13 subjects exhibited normal language or were able to sustain a simple conversation (37,1%, median: 19 years, range: 14-32).
Autistic features were encountered in 8 individuals (22.9%, median: 30 years, range: 13-46).
Complete data are shown in Table 1.

VABS-II Questionnaire Results

Standard Scores (SS)

The adaptive behaviour composite score was in the low range for 32 out of 35 individuals. In particular, 5 out of 10 adolescents and 21 out of 25 adults obtained the minimum composite score (SS=20. 26/35 subjects, 74% of the total population). No significant difference was found between the average composite scores of adolescents (SS=29.6) and adults (SS=27.9) (p>.05) (Figure 1a).
Communication scores were in the low range for 31 out of 35 participants (mean SS=31.8; min: 20. max:109) (Figure 1b). In the Daily-Living and Socialization domains, 32 out of 35 individuals rated in the low range (Daily-Living mean SS=34.4, min: 20. max:103; Socialization mean SS=30.2; min: 20. max:100)(Figure 1c-d).
No significant differences emerged between the subscale scores for adolescents and adults (p>.05).
Adjusted Raw Scores (ARS)
The highest ARS were obtained in the subdomains of Fine Motor ability (median: 74.6%, min: 23,6% max 104%), Gross Motor ability (median: 73,8%, min: 48.8% max 100%), Expressive Communication (median: 72.2%, min: 5.6% max 100%), and Receptive Communication (median: 67.5%, min: 15% max 100%).

The lowest ARS were achieved in Written Communication (median: 33,3%, min: 0% max 100%), Domestic daily-living skills (median: 33,3%, min: 0% max 100%), Community daily-living skills (median: 33%, min: 0% max 92%), and Coping skills (i.e., behavioural and emotional skills utilized across different social situations) (median: 33%, min: 0% max 92%).

Complete results are shown in Table 1.

ID	Age	EPILEPSY FEATURES											NEUROLOGICAL FEATURES AT VABS-II QUESTIONNAIRE							VABS-II SCORES															ADAPTIVE BEHAVIOUR PROFILE	EPILEPSY PHENOTYPE
																				STANDARD SCORES				ADJUSTED RAW SCORES (% of the age-expected)															
																									Communication			Daily Living			Socialization			Motricity				
		Age at epilepsy onset (months)	First seizure during fever	Age at first afebrile seizure (months)	Convulsive Status Epilepticus	Hemiclonic Seizures	Non-motor Focal Onset Seizures	Absence Seizures	Absence Status Epilepticus	Myoclonic Seizures	Photosensitivity	Self-induced seizures	Cortical Myoclonus	Ataxic Gait	Pyramidal Signs	Parkinsonism/ Bradikinesia	Autistic Traits	Seizure Type	Language	Seizure frequency	Composite	Communicaton	Day-Living	Socialization	Composite	Receptive	Expressive	Written	Person	Domestic	Community	Interpersonal relationships	Play and Leisure	Coping Skills	Gross motor skills	Fine motor skills		
1	12y 8m	3	Y	4	Y	Y	Y	Y	Y	Y	Y							F, TC	oo	●●	58	46	70	72	79	68	74	65	85	93	55	70	84	85	92	100	High	Complete
2	13	5	Y	5	Y					Y							Y	-	oo	-	32	45	44	42	76	78	84	67	79	50	63	64	74	65	90	99	High	Incomplete
3	13y 1m	2		2	Y	Y	Y	Y	Y	Y				Y		Y		F	oo	●●	20	25	27	21	58	75	60	33	61	60	34	51	76	32	76	75	Low	Complete
4	13y 2m	3		3	Y	Y											Y	TC	-	●●	20	20	36	32	52	80	51	9	55	100	19	67	60	33	60	57	Low	Incomplete
5	14y 7m	4	Y	7	Y	Y	Y	Y					Y					TC	ooo	●●	25	31	47	35	67	78	74	38	76	70	56	64	74	41	75	78	High	Incomplete
6	14y 9m	4		4		Y	Y	Y		Y	Y		Y	Y				TC	oo	●●	20	20	20	20	52	58	53	22	49	47	22	41	69	24	84	86	Low	Complete
7	15y 9m	3		3	Y		Y	Y	Y	Y	Y	Y	Y	Y			Y	A, M, TC	oo	●●●●	20	20	20	20	57	70	75	52	56	33	37	41	40	58	79	67	Low	Complete
8	16y	4	Y	10		Y	Y											TC	ooo	●●	31	43	52	36	76	88	82	70	76	69	65	62	76	47	86	102	High	Incomplete
9	17y 6m	4		4	Y	Y	Y						Y	Y				-	ooo	-	50	71	70	34	87	100	98	81	89	100	74	92	76	44	94	104	High	Incomplete
10	17y	10	Y	15	Y	Y	Y	Y	Y									F	ooo	●●	20	31	30	20	79	80	85	83	83	18	85	60	85	64	95	102	High	Incomplete
11	18y	6	Y	14	Y	Y	Y	Y	Y	Y	Y		Y	Y		Y		TC	ooo	●●●	20	20	20	20	54	55	74	12	60	27	30	58	58	9	80	82	Low	Complete
12	18y 6m	4		4	Y	Y	Y							Y		Y		TC	ooo	●●	79	88	61	100	85	100	98	88	93	44	56	93	92	34	76	100	High	Incomplete
13	19	4	Y	5	Y	Y	Y	Y	Y	Y	Y		Y	Y		Y		TC	oo	●●	20	20	20	20	51	80	58	26	44	6	16	49	73	22	73	75	Low	Complete
14	19y 6m	9		9		Y	Y						Y	Y				-	ooo	-	20	34	20	20	51	80	78	70	65	6	41	29	40	1	58	71	Low	Incomplete
15	20y	8	Y	10	Y		Y	Y			Y		Y	Y				F, TC	oo	●●●	20	23	29	20	66	93	71	52	73	48	33	59	71	23	83	89	Low	Incomplete
16	20y 1m	5		5	Y	Y	Y	Y	Y	Y			Y	Y		Y		TC	oo	●●	20	20	21	20	59	60	71	50	66	15	44	53	68	19	68	83	Low	Complete
17	20y 2m	6	Y	34		Y	Y											-	ooo	-	72	109	67	58	90	100	100	100	95	56	67	86	94	31	100	100	High	Incomplete
18	20y 3m	6		6	Y	Y	Y											TC	ooo	●	20	20	20	20	59	55	78	56	76	6	40	70	66	15	63	71	Low	Incomplete
19	21y	6	Y	10	Y	Y	Y	Y	Y	Y	Y	Y		Y				TC	-	●	20	20	20	20	30	60	19	12	35	21	15	37	10	7	55	49	Low	Complete
20	24y	3	Y	4	Y	Y	Y	Y	Y	Y	Y	Y		Y				-	ooo	-	20	20	50	20	73	88	78	60	85	58	58	61	58	21	100	90	High	Complete
21	26y	5		5	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		TC	o	●●●●	20	20	20	20	36	48	41	8	52	19	6	47	21	5	70	54	Low	Complete
22	26y 1m	5	Y	5	Y		Y	Y	Y	Y	Y		Y	Y			Y	TC	-	●●●	20	20	20	20	14	15	6	0	12	4	0	21	7	0	55	24	Low	Complete
23	27y	6	Y	14		Y	Y						Y					TC	ooo	●	20	20	34	20	68	95	82	52	76	42	64	57	37	29	74	81	Low	Incomplete
24	29y	7	Y	84														-	ooo	-	55	49	103	61	94	100	97	84	94	96	92	84	100	30	100	100	High	Incomplete
25	29y	4		4	Y	Y	Y	Y	Y	Y		Y	Y					TC	oo	●●●	20	20	20	20	48	55	75	4	62	10	8	68	29	12	83	64	Low	Complete
26	31y	4	Y	5	Y		Y	Y		Y			Y	Y	Y			TC	-	●●●	20	20	20	20	41	65	44	8	59	38	13	40	34	9	69	58	Low	Incomplete
27	32y	11	Y	60	Y		Y											-	ooo	-	72	77	82	87	93	100	98	90	90	69	91	86	100	36	99	100	High	Incomplete
28	32y 6m	9		9	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y				TC	oo	●●	20	20	20	20	44	65	58	0	45	29	15	46	52	17	65	49	Low	Complete
29	34y	3		3	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y	TC	oo	●●●	20	20	20	20	39	65	55	6	66	35	5	37	13	7	68	43	Low	Complete
30	34y 6m	3		3	Y	Y	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y	TC	-	●●●●	20	20	20	20	23	28	17	0	29	8	1	29	10	4	63	49	Low	Complete
31	36y 7m	6		6	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y	Y		TC	oo	●●●	20	20	20	20	34	63	49	6	42	0	0	32	39	4	66	49	Low	Complete
32	36y 7m	8	Y	15	Y	Y		Y	Y	Y	Y		Y	Y				TC	oo	●●●	20	20	20	20	52	68	71	20	72	29	25	53	19	21	64	71	Low	Complete
33	37y 6m	4		4	Y	Y	Y	Y		Y	Y	Y	Y	Y				TC	oo	●●	20	20	20	20	42	68	72	18	39	21	25	45	31	10	49	46	Low	Complete
34	38y	9		9	Y	Y	Y	Y	Y	Y		Y			Y		Y	TC	oo	●●●	20	20	20	20	44	55	59	12	63	4	26	53	19	15	55	63	Low	Complete
35	46y	5	Y	8	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y		Y	TC	oo	●●	20	20	20	20	40	60	63	16	49	8	16	46	32	16	54	38	Low	Complete
			51%		83%	80%	89%	69%	54%	63%	46%	34%	60%	60%	20%	23%	23%																					
Median	20y	5		5																	Median	20	20	20	20	53,9	68	72	33	65	33	33	53	58	21	74	75		
Q1	17y	4		4																	Q1	20	20	20	20	43	60	56	10	51	13	15	43	31	10	63	55		
Q3	31y	6		10																	Q3	23	33	46	33	74	84	80	66	78	57	57	66	75	34	85	94		

Table 1. Demographic features, epilepsy characteristics, neurological outcome, Vineland-II scores, and cluster memberships. ' -=Absent; ○ =Single Words; ○ ○ =Short Sentences; ○ ○ ○ =Simple Conversations/Normal; ● =Sporadic; ● ● =Monthly; ● ● ● =Weekly; ● ● ● ● =Daily/almost daily; A=Absences; F=Focal; M=Myoclonic; m=months; Q1=lower quartile; Q3=upper quartile; T=Tonic/Tonic-clonic/Tonic-vibratory; Y=yes; y=years

Figure 1. Standard scores obtained by each patient in the different domains. (a) Twenty-two individuals (63%) obtained the minimum score in the communication subscale. Nineteen (54%) and twenty-four (69%) participants displayed the minimum score the day-living (b) and socialization (c) subscale respectively. Twenty-six (74%) obtained the minimum composite score.



Adaptive Behaviour Profile (ABP) Cluster Analysis

The two-step cluster analysis based on the ARS identified two subgroups with different ABPs (cluster memberships are shown in Table 1). The first group (n=11), which we named "higher ABP", demonstrated a higher level of functioning and the mean composite adjusted raw score was 81.8% (min: 67.1% max: 93,6%). The second group (n=24), named "lower ABP", was composed of individuals with a lower level of functioning (mean composite adjusted raw score=46.3%; min:13,7% max: 67.9%).

Adjusted raw scores were significantly different across all subscales between these two groups (p<.001 in all subscales) (Figure 2). Main differences emerged in Communication-Written (mean 75% *versus* 22.7%), Daily Living-Community (69.2% *versus* 22.2%), Socialization-Play and Leisure (83% *versus* 40.5%), Daily Living-Domestic (65.7% *versus* 25.7%) and Fine Motor Skills (97.6% *versus* 62.1%).

No differences were found between the two groups with regard to gender (p=.43), while seizure frequency (p=.004) and age were higher in the lower ABP group (median 26 years versus 17.5 years, p=.036).

Epilepsy Phenotype Cluster Analysis

A data-driven two-step cluster analysis based on epilepsy features revealed two distinct subgroups. Cluster memberships are shown in Table 1.

- The first group included 20 subjects. All had myoclonic seizures (100%) and absence seizures (100%). Most had absence status epilepticus (90%), seizures triggered by flashing lights (75%), and self-induced seizures (60%). This group also exhibited a lower age at first seizure (mean 4.85 months) and at first afebrile seizure (mean 6.1 months), and a higher probability of first seizure without fever (60%).
- The second group included 15 subjects. A small percentage had myoclonic seizures (13,3%, p<.001), absence seizures (26.7%, p<.001), absence status epilepticus (6.7%, p<.001), seizures triggered by flashing lights (6.7%, p<.001) and self-induced seizures (0%, p=.002). When compared to the first group, this group had a slightly higher age at first seizure (mean 6.07 months, p=.117) and at first afebrile seizure (mean 18 months, p=.033), with a lower probability of first seizure without fever (34.3%, p=.191).

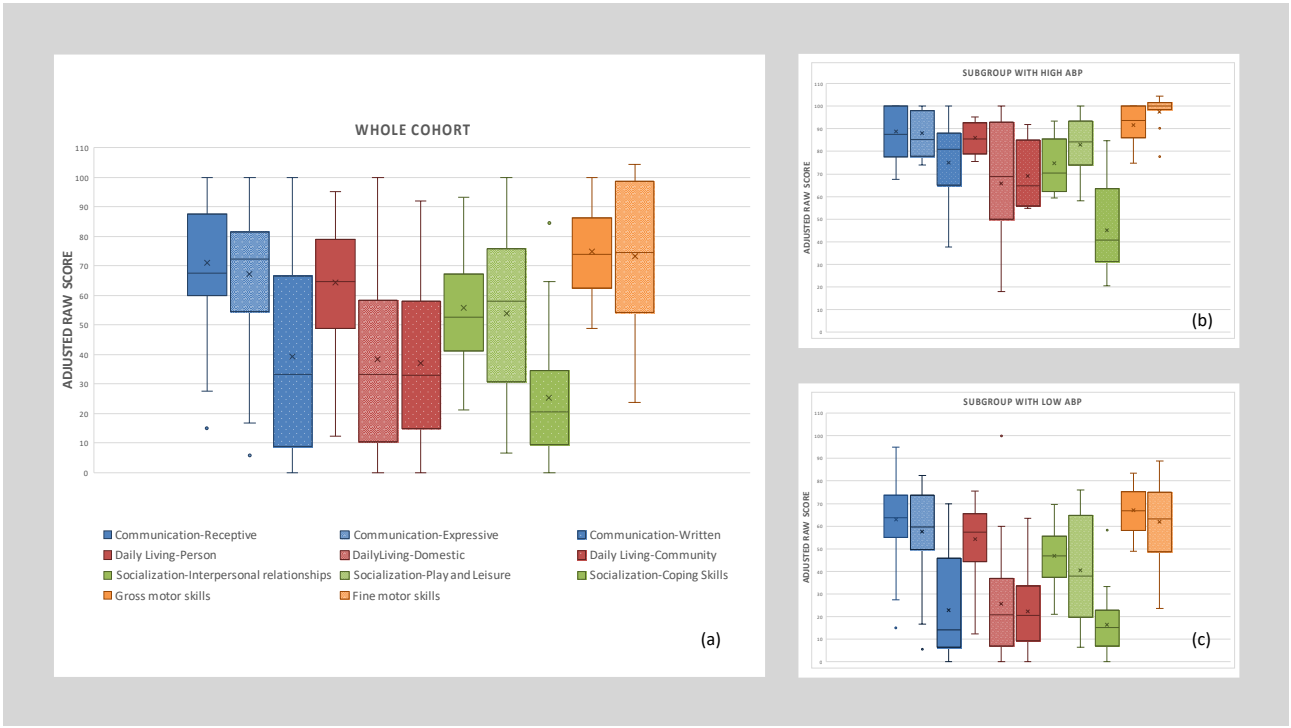
The differences that emerged between these two groups are consistent with the division of DS into two subtypes proposed by some authors: one subtype with a “complete phenotype”, which corresponds to our first subgroup, and another subtype in which key features, such as myoclonic seizures or generalized spike-and-wave activity, are missing (“incomplete phenotype”), corresponding to our second group.[15]

No significant differences were found between the two groups for focal seizures, convulsive status epilepticus, and hemiclonic seizures (p>.05).

The “complete phenotype” was related to worse outcomes at last assessment, including major language impairment (p=.004), higher seizure frequency (p=.002), and ataxic gait (p=.019).

Furthermore, individuals with the “complete phenotype” also demonstrated worse outcomes in terms of ABP: 18 out of 20 subjects belonged to the lower ABP group, while 11 out of 15 subjects with the “incomplete phenotype” fell within the higher ABP group (p=.003).

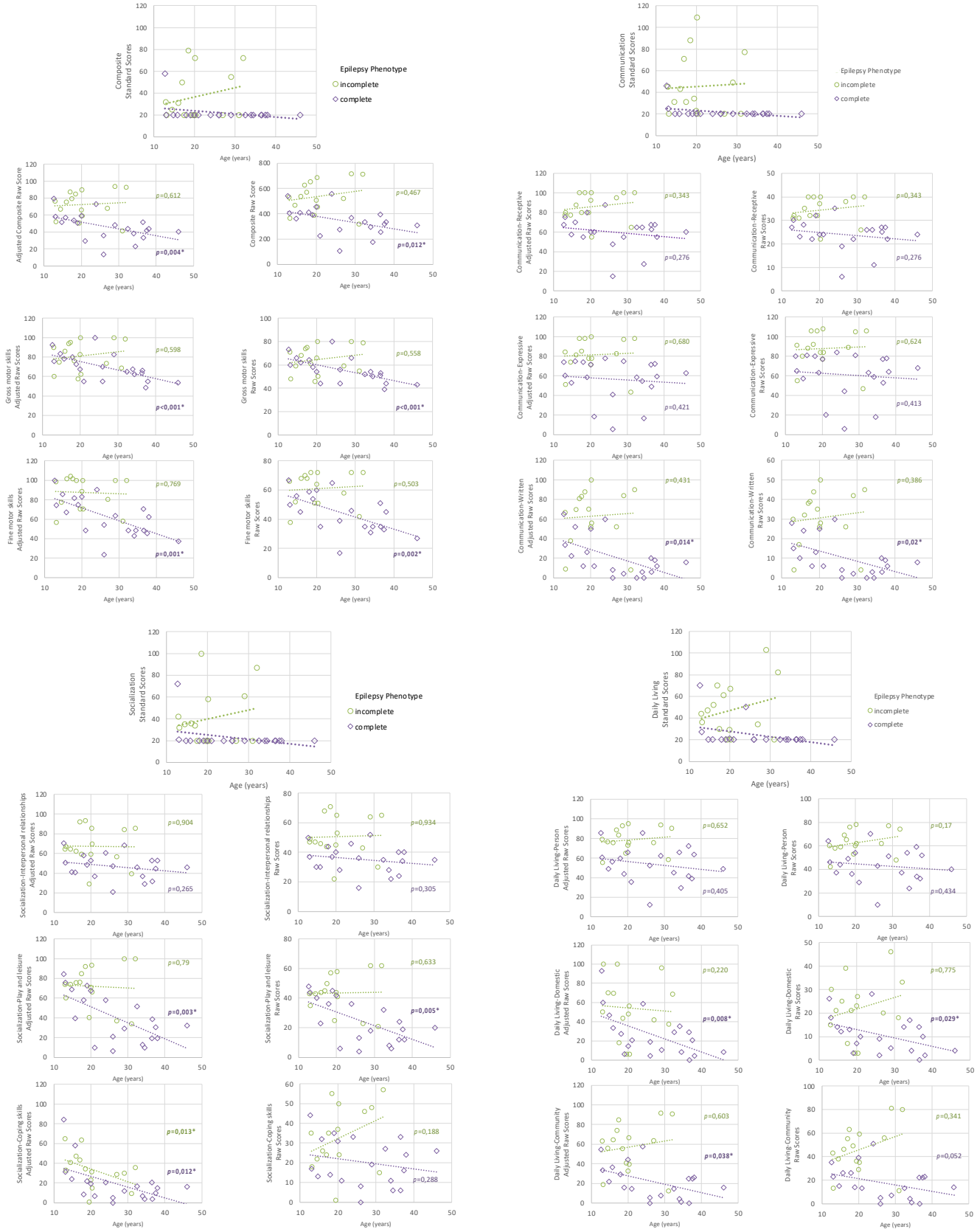
Figure 2. Adjusted raw score distribution for the whole cohort (a) and in subgroups with high (b) and low (c) adaptive behaviour profiles (ABPs).



Correlation with age

There was a significant correlation between age and the composite raw score (p=.006, Spearman's coefficient= -0.45), as well as between age and the adjusted composite raw score (p=.002, Spearman's coefficient= -0.51). In particular, this worsening trend of both composite raw scores (p=.012, Spearman's coefficient= -0.55) and adjusted composite raw scores (p=.004, Spearman's coefficient= -0.61) was present in individuals with the "complete phenotype". For this subsample, a significant inverse relationship between age and raw-scores emerged across several subscales: "Communication-Written", "Daily Living-Domestic", "Daily Living-Community", "Socialization-Play and Leisure", "Socialization-Coping Skills", "Gross Motor Skills", and "Fine Motor Skills" (p<.05). Conversely, no significant decreases were noted in the raw scores of individuals with the "incomplete phenotype". (Figure 3).

Figure 3. Distribution and trends by age of standard and adjusted raw scores in the different domains and subdomains. Patients with the “complete phenotype” are shown in violet and those with the “incomplete phenotype” are represented in green. A significant inverse relationship between age and raw-scores emerged across several subscales (p value in bold).



Most common vulnerabilities and well-mastered skills

A descriptive analysis of the ABPs of adults was performed, dividing individuals with “complete” and “incomplete phenotypes”. For this analysis, an individual item on the VABS-II was considered as “passed” if the participant received a score of 1 (“Behaviour is sometimes or partially performed”) or 2 (“Behaviour is usually or habitually performed”). “Vulnerabilities” were defined as abilities unconsolidated by 90% or more individuals, while “Well-mastered skills” were defined as abilities consolidated by 90% or more individuals. Results are shown in Appendix S1.

Discussion

With the current study’s goal of expanding knowledge about the natural history of DEEs, we investigated the ABPs of a sample of adolescents and adults with DS, the most well-known and studied DEE.

The analysis of adaptive behaviour through the VABS-II was useful to outline an operating profile and to identify treatment targets for rehabilitative interventions. However, as already suggested by Berg et al., this measure shows some limitations when applied to individuals with DEEs, mostly related to poor sensitivity in measuring fine clinical details and capturing small longitudinal changes these subjects can show in response to shifts in treatment or disease course.[11]

Most of the subjects, and in particular the adults, obtained the minimum score for composite standard scores and subscales (score: 20), producing the so-called "floor-effect." This finding confirms what has been previously reported in the literature regarding the poor long-term outcome of DS,[3][16][17][18][19][20][21] but provides little information regarding inter- and intra-individual variability and does not permit the delineation of a functional profile. Conversely, a raw score analysis outlines some differences among skills, with greater abilities in the areas of motricity and verbal communication, and major problems in the areas of daily-living skills, written communication, and behavioural and emotional control. This profile partially diverges from those previously reported for children with DS, in which communication capacity were found to be lower than socialization capacity.[8] Similar to our findings, daily-life abilities were reported to be considerably impaired in childhood.[8] Further studies should be conducted with adolescents and adults affected by other genetic DEEs to continue investigating the specificity of this profile.

It should be noted, however, that the profile in the current study is characterized by a degree of interindividual variability. That is, we found two subpopulations exhibiting gross differences in their scores: one with a lower level of functioning (“lower ABP”) and one with a higher level of functioning (“higher ABP”). These two groups correspond respectively, with few exceptions, to individuals

exhibiting epilepsy within the “complete phenotype” of DS (i.e., presence of myoclonic seizures or spike-and-wave activity) versus the “incomplete phenotype” of DS (i.e., absence of myoclonic seizures or spike-and-wave activity).[15]

This finding is consistent with what is already reported in the literature regarding worse outcomes for individuals with DS with the “complete phenotype”.[21][22] In fact, in the current cross-sectional study, there was no correlation between age at test administration and VABS-II raw scores for individuals with the "incomplete phenotype." This is comparable to other genetic DEEs, where a longitudinal decline of standardized scores is absent or due to a slower progression.[12][23] Conversely, in the subpopulation with the “complete phenotype”, an inverse relationship between age and both raw-and standard scores emerged. Even in the motricity domains, in which a ceiling effect is seen over 6 years in the general population, the raw scores of older individuals with the “complete phenotype” were lower than the raw scores of younger individuals. This data suggests a possible deterioration of fine-motor skills in this sub-cohort, similarly to what has been previously reported for gait in DS.[24]

The differences we found between the two phenotypes can help guide forward-looking rehabilitative planning by focusing on realistically attainable goals.
For example, for subjects with the “complete phenotype”, treatment could aim to expand lexiconic and phonological abilities; in subjects with the “incomplete phenotype”, who are usually able to produce complete sentences and tell stories, treatment could instead focus on understanding more complex instructions, reading and writing of simple passages, and practicing reading comprehension. As for personal autonomy skills, in subjects with the “complete phenotype”, rehabilitation could aim to improve personal hygiene skills, dressing/undressing, using simple kitchen utensils and appliances such as a microwave, and practicing using money for small purchases. In subjects with the “incomplete phenotype” the following interventions can be targeted: cleaning and tidying tasks, cooking simple foods on the stove, using money with the support of a calculator, and using public transportation for short trips.

While several studies report the potential efficacy of new anti-seizure treatments started in adulthood and their possible effects on daily life,[25][26] data regarding implications of rehabilitation interventions in adulthood are lacking. This is a crucial issue since, despite being typically diagnosed in childhood, late diagnosis of DS (and other genetic DEEs) has been increasing in recent years.[27] Similarly, the impact of rehabilitation withdrawal during adulthood is not well studied. Possibly some cognitive and adaptive skills may be lost over time due to several reasons: school attendance stoppage

with subsequent decrease in cognitive and operational stimuli, reduction in opportunities for socialization, interruption of sports and other recreational activities. Rehabilitation interventions are actually usually reduced during adulthood, and most adult patients usually live in long-term-care residential centres.[28][29][30]

Findings highlight the need for holistic care management for patients with DEEs. Quality of life was seen to be adversely influenced by seizures and multiple comorbidities and which differed in quality and severity, and in some cases were seen to be disease specific and may also vary considerably with age.[31] Therefore, in individuals with a DEE, the assessment of different functions should not be carried out in comparison to the general population. Rather, it is recommended that assessment should be specific for the underlying condition and tailored for the individual over time. This scenario implies the need for disease registries to define the natural history of each of these rare conditions and evaluate the response to pharmacological and/or rehabilitative therapies. These recommendations are also aligned with the requests of families and international agencies involved with DEEs.[4][32]

Study Limitations

The main limitation is the relatively small sample. Although the cohort for the current study has been followed longitudinally, only cross-sectional data for the ABP and neurological features have been presented in this paper.

Notably, the possible role of different pharmacological approaches that patients received over the course of their lives cannot be excluded in the finding of lower scores in older adults.

Furthermore, our data does not address other potential causes of impairment (e.g., crouched gait, slowed movements, perseveration, and psychiatric symptoms) which may highly impact everyday functioning and may be specific treatment areas. Thus, the possible impact of different environmental factors upon these subjects is difficult to characterize and has not been taken into account as it was beyond the scope of the current study.

Conclusion

With the rising emerging potential for precision medicine, acknowledgement of the natural history of rare diseases is crucial to identify targeted interventions and novel treatment options.[31][32][33]

This study shows that using VABS-II standard scores (that is, comparison with the general population) in the assessment of global functioning of individuals with DEEs is not the optimal approach to outline the characteristics of these subjects, neither in terms of interindividual differences

nor in the evaluation of changes over time. Conversely, the current study's deeper raw score analysis allowed for exploration of both aims. Overall, there is a need for disease-specific tools, including a multidimensional approach in which interviews are accompanied by clinical evaluations to provide sufficient granular detail.

The current study confirms that, in DS, adaptive behaviour outcomes are generally poor; it also underlines differences between the two clinical phenotypes. These findings can help guide forward-looking rehabilitation programs for individuals with DS by establishing realistic and reachable treatment goals.

The real impact of rehabilitation treatments in adults with DEEs is not currently known and should be the object of further studies. Furthermore, rehabilitation intervention measures and treatment targets should be included in transitional programs for young patients with DEEs to the adult health-care system.[33]

References

[1] Scheffer IE, Berkovic S, Capovilla G, et al. ILAE classification of the epilepsies: Position paper of the ILAE Commission for Classification and Terminology. Epilepsia 2017; 58(4):512-521.

[2] Ho NT, Kroner B, Grinspan Z, et al. Comorbidities of Rare Epilepsies: Results from the Rare Epilepsy Network. J Pediatr 2018; 203:249-258.

[3] Nabbout R, Chemaly N, Chipaux M, et al. Encephalopathy in children with Dravet syndrome is not a pure consequence of epilepsy. Orphanet J Rare Dis 2013; 8:176.

[4] Nabbout R, Auvin S, Chiron C, et al. Zogenix and Adelphi Values study group. Perception of impact of Dravet syndrome on children and caregivers in multiple countries: looking beyond seizures. Dev Med Child Neurol 2019; 61(10):1229-1236.

[5] Chemaly N, Kuchenbuch M, Teng T, et al. A European pilot study in Dravet Syndrome to delineate what really matters for the patients and families. Epilepsia Open 2021; 7. Epub ahead of print.

[6] Spiridigliozzi GA, Goeldner C, Edgin J, et al. Adaptive behavior in adolescents and adults with Down syndrome: Results from a 6-month longitudinal study. Am J Med Genet A 2019; 179(1):85-93.

[7] Sparrow SS, Cicchetti DV. The Vineland adaptive behavior scales. Allyn & Bacon; 1989.

[8] Villeneuve N, Laguitton V, Viellard M, et al. Cognitive and adaptive evaluation of 21 consecutive patients with Dravet syndrome. Epilepsy Behav 2014; 31:143-8.

[9] Ouss L, Leunen D, Laschet J, et al. Autism spectrum disorder and cognitive profile in children with Dravet syndrome: Delineation of a specific phenotype. Epilepsia Open 2018; 4(1):40-53.
[10] Fisher RS, Cross JH, French JA, et al. Operational classification of seizure types by the International League Against Epilepsy: Position Paper of the ILAE Commission for Classification and Terminology. Epilepsia. 2017;58(4):522-530.
[11] Berg AT, Palac H, Wilkening G, et al. SCN2A-Developmental and Epileptic Encephalopathies: Challenges to trial-readiness for non-seizure outcomes. Epilepsia 2021; 62(1):258-268.
[12] Bar C, Breuillard D, Kuchenbuch M, et al. Adaptive behavior and psychiatric comorbidities in KCNB1 encephalopathy. Epilepsy Behav 2022; 126:108471.
[13] IBM-SPSS Statistics, version 25.0 - TwoStep Cluster Analysis https://www.ibm.com/docs/en/spss-statistics/25.0.0?topic=features-twostep-cluster-analysis (Accessed on 1 October 2022)
[14] Kent P, Jensen RK, Kongsted A. A comparison of three clustering methods for finding subgroups in MRI, SMS or clinical data: SPSS TwoStep Cluster analysis, Latent Gold and SNOB. BMC Med Res Methodol. 2014 Oct 2;14:113.
[15] Guerrini R, Oguni H. Borderline Dravet syndrome: a useful diagnostic category? Epilepsia 2011; 52 Suppl 2:10-2.
[16] Jansen FE, Sadleir LG, Harkin LA, et al. Severe myoclonic epilepsy of infancy (Dravet syndrome): recognition and diagnosis in adults. Neurology 2006; 67:2224–6.
[17] Akiyama M, Kobayashi K, Yoshinaga H, et al. A long-term follow-up study of Dravet syndrome up to adulthood. Epilepsia 2010; 51:1043–52.
[18] Genton P, Velizarova R, Dravet C. Dravet syndrome: the long-term outcome. Epilepsia. 2011; 52:44–9.
[19] Takayama R, Fujiwara T, Shigematsu H, et al. Long-term course of Dravet syndrome: a study from an epilepsy center in Japan. Epilepsia 2014; 55:528–38.
[20] Lagae L, Brambilla I, Mingorance A, et al. Quality of life and comorbidities associated with Dravet syndrome severity: a multinational cohort survey. Dev Med Child Neurol 2018; 60:63–72.
[21] Darra F, Battaglia D, Dravet C, et al. Dravet syndrome: Early electroclinical findings and long-term outcome in adolescents and adults. Epilepsia 2019; 60 Suppl 3:S49-S58.
[22] Ragona F, Granata T, Dalla Bernardina B, et al. Cognitive development in Dravet syndrome: a retrospective, multicenter study of 26 patients. Epilepsia 2011; 52(2):386-92.
[23] Berg AT, Mahida S, Poduri A. KCNQ2-DEE: developmental or epileptic encephalopathy? Ann Clin Transl Neurol 2021; 8(3):666-676.
[24] Rodda JM, Scheffer IE, McMahon JM, et al. Progressive gait deterioration in adolescents with Dravet syndrome. Arch Neurol 2012; 69(7):873-8.

[25] Catarino CB, Liu JY, Liagkouras I, Gibbons VS, et al. Dravet syndrome as epileptic encephalopathy: evidence from long-term course and neuropathology. Brain 2011; 134(Pt 10):2982-3010.
[26] Bishop KI, Isquith PK, Gioia GA, et al. Improved everyday executive functioning following profound reduction in seizure frequency with fenfluramine: Analysis from a phase 3 long-term extension study in children/young adults with Dravet syndrome. Epilepsy Behav 2021; 121(Pt A):108024
[27] Silvennoinen K, Puvirajasinghe C, Hudgell K et al. Late diagnoses of Dravet syndrome: How many individuals are we missing? Epilepsia Open 2021; 6(4):770-776.
[28] Villas N, Meskis MA, Goodliffe S. Dravet syndrome: Characteristics, comorbidities, and caregiver concerns. Epilepsy Behav 2017; 74:81-86.
[29] Ceulemans B. Overall management of patients with Dravet syndrome. Dev Med Child Neurol 2011; 53 Suppl 2:19-23.
[30] Kuchenbuch M, Chemaly N, Chiron C, et al. Transition and transfer from pediatric to adult health care in epilepsy: a families' survey on Dravet syndrome. Epilepsy Behav 2013; 29(1):161-5.
[31] Berg AT, Gaebler-Spira D, Wilkening G, et al. Nonseizure consequences of Dravet syndrome, KCNQ2-DEE, KCNB1-DEE, Lennox-Gastaut syndrome, ESES: A functional framework. Epilepsy Behav 2020; 111:107287.
[32] FDA. Rare Diseases: Natural History Studies for Drug Development Guidance for Industry; 2019. Available from https://www.fda.gov/media/122425/download (Accessed on 1 October 2022)
[33] Cardenal-Muñoz E, Auvin S, Villanueva V, Cross JH, Zuberi SM, Lagae L, Aibar JÁ. Guidance on Dravet syndrome from infant to adult care: Road map for treatment planning in Europe. Epilepsia Open. 2021 Dec 9. doi: 10.1002/epi4.12569. Epub ahead of print. PMID: 34882995

Appendix

Detailed description of most common vulnerabilities and well-mastered skills found in subgroups with “complete” and “incomplete phenotypes”.

Complete phenotype

	Well-mastered skills	Vulnerabilities
Communication - Receptive	Understanding of 'yes' and 'no'; indicating body parts on request; executing commands that require an action and an object (e.g., “Take the shoes”).	Few difficulties; when present, difficulties were more related to listening span and understanding of idioms.

Communication - Expressive	Saying and repeating single words; making a pointing gesture.	Various difficulties related to conversation skills, describing experiences and objectives in detail, and giving instructions to others.
Communication - Written	None	Major difficulties even in basic skills, such as writing their own name and reading/writing simple words.
Daily Living - Personal	Discreet toilet management (with good daytime urination and defecation control); basic eating and drinking skills; ability to take off clothes that open in the front.	Difficulty with basic personal hygiene skills and major difficulty in managing personal health, even for relatively simple skills such as using a thermometer.
Daily Living – Domestic	Paying attention to hot objects.	Difficulty with basic skills involving cleaning, cooking, washing dishes and clothes, and performing simple maintenance such as changing a light bulb.
Daily Living - Community	Understanding the function of a telephone.	Notable difficulty with basic skills involving money management, traveling independently both on foot and by using other means of transport, using a clock, and finding and maintaining a job.
Socialization - Social Interaction	Basic skills such as showing interest in peers, imitating simple movements, and showing affection.	Notable difficulty with basic conversation skills, emotional expression, understanding inexplicit aspects of communication, and the creation and management of sentimental relationships.
Socialization - Play and Leisure	Basic skills such as showing interest in surroundings, accepting jokes, and participating in simple games.	Some difficulty, especially with games or sports that have articulated rules and scores, as well as in planning recreational activities and going out with friends.
Socialization - Coping Skills	Basic skills such as changing behavior depending on familiarity with the other person and apologizing for unintentional negative behavior.	Some difficulty with more complex skills, such as understanding the risks of certain social situations, consideration of consequences when making decisions, and respect for colleagues.
Gross Motor Skills	Well-maintained standing position; autonomous walk, discreetly performed the stairs, able to throw a ball.	Notable difficulty with jumping, using a bicycle, and catching a small ball.

Fine Motor Skills	Basic skills to browse a book, open a door using the handle, unwrap objects.	Difficulty with cutting complex figures with scissors, using a ruler, tying a bow, and using a keyboard to type text.

Incomplete phenotype

	Well-mastered skills	Vulnerabilities
Communication – Receptive	Understanding of 'yes' and 'no'; indicating body parts on request; executing commands that require one action and one object (e.g., "Take the shoes"); fair listening spans.	None
Communication - Expressive	Fairly broad vocabulary; use of possessives; naming of colors; use of negation and verbal forms to the past, ability to tell in detail the experiences and to converse, report date of birth and phone number.	None
Communication – Written	Able to read and write at least a small number of words independently (not texts).	None
Daily Living - Personal	Good toilet management; good table skills (including use of a knife); greater autonomy in bathing, dressing (including use of zippers and buttons), and putting on shoes.	Difficulty with more complex personal skills, such as scheduling medical appointments and managing medications.
Daily Living - Domestic	Minimum skills, such as collaborating in simple housework and clearing their place at the table.	Difficulty with complex cooking skills, such as planning and cooking main meals; difficulty with domestic cleaning.
Daily Living - Community	Making phone calls; simple use of some digital devices; understanding the function of money; compliance with certain road safety rules.	Difficulty with more complex skills, such as owning and managing a credit card.
Socialization - Social Interaction	Basic skills, such as expressing interest in peers, imitating simple movements, and showing love.	Difficulty with more complex skills, such as dating.
Socialization - Play and Leisure	Basic skills, such as expressing interest in their surroundings, accepting jokes, making simple interaction games and collaborative games, and sharing materials even without supervision.	None
Socialization - Coping Skills	Basic skills, such as changing behavior depending on familiarity with another person and apologizing for unintentional negative behavior.	Some difficulty with more complex skills such as understanding the risks of certain social situations, considering consequences of decisions, and respect for colleagues.
Gross Motor Skills	Well-maintained standing position; autonomous walking; use of stairs; can throw a ball, climb, and run.	None
Fine Motor Skills	Basic skills such as browsing a book, opening a door using the handle, unwrapping objects, building three-dimensional structures, coloring without leaving the edges, drawing simple shapes, opening locks with a key, writing recognizable letters and numbers, and using a keyboard.	None

Conclusion and future perspectives

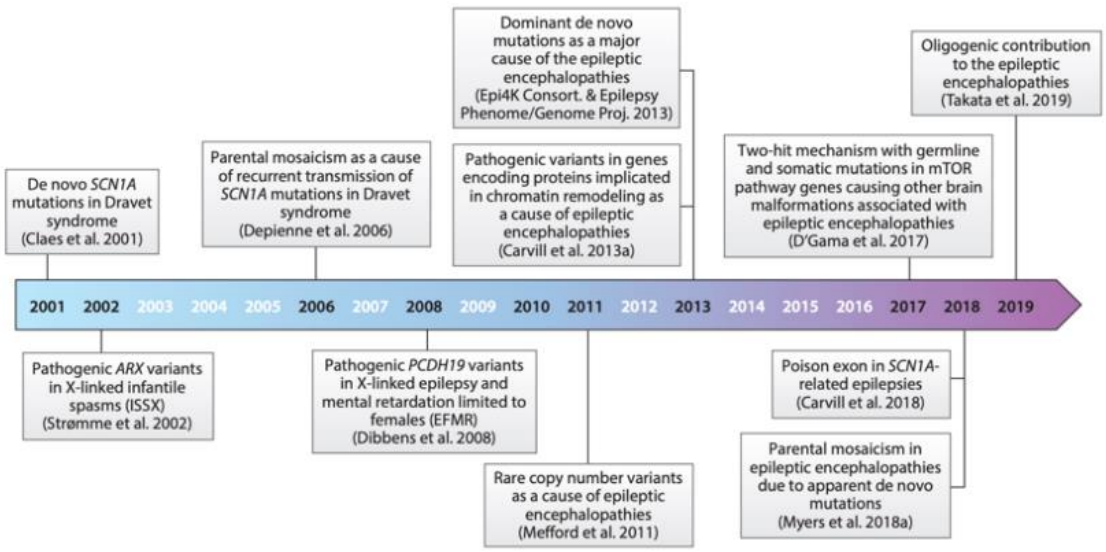
Although the nosological identification of diseases characterized by psychomotor delay, various kinds of comorbidities, and clinical worsening caused by phases of epileptic encephalopathies is certainly not new, the concept of developmental and epileptic encephalopathies (DEEs) has been introduced quite recently.[1] The introduction of this new definition—and in particular, the differentiation from epileptic encephalopathies in the strict sense—implies an outcome-related connotation, which is worse in DEEs. In addition, this new concept and the advances in neurogenetics allow for better definitions of several conditions previously only assignable within the framework of an epileptic syndrome (e.g., *CDKL5*-related DEE versus West syndrome/Lennox-Gastaut syndrome) with implications in terms of outcome definition and targeted treatment.[2][3][4]

However, at the moment, the mere genetic (monogenic) definition of the cause of epilepsy is still not sufficient alone to define prognosis. Indeed, in most cases, establishing definite genotype-phenotype correlations is not possible and mutations of the same gene can determine conditions from the most severe to the mildest, sometimes also being the same mutation.[5][6][7]
In the coming years, research in the field of neurogenetics (including the epigenetic regulation, complex inheritance, and functional testing of variants) will likely provide additional knowledge and guide targeted treatments.[8][9][10][11][12][13][14] To date, the most advanced approach in terms of outcome definition and treatment indication is represented by the syndromic classification of epilepsy associated with the etiological definition.[1]

Dravet syndrome (DS) is the archetype of DEEs and was clinically defined in 1978. Twenty years later, the gene responsible for DS was discovered. This advent of genetic diagnostics allowed clinicians to determine a monogenic etiology in more than 85% of cases (*SCN1A*). [15][16]
The natural history of DS has been outlined over the years, mostly through clinical data collections from third-level clinical centers predominantly interested in pediatric patients. It is known that the onset is in the first year of life in children with normal psychomotor development who start to present seizures, often prolonged and triggered by fever (first stage). Subsequently, a drug-resistant epilepsy establishes with polymorphic focal or generalized seizures, and there is an increased risk of SUDEP. A delay of psychomotor development and consequent cognitive impairment becomes evident with several types of comorbidities involving language, movement, behavior, socialization, sleep, eating, and autonomic functions (second phase). Epilepsy severity progressively decreases from childhood to adolescence and throughout adulthood. Comorbidities have a greater impact on the quality of life, and in some cases, there is a gait deterioration (third phase).[15][16]

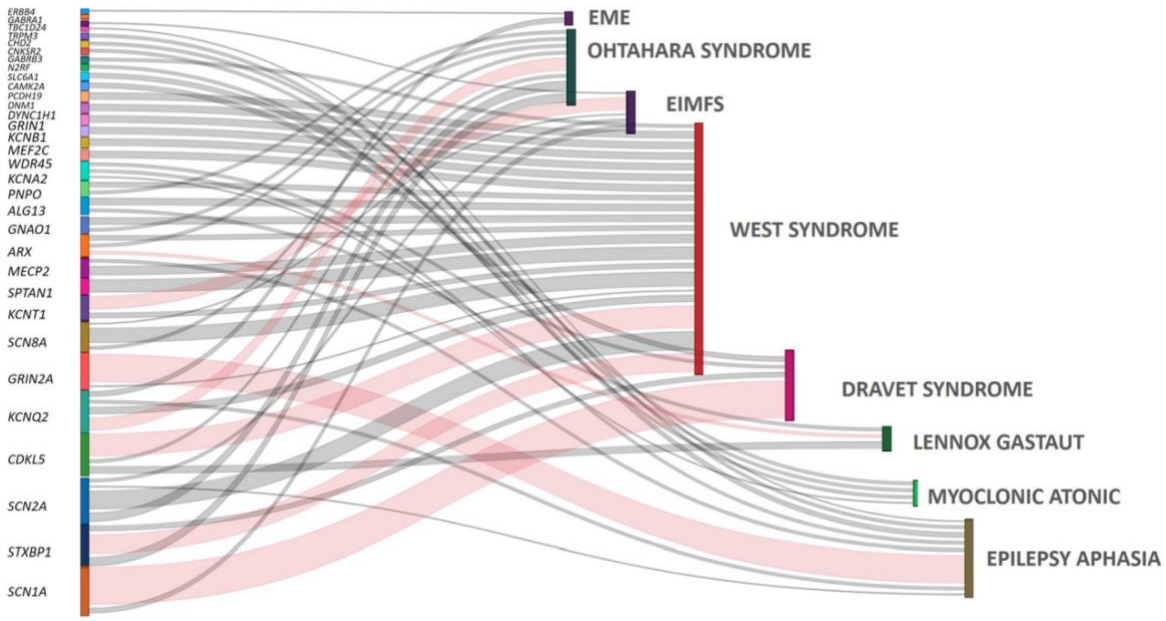
Recent studies are better delineating several facets of the natural history. For example, to date it is known that first seizure does not necessarily appear during fever,[15] that cognitive course is not a pure consequence of epilepsy,[17] that first motor milestones could be reached with slight delay,[18] that outcome is worse in patients who show myoclonic seizures and absences,[19][20] and that autonomous and peripheral nervous systems are probably involved in the pathogenesis.[21][22] Thus far, several clinical aspects, especially those concerning adulthood, are still unknown. For example, while several studies reported the efficacy of new antiseizure treatments started in adulthood and their effects on daily life,[23][24] data regarding the implications of rehabilitation interventions in adulthood are lacking. This is a crucial issue since it is known that while rehabilitation interventions are usually reduced for adults with DS, late diagnosis of DS and other genetic DEEs has been increasing in recent years.[25][26]

Figure 1. Timeline of major molecular discoveries in the DEEs (from Perucca et al. [3])



Despite that disease course may be quite variable, DS can be considered a kind of homogenous DEE. Most epileptic syndromes defined as DEE share common electroclinical features but are very heterogeneous for etiologies and outcomes (e.g., Lennox-Gastaut syndrome, Infantile Epileptic Spasms syndrome)(Figure 2). [12][27][28]

Figure 2. Genotype-phenotype correlation of genes associated with DEE (from Syrbe 2022 [28])



Considering the emerging approaches to diagnosis and treatment—and in the spirit of precision medicine—it is necessary to rapidly increase knowledge around new genetic frameworks to develop a characterization of the natural history and, subsequently, of the outcome and treatment. In this thesis, we applied different strategies of phenotyping and diagnosis, complementary to the conventional, established ones, to DS with the aim of exploring their feasibility and considering their applicability to other new DEEs.

Disease registries are organized systems that use observational study methods to collect uniform data to evaluate specified outcomes.[29][30] They represent an ideal tool for research and improving knowledge in the field of rare diseases where data are often scarce, heterogeneous, and fragmented.[31][32]

In the field of DEEs, some clinical registries have emerged in last years, mainly driven by patients' associations with networks of third-level epilepsy centers. [33][34][35][36][37] The Residras registry (**chapter 1**), founded in 2014, represents a collection of standardized national data of patients with DS that relies on a shareable and interoperable framework, promoting multicenter high-quality data collection. Additionally, since May 2020. EpiCARE, an ERN (European Reference Network) collaborative platform that includes third-level European centers highly experienced in rare and complex epilepsies, has been promoting the EpiCARE Registry project with the aim of better understanding the epidemiology of rare and complex epilepsies and determining their natural histories.[38] Using the EpiCARE registry would lead to providing guidelines for early recognition of the diseases, localize patients, and facilitate their participation in therapy trials, as well as reduce inequality in patients care.

The use of electronic data (like disease registries) is promoting the emergence of several data mining techniques to study diverse facets of epilepsy, including risk of comorbidities and complications, imaging analysis and pre-surgical planning, and medication response and outcome.[39][40][41][42][43][44][45][46][47][48] In **chapters 2 and 3**, we explored the feasibility of an innovative informatics technique—natural language processing (NLP)—to extrapolate diagnostic markers and natural history from reports of patients with DS. The extrapolation of unstructured data is particularly appropriate in the field of rare diseases: free text reports are produced without constraints and may be used to express nuanced, unexpected, and unexplained signs or symptoms regarding patient cases.[49] We showed that narrative medical reports of individuals younger than two years with febrile seizures contain different words depending on whether they have or will develop clinical phenotype of DS. With the same approach, we demonstrated that concepts from reports of children with DS and the appearance of these terms in the reports during the clinical course follow the same outline as what is known for DS. These results prove the feasibility of this technique for these aims and can be transposed to investigate the natural history of other rare, lesser-known DEEs.[50]

The acknowledgment of the natural history and the easy access to cohorts of patients who may benefit from emerging novel therapies (like diseases registries) are essential steps for developing clinical trials. In epilepsy, evaluation of clinical trials is mainly based on primary outcome measure targeting efficacy and safety. In our study (**chapter 4**), we showed that desirable patient-centered outcome measures, collected through a caregiver survey, change over time in patients with DS, and items related to daily life (sleep, communication, behavior, language, and motor skills) had the highest rank compared with items about seizures. These findings suggest that secondary outcomes of clinical trials should consider additional aspects beyond seizures that equally contribute to determine quality of life, and reinforce the benefit of exploring patients' concerns and caregivers in defining natural history and unmet needs.[51][52][53][54][55][56][57]

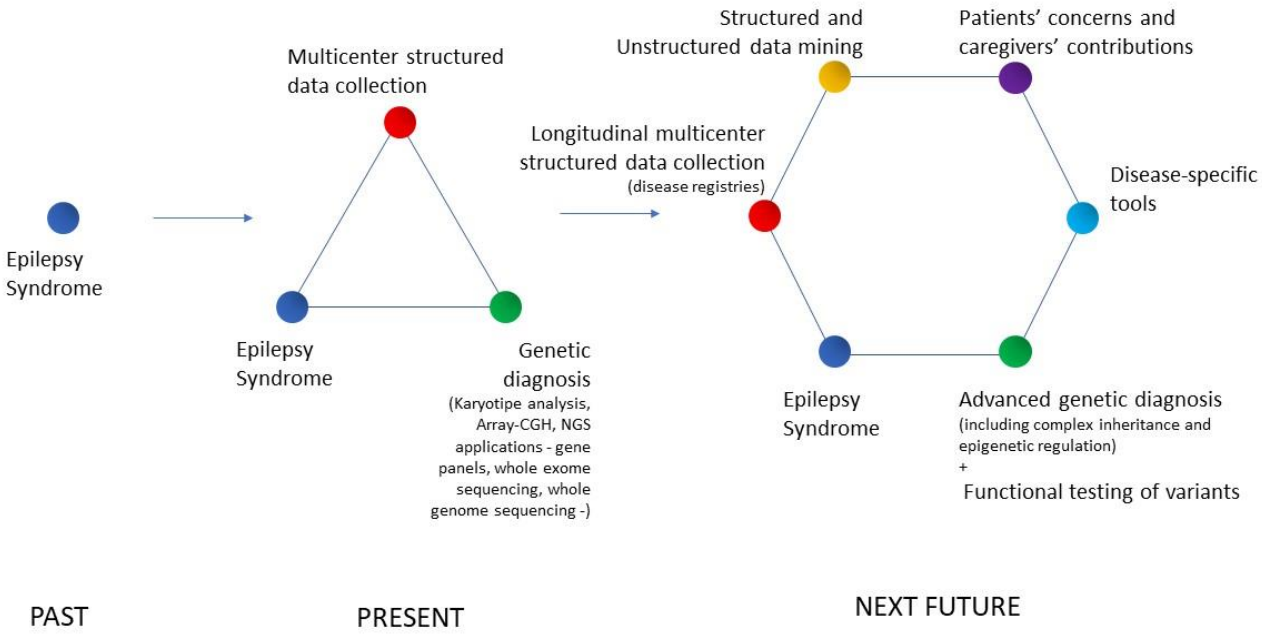
The findings also suggest the necessity for continuously monitoring patients with DEEs since the disease course may not be stable by the time comorbidities appear and change, and individuals' needs require reassessment. Tools commonly available to monitor the progress of DEEs are, in most cases, not sufficiently sensitive for these purposes because they are calibrated on standard populations.[58] As such, in **chapter 5**, we explored the feasibility of using an adaptive behavior profile (ABP) assessment generated from a common tool as an instrument for outcome measures in adolescents and adults with DS. We showed the need to adopt a deep raw data analysis to extrapolate operable results.

With this approach, it also became possible to enrich knowledge about DS in adulthood and reinforce the hypothesis of two distinct outcome trajectories strongly linked to the epileptic phenotype. Difficulties experienced underline the need for disease-specific tools.[59][60]

These demonstrated approaches, which are complementary to those conventional ones, are extendable to other DEEs whose natural history is to date less known than DS. We can therefore prospect that in the near future defining the natural history of a DEE and the development of targeted treatments, would include the following:

- definition of the epilepsy syndrome and the underlying etiology,
- cross-sectional and longitudinal multicentric structured data collections (disease registries),
- findings from structured and unstructured data mining (regarding written medical reports, genetic/epigenetic data, medical images, and biochemical data),
- patients' concerns and caregivers' contributions to define comorbidities, patients' needs and their changes during the diseases course, and
- disease-specific tools to monitor course and evolution.

Figure 3. Evolution in determining natural history of genetic DEEs.



In conclusion, scientific and technological progress, the possibility of rapidly disseminating data and knowledge, and the increasing active involvement of patients and their families in research will allow the identification of novel strategies to study rare diseases, including DEEs. These processes will

allow clinicians to understand natural history and individuate targeted therapies more quickly than in the past, improving quality of life for patients and their families.

This work shows that the use of these complementary tools provides a more complete view of the disease. Future studies aimed at investigating the natural history of "new" genetic conditions associated with a DEE, should encourage the structured collection of data through diseases registries, allowing to define the longitudinal evolution of the disease. This data should be periodically integrated with those coming from caregivers, allowing to bring to light problems that could be unnoticed or underestimated by clinicians. Considering the rarity of these conditions, large-scale data sharing between specialized centers is indispensable. Based on this data, mining and artificial intelligence techniques will help to further define natural history. Finally, to ensure adequate care, specific severity assessment tools should be developed basing on the natural history.

References

[1] Scheffer IE, Berkovic S, Capovilla G, Connolly MB, Guilhoto L, Hirsch E, et al. ILAE Classification of the Epilepsies Position Paper of the ILAE: Commission for Classification and Terminology. Epilepsia. 2017;58:512–21

[2] Knight EMP, Amin S, Bahi-Buisson N, Benke TA, Cross JH, Demarest ST, Olson HE, Specchio N, Fleming TR, Aimetti AA, Gasior M, Devinsky O; Marigold Trial Group. Safety and efficacy of ganaxolone in patients with CDKL5 deficiency disorder: results from the double-blind phase of a randomised, placebo-controlled, phase 3 trial. Lancet Neurol. 2022 May;21(5):417-427.

[3] Perucca P, Bahlo M, Berkovic SF. The Genetics of Epilepsy. Annu Rev Genomics Hum Genet. 2020 Aug 31;21:205-230. doi: 10.1146/annurev-genom-120219-074937. Epub 2020 Apr 27. PMID: 32339036.)

[4] Ko A, Youn SE, Kim SH, et al. Targeted gene panel and genotype-phenotype correlation in children with developmental and epileptic encephalopathy. Epilepsy Res. 2018;141:48-55.

[5] Goldberg-Stern, H. et al. Broad phenotypic heterogeneity due to a novel SCN1A mutation in a family with genetic epilepsy with febrile seizures plus. J.Child Neurol. 29, 221–226 (2014).

[6] Helbig, I. & Tayoun, A. A. Understanding genotypes and phenotypes in epileptic encephalopathies. Mol. Syndr. 7, 172–181 (2016).

[7] Wolff M, Johannesen KM, Hedrich UBS, Masnada S, Rubboli G, et al. Genetic and phenotypic heterogeneity suggest therapeutic implications in SCN2A-related disorders. Brain. 2017 May 1;140(5):1316-1336.

[8] Conboy K, Henshall DC, Brennan GP. Epigenetic principles underlying epileptogenesis and epilepsy syndromes. Neurobiol Dis. 2021 Jan;148:105179.
[9] Takata A, Nakashima M, Saitsu H, Mizuguchi T, Mitsuhashi S, Takahashi Y, et al. Comprehensive analysis of coding variants highlights genetic complexity in developmental and epileptic encephalopathy. Nat Commun. 2019 Jun 7;10(1):2506 .
[10] Zhu Y, Huang D, Zhao Z, Lu C. Bioinformatic analysis identifies potential key genes of epilepsy. PLoS One. 2021 Sep 23;16(9):e0254326.
[11] Clatot J, Parthasarathy S, Cohen S, McKee JL, Massey S, Somarowthu A, Goldberg EM, Helbig I. SCN1A gain-of-function mutation causing an early onset epileptic encephalopathy. Epilepsia. 2022 Oct 26. Epub ahead of print.
[12] Bayat A, Bayat M, Rubboli G, Møller RS. Epilepsy Syndromes in the First Year of Life and Usefulness of Genetic Testing for Precision Therapy. Genes (Basel). 2021;12(7):1051. Published 2021 Jul 8.
[13] Specchio N, Pietrafusa N, Perucca E, Cross JH. New paradigms for the treatment of pediatric monogenic epilepsies: Progressing toward precision medicine. Epilepsy Behav. 2022;131(Pt B):107961.
[14] Ellis CA, Petrovski S, Berkovic SF. Epilepsy genetics: clinical impacts and biological insights. Lancet Neurol. 2020;19(1):93-100. doi:10.1016/S1474-4422(19)30269-8
[15] Dravet C. The core Dravet syndrome phenotype. Epilepsia. 2011;52 Suppl 2:3-9.
[16] Gataullina S, Dulac O. From genotype to phenotype in Dravet disease. Seizure. 2017;44:58-64
[17] Nabbout R, Chemaly N, Chipaux M, Barcia G, Bouis C, Dubouch C, Leunen D, Jambaqué I, Dulac O, Dellatolas G, Chiron C. Encephalopathy in children with Dravet syndrome is not a pure consequence of epilepsy. Orphanet J Rare Dis. 2013 Nov 13;8:176.
[18] Verheyen K, Verbecque E, Ceulemans B, Schoonjans AS, Van De Walle P, Hallemans A. Motor development in children with Dravet syndrome. Dev Med Child Neurol. 2019 Aug;61(8):950-956.
[19] Ragona F, Granata T, Dalla Bernardina B, Offredi F, Darra F, Battaglia D, et al. Cognitive development in Dravet syndrome: a retrospective, multicenter study of 26 patients. Epilepsia. 2011;52:386–92.
[20] Darra F, Battaglia D, Dravet C, Patrini M, Offredi F, Chieffo D, et al. Dravet syndrome: Early electroclinical findings and long-term outcome in adolescents and adults. Epilepsia. 2019;60 Suppl 3:S49–58.
[21] Gitiaux C, Chemaly N, Quijano-Roy S, Barnerias C, Desguerre I, Hully M, Chiron C, Dulac O, Nabbout R. Motor neuropathy contributes to crouching in patients with Dravet syndrome. Neurology. 2016 Jul 19;87(3):277-81.

[22] Myers KA, Bello-Espinosa LE, Symonds JD, Zuberi SM, Clegg R, Sadleir LG, Buchhalter J, Scheffer IE. Heart rate variability in epilepsy: A potential biomarker of sudden unexpected death in epilepsy risk. Epilepsia. 2018 Jul;59(7):1372-1380.

[23] Catarino CB, Liu JY, Liagkouras I, Gibbons VS, et al. Dravet syndrome as epileptic encephalopathy: evidence from long-term course and neuropathology. Brain 2011; 134(Pt 10):2982–3010.

[24] Bishop KI, Isquith PK, Gioia GA, et al. Improved everyday executive functioning following profound reduction in seizure frequency with fenfluramine: Analysis from a phase 3 long-term extension study in children/young adults. Epilepsy Behav. 2021;121(Pt A):108024.

[25] Silvennoinen K, Puvirajasinghe C, Hudgell K et al. Late diagnoses of Dravet syndrome: How many individuals are we missing? Epilepsia Open 2021; 6(4):770-776.

[26] Isom LL, Knupp KG. Dravet Syndrome: Novel Approaches for the Most Common Genetic Epilepsy. Neurotherapeutics. 2021;18:1524–34.

[27] Scheffer IE, Liao J. Deciphering the concepts behind "Epileptic encephalopathy" and "Developmental and epileptic encephalopathy". Eur J Paediatr Neurol. 2020 Jan;24:11-14. Epub 2019 Dec 31.

[28] Syrbe, Steffen. Developmental and epileptic encephalopathies – therapeutic consequences of genetic testing. Medizinische Genetik, vol. 34, no. 3, 2022, pp. 215-224.

[29] Gliklich R, Dreyer N, Leavy M. Registries for evaluating patient outcomes: a user's guide. Agency for Healthcare Research and Quality. 2014. https://www.ncbi.nlm.nih.gov/books/NBK208616/. Accessed 20 Dec 2022.

[30] U.S. Food and Drug Administration. Framework for FDA's Real-World Evidence Program. 2018. https://www.fda.gov/ media/120060/download. Accessed 20 Dec 2022.

[31] Boulanger V, Schlemmer M, Rossov S, Seebald A, Gavin P. Establishing Patient Registries for Rare Diseases: Rationale and Challenges. Pharmaceut Med. 2020;34(3):185-190. doi:10.1007/s40290-020-00332-1;

[32] Rare disease registries in Europe. Orphanet Report Series, September 2020 [Internet]. [cited 2022 Dec 10]. Available from: https://www.orpha.net/orphacom/cahiers/docs/GB/Registries.pdf

[33] CDKL5 Registry [Internet]. [cited 2022 Dec 10]. Available from: https://www.cdkl5.com/connect-cdkl5/

[34] Krey I, Heine C, Frömming M, Herrmann J, Møller RS, Weckhuysen S, Courage C, Beblo S, Syrbe S, Lemke JR. The Angelman Syndrome Online Registry - A multilingual approach to support global research. Eur J Med Genet. 2021 Dec;64(12):104349

[35] Ha A, Kim SH, Baek SU, Kim JS, Yoon HJ, Kim YK. Incidence of Sturge-Weber Syndrome and Risk of Secondary Glaucoma: A Nationwide Population-based Study Using a Rare Disease

Registry [published online ahead of print, 2022 Nov 11]. Am J Ophthalmol. 2022;S0002-9394(22)00431-7.

[36] ENRAH: European alternating hemiplegia and rare epilepsies registry in childhood [Internet]. [cited 2022 Dec 10]. Available from: http://www.enrah.net/

[37] Encinas AC, Moore IKM, Watkins JC, Hammer MF. Influence of age at seizure onset on the acquisition of neurodevelopmental skills in an SCN8A cohort. Epilepsia. 2019;60(8):1711-1720.

[38] EpiCare Registry Project [Internet]. [cited 2022 Dec 10]. Available from: https://epi-care.eu/epicares-activities/registry-project/

[39] Bensken WP, Fernandez-Baca Vaca G, Jobst BC, et al. Burden of Chronic and Acute Conditions and Symptoms in People With Epilepsy. Neurology. 2021;97(24):e2368-oke2380. doi:10.1212/WNL.0000000000012975

[40] Chacon LM, Garcia LG, Bosch-Bayard J, et al. Relation of Brain Perfusion Patterns to Sudden Unexpected Death Risk Stratification: A Study in Drug Resistant Focal Epilepsy. Behav Sci (Basel). 2022;12(7):207. Published 2022 Jun 24.

[41] Barbour K, Hesdorffer DC, Tian N, et al.: Automated detection of sudden unexpected death in epilepsy risk factors in electronic medical records using natural language processing. Epilepsia 60:1209–1220. 2019

[42]Abbasi B, Goldenholz DM. Machine learning applications in epilepsy. Epilepsia. 2019;60(10):2037-2047

[43] Fernandes M, Donahue MA, Hoch D, et al. A replicable, open-source, data integration method to support national practice-based research & quality improvement systems. Epilepsy Res. 2022;186:107013.

[44] An K. Psychosocial Factors and Psychological Characteristics of Personality of Patients with Chronic Diseases Using Artificial Intelligence Data Mining Technology and Wireless Network Cloud Service Platform. Comput Intell Neurosci. 2022;2022:8418589. Published 2022 Apr 13.

[45] Breitenstein PS, Mahmoud I, Al-Azzawi F, Shakibfar S, Sessa M. A machine-learning guided method for predicting add-on and switch in secondary data sources: A case study on anti-seizure medications in Danish registries. Front Pharmacol. 2022;13:954393. Published 2022 Nov 10.

[46] Hully M, Lo Barco T, Kaminska A, et al. Deep phenotyping unstructured data mining in an extensive pediatric database to unravel a common KCNA2 variant in neurodevelopmental syndromes. Genet Med. 2021;23(5):968-971. doi:10.1038/s41436-020-01039-z

[47] Koivusilta L, Kaltiala R, Myöhänen A, Hotulainen R, Rimpelä A. A Chronic Disease in Adolescence and Selection to an Educational Path-A Longitudinal Study. Int J Environ Res Public Health. 2022;19(21):14407. Published 2022 Nov 3.

[48] He K, Hong N, Lapalme-Remis S, et al. Understanding the patient perspective of epilepsy treatment through text mining of online patient support groups. Epilepsy Behav. 2019;94:65-71.
[49] Next generation phenotyping using narrative reports in a rare disease clinical data warehouse. Orphanet J Rare Dis. 2018; 13(1):85.
[50] Yew ANJ, Schraagen M, Otte WM, van Diessen E. Transforming epilepsy research: a systematic review on natural language processing applications [published online ahead of print, 2022 Dec 3]. Epilepsia. 2022Garcelon N, Neuraz A, Salomon R, et al.
[51] Lagae L, Brambilla I, Mingorance A, Gibson E, Battersby A. Quality of life and comorbidities associated with Dravet syndrome severity: a multinational cohort survey. Dev Med Child Neurol. 2018;60(1):63–72.
[52] Lo Barco T, De Gaetano L, Santangelo E, et al. SYNGAP1-related developmental and epileptic encephalopathy: The impact on daily life. Epilepsy Behav. 2022;127:108500.
[53] Berg AT, Mahida S, Poduri A. KCNQ2-DEE: developmental or epileptic encephalopathy?. Ann Clin Transl Neurol. 2021;8(3):666-676.
[54] Cutts A, Savoie H, Hammer MF, et al. Clinical characteristics and treatment experience of individuals with SCN8A developmental and epileptic encephalopathy (SCN8A-DEE): Findings from an online caregiver survey. Seizure. 2022;97:50-57.
[55] Aledo-Serrano Á, Del Ser T, Gil-Nagel A. Antiseizure medication withdrawal in seizure-free patients with PCDH19-related epilepsy: A multinational cohort survey. Seizure. 2020;80:259-261.
[56] Juandó-Prats C, James E, Bilder DA, et al. DRAVET ENGAGE. Parent caregivers of children with Dravet syndrome: Perspectives, needs, and opportunities for clinical research. Epilepsy Behav. 2021;122:108198.
[57] Wright D, Kenny A, Eley S, McKechanie AG, Stanfield AC. Clinical and behavioural features of SYNGAP1-related intellectual disability: a parent and caregiver description. J Neurodev Disord. 2022;14(1):34. Published 2022 Jun 2.
[58] Berg AT, Palac H, Wilkening G, et al. SCN2A-Developmental and Epileptic Encephalopathies: Challenges to trial-readiness for non-seizure outcomes. Epilepsia 2021; 62(1):258-268
[59] De Vries PJ, Whittemore VH, Leclezio L, Byars AW, Dunn D, Ess KC, et al. Tuberous sclerosis associated neuropsychiatric disorders (TAND) and the TAND checklist. Pediatr Neurol. 2015;52(1):25–35
[60] Demarest S, Pestana-Knight EM, Olson HE, et al. Severity Assessment in CDKL5 Deficiency Disorder. Pediatr Neurol.2019;97:38-42.