

## Common terms for rare epilepsies: Synonyms, associated terms, and links to structured vocabularies

\*Zachary M. Grinspan, †Niu Tian, ‡Elissa G. Yozawitz, §Patricia E. McGoldrick, §Steven M. Wolf, ¶Tiffani L. McDonough, #Aaron Nelson, \*Baria Hafeez, \*Stephen B. Johnson, and ¶Dale C. Hesdorffer

*Epilepsia Open*, 3(1):91–97, 2018  
doi: 10.1002/epi4.12095

### SUMMARY

Identifying individuals with rare epilepsy syndromes in electronic data sources is difficult, in part because of missing codes in the International Classification of Diseases (ICD) system. Our objectives were the following: (1) to describe the representation of rare epilepsies in other medical vocabularies, to identify gaps; and (2) to compile synonyms and associated terms for rare epilepsies, to facilitate text and natural language processing tools for cohort identification and population-based surveillance. We describe the representation of 33 epilepsies in 3 vocabularies: Orphanet, SNOMED-CT, and UMLS-Metathesaurus. We compiled terms via 2 surveys, correspondence with parent advocates, and review of web resources and standard vocabularies. UMLS-Metathesaurus had entries for all 33 epilepsies, Orphanet 32, and SNOMED-CT 25. The vocabularies had redundancies and missing phenotypes. Emerging epilepsies (SCN2A-, SCN8A-, KCNQ2-, SLC13A5-, and SYNGAP-related epilepsies) were underrepresented. Survey and correspondence respondents included 160 providers, 375 caregivers, and 11 advocacy group leaders. Each epilepsy syndrome had a median of 15 (range 6–28) synonyms. Nineteen had associated terms, with a median of 4 (range 1–41). We conclude that medical vocabularies should fill gaps in representation of rare epilepsies to improve their value for epilepsy research. We encourage epilepsy researchers to use this resource to develop tools to identify individuals with rare epilepsies in electronic data sources.

**KEY WORDS:** Rare epilepsy, Terminology, Classification, Natural language processing, Synonyms.



Zachary M. Grinspan is director of pediatric epilepsy at Weill Cornell Medicine in New York City.

One obstacle to epilepsy research is limited representation in diagnostic coding systems like the International Classification of Diseases (ICD) system. Only a few

Accepted November 14, 2017.

\*Weill Cornell Medicine, New York, New York, U.S.A.; †Centers for Disease Control and Prevention, Atlanta, Georgia, U.S.A.; ‡Albert Einstein College of Medicine, Bronx, New York, U.S.A.; §Mount Sinai Health System, New York, New York, U.S.A.; ¶Columbia University Medical Center, New York, New York, U.S.A.; and #New York University Langone Medical Center, New York, New York, U.S.A.

Address correspondence to Zachary Grinspan, Room LA 220, 402 East 67th Street, New York, NY 10065, U.S.A. E-mail: zag9005@med.cornell.edu

© 2017 The Authors. *Epilepsia Open* published by Wiley Periodicals Inc. on behalf of International League Against Epilepsy.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

entities appear in versions 9 and 10, such as tuberous sclerosis (ICD-9-CM 759.5, ICD-10 Q85.1) and infantile spasms (ICD-9-CM 345.6, ICD-10 G40.82). Although version 10 adds juvenile myoclonic epilepsy (G40.B), absence epilepsy (G40.A), and Lennox-Gastaut syndrome (G40.81), many rare epilepsies appear in neither system. For example, Dravet syndrome is coded with the non-specific ICD-9 345.1 “Generalized Convulsive Epilepsy” or ICD-10 G40.4 “Other generalized epilepsy and epileptic syndromes.” This limits the utility of large administrative and clinical datasets for epilepsy research, despite the clear value of such data for research in epidemiology, comparative effectiveness, health services, and quality improvement.<sup>1</sup>

Given the increasing availability of electronic health records for research, computerized analysis of clinical

notes may be useful for finding patients with specific epilepsy syndromes. For example, text processing can identify children with complex febrile seizures,<sup>2</sup> and natural language processing can identify candidates for epilepsy surgery.<sup>3</sup>

One challenge for using these techniques is the diversity of terms for epilepsies. Dravet syndrome, for example, is “severe myoclonic epilepsy of infancy” for some clinicians and “intractable childhood epilepsy with generalized tonic clonic seizures” for others. Although existing standardized medical vocabularies document some synonyms, many vocabularies have gaps. For example, OMIM (Online Mendelian Inheritance in Man) lacks well-defined entries for infantile spasms, Lennox-Gastaut syndrome, and migrating partial seizures of infancy.

To facilitate the development of text and natural language processing tools for epilepsy, we compiled synonyms and associated terms for 33 rare epilepsies, with links to 3 standard medical vocabularies.

## METHODS

### Study design

We compiled synonyms and associated terms for rare epilepsies via 6 sources: a survey of pediatric neurology clinicians (“provider survey”), a survey of caregivers of individuals with rare epilepsies (“caregiver survey”), manual review of online resources, a manual review of 3 structured medical vocabularies, correspondence with leaders of rare epilepsy advocacy groups, and independent clinician review. We selected 33 rare epilepsies based on the parent-led advocacy groups that constitute the Rare Epilepsy Network (REN; ren.rti.org), an umbrella organization that fosters research collaboration.

### Surveys

For each survey, we designed, piloted, and iteratively refined questions using online software (SurveyMonkey, Inc., San Mateo, CA, U.S.A.). The “provider survey” was distributed to members of the Child Neurology Society (CNS) via email. Five versions were distributed, each asking respondents to list synonyms for 5–6 of the 33 epilepsies. The “caregiver survey” was distributed to REN participants. It asked respondents to name the epilepsy affecting the individual, and then list synonyms used by clinicians and family members.

### Websites

One investigator (ZG) manually reviewed several websites for more synonyms and associated terms. These included websites focused on epilepsy (epilepsy.com, dup15q.org), rare diseases (rarediseases.org, rarechromo.org, and ghr.nlm.nih.gov), and general knowledge (wikipedia.com).

### Vocabularies

Two investigators (ZG and BH) manually reviewed 3 standardized vocabularies: OrphaNet ([www.orpha.net](http://www.orpha.net)), Unified Medical Language System (UMLS) Metathesaurus ([utslogin.nlm.nih.gov/cas/login](http://utslogin.nlm.nih.gov/cas/login)), and Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT; [browser.ihtsdotools.org](http://browser.ihtsdotools.org)). In SNOMED-CT and UMLS, we included terms and subterms (ie, “infantile spasms” and “refractory infantile spasms”) as well as pathognomonic physical examination findings (“Ash leaf spot, tuberous sclerosis”). We limited the search to terms in English.

### Advocacy groups

We individually contacted advocacy group leaders via email. Each email contained a working list of synonyms and associated terms for the specified epilepsy, and asked the group leader to review, amend, and/or expand the list.

### Clinical experts

Five clinical pediatric epilepsy specialists (EY, TM, PM, AN, and SW) reviewed the drafted compendium and made additional recommendations and edits.

### Synonyms and associated terms

We reviewed terms that were synonymous with a rare epilepsy (eg, “Dimitri disease” for Sturge-Weber), as well as terms specific to a rare epilepsy, but better characterized as an “associated term.” For example, “shagreen patch” is specific to tuberous sclerosis, but is not a synonym. In addition, we reviewed terms that were not specific to one type of epilepsy (“intractable”). We sorted terms into 3 categories (synonyms, associated terms, and nonspecific) as determined jointly by 2 of the authors (ZG and DH). Disagreements were rare, and were resolved through discussion. We included only synonyms and associated terms.

## RESULTS

There were 160 respondents to the provider survey of the 1982 emailed CNS members (response rate 8%). These included 107 (67%) pediatric neurologists, 40 (25%) pediatric epilepsy specialists, and 13 (8%) other clinicians (ie, nurse, nurse practitioners, adult epilepsy specialist, physician assistant, EEG technician, or medical student). Twenty-six (16%) were physician trainees (ie, residents or fellows). Most worked in the United States (153; 96%), including responses from 33 states and the District of Columbia. The 7 international responses were from Canada (5), Lebanon (1), and the United Kingdom (1).

There were 375 respondents to the caregiver survey, of 1162 members of the REN (response rate 32%). Most (356; 95%) were parents of an individual with epilepsy. The remainder included 2 individuals with rare epilepsy (both progressive myoclonic epilepsy), 4 unspecified caregivers, and 13 unknown (left blank). The majority lived in the

United States (319; 85%), including 48 states (none from Alaska or North Dakota). There were 55 responses from 20 countries, and 1 unknown.

Leaders of 11 advocacy groups added more terms for the following disorders: Aicardi syndrome, Doose syndrome, Dravet syndrome, hypothalamic hamartoma, Lennox-Gastaut syndrome, PCDH19, Phelan-McDermid syndrome, Ring chromosome 20, *SCN8A*, *SYNGAP*, and tuberous sclerosis.

The final list included all 33 rare epilepsies (Table 1), including 16 defined by phenotype and 17 by genotype. Across all 33 rare epilepsies, there was a median of 14 synonyms (range 6–28). Fifteen had no associated terms. The remaining 19 epilepsies had a median of 4 associated terms (range 1–41). For example, Aicardi syndrome had 6 synonyms (AS, Aicardi's syndrome, Aicardi disease, Aicardi's disease, Aicardi, and Aicardi's) and 12 associated terms (microcephaly, retinal lacunae, agenesis of corpus callosum, absent corpus callosum, infantile spasms, polymicrogyria, porencephalic cysts, coloboma, optic disc, ACC, retinal lesions, and lacunae) (Table S1).

The UMLS-Metathesaurus had entries for all 33 epilepsies. Half (15) had 1 entry; the remaining had multiple entries, including 11 for Lennox-Gastaut syndrome, 18 for infantile spasms, and 25 for neuronal ceroid lipofuscinosis. Several entries were redundant—for example, there were 7 entries with overlapping descriptions of “refractory infantile spasms,” and 2 nearly identical entries for nonrefractory Lennox-Gastaut syndrome (Table 2). Five epilepsies (*KCNQ2*, *SCN2A*, *SCN8A*, *SYNGAP*, and *SLC13A5*) were not linked to disease entities, but rather to entries entitled “Caused by a mutation in [gene].”

OrphaNet had entries for 32 epilepsies. The missing epilepsy (*SLC13A5*) was linked with “ORPHA442835 Undetermined early-onset epileptic encephalopathy” but did not have its own entry. Three contained information about the relevant gene (*SCN2A*, *SCN8A*, and *SYNGAP*) only, without phenotype information (Table 2). Thirty had 1 entry, 2 had 7 entries: Angelman syndrome and neuronal ceroid lipofuscinosis.

SNOMED-CT had entries for 25 rare epilepsies (missing: *CDKL5*, Dup15q Syndrome, Glut1 deficiency, *KCNQ2*, *SCN2A*, *SCN8A*, *SLC13A5*, and *SYNGAP*; Table 2). Of these 25 rare epilepsies, 2 had multiple entries (6 for infantile spasms and 4 for tuberous sclerosis).

## DISCUSSION

### Summary

We compiled synonyms and associated terms for 33 rare epilepsies, including links to 3 standardized vocabularies. The compilation is available freely online. There are gaps in these vocabularies, including poor representation of emerging epilepsies (eg, *SCN2A*, *SCN8A*, *SYNGAP*, *KCNQ2*, and

*SLC13A5*), and redundancies of clinical concepts (treatment resistance).

### Significance

This work builds an important resource to help clinicians and researchers find epilepsy subpopulations in electronic health record systems. Currently, to collect large cohorts of children with epilepsy, clinical researchers are building registries via recruitment of patients one by one in clinical settings.<sup>4</sup> In contrast, text processing and/or natural language processing tools can quickly identify cohorts for observational research and clinical trials. Such tools may also support care coordination and quality improvement initiatives.

Three examples follow. First, case finding may be possible with simple text searches of clinical notes. This approach was used successfully (in combination with chart review) to identify children with febrile seizures.<sup>2</sup> However, simple text search may fail when terms are ambiguous (“*SCN1A* pending”), negated (“*SCN1A* testing unrevealing”), or included in boilerplate text (“Our epilepsy gene panel includes *SCN1A*, *SCN1B*, *SCN2A*...”). Second, customized natural language processing algorithms may be developed for specific populations. For example, in epilepsy, recent work identified surgical candidates by applying support vector machines to the text of physician notes.<sup>3</sup> Similar work in cardiology can identify individuals with heart failure.<sup>5</sup> However, such algorithms are often tailored to a single center, and may be difficult to disseminate. Third, published epilepsy classification systems<sup>6,7</sup> and ontologies<sup>8</sup> could be integrated into existing natural language processing systems.<sup>9</sup> This approach would require evaluation of baseline performance, identification of gaps (as we have done here), and ongoing review and maintenance.

The specific deficiencies in currently available vocabularies merit additional commentary. The phenotype of *SCN2A* encephalopathy<sup>10</sup> and *SCN8A* encephalopathy<sup>11</sup> are only beginning to emerge, largely based on international registry data. These phenotypes should be incorporated into the existing medical informatics infrastructure to support ongoing research into these diseases. In contrast, the epilepsies associated with *SLC13A5*<sup>12</sup> and *SYNGAP1*<sup>13</sup> are described only through small case series. Their full phenotype is not well understood.

Finally, redundancies in the epilepsy vocabularies can reduce their utility. Redundancies make it harder to use and maintain a structured vocabulary, and erode the premise that a one-to-one relationship ties each clinical concept to a term.<sup>14</sup> For example, the redundant entries for “treatment resistance” is particularly important, given that (1) research often focuses on this population, and (2) treatment-resistant patients may benefit from referral to a comprehensive epilepsy center.

**Table 1. Number of synonyms, associated terms, and entries in standardized medical vocabularies for 33 rare epilepsies**

Category	Rare epilepsy	Synonyms	Associated terms	SNOMED-CT entries	OrphaNet entries	UMLS meta-thesaurus
Epilepsies primarily defined by phenotype	Aicardi syndrome	6	12	1	1	1
	Doose syndrome	20	0	1	1	1
	Dravet syndrome	27	0	1	1	3
	Holoprosencephaly	10	5	1	1	1
	Hypothalamic hamartoma	12	13	1	1	2
	Infantile spasms	9	8	6	1	18
	Landau-Kleffner	18	0	1	1	3
	Lennox-Gastaut syndrome	8	2	1	1	11
	Migrating partial seizures of infancy	15	0	1	1	3
	Myoclonic epilepsy with ragged red fibers	20	2	1	1	3
	Neuronal ceroid lipofuscinosis	25	1	1	7	25
	Ohtahara syndrome	8	0	1	1	3
	Rasmussen syndrome	9	1	1	1	3
	Rett syndrome	17	2	1	1	5
	Sturge-Weber	16	6	1	1	2
	Tuberous sclerosis	10	41	1	4	7
	Epilepsies primarily defined by genotype	Alpers disease (POLG mutations)	25	1	1	1
Angelman syndrome		13	0	1	7	1
Epilepsy due to CDKL5 mutations		19	0	0	0	3
Dup15q		13	0	0	0	2
Fragile X syndrome		17	8	1	1	1
Glut 1 deficiency		17	0	0	0	1
KCNQ2 Encephalopathy		28	0	0	0	2
PCDH19		21	0	0	1	2
Phelan-McDermid		13	25	1	1	1
Prader-Willi		15	3	1	1	1
Ring chromosome 14		6	0	1	1	1
Ring chromosome 20		15	0	1	1	1
SCN2A encephalopathy		18	4	0	0	2
SCN8A encephalopathy		13	0	0	0	2
Epilepsy due to SLC13A5 mutations		8	0	0	0	1
Epilepsy due to SYNGAP mutations		9	17	0	0	1
Unverricht-Lundborg Disease (CSTB mutations)	14	0	0	1	1	

Table 2. Examples of deficiencies in standard medical vocabularies		
Deficiency	Clinical concept	Terms
	Vocabulary	
Redundancy	Lennox-Gastaut syndrome, not intractable Infantile spasms, Intractable	C3648103 "Lennox-Gastaut syndrome, not intractable" C3494904 "Lennox-Gastaut syndrome, non-refractory" C0154716 "Infantile spasms, with intractable epilepsy" C1827396 "Refractory infantile spasms" C2712779 "Infantile spasms, poorly controlled" C2712780 "Infantile spasms, refractory (medically)" C2712781 "Infantile spasms, pharmacologically resistant" C2712782 "Infantile spasms, treatment resistant" C3648801 "Infantile spasms with intractable epilepsy with status epilepticus" C3837134 "Infantile spasms, with intractable epilepsy, without status epilepticus" ORPHA118500 "SCN2A—sodium voltage-gated channel alpha subunit 2" C3277014 "Caused by mutation in the alpha-1 subunit of the voltage-gated type II sodium channel gene (SCN2A, 182390.0002)" C3279128 "Caused by mutation in the voltage-gated, type II sodium channel, alpha subunit (SCN2A, 182390.0008)" C3280426 "Caused by mutation in the voltage-gated sodium channel, type VIII, alpha subunit gene (SCN8A, 600702.0001)" C3553209 "Caused by mutation in the voltage-gated sodium channel, type VIII, alpha subunit gene (SCN8A, 600702.0002)" C3808213 "Caused by mutation in the synaptic Ras GTPase activating protein 1 gene (SYNGAP1, 603384.0001)" C4014623 "Caused by mutation in the solute carrier family 13 (sodium-dependent citrate transporter), member 5 gene (SLC13A5, 608305.0001)" C1852593 "Caused by mutation in the potassium voltage-gated channel, KQT-like subfamily, member 2 gene (KCNQ2, 602235.0001)" C3279124 "Caused by mutation in the voltage-gated potassium channel, KQT-like subfamily, member 2 gene (KCNQ2, 602235.0007)"
No phenotype	SCN2A encephalopathy	OrphaNet UMLS Metathesaurus
		UMLS Metathesaurus
	SCN8A encephalopathy	UMLS Metathesaurus
	Epilepsy due to SYNGAP	UMLS Metathesaurus
	Epilepsy due to SLC13A5	UMLS Metathesaurus
	KCNQ2 encephalopathy	UMLS Metathesaurus
No entry	CDKL5 mutations Dup15q syndrome Glut1 deficiency KCNQ2 encephalopathy SCN2A encephalopathy SCN8A encephalopathy Epilepsy due to SYNGAP Epilepsy due to SLC13A5 Epilepsy due to SLC13A5	SNOMED-CT SNOMED-CT SNOMED-CT SNOMED-CT SNOMED-CT SNOMED-CT SNOMED-CT SNOMED-CT SNOMED-CT SNOMED-CT OrphaNet

## Limitations

First, our selection of rare epilepsies was guided by patient advocacy groups, each of which is unified by a specific feature of the affected individuals; for some a gene, and others a clinical phenotype. This may cause some individuals to be classified in multiple groups. *PCDH19* mutations, for example, may cause the clinical phenotype of Dravet syndrome. Second, the survey respondents were largely US based, and the terms were limited to the English language. Further work is required to add non-English-language terms. Third, language evolves over time. Thus the compendium will need updating as terms arise or fade. Fourth, we selected only a subset of several hundred rare epilepsies.<sup>15</sup> Fifth, the survey response rates were low. However, we generated a robust list of terms by supplementing the 2 surveys with other sources, thereby meeting the study objectives. Finally, our work is silent on how terms are used in clinical practice. Further work is needed to describe how terms appear in clinical notes to understand their value for case finding.

## CONCLUSIONS

Epilepsy terms in structured medical vocabularies have gaps and redundancies, which should be addressed. These collected terms may help researchers and clinicians find individuals with rare epilepsies in electronic data sources. Further work is needed to evaluate their utility in identifying affected individuals.

## FUNDING

This project was supported by Centers for Disease Control and Prevention Cooperative Agreement number U01DP006089.

## ACKNOWLEDGMENTS

We are grateful to the Centers for Disease Control and Prevention for supporting this work (U01 DP006089).

## DISCLOSURES

Grinspan receives research support from the Centers for Disease Control and Prevention, the Pediatric Epilepsy Research Foundation, the BAND foundation, the Nanette Laitman Research Scholars Program, and the Patient Centered Outcomes Research Institute.

Tian has no disclosures.

Yozawitz has no disclosures.

Patricia McGoldrick serves as a speaker for Lundbeck, Supernus, and Sunovion Pharmaceuticals, and receives research support from GW Pharmaceuticals and the Centers for Disease Control and Prevention.

Wolf is on the Speaker's Bureau for Eisai, Sunovion, Lundbeck, and Supernus, and receives research support from GW Pharmaceuticals and the Centers for Disease Control and Prevention.

McDonough has no disclosures.

Nelson has no disclosures.

Ms. Hafeez receives research support from the Centers for Disease Control and Prevention and the Patient Centered Outcomes Research Institute.

Johnson has no disclosures.

Hesdorffer serves as a consultant at the Mount Sinai Medical Center Injury Prevention Center. She received research support from Centers for Disease Control and Prevention, National Institute of Neurological Disease and Stroke, Patient-Centered Outcomes Research Institute, and the Epilepsy Study Consortium. She serves as Associate Editor for *Epilepsia* and *Epilepsy & Behavior*.

We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

## DISCLAIMER

The findings and conclusions of this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

## REFERENCES

- Jette N, Beghi E, Hesdorffer D, et al. ICD coding for epilepsy: past, present, and future – a report by the International League Against Epilepsy Task Force on ICD codes in epilepsy. *Epilepsia* 2015;56:348–355.
- Kimia A, Ben-Joseph EP, Rudloe T, et al. Yield of lumbar puncture among children who present with their first complex febrile seizure. *Pediatrics* 2010;126:62–69.
- Cohen KB, Glass B, Greiner HM, et al. Methodological issues in predicting pediatric epilepsy surgery candidates through natural language processing and machine learning. *Biomed Inform Insights* 2016;8:11–18.
- Knupp KG, Coryell J, Nickels KC, et al. Response to treatment in a prospective national infantile spasms cohort. *Ann Neurol* 2016;79:475–484.
- Wang Y, Luo J, Hao S, et al. NLP based congestive heart failure case finding: a prospective analysis on statewide electronic medical records. *Int J Med Inform* 2015;84:1039–1047.
- 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;58:522–530.
- 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;58:512–521.
- Sahoo SS, Lhatoo SD, Gupta DK, et al. Epilepsy and seizure ontology: towards an epilepsy informatics infrastructure for clinical research and patient care. *J Am Med Inform Assoc* 2014;21:82–89.
- Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *J Am Med Inform Assoc* 2010;17:507–513.
- Howell KB, McMahon JM, Carvill GL, et al. SCN2A encephalopathy: a major cause of epilepsy of infancy with migrating focal seizures. *Neurology* 2015;85:958–966.
- Larsen J, Carvill GL, Gardella E, et al. The phenotypic spectrum of SCN8A encephalopathy. *Neurology* 2015;84:480–489.

12. Hardies K, de Kovel CG, Weckhuysen S, et al. Recessive mutations in SLC13A5 result in a loss of citrate transport and cause neonatal epilepsy, developmental delay and teeth hypoplasia. *Brain* 2015;138:3238–3250.
13. Parker MJ, Fryer AE, Shears DJ, et al. De novo, heterozygous, loss-of-function mutations in SYNGAP1 cause a syndromic form of intellectual disability. *Am J Med Genet A* 2015;167A:2231–2237.
14. Grimm S, Wissmann J. Elimination of redundancy in ontologies. *Semantic Web: Research and Applications* 2011;6643:260–274.
15. Ran X, Li J, Shao Q, et al. EpilepsyGene: a genetic resource for genes and mutations related to epilepsy. *Nucleic Acids Res* 2015;43:D893–D899.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

**Table S1.** The Compendium of Terms.