HHS Public Access

Author manuscript

Epilepsia. Author manuscript; available in PMC 2021 August 30.

Published in final edited form as:

Epilepsia. 2021 June ; 62(6): 1293–1305. doi:10.1111/epi.16908.

Modeling seizures in the Human Phenotype Ontology according to contemporary ILAE concepts makes big phenotypic data tractable

David Lewis-Smith[1,2], **Peter D. Galer**[3,4,5,6], **Ganna Balagura**[7], **Hugh Kearney**[8,9], **Shiva Ganesan**[3,4,5], **Mahgenn Cosico**[3,4], **Margaret O'Brien**[3,4], **Priya Vaidiswaran**[3,4], **Roland Krause**[10], **Colin A. Ellis**[4,6], **Rhys H. Thomas**[1,2], **Peter N. Robinson**[11,12], **Ingo Helbig**[3,4,5,6]

[1]Translational and Clinical Research Institute, Newcastle University, Newcastle-upon-Tyne, UK

[2]Department of Clinical Neurosciences, Royal Victoria Infirmary, Newcastle-upon-Tyne, UK

[3]Division of Neurology, Children's Hospital of Philadelphia, Philadelphia, PA, USA

[4]The Epilepsy NeuroGenetics Initiative (ENGIN), Children's Hospital of Philadelphia, Philadelphia, PA, USA

[5]Department of Biomedical and Health Informatics (DBHi), Children's Hospital of Philadelphia, Philadelphia, PA, USA

[6]Department of Neurology, University of Pennsylvania, Philadelphia, PA, USA

[7]Medical Genetics Unit, IRCSS Giannina Gaslini Institute, Genoa, Italy

[8]FutureNeuro the SFI Research Centre for Chronic and Rare Neurological Diseases, Royal College of Surgeons in Ireland, Dublin, Ireland

[9]Department of Neurology, Beaumont Hospital, Dublin, Ireland

[10]Luxembourg Centre for Systems Biomedicine, Université du Luxembourg, Esch-sur-Alzette, Luxembourg

[11]The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA

[12]Institute for Systems Genomics, University of Connecticut, Farmington, CT, USA

Abstract

Correspondence David Lewis-Smith, Translational and Clinical Research Institute, The Medical School, Newcastle University, Newcastle-upon-Tyne, NE2 4HH, UK., david.lewis-smith@newcastle.ac.uk, Ingo Helbig, Division of Neurology, Children's Hospital of Philadelphia, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA., helbigi@email.chop.edu.

CONFLICTS OF INTEREST

H.K. has received speaking honoraria from Biogen, Roche, and Teva, and was funded by Congenica. R.H.T. has received honoraria from Arvelle, Eisai, GW Pharma, Sanofi, UCB Pharma, and Zogenix, and meeting support from Bial, LivaNova, Novartis, and UCB Pharma. I.H. has served as a paid consultant for Biogen. D.L.-S., P.D.G., G.B., S.G., M.C., M.O., P.V., C.A.E., R.K., and P.N.R. have no conflicts of interest to disclose. 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.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

Objective: The clinical features of epilepsy determine how it is defined, which in turn guides management. Therefore, consideration of the fundamental clinical entities that comprise an epilepsy is essential in the study of causes, trajectories, and treatment responses. The Human Phenotype Ontology (HPO) is used widely in clinical and research genetics for concise communication and modeling of clinical features, allowing extracted data to be harmonized using logical inference. We sought to redesign the HPO seizure subontology to improve its consistency with current epileptological concepts, supporting the use of large clinical data sets in high-throughput clinical and research genomics.

Methods: We created a new HPO seizure subontology based on the 2017 International League Against Epilepsy (ILAE) Operational Classification of Seizure Types, and integrated concepts of status epilepticus, febrile, reflex, and neonatal seizures at different levels of detail. We compared the HPO seizure subontology prior to, and following, our revision, according to the information that could be inferred about the seizures of 791 individuals from three independent cohorts: 2 previously published and 150 newly recruited individuals. Each cohort's data were provided in a different format and harmonized using the two versions of the HPO.

Results: The new seizure subontology increased the number of descriptive concepts for seizures 5-fold. The number of seizure descriptors that could be annotated to the cohort increased by 40% and the total amount of information about individuals' seizures increased by 38%. The most important qualitative difference was the relationship of focal to bilateral tonic-clonic seizure to generalized-onset and focal-onset seizures.

Keywords

big data; classification; epilepsy; genetics

1 | INTRODUCTION

Information about the clinical features of research participants is essential for diagnostic interpretation of genetic findings and discovery research in epilepsy genetics.[1,2] To fully exploit these data, information must be harmonized to allow reliable comparison of individuals to diagnostic criteria, or to each other.[3] Interpretation of phenotypic data is challenging in epilepsy, particularly because the central clinical features—seizures— are diverse, dynamic, and can occur in different combinations and in different developmental and comorbidity contexts.[4–9] Crucially, the same seizure can be classified at various levels of detail according to clinical need, or to accommodate incomplete information.[5,6,8]

Even a single individual's clinical data can require harmonization between different formal classifications because of the preferences of their various health care providers, the design of study-specific research data collection forms, or because their history may span several decades.[6,8,10] For example, different sources may describe the same seizure as a focal aware non-motor seizure,[6] focal seizure,[6,9,11] aura,[8,9,12] sensory seizure,[9] or simple partial seizure.[10]

Harmonization traditionally occurs manually during data entry, dependent on the clinical expertise of the contributor completing a form specific to a particular diagnostic service or

research study. Manual harmonization of data places a large burden on clinicians and risks overly constraining the depth and quality of the phenotypic details provided, or even case ascertainment. In addition, the format of clinical data collected for one study may not map onto the variables required for another study, limiting the ease and yield of reuse. The need for rapid and accurate automated phenotypic harmonization of large multicenter data sets is increasingly important as diagnostic services and research studies grow to analyze data from tens of thousands of individuals.[1,3,13–19]

Ontologies formally model the concepts and relationships that exist in a particular expert domain.[20] The Human Phenotype Ontology (HPO) is a tool that has become the *lingua franca* for clinical and research geneticists for concise communication and harmonization of clinical features (human-phenotype-ontology.org).[21] The HPO is central to data capture in the Deciphering Developmental Disorders study [13] and the National Institutes of Health Research BioResource for the 100 000 Genomes Project,[14] and for the upcoming National Health Service England Genomic Medicine Service clinical genome sequencing service.[19] Through incremental development with contributions from experts across the breadth of clinical domains since 2008, the HPO has come to include 15 247 terms describing distinct clinical phenotypic concepts.[21] Each concept has a name, unique identifier (eg, HP:0001250), and definition, and may have synonyms (by which it may be searched for), comments, cross-referenced concepts from other databases such as OMIM (Online Mendelian Inheritance in Man, omim.org)[22] or SNOMED-CT (Systematized Nomenclature of Medicine – Clinical Terms, snomed.org),[23] and PubMed identifiers of supporting references. Current projects include translation of the HPO into multiple languages and improvements in the cross-referencing to clinical diagnostic and research resources such as Orphanet (http://www.orphadata.org) and ontologies of model organism phenotypes.[21]

Each relevant clinical feature of an individual is annotated as the most conceptually specific HPO terms necessary to describe it fully. To distinguish HPO terms from formal seizure types or semiological descriptions within the current manuscript, we refer to HPO concepts in italics by their capitalized name and HPO identifier. For example, if an individual has focal aware motor seizures, these should be coded as *Focal aware motor seizure (HP:0020217)*. Automated reasoning algorithms can then infer from the *is_a* relationship within the HPO that the individual has a type of *Focal aware seizure (HP:0002349)* and *Focal motor seizure (HP:0011153)*, and from either of these that they have a *Focal-onset seizure (HP:0007359)*, and from this a *Seizure (HP:0001250)*. Hence, from a single input datum, an individual's clinical features can be automatically described across the breadth of applicable descriptive terms, terms that span phenotypic depth. Consequently, coupled with natural language processing and thesauruses of synonyms, these algorithms can translate medical records and research data into the same HPO format and compare the clinical features of individuals to each other for research,[24–28] or to the clinical features associated with a particular genetic disorder or syndrome in a database for variant prioritization (Figure 1).[29–31] As is the case with clinical records spanning classifications, data coded as HPO terms may span multiple versions because the HPO is regularly updated to incorporate advances in phenotypic conceptualization. Unique term identifiers map to consistent concepts to facilitate translation from one HPO version to another.

Recognizing the importance of the HPO in genomics, the pertinence of seizure types to the conceptualization of epilepsies,[4,8,32] and the recent revision of seizure classification (ILAE 2017),[6] we have created a new ontology of contemporary seizure concepts for the HPO. Herein we demonstrate harmonization of three cohorts using the HPO, and compare the information captured by our new seizure subontology (release 2020–12-07) to that from a prior iteration (release 2017–12-12).

2 | METHODS

2.1 | Cohorts providing seizure data

The local epilepsy cohort for this study comprises 150 individuals with genetic generalized epilepsy (GGE, n = 65), focal epilepsy (n = 53), and developmental and epileptic encephalopathy (DEE, n = 32) recruited from Children's Hospital of Philadelphia for the Epi25 Collaborative study.[18] We used seizure data within the Epi25 format but did not use primary data of the Epi25 Collaborative. Informed consent for participation was obtained from participants themselves or, where necessary, their parents in agreement with the Declaration of Helsinki, and the study was completed per protocol using de-identified data with local approval by the Children's Hospital of Philadelphia Institutional Review Board (IRB 15–12226). Data from two further and independent cohorts with DEE were reused to demonstrate harmonization of distinct data sets. These have been published previously: Helbig et al. 2019[24] comprising 306 individuals and a subcohort from Galer et al. 2000[25] comprising 335 individuals from the Epilepsy Phenome/Genome Project (EPGP). We included only unique individuals documented as having any form of epileptic seizure.

The local cohort's data were in the form of categorical seizure variables from three Epi25 Collaborative data collection forms mapping to each of GGE, focal epilepsy, and DEE (epi-25.org/epi25-forms). Data from Helbig et al. 2019 were based on HPO release 2017–12-12 and those from Galer et al. 2020 were based on HPO release 2018–11-08.

2.2 | Harmonization of data using the HPO

For each cohort, we created a dictionary translating seizure data from their raw format into the specified version of the HPO. Once data had been translated into HPO annotations the *is_a* relationship between terms of the HPO were used to infer the other, conceptually less specific HPO seizure descriptors that must apply to the individual, as described with focal aware motor seizures above. We call this process of capturing all the terms applicable to an initial annotation, comprehensively describing a phenotypic feature at different levels of detail, propagation.[33] The set of propagated terms includes both the initial term annotated through translation using a dictionary and those inferred by logic within the ontology.

2.3 | Statistical procedures

All statistical tests and ontology-based automatic inferences were performed using the R Statistical Framework,[34] including the *tidyverse* collection of packages[35] and *ontologyIndex*.[36]

2.4 | The process of revising the HPO seizure subontology

We set up an international collaboration to build a seizure subontology for the HPO based on contemporary seizure concepts at the EMBO workshop on *Phenotyping neurological syndromes for systems genetics*, in Luxembourg, October 4–10, 2018. Our group met fortnightly by videoconference to coordinate ontology design using the online tool WebProtégé (webprotege.stanford.edu),[37] under the supervision of I.H. and R.K. (Epilepsiome Task Force of the Genetics Commission of the International League Against Epilepsy) and P.N.R. (creator of the HPO). We considered feedback and comments provided by HPO users on the HPO GitHub page (github.com/obophenotype/human-phenotype-ontology).

3 | RESULTS

3.1 | Human Phenotype Ontology–based inference harmonizes seizure descriptions across different levels of precision, semiologies, and classifications

We translated categorical seizure data in Epi25 format from 150 people in our local cohort and HPO release 2019–11-08 seizure annotations of 335 participants in EPGP (Galer 2020 cohort)[25] into HPO release 2017–12-12 (HPO 2017) terms. This harmonized their data with those already coded according to HPO 2017 from 306 individuals (Helbig 2019 cohort).[24] We compared the number of seizure descriptors that could be annotated to each of these 791 individuals prior to and following propagation using this version of the HPO (Figure 2A). We found that the median number of terms by which each individual's seizures could be described increased from 2 to 6 as a result of propagation, with a median increase of 3 and Wilcoxon signed rank test two-sided p-value <.001.

As a consequence of having harmonized data from these three independent cohorts onto the same conceptual model, we were able to compare them by the proportion of individuals annotated with particular seizure descriptions (Figure 2B). The relative abundance of seizure types in each cohort reflects the phenotypic focuses of each study. A greater proportion of individuals in the previously published DEE cohorts than in the local cohort had *Generalized myoclonic seizures (HP:0002123)* and *Generalized tonic seizures (HP:0010818)*. The Helbig 2019 cohort had a high burden of *Status epilepticus (HP:0002133)* and a prevalence of *Focal seizures (HP:0007359)* similar to the mixed cohort, but a low prevalence of *Generalized tonic-clonic seizures without focal onset (HP:0025190). Epileptic spasms (HP:0011097), Infantile spasms (HP:0012469)*, and *Atypical absences (HP:0007270)* were particularly frequently annotated in the Galer 2020 cohort. Conversely, more individuals in the mixed cohort had *Dialeptic seizures (HP:0011146)*, both *Typical absences (HP:0011147)* and *Focal seizures with impairment of awareness (HP:0002384)*.

3.2 | Creation of a new subontology of contemporary seizure concepts for the Human Phenotype Ontology

At the time that the ILAE 2017 was published, the seizure subontology of the HPO lacked the lexicon to distinguish between some seizures types yet included redundant terms and poorly integrated concepts. For example, *Generalized tonic-clonic seizures with focal onset (HP:0007334)* (a focal to bilateral tonic-clonic seizure) was implied to be a type of

generalized-onset rather than focal-onset seizure. Recognizing such existing discrepancies, we generated a new ontology of seizure types and then mapped pre-existing HPO terms across synonymous concepts where possible to allow historical HPO data to be mapped to modern concepts for reuse.

The core of the new ontology was based on the ILAE 2017, with subsequent incorporation of other classifications and commonly used concepts (Table 1). We selected preferred names for seizure concepts that describe their features concisely for visualization in ontology browsers such as hpo.jax.org. We generally favored terminology from ILAE classifications where applicable and followed precedence where not. However, we preferred the name "Bilateral tonic-clonic seizure with generalized onset" to "Generalized tonic-clonic seizure" to minimize misinterpretation by people who use the latter irrespective of onset, contrary to the ILAE 2017. HPO terms can be searched for by preferred name, identifier, or synonym. We provided ILAE and historical classification terminology as synonyms where necessary.

Although it is clinically important for classifications to include explicit terms to denote diagnostic uncertainty, such as unknown onset tonic-clonic seizure, this explicit uncertainty does not pertain to the seizure itself (the tonic-clonic seizure must either be generalized or focal in onset), but rather to the clinician's ability to classify the seizure based on the available evidence: "the term 'unknown onset' is a placeholder – not a characteristic of the seizure, but of ignorance." [6] HPO terms represent concepts of phenotypes themselves, rather than the limitations of classification in a particular instance. Accordingly, within the HPO a seizure that is classified as unknown onset tonic-clonic seizure should be mapped to an HPO term that implies that this seizure is either a *Bilateral tonic-clonic seizure with generalized onset (HP:0025190)* or *Bilateral tonic-clonic seizure with focal onset (HP:0007334)*—generalized tonic-clonic seizure or focal to bilateral tonic-clonic seizure in the ILAE 2017—rather than a third, intrinsically different seizure type, such as an absence seizure. Hence, the seizure subontology includes additional terms such as *Bilateral tonic-clonic seizure (HP:0002069)* without specifying onset. This term applies to the ILAE 2017 concepts of unknown onset tonic-clonic seizure, generalized tonic-clonic seizure, and focal to bilateral tonic-clonic seizure. The onset-specific forms have their respective onset-specific *is_a* relationships to *Generalized-onset motor seizure (HP:0032677)* and *Focal-onset seizure (HP:0007359)*, and in addition have a second *is_a* relationship to this onset-agnostic term. Consequently, individuals with unknown onset tonic-clonic seizures are annotated with the onset-agnostic term in the absence of the onset-specific terms. This arrangement of concepts is analogous to the ILAE 2017 classification of focal-onset seizures, which allows a particular instance of a focal-onset seizure to be classified by either awareness or motor versus non-motor onset independently without explicit reference to the other when necessary, rather than requiring all focal seizures to be classified according to both of these dimensions. In addition, it allows optimal integration of the proposed classification of convulsive status epilepticus.[38]

So-called "dual parentage" can also help to mitigate misclassification in diagnostic decision support tools using HPO-based algorithms. In some circumstances, distinguishing between atypical absence seizures and focal impaired awareness seizures can be difficult, and nonspecialists often mislabel focal impaired awareness seizures as "absences" or

even "petit-mal." The pre-existing HPO term *Dialeptic seizure (HP:0011146)* recognizes the semiological similarity between *Absence seizure (HP:0002121)* and *Focal impaired awareness seizure (HP:0002384)*.[12] Without a parent term in the HPO spanning electrographic onsets, these two concepts would appear no more similar to each other than *Absence seizure (HP:0002121)* does to *Bilateral tonic-clonic seizure with focal onset (HP:0007334)*: their most specific common ancestor would be *Seizure (HP:0001250)*. Thus, were an individual with a focal impaired awareness seizure to be annotated incorrectly with *Absence seizure (HP:0002121)*, an algorithm comparing this to a reference of *Focal impaired awareness seizure (HP:0002384)* can use "fuzzy matching" to recognize that a correct (yet imprecise) match would have been made, if the individual was coded with a parent of their actual annotation: *Dialeptic seizure (HP:0011146)*.

We extended this strategy of creating dual parentage independent of electrographic onset beyond its clinical scope for mitigating uncertainty and misclassification to handle those semiologies that can occur with either focal or generalized onset. Researchers may be interested in identifying individuals whose seizures share semiological features independent of electrographic onset in order to investigate shared associations and mechanisms. Dual parentage allows the ontology to recognize that a focal tonic seizure is similar to other types of focal motor seizure but also that it is more similar to a generalized tonic seizure than it is to a generalized myoclonic seizure by introducing a second parent term, *Tonic seizure (HP:0032792)*, which is indifferent to onset. Users not wishing to use these or other concepts, need not do so simply by overlooking them in their data dictionaries and sets of propagated terms.

A major component of the phenotypic bottleneck in clinical genetic diagnostics and research is data entry. We provide very specific seizure concepts and relationships that allow the ontology to infer as much information as possible about the seizures from the provision of the minimum number of input terms. For example, on provision of the single highly specific term *Focal aware cognitive seizure with deja vu/jamais vu (HP:0032883)* a reasoning algorithm can use the ontology to infer the applicability of 12 other ontological seizure descriptors capturing its focal onset, cognitive ictal manifestation, and the retention of awareness throughout in different combinations and at different degrees of granularity.

Pre-existing HPO terms were retired only if they could not be mapped among current seizure concepts or were redundant (Table S1). Otherwise, terms were integrated into the new ontology, with modification of their definition or relationships if required, or as a synonym for a new term.

3.3 | The 2020 HPO seizure subontology captures more seizure information than the 2017 version

After completing our integration of the new seizure subontology into the HPO release 2020–12-07 (HPO 2020), we compared this to the version released 2017–12-12 (HPO 2017) prior to commencement of our revision according to the number of terms by which seizures could be described and the relationships between these concepts (Table 2). The HPO 2020 contained 5-fold as many terms and 7-fold as many relationships.

Applying this, we harmonized categorical data from our local cohort and HPO 2017 data from the previously published DEE cohorts into the HPO 2020 subontology and then compared the annotated seizure data from the combined cohort of 791 individuals by HPO version (Table 2). Several new terms depict novel concepts and thus were not captured by the HPO 2017. *Focal-onset seizures (HP:0007359)* appear more common and *Generalized-onset seizures (HP:0002197)* appear less so in the HPO 2020. This is a result of revised relationships to make *Bilateral tonic-clonic seizure with focal onset (HP:0007334)* a form of *Focal-onset seizure (HP:0007359)* rather than *Generalized-onset seizure (HP:0002197)*. The increase in frequency of *Bilateral tonic-clonic seizure (HP:0002069)*, which is indifferent to onset, results from the addition of a specific term for *Convulsive status epilepticus (HP:0032660)* as a prolonged form of the former.

Phenotypic depth can be conceptualized as the amount of information provided about an individual's or cohort's phenotypes. We compared the depth of phenotypic information about our combined cohort encoded using each seizure subontology (Table 3). The number of seizure terms annotated to an individual after translation is a measure of the concepts within the raw data that can be captured by the ontology without inference. The number of terms annotated after propagation measures the total number of potential ways that the seizures in the raw data can be described, or categorized, by the ontology after inference. We found a modest detectable increase in annotations to the cohort after translation: the HPO 2020 increased the number of annotations in 46 individuals and reduced the number of annotations in 13. Compared to the HPO 2017, the HPO 2020 increased the number of propagated annotations in the combined cohort by 40%. The median increase in the number of propagated terms per individual was 50% (Table 3 and Figure 3A).

However, the number of annotations does not distinguish between common and rare features of seizures. Types of seizure that are encountered more rarely are intuitively of greater value for distinguishing an individual from the rest of the cohort than common types of seizure. In other words, the presence of a rare seizure type is more informative than that of a common seizure type. In information theory this can be quantified as information content, usually measured in bits and defined as the negative logarithm of their frequency in the cohort *(*$IC = -log_2(f)$*)*. This gives annotations of rare seizure terms more bits of information than those that are more common. Because annotations are propagated from conceptually narrow to broader parent terms, information content tends to increase as one goes from broad to more detailed terms and thus is a good measure of specificity.

Accordingly, after propagation one can measure the true frequency of each seizure term and calculate the information content of each term annotated in a cohort. The total amount of information captured by the ontology describing an individual's seizure repertoire can be calculated as the sum of the information content of each of their seizure annotations. We found that the total amount of information about the cohort's seizures increased by 38% when using HPO 2020 (Table 3 and Figure 3B). The information about each individual's seizures increased for 686 individuals and decreased for 86, demonstrating a large effect size ($|z/\sqrt{N}| = 0.80$). Although there was a small reduction in the quantity of information about 86 individuals, this was attributable to the changes described earlier, correcting the relationships of the concept of *Bilateral tonic-clonic seizure with focal onset (HP:0007334*)

and introducing *Convulsive status epilepticus (HP:0032660)* as a type of *Bilateral tonic-clonic seizure (HP:0002069)*. Hence, despite a reduction in quantity of information in a minority of individuals, the quality of information captured by HPO 2020 is greater.

4 | DISCUSSION

We describe the revision of the seizure domain within the HPO to align it with contemporary classification concepts, and the resulting effect on precise seizure description in 791 individuals. We demonstrate how seizure data can be encoded in HPO terms, automatically classifying each seizure according to multiple conceptually broader descriptors, and how it can harmonize data provided in different formats. We show that compared to a previous iteration, the new subontology increased the amount of information that could be encoded in HPO terms about an individual's seizure repertoire, thereby facilitating combination with other data sets, or analysis, based on automated inference of seizure concepts. Our newly developed seizure subontology is fully implemented in the official HPO release as of 2020–12-07 and freely available at hpo.jax.org.

The new subontology encompasses the breadth of epileptic seizure types encountered in clinical practice, taking evidence from formal classifications, consensus statements, and pertinent peer-reviewed literature that describe particular seizure types that define clinical entities. It is not a clinical classification and serves a different purpose: to facilitate automated inference of clinical seizure data for harmonized analysis or comparison. Although we did not amend classification concepts themselves, we used ontological approaches to retain as much information as possible about seizures that can be classified only incompletely. This allows semiological descriptions of seizures in medical records that do not specify electroclinical onset (for example, "she continues to have tonic seizures") to be coded without assuming the electrographic onset or losing all description of the clinical manifestation. Similarly, the subontology includes semiological terms that can code seizures without 80% certainty of onset.[6] For example, the term *Non-motor seizure (HP:0033259)* may be relevant for people with epilepsy who are unable to disclose sufficient subjective experience to confirm a specific type of focal non-motor seizure, and in whom insufficient supporting evidence can be obtained to differentiate between a focal and absence seizure, thus including the ILAE 2017 concept of an unknown onset behavior arrest seizure. The ability to code seizures informatively according to only a limited patient or witness history optimizes harmonization of data emanating from health care settings with limited access to resources such as video-electroencephalography.

The HPO continues to include older concepts such as *Epileptic aura (HP:0033348)* and *Dialeptic seizure (HP:0011146)* that some epileptologists find useful, and we have integrated these with ILAE 2017 concepts where possible.[5,8,9,12,32] This maximizes the precision of information from historical data that can be harmonized with current ILAE concepts using the HPO.

A priority when developing the new subontology was to try to make the encoding of seizure types as efficient as possible to reduce the burden of manual data entry and dictionary design. Single input terms allow harmonization to seizure types and descriptors at different

levels of detail and also unambiguously bind together features such as level of awareness and initial manifestation for a particular instance of a seizure. An exception to this is the concept of a focal to bilateral tonic-clonic seizure, which the ILAE 2017 considers a propagation pattern rather than unitary seizure type.[5] Similarly, we have represented this concept as a single independent term, *Bilateral tonic-clonic seizure with focal onset (HP:0007334)*, which avoids making the number of terms in the subontology unwieldy.

Consistent with the ILAE 2017, the structure of the new HPO subontology was based on electrographic onset and prioritization of initial manifestations. An ontology-based clinical diagnostic support tool using ILAE 2017 terms to suggest syndromic or genetic diagnoses has been developed.[39] Although the list of seizure concepts in this ontology was published, the relationships between specific concepts are not openly available for comparison with our subontology.

The semiological evolution of a focal seizure adds granularity that is important for mapping its spread across eloquent cortex and can be recorded using descriptors from the ILAE 2017,[5] ILAE glossary,[9] or a sequence of components within the semiology classification of the Four-dimensional Epilepsy Classification.[12] Beyond loss of awareness and bilateral propagation, different semiological evolutions of the same unitary focal seizure type are indistinguishable in the HPO. This may be insufficient for detailed surgical and network analyses. However, the HPO is predominantly a tool of genomics, and phenotypic analysis in epilepsy genetics has focused on seizure types rather than differences in the evolution of seizures of the same type. The Epilepsy Syndrome Seizure Ontology (version 1.03, bioportal.bioontology.org/ontologies/ESSO) and Epilepsy and Seizure Ontology (version 1.0, bioportal.bioontology.org/ontologies/EPSO) predate the ILAE 2017, and their multidimensional representation of seizures (aligned to the Four-dimensional Epilepsy Classification) may be better suited to such tasks.[8,40] Another ontological approach to semiology prior to ILAE 2017 has shown that analysis of combinations of semiological features can be used for classification of temporal lobe and extra-temporal lobe epilepies.[41] Advances in understanding and classification of focal seizures may facilitate prioritization of important sequences of semiological components for inclusion in the HPO.

Currently, as in this study, HPO terms are usually annotated manually on review of the medical records or through bulk translation of structured categorical data from research or clinical genetic clinical data forms. However, the burden of HPO coding will decrease as the availability of electronic medical record systems with integrated HPO annotation modules increases, and with advances in natural language processing.

Although this coordinated revision of the HPO seizure subontology has improved the integration of its terms and its consistency with current ILAE classifications, scope remains for refinement, including integration of concepts from other classifications to maximize its value for harmonization of non-ILAE 2017 data, such as that from historical cohorts. Continuing expert engagement (we encourage suggestions via https://github.com/obophenotype/human-phenotype-ontology/issues) will lead to incremental improvements in HPO domains pertinent to epilepsy that will benefit clinical diagnostic algorithms and research for people with epilepsy over years to come.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank the attendees who contributed to early discussions at the EMBO workshop on *Phenotyping neurological syndromes for systems genetics*, in Luxembourg, October 4–10, 2018; HPO users who provided feedback via GitHub; and the Epilepsiome Task Force of the International League Against Epilepsy for facilitating this project.

This research was funded in whole, or in part, by the Wellcome Trust [203914/Z/16/Z]. D.L.-S. was also supported by the Guarantors of Brain. I.H. was supported by The Hartwell Foundation through an Individual Biomedical Research Award and by the International League Against Epilepsy. This work was also supported by the National Institute of Neurological Disorders and Stroke (K02 NS112600), including support through the Center Without Walls on ion channel function in epilepsy ("Channelopathy-associated Research Center," U54 NS108874), the Eunice Kennedy Shriver National Institute of Child Health and Human Development (U54 HD086984) through the Intellectual and Developmental Disabilities Research Center (IDDRC) at Children's Hospital of Philadelphia and the Perelman School of Medicine, University of Pennsylvania, and by intramural funds of the Children's Hospital of Philadelphia through the Epilepsy NeuroGenetics Initiative (ENGIN). Research reported in this publication was also supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR001878.

This project was also supported in part by the Institute for Translational Medicine and Therapeutics (ITMAT) Transdisciplinary Program in Translational Medicine and Therapeutics at the Perelman School of Medicine of the University of Pennsylvania, the EuroEPINOMICS-Rare Epilepsy Syndrome (RES) Consortium by the German Research Foundation (HE5415/3-1 to I.H.) within the EuroEPINOMICS framework of the European Science Foundation, by the German Research Foundation (DFG; HE5415/-1, HE5415/6-1 to I.H) and by the DFG/FNR INTER Research Unit FOR2715 (He5415/7–1 to I.H, INTER/ DFG/17/11583046 to R.K.), and in part by a research grant from Science Foundation Ireland (SFI) under Grant Number 16/RC/3948 and co-funded under the European Regional Development Fund and by FutureNeuro industry partner Congenica, and by the European Commission (Solve-RD; 779257).

Funding information

Children's Hospital of Philadelphia; Perelman School of Medicine, University of Pennsylvania; Wellcome, Grant/ Award Number: 203914/Z/16/Z; Hartwell Foundation; Deutsche Forschungsgemeinschaft, Grant/Award Number: HE5415/3-1, HE5415/6-1, HE5415/7-1, INTER/DFG/17/11583046 and HE5415/5-1; Guarantors of Brain; European Commission, Grant/Award Number: Solve-RD and 779257; National Institute of Neurological Disorders and Stroke, Grant/Award Number: U54 NS108874; Eunice Kennedy Shriver National Institute of Child Health and Human Development, Grant/Award Number: U54 HD086984; Science Foundation Ireland, Grant/Award Number: 16/RC/394816/RC/3948; National Center for Advancing Translational Sciences, Grant/Award Number: UL1TR001878; European Regional Development Fund; International League Against Epilepsy; Congenica Ltd.

REFERENCES

1. Helbig I, Riggs ER, Barry C-A, Klein KM, Dyment D, Thaxton C, et al.The ClinGen Epilepsy Gene Curation Expert Panel—Bridging the divide between clinical domain knowledge and formal gene curation criteria. Hum Mutat. 2018;39:1476–84. [PubMed: 30311377]
2. Greenberg DA, Subaran R. Blinders, phenotype, and fashionable genetic analysis: a critical examination of the current state of epilepsy genetic studies. Epilepsia. 2011;52:1–9.
3. Lhatoo SD, Bernasconi N, Blumcke I, Braun K, Buchhalter J, Denaxas S, et al.Big data in epilepsy: clinical and research considerations. Report from the Epilepsy Big Data Task Force of the International League Against Epilepsy. Epilepsia. 2020;61:1869–83. [PubMed: 32767763]
4. Scheffer IE, Berkovic S, Capovilla G, Connolly MB, French J, Guilhoto L, et al.ILAE classification of the epilepsies: position paper of the ILAE Commission for Classification and Terminology. Epilepsia. 2017;58:512–21. [PubMed: 28276062]
5. Fisher RS, Cross JH, D'Souza C, French JA, Haut SR, Higurashi N, et al.Instruction manual for the ILAE 2017 operational classification of seizure types. Epilepsia. 2017;58:531–42. [PubMed: 28276064]

6. Fisher RS, Cross JH, French JA, Higurashi N, Hirsch E, Jansen FE, 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;58:522–30. [PubMed: 28276060]
7. Fisher RS, Acevedo C, Arzimanoglou A, Bogacz A, Cross JH, Elger CE, et al.ILAE official report: a practical clinical definition of epilepsy. Epilepsia. 2014;55:475–82. [PubMed: 24730690]
8. Lüders H, Vaca GFB, Akamatsu N, Amina S, Arzimanoglou A, Baumgartner C, et al.Classification of paroxysmal events and the four-dimensional epilepsy classification system. Epileptic Disord. 2019;21:1–29.
9. Blume WT, Luders HO, Mizrahi E, Tassinari C, Van Emde Boas W, Engel J. Glossary of descriptive terminology for ictal semiology: report of the ILAE Task Force on Classification and Terminology. Epilepsia. 2001;42:1212–8. [PubMed: 11580774]
10. Commission on Classification and Terminology of the International League Against Epilepsy. Proposal for revised clinical and electro-encephalographic classification of epileptic seizures. Epilepsia. 1981;22:489–501. [PubMed: 6790275]
11. Berg AT, Berkovic SF, Brodie MJ, Buchhalter J, Cross JH, van Emde Boas W, et al.Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005–2009. Epilepsia. 2010;51:676–85. [PubMed: 20196795]
12. Lüders H, Acharya J, Baumgartner C, Benbadis S, Bleasel A, Burgess R, et al.Semiological seizure classification. Epilepsia. 1998;39:1006–13. [PubMed: 9738682]
13. Akawi N, McRae J, Ansari M, Balasubramanian M, Blyth M, Brady AF, et al.Discovery of four recessive developmental disorders using probabilistic genotype and phenotype matching among 4,125 families. Nat Genet. 2015;47:1363–9. [PubMed: 26437029]
14. Turro E, Astle WJ, Megy K, Gräf S, Greene D, Shamardina O, et al.Whole-genome sequencing of patients with rare diseases in a national health system. Nature. 2020;583:96–102. [PubMed: 32581362]
15. Lindy AS, Stosser MB, Butler E, Downtain-Pickersgill C, Shanmugham A, Retterer K, et al.Diagnostic outcomes for genetic testing of 70 genes in 8565 patients with epilepsy and neurodevelopmental disorders. Epilepsia. 2018;59:1062–71. [PubMed: 29655203]
16. Truty R, Patil N, Sankar R, Sullivan J, Millichap J, Carvill G, et al.Possible precision medicine implications from genetic testing using combined detection of sequence and intragenic copy number variants in a large cohort with childhood epilepsy. Epilepsia Open. 2019;4:397–408. [PubMed: 31440721]
17. International League Against Epilepsy Consortium on Complex Epilepsies. Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. Nat Commun. 2018;9:5269. [PubMed: 30531953]
18. Epi25 Collaborative. Ultra-rare genetic variation in the epilepsies: a whole-exome sequencing study of 17,606 individuals. Am J Hum Genet. 2019;105:267–82. [PubMed: 31327507]
19. NHS Health Education England. Requesting whole genome sequencing: information for clinicians. Available at: https://www.genomicseducation.hee.nhs.uk/supporting-the-nhs-genomic-medicine-service/requesting-whole-genome-sequencing-information-for-clinicians/.Accessed Dec 2020.
20. Gruber TR. A translation approach to portable ontology specifications. Knowledge Acquisition. 1993;5:199–220.
21. Kohler S, Gargano M, Matentzoglu N, Carmody LC, Lewis-Smith D, Vasilevsky NA, et al.The human phenotype ontology in 2021. Nucleic Acids Res. 2021;49:D1207–17. [PubMed: 33264411]
22. McKusick-Nathans Institute of Genetic Medicine JHUB, MD. Online Mendelian Inheritance in Man, OMIM®. Available at: https://omim.org/.Accessed 10 Dec 2020.
23. SNOMED International. SNOMED Clinical Terms. Available at: https://www.snomed.org/snomed-ct/sct-worldwide.Accessed 10 Dec 2020.
24. Helbig I, Lopez-Hernandez T, Shor O, Galer P, Ganesan S, Pendziwiat M, et al.A recurrent missense variant in AP2M1 impairs clathrin-mediated endocytosis and causes developmental and epileptic encephalopathy. Am J Hum Genet. 2019;104:1060–72. [PubMed: 31104773]

25. Galer PD, Ganesan S, Lewis-Smith D, McKeown SE, Pendziwiat M, Helbig KL, et al.Semantic similarity analysis reveals robust gene-disease relationships in developmental and epileptic encephalopathies. Am J Hum Genet. 2020;107:683–97. [PubMed: 32853554]
26. Ganesan S, Galer PD, Helbig KL, McKeown SE, O'Brien M, Gonzalez AK, et al.A longitudinal footprint of genetic epilepsies using automated electronic medical record interpretation. Genet Med. 2020;22:2060–70. [PubMed: 32773773]
27. Son JH, Xie G, Yuan C, Ena L, Li Z, Goldstein A, et al.Deep phenotyping on electronic health records facilitates genetic diagnosis by clinical exomes. Am J Hum Genet. 2018;103:58–73. [PubMed: 29961570]
28. Shen F, Wang L, Liu H. Phenotypic analysis of clinical narratives using human phenotype ontology. Stud Health Technol Inform. 2017;245:581–5. [PubMed: 29295162]
29. Kohler S, Schulz MH, Krawitz P, Bauer S, Dölken S, Ott CE, et al.Clinical diagnostics in human genetics with semantic similarity searches in ontologies. Am J Hum Genet. 2009;85:457–64. [PubMed: 19800049]
30. Zemojtel T, Kohler S, Mackenroth L, Jäger M, Hecht J, Krawitz P, et al.Effective diagnosis of genetic disease by computational phenotype analysis of the disease-associated genome. Sci Transl Med. 2014;6(252):252ra123.
31. Girdea M, Dumitriu S, Fiume M, Bowdin S, Boycott KM, Chénier S, et al.PhenoTips: patient phenotyping software for clinical and research use. Hum Mutat. 2013;34:1057–65. [PubMed: 23636887]
32. Rosenow F, Akamatsu N, Bast T, Bauer S, Baumgartner C, Benbadis S, et al.Could the 2017 ILAE and the four-dimensional epilepsy classifications be merged to a new "Integrated Epilepsy Classification"?Seizure. 2020;78:31–7. [PubMed: 32155575]
33. Schulz MH, Kohler S, Bauer S, Robinson PN. Exact score distribution computation for ontological similarity searches. BMC Bioinformatics. 2011;12:441. [PubMed: 22078312]
34. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. 2017. Available at: https://www.R-project.org.
35. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al.Welcome to the tidyverse. Journal of Open Source Software. 2019;4(43):1686. doi: 10.21105/joss.01686.
36. Greene D, Richardson S, Turro E. ontologyX: a suite of R packages for working with ontological data. Bioinformatics. 2017;33:1104–6. [PubMed: 28062448]
37. Tudorache T, Nyulas C, Noy NF, Musen MA. WebProtégé: A collaborative ontology editor and knowledge acquisition tool for the web. Semant Web. 2013;4:89–99. [PubMed: 23807872]
38. Trinka E, Cock H, Hesdorffer D, Rossetti AO, Scheffer IE, Shinnar S, et al.A definition and classification of status epilepticus–Report of the ILAE Task Force on Classification of Status Epilepticus. Epilepsia. 2015;56:1515–23. [PubMed: 26336950]
39. Chiang K-L, Huang C-Y, Hsieh L-P, Chang K-P. A propositional AI system for supporting epilepsy diagnosis based on the 2017 epilepsy classification: Illustrated by Dravet syndrome. Epilepsy Behav. 2020;106:107021.
40. Sahoo SS, Lhatoo SD, Gupta DK, Cui L, Zhao M, Jayapandian C, et al.Epilepsy and seizure ontology: towards an epilepsy informatics infrastructure for clinical research and patient care. J Am Med Inform Assoc. 2014;21:82–9. [PubMed: 23686934]
41. Kassahun Y, Perrone R, De Momi E, Berghöfer E, Tassi L, Canevini MP, et al.Automatic classification of epilepsy types using ontology-based and genetics-based machine learning. Artif Intell Med. 2014;61:79–88. [PubMed: 24743020]
42. International League Against Epilepsy. EpilepsyDiagnosis.org. Available at: https://www.epilepsydiagnosis.org.Accessed 10 Dec 2020.
43. Pressler RM, Cilio MR, Mizrahi EM, Moshé SL, Nunes ML, Plouin P, et al.The ILAE classification of seizures and the epilepsies: Modification for seizures in the neonate. Position paper by the ILAE Task Force on Neonatal Seizures. Epilepsia. 2021;62(3):615–28. [PubMed: 33522601]
44. Nelson KB, Ellenberg JH. Predictors of epilepsy in children who have experienced febrile seizures. N Engl J Med. 1976;295:1029–33. [PubMed: 972656]

45. Steering Committee on Quality Improvement and Management, Subcommittee on Febrile Seizures American Academy of Pediatrics. Clinical practice guideline for the long-term management of the child with simple febrile seizures. Pediatrics. 2008;121:1281–6. [PubMed: 18519501]

46. Scheffer IE, Berkovic SF. Generalized epilepsy with febrile seizures plus. A genetic disorder with heterogeneous clinical phenotypes. Brain. 1997;120(Pt 3):479–90. [PubMed: 9126059]

47. Uemura N, Okumura A, Negoro T, Watanabe K. Clinical features of benign convulsions with mild gastroenteritis. Brain Dev. 2002;24:745–9. [PubMed: 12453597]

48. Caraballo RH, Ganez L, Santos Cde L, Espeche A, Cersósimo R, Fejerman N. Benign infantile seizures with mild gastroenteritis: study of 22 patients. Seizure. 2009;18:686–9. [PubMed: 19828335]

49. Chen B, Cheng M, Hong S, Liao S, Ma J, Li T, et al.Clinical outcome of recurrent afebrile seizures in children with benign convulsions associated with mild gastroenteritis. Seizure. 2018;60:110–4. [PubMed: 29935410]

50. Engel J Jr. A proposed diagnostic scheme for people with epileptic seizures and with epilepsy: report of the ILAE Task Force on Classification and Terminology. Epilepsia. 2001;42:796–803. [PubMed: 11422340]

Key Points

- The Human Phenotype Ontology (HPO) formally models the relationships between clinical concepts to facilitate concise communication, harmonization, and algorithmic inference
- Our new HPO seizure subontology increases the amount of information that can be encoded and made available for automated reasoning, making clinical data tractable in big data sets

Significance:

We have generated a detailed contemporary conceptual map for harmonization of clinical seizure data, implemented in the official 2020–12-07 HPO release and freely available at hpo.jax.org. This will help to overcome the phenotypic bottleneck in genomics, facilitate reuse of valuable data, and ultimately improve diagnostics and precision treatment of the epilepsies.

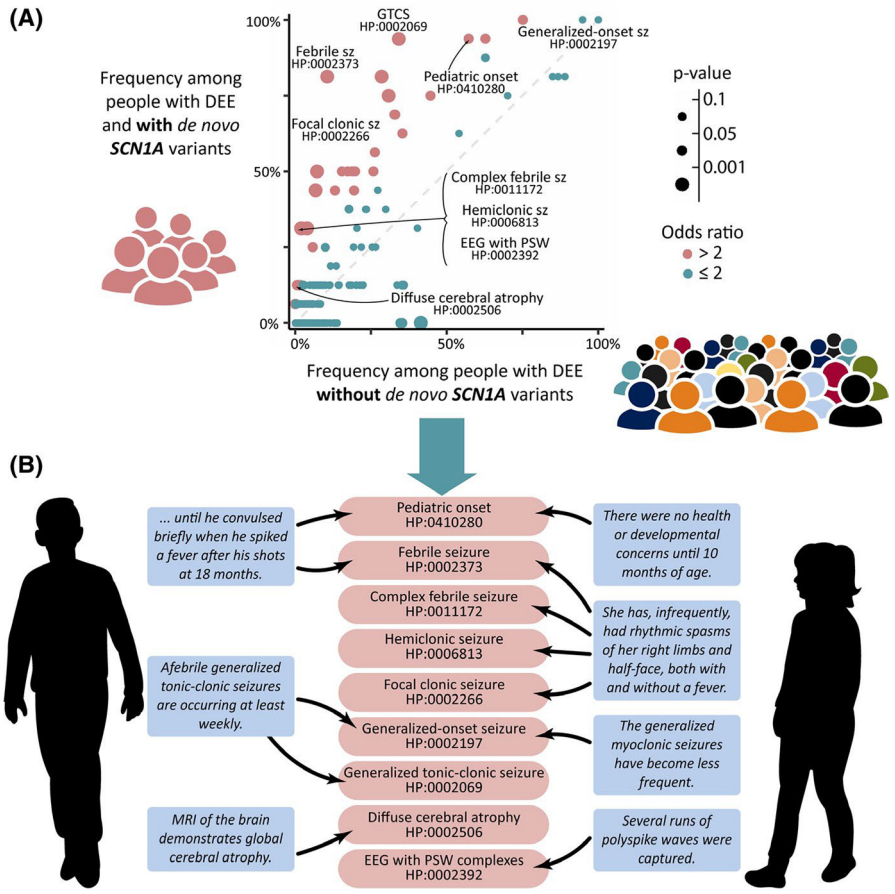

FIGURE 1.

(A) The Human Phenotype Ontology (HPO) can harmonize phenotypic data from large cohorts to identify phenotypic associations with a particular genetic (or other) categorical factor, in this case de novo variants in *SCN1A*. Data from Galer et al. 2020;[25] *p-values* and odds ratio obtained using Fisher's exact test. (B) Once the set of HPO terms associated with this form of epilepsy is known, the clinical features of patients can be translated into HPO terms to assess how closely they match phenotypically. GTCS, generalized tonic-clonic seizure; PSW, polyspike-wave; sz, seizure

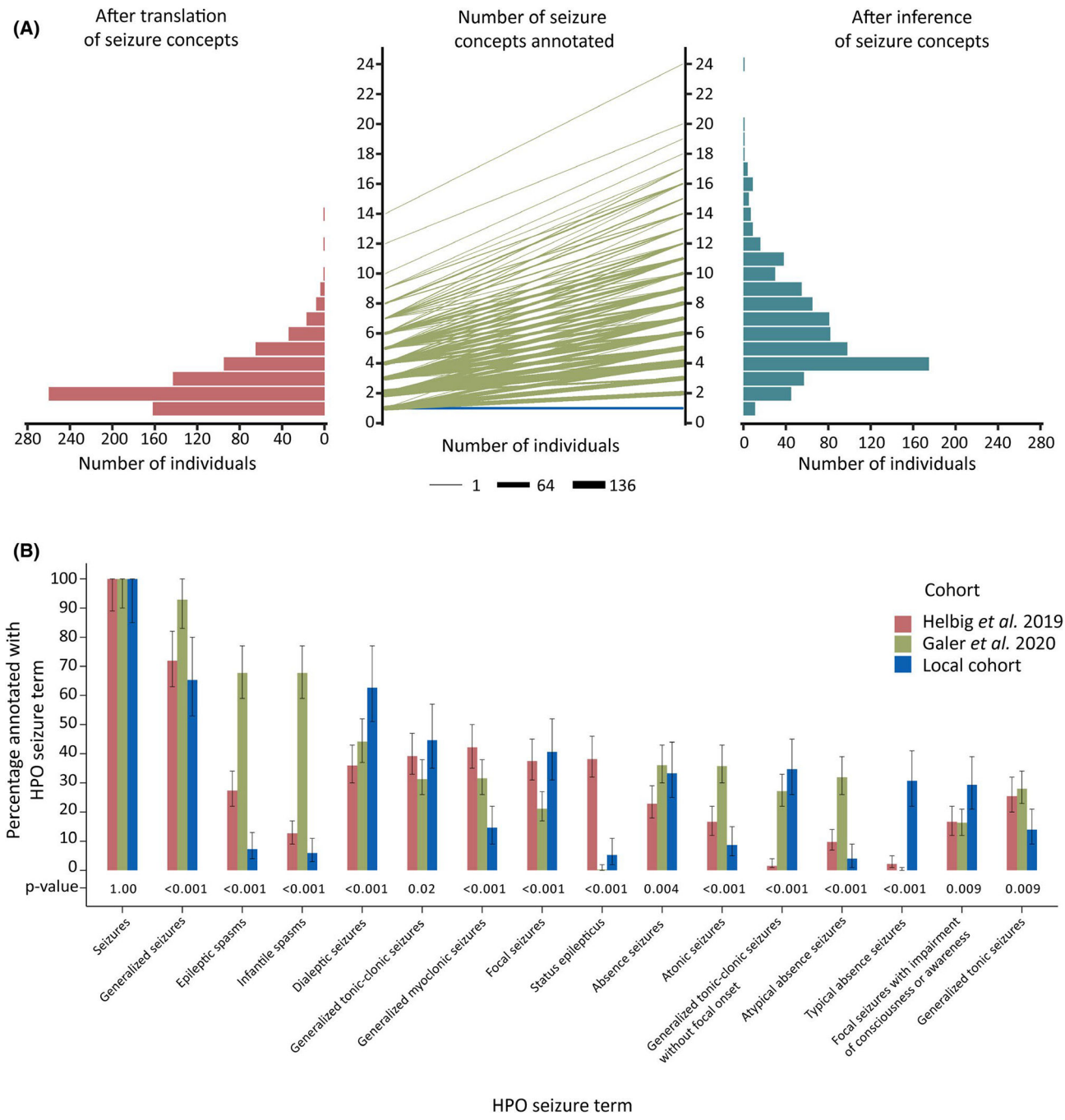

FIGURE 2.

(A) The increase in the number of seizure concepts by which data collected from 791 individuals can be described as a result of automated inference using the Human Phenotype Ontology (HPO) version release 2017–12-12. Green lines indicate an increase in the number of seizure descriptors annotated by inference after translation. The blue line indicates eight individuals whose only HPO term translated from input data was *Seizure (HP:0001250)*, which is insufficient to infer descriptions. (B) Comparison of the three independent cohorts according to the percentage of individuals annotated with each of 16 HPO seizure terms,

comprising the 10 most common seizure concepts in each cohort. Note that in this HPO version, *Generalized tonic-clonic seizure (HP:0002069)* is indifferent to onset. *p-values* are two-sided from Fisher's exact test with Holm-Bonferroni adjustment for 16 comparisons; 95% confidence intervals were calculated from the Poisson distribution

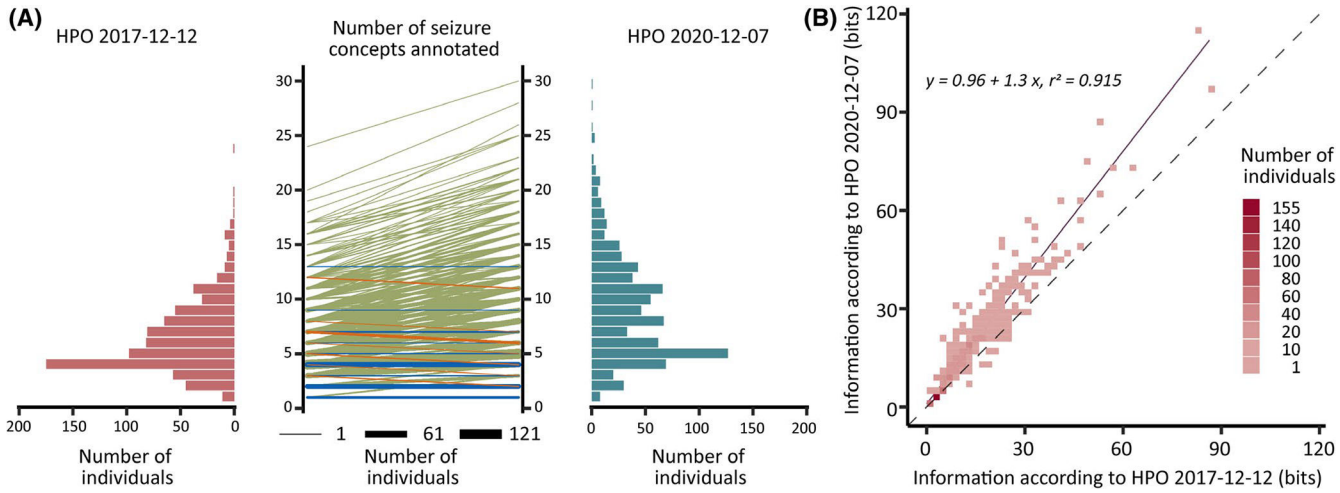

FIGURE 3.

Comparison of the phenotypic information encoded from the same data from 791 individuals using the two seizure subontologies. (A) Green lines indicate an increase, red lines a decrease, and blue lines no difference in the number of seizure descriptors annotated after inference. (B) The total amount of information encoded about each individual's seizure types according to each Human Phenotype Ontology (HPO) seizure subontology. Numbers of individuals shown correspond to those falling within 2 bit by 2 bit bins, the dashed gray line indicates equality, and the regression line is shown in purple

TABLE 1

Classifications and definitions consulted to generate the HPO 2020–12-07 seizure subontology

Domain	Reference
Seizure types and semiological descriptions	Operational classification of seizure types by the ILAE: Position Paper of the ILAE Commission for Classification and Terminology. Fisher et al. 2017[6] Instruction manual for the ILAE 2017 operational classification of seizure types. Fisher et al. 2017[5] EpilepsyDiagnosis.org[42] Glossary of descriptive terminology for ictal semiology: Report of the ILAE Task Force on Classification and Terminology. Blume et al. 2001[9] Semiological seizure classification. Lüders et al. 1998[12] Classification of paroxysmal events and the four-dimensional epilepsy classification system. Lüders et al. 2019[8]
Status epilepticus	A definition and classification of status epilepticus: Report of the ILAE Task Force on Classification of Status Epilepticus. Trinka et al. 2015[38]
Neonatal seizure types	The ILAE classification of seizures and the epilepsies: Modification for seizures in the neonate. Position Paper by the ILAE Task Force on Neonatal Seizures. Pressler et al. 2021[43]
Febrile and mild gastroenteritis-associated seizures	Predictors of epilepsy in children who have experienced febrile seizures. Nelson and Ellenberg 1976[44] Febrile seizures: Clinical practice guideline for the long-term management of the child with simple febrile seizures. American Academy of Pediatrics 2008[45] Generalized epilepsy with febrile seizures plus. A genetic disorder with heterogeneous clinical phenotypes. Scheffer and Berkovic 1997[46] Clinical features of benign convulsions with mild gastroenteritis. Uemura et al. 2002[47] Benign infantile seizures with mild gastroenteritis: Study of 22 patients. Caraballo et al. 2009[48] Clinical outcome of recurrent afebrile seizures in children with benign convulsions associated with mild gastroenteritis. Chen et al. 2018[49]
Reflex seizures	A proposed diagnostic scheme for people with epileptic seizures and with epilepsy: Report of the ILAE Task Force on Classification and Terminology. Engel Jr 2001[50]

Epilepsia. Author manuscript; available in PMC 2021 August 30.

TABLE 2

Description of the HPO 2017 and the HPO 2020 seizure subontologies and the frequency of the 32 most frequent seizure concepts in the combined cohort of 791 individuals according to each. Terms with frequency "NA" were not conceptualized

	HPO subontology	
Metadata of the seizure subontologies	**2017 release**	**2020 release**
Number of seizure terms	71	347
Number of is_*a* relationships between two seizures terms	73	518
Number of is_*a* relationships between a seizure and non-seizure term	3	4
Frequency of different seizure concepts captured from the cohort of 791 individuals		
Term identifier	**Seizure descriptor from 2020 release**	**Frequency in 2017 release**	**Frequency in 2020 release**
HP:0001250	Seizure	1	1
HP:0020219	Motor seizure	NA	0.8078
HP:0002197	Generalized-onset seizure	0.7952	0.7320
HP:0032677	Generalized-onset motor seizure	NA	0.4994
HP:0011146	Dialeptic seizure	0.4450	0.4450
HP:0011097	Epileptic spasm	0.4071	0.4071
HP:0002069	Bilateral tonic-clonic seizure	0.3692	0.3704
HP:0012469	Infantile spasms	0.3477	0.3477
HP:0033259	Non-motor seizure	NA	0.3350
HP:0007359	Focal-onset seizure	0.3123	0.3211
HP:0032794	Myoclonic seizure	NA	0.3148
HP:0002123	Generalized myoclonic seizure	0.3249	0.3135
HP:0002121	Generalized non-motor (absence) seizure	0.3047	0.3047
HP:0032792	Tonic seizure	NA	0.2617
HP:0010818	Generalized tonic seizure	0.2440	0.2440
HP:0010819	Atonic seizure	0.2326	0.2326
HP:0002384	Focal impaired awareness seizure	0.1896	0.1896
HP:0025190	Bilateral tonic-clonic seizure with generalized onset	0.1871	0.1871
HP:0007270	Atypical absence seizure	0.1808	0.1795
HP:0002133	Status epilepticus	0.1593	0.1593

HP:0011153	Focal motor seizure	0.1315	0.1315	
HP:0032892	Infection-related seizure	NA	0.1302	
HP:0032894	Seizure precipitated by febrile infection	NA	0.1302	
HP:0002373	Febrile seizure (within age range 3 months – 6 years)	0.1226	0.1226	
HP:0020221	Clonic seizure	NA	0.0872	
HP:0007334	Bilateral tonic-clonic seizure with focal onset	0.0670	0.0670	
HP:0011147	Typical absence seizure	0.0670	0.0670	
HP:0002266	Focal clonic seizure	0.0632	0.0632	
HP:0011170	Generalized myoclonic-atonic seizure	0.0493	0.0493	
HP:0032679	Focal non-motor seizure	NA	0.0442	
HP:0011175	Focal motor seizure with version	0.0341	0.0341	
HP:0002349	Focal aware seizure	0.0152	0.0303	

TABLE 3

Comparison of the amount of phenotypic information captured about the seizures of 791 individuals using the HPO 2017 and the HPO 2020

Comparison	HPO subontology		Difference (% change or median, *p-value*)
	2017 release	2020 release	
Number of input annotations			
Entire cohort (total)	2275	2308	Increase by 1.5%
Per individual (median)	2	2	0, $p < .001$
Number of annotations after propagation (inference)			
Entire cohort (total)	5029	7063	Increase by 40%
Per individual (median)	6	8	3, $p < .001$
Total information content (bits)			
Entire cohort (total)	8280	11 403	Increase by 38%
Per individual (median)	7.90	11.43	3.15, $p < .001$

Differences between total cohort measures are given as percentages of the HPO 2017 figures; paired differences between individuals are given as median absolute differences; *p-values* are two-sided from a Wilcoxon signed-rank test.