Received: 25 October 2019 | Revised: 7 July 2020 | Accepted: 8 July 2020

DOI: 10.1111/epi.16633

SPECIAL REPORT

Epilepsia®

Big data in epilepsy: Clinical and research considerations. Report from the Epilepsy Big Data Task Force of the International League Against Epilepsy

Samden D. Lhatoo[1] | **Neda Bernasconi**[2] | **Ingmar Blumcke**[3] | **Kees Braun**[4] | **Jeffrey Buchhalter**[5] | **Spiros Denaxas**[6] | **Aristea Galanopoulou**[7] | **Colin Josephson**[8] | **Katja Kobow**[3] | **Daniel Lowenstein**[9] | **Philippe Ryvlin**[10] | **Andreas Schulze-Bonhage**[11] | **Satya S. Sahoo**[12] | **Maria Thom**[13] | **David Thurman**[14] | **Greg Worrell**[15] | **Guo-Qiang Zhang**[1] | **Samuel Wiebe**[8]

[1]University of Texas Health Sciences Center at Houston, Houston, Texas

[2]Neuroimaging of Epilepsy Laboratory, Montreal Neurological Institute and Hospital, McGill University, Montreal, Quebec, Canada

[3]Friedrich-Alexander University Erlangen-Nürnberg, University Hospital Erlangen, Erlangen, Germany

[4]Department of Child Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, the Netherlands

[5]Department of Neurology, St Joseph's Hospital and Medical Center, Phoenix, Arizona

[6]Institute of Health Informatics, University College London, London, UK

[7]Saul Korey Department of Neurology, Dominick P. Purpura Department of Neuroscience, Albert Einstein College of Medicine, Bronx, New York

[8]Department of Clinical Neurosciences, University of Calgary, Calgary, Canada

[9]Department of Neurology, University of California, San Francisco, San Francisco, California

[10]Department of Neurosciences, University of Lausanne, Lausanne, Switzerland

[11]Epilepsy Center, University Hospital Freiberg, Freiberg, Germany

[12]Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH, USA

[13]Institute of Neurology, University College London, London, UK

[14]Emory University, Atlanta, Georgia

[15]Department of Neurology, Mayo Clinic, Rochester, Minnesota

Correspondence
Samden Lhatoo, Texas Comprehensive Epilepsy Program, University of Texas Health Sciences Center, Houston, TX 77030.
Email: slhatoo@aol.com

Abstract

Epilepsy is a heterogeneous condition with disparate etiologies and phenotypic and genotypic characteristics. Clinical and research aspects are accordingly varied, ranging from epidemiological to molecular, spanning clinical trials and outcomes, gene and drug discovery, imaging, electroencephalography, pathology, epilepsy surgery, digital technologies, and numerous others. Epilepsy data are collected in the terabytes and petabytes, pushing the limits of current capabilities. Modern computing firepower and advances in machine and deep learning, pioneered in other diseases, open up exciting possibilities for epilepsy too. However, without carefully designed approaches

[Correction added on Aug 30, 2020, after first online publication: Affiliation of author "Satya S. Sahoo" was changed and the rest of the affiliations were renumbered accordingly.]

to acquiring, standardizing, curating, and making available such data, there is a risk of failure. Thus, careful construction of relevant ontologies, with intimate stakeholder inputs, provides the requisite scaffolding for more ambitious big data undertakings, such as an epilepsy data commons. In this review, we assess the clinical and research epilepsy landscapes in the big data arena, current challenges, and future directions, and make the case for a systematic approach to epilepsy big data.

KEYWORDS

big data, epilepsy, epilepsy informatics, epilepsy ontology

1 | INTRODUCTION

Big data is an intuitive, colloquially used term[1]—first in business, and latterly in science and health care. MetaGroup's 2014 definition describes big data as high-volume, high-velocity, and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision-making. In addition to these "3 V's,"[2] the fourth "V" of data veracity is particularly pertinent, because suspect data produce suspect conclusions (Figure 1). In an era of unprecedented collaboration and resource pooling, big data's promise is both inviting and challenging, especially in epilepsy due to its inherent heterogeneity and the vast array of scientific disciplines it involves. This review examines big data aspects specific to epilepsy and describes the current state of the art as well as future directions.

Key Points

- Epilepsy data are multimodal and require big data principles for proper handling
- Big data approaches provide both clinical and research opportunities
- Structured and principled approaches to epilepsy big data are necessary for maximum impact

2 | THE MEANING OF BIG DATA

In epilepsy, a plethora of disparate data drive variety (phenotype, genotype, video-electroencephalographic [EEG], extracranial and intracranial physiological signal, structural

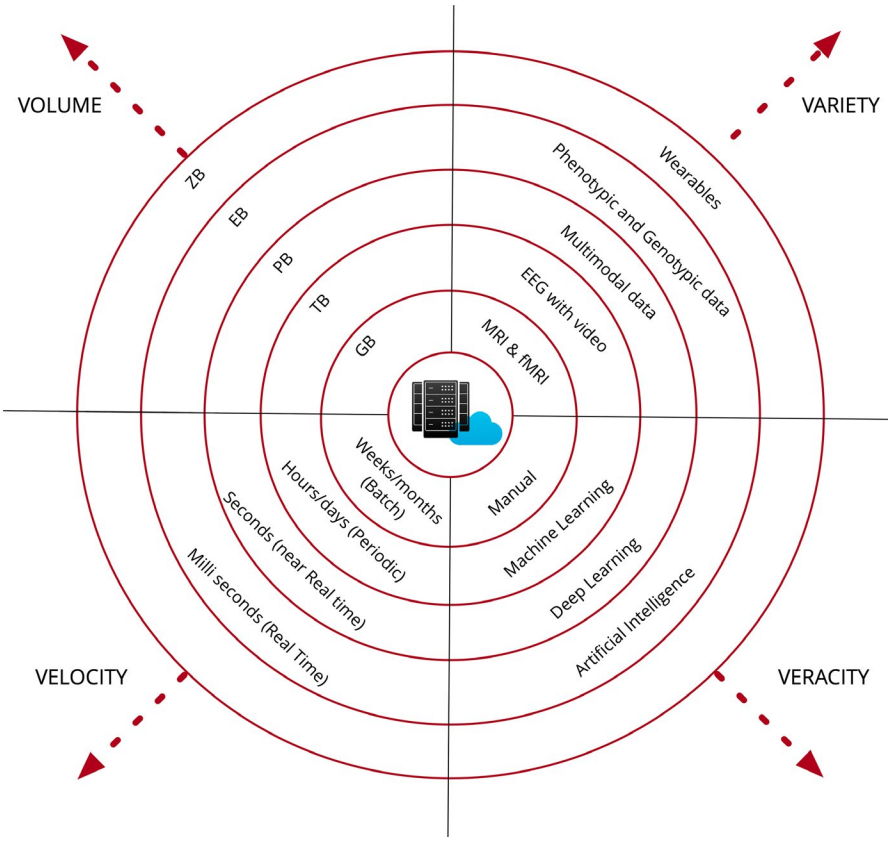

FIGURE 1 Big data "4 V's" of volume, variety, velocity, and veracity as they apply to the epilepsy domain (Lhatoo 2017). EB, Exabyte; EEG, electroencephalography; fMRI, functional MRI; GB, Gigabyte; MRI, magnetic resonance imaging; PB, Petabyte; TB, Terabyte; ZB, Zettabyte

and functional imaging, metabolomics, wearables), which in turn drives volume (currently terabytes), challenges veracity (data acquisition, standardization), and highlights current deficiencies in managing data generated with high velocity (Figure 1). However, as big data becomes increasingly commoditized, it may be more helpful to think of "big data" as a frame of mind. This allows perception of the scientific landscape in a more ambitious data scale, and enables bigger questions. To scale and accelerate scientific progress, a big data frame of mind drives research in three new directions.

2.1 | Collaboration

The past century has seen tremendous progress in health care delivery and research, using traditional approaches, whether pharmaceutical randomized controlled trials or basic science. The field, however, is poised to enter a new era enabling unprecedented collaborative possibilities. The sudden unexpected death in epilepsy (SUDEP) exemplar illustrates big data opportunities. Here, identifying a sufficiently powered cohort of patients requires meticulous, prospective follow-up of large at-risk cohorts in the epilepsy monitoring unit (EMU).[3] Multiple EMUs collaborate, generating several hundred gigabytes of data per patient. Epi25 (genetics; http://epi-25.org) and ENIGMA (neuroimaging; http://enigma.ini.usc.edu/ongoing/enigma-epilepsy) are examples of similar highly successful, domain-specific collaborations that allow for validation of promising ideas at an accelerated pace.

2.2 | Data resource infrastructure (data commons)

New challenges emerge as soon as data are put in the hands of a community of investigators, rather than in individual laboratories. The 2010 Institute of Medicine report,[4]"Elements of an Integrated National Strategy to Accelerate Research and Product Development for Rare Diseases," recommended a national strategy that "shares research resources and infrastructure to make good and efficient use of scarce funding, expertise, data, and biological specimens." This recommendation is especially relevant to the epilepsy community and underscores the need to make data findable, accessible, interoperable, and reusable (the FAIR principles).[5]

The National Institutes of Health (NIH) Data Commons (or Commons) aims at a shared virtual space for digital objects to be found, stored, and commented and computed upon by the scientific community. Four components are considered integral parts of the Commons: a computing resource for accessing and processing digital objects; a "digital object compliance model" that enables digital objects to be FAIR; datasets adhering to the compliance model; and data access services. An Epilepsy Commons, following the sleep exemplar,[6] would greatly facilitate epilepsy research, enhance the efficiency of resource utilization, and ensure the rigor and reproducibility of epilepsy research.

2.3 | New modes for interacting with waveform data

A big data vision requires new modes of management of datasets generated from epilepsy research. One such opportunity is the signal data format called the "file wall" challenge. Because simply expanding storage or adding computing power will not cope with the volume and data complexity, data organization challenges must be addressed. In existing cloud storage/processing systems, large signal datasets are typically stored as identified unstructured "blobs." Traditional distributed file systems present "file wall" barriers that make data access, transmission, processing, and analysis more difficult. There is an immediate need in epilepsy for research into ontology-driven, cloud-based data representation and management methods. Several initiatives have begun to address multimodal interaction, including the Brain Imaging Data Structure (https://bids.neuroimaging.io) and Fast Healthcare Interoperability Resources (https://www.hl7.org), which it may be possible to extrapolate to epilepsy. A significant challenge with waveform data currently lies in a neurophysiological data format that allows interoperability of video-EEG data. MEF3 (Multiscale Electrophysiology Format version 3) is one format that has been proposed as a universal standard, addressing an urgent need in both research and clinical domains, allowing easy exchange of data.[7] There is also an increasing shift toward machine learning, deep learning, and artificial intelligence in epilepsy big data, particularly in its EEG aspects.[8] These include EEG spike detection, and automated surface EEG and intracranial EEG seizure detection, some of which already have impactful clinical applications, such as with closed loop responsive neural stimulation.

2.4 | Data safety and privacy

Increasing data innovation creates inevitable conflict with informational privacy. The application of fair information practice principles is paramount; these include individual control, transparency, respect for context, security, access, accuracy, focused collection, and accountability. These are challenged by big data paradigms, and careful attention to existing regulations (from institutional review boards to national regulations, eg, the Health Insurance Portability and Accountability Act of 1996 [HIPAA]) is essential. For example, there is increasing recognition that data deidentification

and anonymization are not the silver bullets to data privacy issues they once were. Nowhere is this more relevant than in genomics arena data, where subject "reidentification" is a major concern,[9] and as yet incompletely addressed in the legal arena.[10]

3 | GETTING THE BASICS RIGHT: EPILEPSY ONTOLOGY, CLASSIFICATION, AND COMMON DATA ELEMENTS

Biomedical ontologies are widely used to achieve three data management objectives: (1) management of multidimensional knowledge, (2) integration of disparate data, and (3) automated reasoning for decision support and knowledge discovery.[11] For example, the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) is one of the most comprehensive and widely used biomedical ontologies that serves as the de facto standard for encoding clinical information in electronic health record (EHR) systems. SNOMED CT together with several other biomedical ontologies, such as the Human Phenotype Ontology and RxNorm Gene Ontology, are predicted to have a central role in clinical big data applications, including data-driven classification of diseases as part of the precision medicine initiative.[12] Formal languages for modeling are used, such as the Ontology Web Language (OWL), based on description logic.[13] OWL-modeled ontologies accurately model a domain of interest and support the use of automated tools called "reasoners" to discover implicit knowledge from a big data repository. An ontology structure is inherently a knowledge graph that can be used to infer implicit knowledge from large datasets.

The generic informatics definition of ontology is "a formal specification of terms in the domain (eg, epilepsy) and relations among them (eg, focal impaired awareness seizure is a type of focal seizure, which is a type of seizure)." One key feature is the multiaxial categorization of a concept, which enables it to be applied to various organizations of information within that domain. This significantly simplifies tasks of organizing and exploring large datasets. Ontology-supported queries together with reasoning tools allow users to explore hitherto undiscovered knowledge about how data relate to each other.

Multiple International League Against Epilepsy (ILAE) classification and terminology task forces and work groups have been challenged by the complexities inherent in characterization of seizures/epilepsies. The factors that require consideration have included locus of onset, location, propagation, age of onset, age of remission, prognosis, EEG and neuroimaging characteristics, biological mechanism, etiologies, comorbidities, and functional impairment. Even as the need for multiple axes has been appreciated, several "structural" issues contribute to the classification problem. These include (1) lack of standardized definitions for core terms (concepts), (2) the evidence base to determine how these various factors relate to each other, (3) the reality that information available in one setting may not be available in another, and (4) use of uniaxial classification hierarchies that by their structure do not easily allow a term in one branch of a classification to be incorporated into another. The need for "assembly" of various concepts in different hierarchies (eg, based upon age of onset) for different purposes (eg, clinical care, epidemiology) has been recognized,[14,15] but implementation has been impeded due to lack of a computable modeling framework such as OWL.

Currently, there are three public domain seizure/epilepsy ontologies available, hosted by the NIH-funded BioPortal site (URL). We discuss two. The Epilepsy Syndrome Seizure Ontology was the first attempt to harmonize existing seizure/epilepsy classifications to allow common definitions, but most importantly to enable organization by available information. Starting with semiology, demographic and testing (EEG, genetic testing) factors can be added if known (Figure 2). Another robust example is the Epilepsy and Seizure Ontology (EpSO), which models multidimensional information, including seizures, seizure features, and etiology (including gene IDs mapped to gene ontology,[16] and drug information). EpSO (Figure 3) is currently used in a variety of informatics tools.[17–26]

Like all disruptive technologies, the concepts related to ontologies and OWL will require education and familiarity prior to adoption in epilepsy that can lead to meaningful basic and clinical experiments. Challenges for adoption include (1) consensus on meanings of terms (concepts), (2) evidence to provide connections (relationships) between terms, and (3) the necessity for the epilepsy community to embrace the reality that knowledge related to epilepsy is truly multidimensional, requiring harmonization of its various components. Just as gene sequencing is now considered essential for diagnosis, so will ontologies be for understanding epilepsy.

4 | BIG DATA IN INTRACRANIAL EEG RESEARCH

Intracranial EEG (IEEG) recordings provide a window onto mechanisms of brain function with unique temporospatial resolution. For study of electrophysiological activity, access to databases providing IEEG recordings is critical. Recordings with high numbers of multicontact electrodes with an extended frequency range for local field potentials and single neurons require databases that integrate data at rates of up to a terabyte per day. Useful analyses require valid, extended metadata (eg, behavioral), electrophysiological

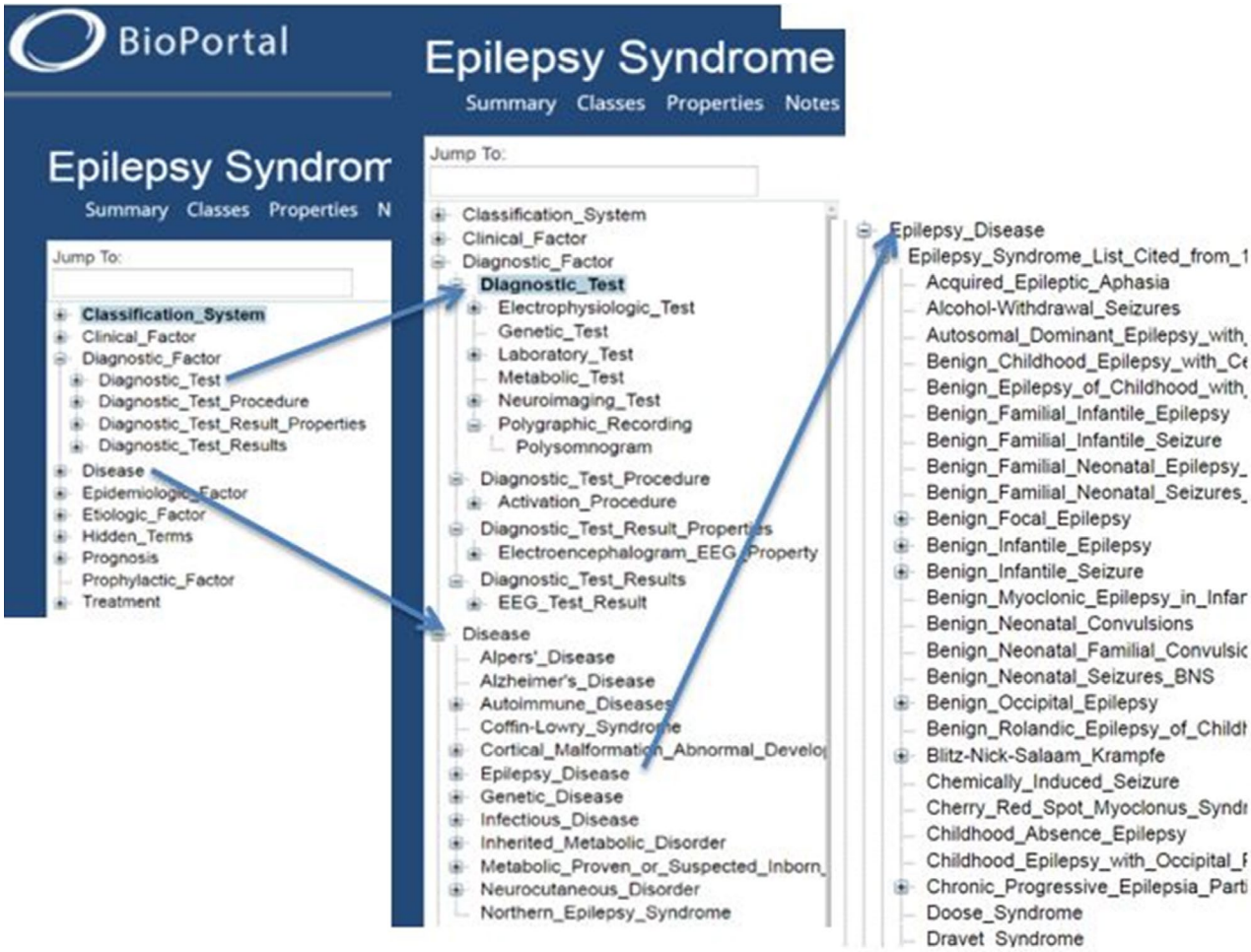

FIGURE 2 A screenshot of the Epilepsy Syndrome Seizure Ontology, showing organization of terms to which testing (electroencephalography, imaging, genetic testing) factors can be added

data, and electrode positions in relation to individual brain structures.

The seizure prediction field highlights the need for large datasets; early results based on limited datasets[27,28] suggested that forecasting was infeasible, whereas later, extensive dataset studies provided the key to success.[29] These efforts[30–33] provide examples of databases that use long-term recordings based on local servers (EU) or commercial cloud services (USA) (Table 1).

Increasingly, centers are offering shared clinical and research datasets.[34,35] Furthermore, multinational collaborations are establishing databases on specific EEG datasets, one example being the F-TRACT database joining information from intracranial evoked corticocortical potentials from 25 European centers to build a network of functional connectivity.[36]

Whereas some technological challenges inherent to data format variability appear solvable, others remain unsolved, including different country-based data safety standards, willingness to share not only data but also algorithms for revalidation, and sustainable funding for maintenance and development beyond the project level.[33]

5 | NEUROPATHOLOGICAL REPOSITORIES AND BIG DATA

Biorepositories should not only offer long-term storage of human brain and blood samples, but also catalog a standardized set of data describing the patient's clinical history and phenotype. Up-to-date patient consent and ethical approval must be in place to allow sharing of biological samples and research data. However, most tertiary epilepsy centers in Europe will select <50 patients per year for epilepsy surgery.[37,38] Very large tertiary centers may operate on >150 patients per year.

The European Epilepsy Brain Bank consortium (EEBB) was established in 2006 as a virtual database aiming to standardize histopathological reporting of specimens obtained during epilepsy surgery and epileptogenic brain lesions.[39] To date, EEBB has collected diagnoses from 9523 children and adults from 36 epilepsy surgery centers in 12 European countries and prompted the ILAE to develop an International Classification of Focal Cortical Dysplasia[40] and of hippocampal sclerosis[41] as well as international

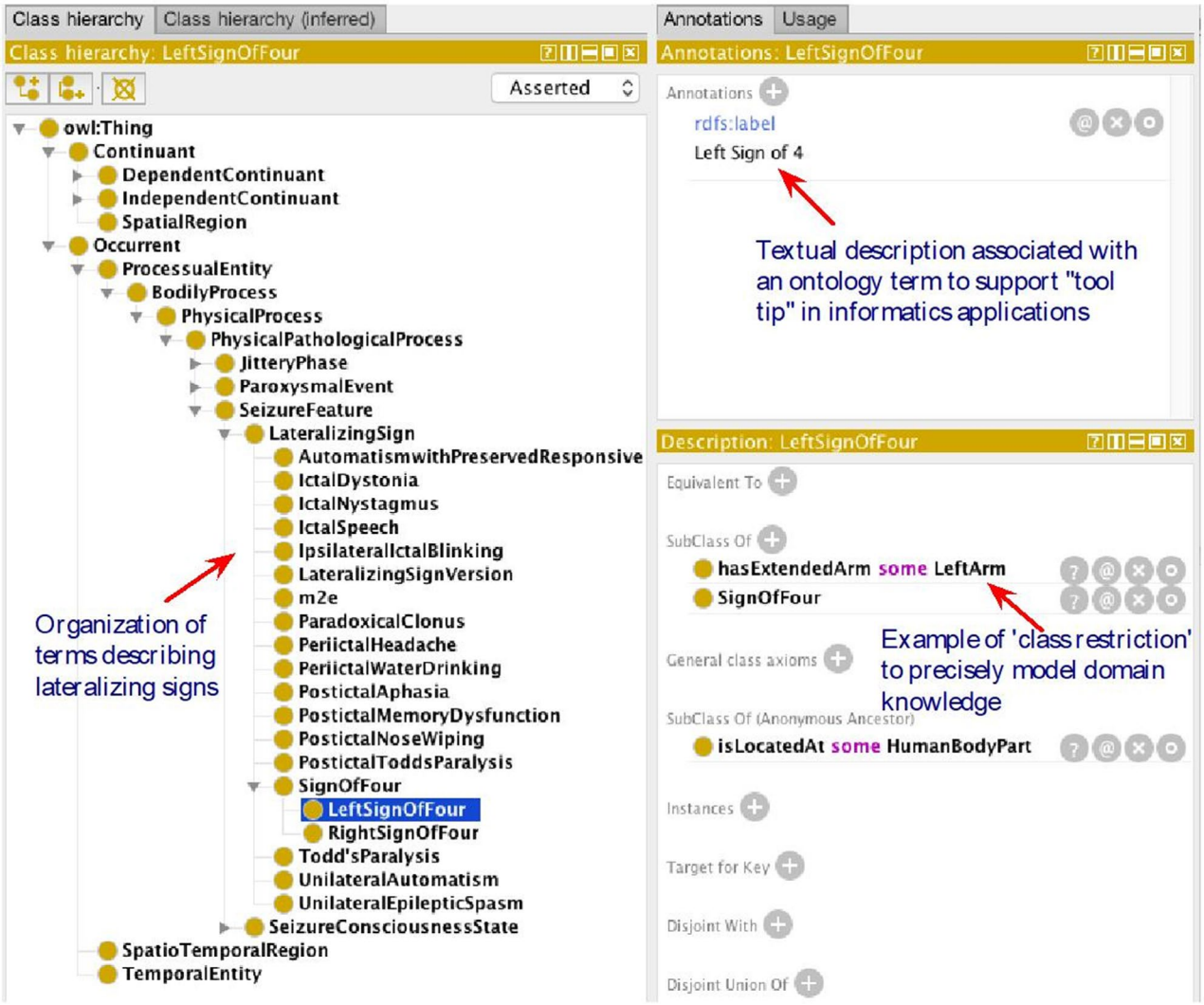

FIGURE 3 A screenshot of the Epilepsy and Seizure Ontology, showing the detailed class hierarchy of lateralizing signs associated with seizures and the use of class-level restrictions to model information at a fine level of granularity. For example, "left sign of 4" is related to left extended arm, and it is a subcategory of "sign of four"

recommendations for the histopathological workup of epilepsy surgery specimens.[42] Disease classifications help to define a disease and also to inform surgical patient management, where decision-making may still rely on small series and randomized trials for difficult-to-treat focal epilepsies.[43,44]

Under the direction of the 7th European Health Framework Program (FP7), EEBB was promoted in 2014 as a biorepository to support clinical trials in epilepsy surgery. The EU framework program Horizon2020 promoted EEBB as a European neuropathology reference center. The biorepository includes long-term storage of paraffin-embedded and fresh frozen brain samples, and matched blood samples with minimum deidentified clinical data.

Limitations of this dataset include its retrospective nature, inability to predict surgical outcomes, medication use, and EEG and magnetic resonance imaging (MRI) biomarkers. The database is encrypted into the web-based SecuTrial platform, which does not allow connection with biomedical "OMICs." International collaboration and data sharing are restricted to partners of the FP7 consortium. These limitations endorse an ILAE mandate to promote international collaboration for big data analysis of human epilepsy brain samples and harmonization of written patient consent, ethical review, and material transfer agreements.

6 | EPILEPSY IMAGING AND BIG DATA

MRI allows for comprehensive analysis of the whole brain. This technique can provide detailed descriptions of structure, function, metabolism, and networks, from single cells to systems in models and humans, which parallels the scope of the genome and other biomes.

In epilepsy, a leading example of collaboration is the Enhancing Neuroimaging Genetics Through Meta-Analysis (ENIGMA) project based on meta- and mega-analyses. Launched in 2015, ENIGMA-Epilepsy uses harmonized quality checks rather than sharing datasets across an international consortium of >50 sites, and it has already produced insightful results.[45] Notably, this approach bypasses challenges related to institutional ethical approvals and the need for high-throughput computing. Other sharing strategies currently not used in epilepsy rely on sharing raw imaging data in repositories, such as the NIH-funded Neuroimaging Informatics Tools and Resources Clearinghouse (www.nitrc.org), a suite of services including a registry, an image repository, and a cloud-based environment.

Beside ethical concerns, data science faces other challenges. For example, variability in clinical assessments, missing data, and variable study cohorts may confound disease severity with study site. A major technical challenge relates to variability in MRI hardware and acquisition, image quality, and parameters, which may lead to differences in data scaling and noise. A mitigation strategy would be the adoption of the newly proposed HARNESS-MRI (Harmonized Neuroimaging of Epilepsy Structural Sequences) protocol entailing a set of acquisitions readily available on most MRI scanners.[46] It is crucial to establish and abide by MRI quality standards and to append standardized phenotypical descriptors based on the most recent classifications. Disease models using multicentric MRI data require consideration of confounders related to unbalanced patient-to-control ratios and measurement variance across sites. A solution would be to first develop models based on a given dataset and test generalizability to others, rather than prioritizing pooling data across sites.

7 | EPILEPSY GENETICS AND BIG DATA

Collective genomic efforts in the past decade have capitalized on access to high-throughput genomic analysis and next generation sequencing, as well as team science, which fosters collaboration and scaling of studies that would never be possible by single investigators. A critical element has been the collection of detailed phenotypic information in addition to other data (eg, EEG records, source documentation from medical records).

The NIH-funded Epilepsy Phenome/Genome Project (EPGP), is an international, collaborative study that collected detailed phenotype data and DNA samples on >4100 subjects (and family members) with specific forms of epilepsy. The repository, with millions of data points, continues to be fully operational for follow-up studies.[47,48] The EPGP cohort, along with European and Australian datasets, were critical to the success of the NIH-funded Epilepsy Center Without Walls entitled "Epi4K: Gene Discovery in 4000 Epilepsy Genomes," which utilized exome sequencing to identify new de novo variants causing epileptic encephalopathy and Lennox-Gastaut syndrome, and ultrarare genetic variation in common forms of epilepsy, among other findings.[49,50] Other examples include the large datasets created through the ILAE Consortium on Complex Epilepsies,[51] the EuroEPINOMICS-RES Consortium,[52] and the massive, international effort currently underway, entitled Epi25, that has a goal of sequencing 25 000 epilepsy exomes (see http://epi-25.org/). Encompassing epilepsy and more, dbGaP (Database of Genotype and Phenotype) is an NIH-maintained database of datasets, which archives and distributes results of studies investigating genotype-phenotype interactions.[53]

The explosion in knowledge of genetic variants associated with human epilepsies and in molecular targets for ictogenesis, epileptogenesis, and comorbidogenesis constitutes rich libraries to search for candidates in molecular pathogenesis and therapy. In animal studies, concerted efforts to generate knockout[54] or conditional knockout mice[55,56,57] and C57 embryonic stem cells have resulted in at least 17 000 knockout mice. Characterization of the phenotype of these genetic models[58] will be offered in a public access database. Although this is invaluable for investigators, significant enhancements will be needed to include endpoints relevant to clinical epilepsy research.

The ILAE/American Epilepsy Society Joint Translational Task Force, in collaboration with NIH National Institute of Neurological Disorders and Stroke, is generating preclinical common data elements for epilepsies and comorbidities to facilitate data input from multiple laboratories into big databases,[59] generate accepted classifications and terminologies for video-EEG studies and seizures in rodents,[60–64] and perform systematic analyses of preclinical studies.[65] An aim is to optimize these products on platforms that could be used in big databases to enhance epilepsy research, including translation from preclinical to clinical arenas.

8 | "RELATIVELY BIG DATA" FROM MULTICENTER COHORT STUDIES

Sharing and assembling results from single-center cohorts through collaborative research efforts have increasingly been applied to address questions in epilepsy. Although these approaches do not comply strictly with definitions of big data, a 2016 *Lancet Neurology* "round-up" comment used the term big data when referring to multicenter cohorts of 1450 surgical patients, 446 children with absences, and as few as 14 neonates.[66] The cooperation of many dedicated physicians

from different centers was considered the key attribute to new research in epilepsy.

For rare epilepsies, for small or heterogeneous populations, and for new, unproven therapies, pooling results from small cohort studies can yield relatively "big data," overcoming their lack of statistical power and sources of bias. Multicenter cohort studies or meta-analyses have merits, even when the quantity of data is much smaller than in big data. In Europe, the European Reference Network EpiCare (http://epi-care.eu/) serves to facilitate multinational collaborations in rare and complex epilepsies. For focused research questions, limited but specific and well-structured "clean" data can be systematically collected retrospectively from available multicenter patient data, enabling multivariate analyses and prediction modeling with sufficient statistical power. For example, the TimeToStop cohort study allowed investigation of whether early withdrawal of antiepileptic drugs (AEDs) after pediatric epilepsy surgery is safe.[67] One researcher collecting data systematically from 766 children across 15 collaborating centers ensured high-quality data and found that early AED withdrawal did not affect seizure outcome.[67]

Alternatively, published single-center cohort studies can be meta-analyzed with either aggregate or individual participant data (IPD). Although IPD is the gold standard for clinical research synthesis,[68] retrieval rate is suboptimal,[69] and it is still underutilized. Nevertheless, IPD meta-analyses have increasingly been applied in epilepsy, for example, to determine AED monotherapy efficacy,[70] to predict seizure outcome after epilepsy surgery in tuberous sclerosis complex,[71] to calculate the chance of seizure recurrence after a first febrile seizure[72] and the risk of ictal asystole in epilepsy,[73] and to produce a prediction model that calculates the individualized risk of seizure relapse following AED withdrawal.[74] Big data research, multicenter cohort studies, and IPD meta-analyses can be considered complementary approaches.

9 | EHRS AND EPILEPSY BIG DATA

Data from EHRs, generated during routine clinical care across multiple settings,[75] are increasingly being linked[76] and used for translational research and for large-scale observational research. EHR data can be classified into three main types[77]:

1. Structured data are mostly used for administrative purposes and annotated using controlled clinical terminologies and statistical classification systems such as SNOMED CT, International Classification of Diseases–10, Logical Observation Identifiers Names and Codes, and RxNorm. These typically include information such as diagnoses, prescriptions, and surgical procedures and interventions during inpatient and outpatient care.
2. Unstructured data are recorded as raw text and typically include a patient's medical history and clinicians' observations and findings.
3. Binary data traditionally include data from imaging procedures and increasingly from personal health care wearable devices or smartphones.

Research platforms such as CALIBER[78] link EHR data from primary care, hospital care, and mortality and offer researchers high-resolution longitudinal data on chronic and acute conditions at a population level. Raw EHR data, however, suffer from multiple challenges[79] and require substantial preprocessing before they can be research-ready for statistical analysis, a process known as *phenotyping*.[80,81]

In the context of epilepsy research, curated EHR data offer substantial advantages compared to traditional methods: (1) EHRs have large sample sizes, enabling scientists to gain accurate measures of incidence and prevalence in populations; (2) linked EHRs can be utilized to quantify health care utilization and costs associated with epilepsy, its treatment, and its comorbidities; (3) high-resolution EHR data can help identify and validate novel epilepsy subtypes using unsupervised machine learning, which can lead to personalized medicine approaches; (4) longitudinal EHR data can help characterize valid phenotypes of disease progression, with unique etiological and prognostic features.

10 | EPIDEMIOLOGICAL OPPORTUNITIES IN BIG DATA

Clinical epidemiological data in the zettabyte (10^{21} bytes or 1 trillion gigabytes)[82] range, representing a large portion of the population of interest, are particularly apt for big data applications, and can detect small but clinically meaningful effect sizes. Such unprecedented statistical power can confer immense precision. However, unless the veracity of the data is ensured, narrow confidence intervals can be erroneously misconstrued as accuracy.[83]

10.1 | Data sources and validation

Large clinical data repositories need not be population-based, but they must be representative of the population of interest. Typically, population-based sources include administrative health data, EHR data, national health surveys, and vital statistics (Table 2).[11] Non–population-based platforms include national and regional clinical registries[84–86] and pooled individual patient data from clinical trials.[69] Frequently, there

TABLE 1 Large multicenter EEG databases

Characteristic	Database name	
	EPILEPSIAE	**IEEG**
Funding	EU	NIH
Year data provided	2013	2016/ongoing
Data providers, n	3	>100
Datasets, n	275	>1000
Content	Scalp and intracranial long-term EEG recordings in epilepsy patients	Intracranial EEG recordings in epilepsy patients and animals
Raw data	Interictal + ictal EEG, 3D MRI	Interictal + ictal EEG, imaging
Metadata	Clinical information	Clinical information
EEG annotations	By human experts	By automated routines

Abbreviations: 3D, three-dimensional; EEG, electroencephalographic; EPILEPSIAE, Evolving Platform for Improving Living Expectation of Patients Suffering from IctAl Events; IEEG, International Epilepsy Electrophysiology Portal; MRI, magnetic resonance imaging; NIH, National Institutes of Health.

TABLE 2 Population-based sources of big data

Data source	Sources	Advantages	Disadvantages
AHRs	• National claims data • Regional claims data	• Often large and population-based • Longitudinal • No selection or recall bias • Cost-effective	• Not collected for clinical purposes • Different methods of coding • Lack granularity • Lack of AHR phenotypes
EHRs	• Clinical Practice Research Datalink • The Health Improvement Network database	• Often large and frequently population-based • Longitudinal • Coded by physicians for clinical purposes • More granular outcome data	• Different methods of coding • Different proprietary EHR software • Patients may move • Lack of EHR phenotypes
Health survey data	• Canadian Community Health Survey • US National Health Interview Survey • WHO World Health Survey	• Population-based • Standardized data collection • Patient-reported outcomes • Comparability across populations of interest	• Self-report may limit diagnostic accuracy • Response rates may vary • Often cross-sectional • Resource intensive
Vital statistics data	• Statistics Canada • UK Office of National Statistics • US National Vital Statistics System	• Population-based • Longitudinal data • Standardized data collection • Cause-specific death	• Inconsistent coding • Variable quality • Incomplete information • Delays in reporting

Abbreviations: AHR, administrative health record; EHR, electronic health record; WHO, World Health Organization.

is a tradeoff between granularity and quality. For instance, clinical registries and pooled trial data are frequently rich and detailed but can have selection bias (trials), and may lack consistency and completeness (voluntary registries).

Valid case definitions for epilepsy now exist for administrative health records[87] and EHRs.[88,89] Although the reported sensitivity and specificity are high (>80%-85%),[87,88] these are often context-specific, and their utility should be quantified when used in different datasets.[90] Likewise, outcome measures, if not validated, can lead to spurious conclusions.[91] Thus, all conditions of interest must be treated with methodological rigor lest results become irrevocably skewed due to misclassification bias.

Analyses from validated epidemiological data have yielded remarkable insights into the incidence[92,93] and prevalence[92,94] of epilepsy, the comorbid profile of epilepsy,[94] the bidirectionality of depression and epilepsy,[95,96] the association between epilepsy and autism,[97,98] and debunking the spurious putative link between AED use and suicide.[99,100] Overall mortality[101,102] and SUDEP[101,103] have been studied with large quantities of EHR and administrative data. Initial endeavors at applying machine learning to big data for the purposes of predicting epilepsy outcomes appear promising.[104] Finally, population-based surveillance of health care access, utilization, and costs is now possible using these large data sources.[105,106]

If it is used prudently, hitherto unforeseen opportunities exist for cost-effective and statistically powerful investigations into the epidemiology of epilepsy using big data. These include disease surveillance, identification of new somatic and psychiatric conditions, precision medicine targets, and health outcome and health care use assessments. However, many of these hypothesis-generating studies will require validation through other methods.

11 | DIGITAL HEALTH, WEARABLE TECHNOLOGY, AND BIG DATA

More than 5 billion human beings currently use a mobile phone, the majority of whom share information on social media. In 2017, close to half of the population used at least one connected care technology to monitor health indicators (Future Health Index 2017, Philips). US hospitals and insurance providers are rapidly transitioning to digital mobile health (mHealth). Concurrently, partnerships between large information technology companies (eg, Apple and Google) and hospitals are developing novel health care ecosystems. Hence, there is huge potential for information technologies to generate big data in any medical field, provided relevant data can be captured and shared.

Most interestingly, connected devices with sensors have already proved useful for detecting generalized tonic-clonic seizures (GTCS),[107–112] and some of these have received US Food and Drug Administration approval (eg, Embrace, Brain Sentinel).[113,114] Others are being developed from nonmedical mainstream wearables (eg, Apple watch), offering potential for major dissemination within the epilepsy population.[113]

GTCS can be reliably detected with sensors measuring body movements during the clonic phase (wrist accelerometer or pressure bed sensor),[115–118] surface electromyography of the arm during the tonic phase,[119–121] and changes in electrodermal activity.[114,122] Changes in heart rate, which can be extracted through photoplethysmography,[123,124] can also be used to detect various seizure types.[125,126] Recently developed multimodal seizure detectors are likely to prove more sensitive and specific than any single-sensor technology.[127–129]

Detecting GTCS with connected devices can enable more precise evaluation of seizure frequency and treatment optimization, as well as timely intervention by triggering alarms. This might help reduce seizure-related adverse events and fatalities.[130,131] Data gathered through connected devices may also provide biomarkers of various comorbidities, for example, AED side effects, risk of SUDEP, and seizure-modulating environmental and internal factors. As already demonstrated by the North-American Brain Initiative and European Human Brain Project, mHealth data are considered complementary to -omics in precision medicine.

This rapidly developing field faces important challenges, including data privacy and merging information from connected devices, usually stored on generic clouds, to EHRs managed by health providers. The research potential of mHealth technologies in epilepsy should be facilitated by the unique collaboration developed by the ILAE and International Bureau for Epilepsy.

12 | BIG DATA FUNDING INITIATIVES IN MEDICAL RESEARCH

Recognizing the transformative opportunities provided by big data, government funding agencies around the world have launched strategic programmatic investment to accelerate big data research as highlighted below.

12.1 | European Union

Horizon 2020 is the leading EU research and innovation program (with nearly €80 billion of funding available over 2014-2020). Funding opportunities under the Horizon 2020 Program (https://ec.europa.eu/programmes/horizon2020/) include topics such as "Big data technologies and extreme-scale analytics" (ICT-12-2018-2020) that are focused on data management, data processing, deep analytics, data protection, data visualization, and user experience.

12.2 | United States

In the USA, the flagship program has been BD2K (Big Data to Knowledge), a trans-NIH initiative launched in 2013 to support the research and development of innovative and transformative approaches and tools to maximize and accelerate the integration of big data and data science into biomedical research. New data science strategy (https://datascience.nih.gov) includes Precision Medicine and the BRAIN initiative, with the National Library of Medicine as the nexus of data science resources. The US National Science Foundation (NSF), another federal funding agency of US, has a standing program on Critical Techniques, Technologies and Methodologies for Advancing Foundations and Applications of Big Data Sciences and Engineering. "Harnessing data for 21st century science and engineering" has been identified as one of the "10 Big Ideas for Future NSF Investments" (https://www.nsf.gov/about/congress/reports/nsf_big_ideas.pdf).

12.3 | China

Big data funding initiatives in China have been represented by investment in precision medicine. Precision medicine is a part of the Chinese government's 5-year plan for 2016-2020 as it works to prioritize genomics to drive better health care outcomes. Investment through programs such as the National High-Tech R&D Program has led its investigator community and infrastructure to the forefront of research involving methodologies of genomics and proteomics, with rapid development in technologies for molecular imaging, drug targets, and big data.

13 | THE ROLE OF THE ILAE TASK FORCE FOR BIG DATA

The digital revolution has opened up tremendous opportunities for large-scale collaborative epilepsy clinical care and research. The ILAE Big Data Task Force comprises epilepsy clinicians and researchers who are engaged in disparate epilepsy research domains and have an interest in large-scale collaborative clinical and research endeavors. Their role is to review past and current epilepsy big data efforts, and over the tenure of the task force, to recommend guidelines and advice that help deliver high-impact big data research that is directly relevant to patient care, while providing a framework of reference for navigating privacy, legal, and ethical issues surrounding such enterprises. Specific laws, such as the General Data Protection Regulation in Europe and HIPAA in the USA, govern such endeavors.

14 | CONCLUSION

Big data, data sharing, and high-performance computing are poised to restructure the way we deliver health care and do research. An overview of these themes in epilepsy shows great opportunities and important challenges. Successful instances where these processes produce new and important knowledge are beginning to emerge and should be strengthened. To harness the full potential of big data will require attention to policies and procedures, secure environments, data quality standards, data platforms, and data science models that can accommodate the large volume and variety of data that are characteristic of epilepsy. Most importantly, it will require a new way of thinking about the evidence derived from big data in its application to health care and research. The greatest chances of success will lie in large national and international multicenter collaborations.

CONFLICT OF INTEREST

None of the authors has any conflict 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.

DISCLAIMER

This report was written by experts selected by the ILAE and was approved for publication by the ILAE. Opinions expressed by the authors, however, do not necessarily represent the policy or position of the ILAE.

ORCID

Samden D. Lhatoo https://orcid.org/0000-0002-0260-3855
Aristea Galanopoulou https://orcid.org/0000-0002-0472-2903
Colin Josephson https://orcid.org/0000-0001-7052-1651
Katja Kobow https://orcid.org/0000-0002-0074-2480
Philippe Ryvlin https://orcid.org/0000-0001-7775-6576
Maria Thom https://orcid.org/0000-0001-7712-2629
Samuel Wiebe https://orcid.org/0000-0002-1061-9099

REFERENCES

1. Lohr S. The origins of "big data": an etymological detective story. New York Times. February 1, 2013.
2. Laney D. 3D Data Management: Controlling Data Volume, Velocity, and Variety, Technical report, META Group. 2001
3. Lhatoo S, Noebels J, Whittemore V. Sudden unexpected death in epilepsy: identifying risk and preventing mortality. Epilepsia. 2015;56:1700–6.
4. Boat TF, Adamson PC, Asbury C. Elements of an Integrated National Strategy to Accelerate Research and Product Development for Rare Diseases. Washington, DC: Institute of Medicine; 2010.
5. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data. 2016;3:160018.
6. Zhang GQ, Cui L, Mueller R, et al. The National Sleep Research Resource: towards a sleep data commons. J Am Med Inform Assoc. 2018;25:1351–8.
7. Stead M, Halford JJ. Proposal for a standard format for neurophysiology data recording and exchange. J Clin Neurophysiol. 2016;33:403–13.
8. Kearney H, Byrne S, Cavalleri GL, Delanty N. Tackling epilepsy with high-definition precision medicine: a review. JAMA Neurol. 2019;76(9):1109.
9. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying personal genomes by surname inference. Science. 2013;339:321–4.
10. Clayton EW, Evans BJ, Hazel JW, Rothstein MA. The law of genetic privacy: applications, implications, and limitations. J Law Biosci. 2019;6:1–36.
11. Bodenreider O. Biomedical ontologies in action: role in knowledge management, data integration and decision support. Yearb Med Inform. 2008;67–79.
12. Haendel MA, Chute CG, Robinson PN. Classification, ontology, and precision medicine. N Engl J Med. 2018;379:1452–62.
13. Hitzler P, Krotszch M, Parsia B, Patel-Schneider PF, Rudolph S. OWL 2 Web ontology language primer. W3C recommendation. 2009.

14. Berg AT, Berkovic SF, Brodie MJ, 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.
15. Berg AT, Cross JH. Towards a modern classification of the epilepsies? Lancet Neurol. 2010;9:459–61.
16. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–9.
17. 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;21:82–9.
18. Cui L, Bozorgi A, Lhatoo SD, Zhang GQ, Sahoo SS. EpiDEA: extracting structured epilepsy and seizure information from patient discharge summaries for cohort identification. AMIA Annu Symp Proc. 2012;2012:1191–200.
19. Cui L, Huang Y, Tao S, Lhatoo SD, Zhang GQ. ODaCCI: ontology-guided data curation for multisite clinical research data integration in the NINDS Center for SUDEP Research. AMIA Annu Symp Proc. 2016;2016:441–50.
20. Cui L, Sahoo SS, Lhatoo SD, et al. Complex epilepsy phenotype extraction from narrative clinical discharge summaries. J Biomed Inform. 2014;51:272–9.
21. Jayapandian C, Wei A, Ramesh P, et al. A scalable neuroinformatics data flow for electrophysiological signals using MapReduce. Front Neuroinform. 2015;9:4.
22. Jayapandian CP, Chen CH, Bozorgi A, Lhatoo SD, Zhang GQ, Sahoo SS. Cloudwave: distributed processing of "big data" from electrophysiological recordings for epilepsy clinical research using Hadoop. AMIA Annu Symp Proc. 2013;2013:691–700.
23. Sahoo SS, Jayapandian C, Garg G, et al. Heart beats in the cloud: distributed analysis of electrophysiological 'Big Data' using cloud computing for epilepsy clinical research. J Am Med Inform Assoc. 2014;21:263–71.
24. Sahoo SS, Wei A, Valdez J, et al. NeuroPigPen: a scalable toolkit for processing electrophysiological signal data in neuroscience applications using Apache Pig. Front Neuroinform. 2016;10:18.
25. Sahoo SS, Zhao M, Luo L, et al. OPIC: ontology-driven patient information capturing system for epilepsy. AMIA Annu Symp Proc. 2012;2012:799–808.
26. Zhang GQ, Cui L, Lhatoo S, Schuele SU, Sahoo SS. MEDCIS: multi-modality epilepsy data capture and integration system. AMIA Annu Symp Proc. 2014;2014:1248–57.
27. Mormann F, Kreuz T, Rieke C, et al. On the predictability of epileptic seizures. Clin Neurophysiol. 2005;116:569–87.
28. Schulze-Bonhage A, Feldwisch-Drentrup H, Ihle M. The role of high-quality EEG databases in the improvement and assessment of seizure prediction methods. Epilepsy Behav. 2011;22(Suppl 1):S88–93.
29. Cook MJ, O'Brien TJ, Berkovic SF, et al. Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study. Lancet Neurol. 2013;12:563–71.
30. Ihle M, Feldwisch-Drentrup H, Teixeira CA, et al. EPILEPSIAE—a European epilepsy database. Comput Methods Programs Biomed. 2012;106:127–38.
31. Klatt J, Feldwisch-Drentrup H, Ihle M, et al. The EPILEPSIAE database: an extensive electroencephalography database of epilepsy patients. Epilepsia. 2012;53:1669–76.
32. Kini LG, Davis KA, Wagenaar JB. Data integration: combined imaging and electrophysiology data in the cloud. Neuroimage. 2016;124:1175–81.
33. Wagenaar JB, Worrell GA, Ives Z, Dumpelmann M, Litt B, Schulze-Bonhage A. Collaborating and sharing data in epilepsy research. J Clin Neurophysiol. 2015;32:235–9.
34. O'Regan ME, Brown JK. Abnormalities in cardiac and respiratory function observed during seizures in childhood. Dev Med Child Neurol. 2005;47:4–9.
35. EEG motor movement/imagery dataset. [cited 2020 July 28]. Available at: https://archive.physionet.org/pn4/eegmmidb/.
36. Trebaul L, Deman P, Tuysenge V, et al. Probabilistic functional tractography of the human cortex revisited. Neuroimage. 2018;181:414–29.
37. Lhatoo SD, Solomon JK, McEvoy AW, Kitchen ND, Shorvon SD, Sander JW. A prospective study of the requirement for and the provision of epilepsy surgery in the United Kingdom. Epilepsia. 2003;44:673–6.
38. Cloppenborg T, May TW, Blümcke I, et al. Trends in epilepsy surgery: stable surgical numbers despite increasing presurgical volumes. J Neurol Neurosurg Psychiatry. 2016;87:1322–9.
39. Blumcke I, Spreafico R, Haaker G, et al. Histopathological findings in brain tissue obtained during epilepsy surgery. N Engl J Med. 2017;377:1648–56.
40. Blumcke I, Spreafico R. An international consensus classification for focal cortical dysplasias. Lancet Neurol. 2011;10:26–7.
41. Blumcke I, Cross JH, Spreafico R. The international consensus classification for hippocampal sclerosis: an important step towards accurate prognosis. Lancet Neurol. 2013;12:844–6.
42. Blümcke I, Aronica E, Miyata H, et al. International recommendation for a comprehensive neuropathologic workup of epilepsy surgery brain tissue: a consensus task force report from the ILAE Commission on Diagnostic Methods. Epilepsia. 2016;57:348–58.
43. Dwivedi R, Ramanujam B, Chandra PS, et al. Surgery for drug-resistant epilepsy in children. N Engl J Med. 2017;377:1639–47.
44. Wiebe S, Blume WT, Girvin JP, Eliasziw M. A randomized, controlled trial of surgery for temporal-lobe epilepsy. N Engl J Med. 2001;345:311–8.
45. Whelan CD, Altmann A, Botía JA, et al. Structural brain abnormalities in the common epilepsies assessed in a worldwide ENIGMA study. Brain. 2018;141:391–408.
46. Bernasconi A, Cendes F, Theodore WH, et al. Recommendations for the use of structural magnetic resonance imaging in the care of patients with epilepsy: a consensus report from the International League Against Epilepsy Neuroimaging Task Force. Epilepsia. 2019;60:1054–68.
47. McGovern K, Stillman N, McKenna K, et al. The epilepsy phenome/genome project. Clin Trials. 2013;10:568–86.
48. Nesbitt G, McKenna K, Mays V, et al. The Epilepsy Phenome/Genome Project (EPGP) informatics platform. Int J Med Inform. 2013;82:248–59.
49. Epi4K Consortium. Epilepsy Phenome/Genome Project. Ultra-rare genetic variation in common epilepsies: a case-control sequencing study. Lancet Neurol. 2017;16:135–43.
50. Epi4K Consortium, Epilepsy Phenome/Genome Project, Allen AS, et al. De novo mutations in epileptic encephalopathies. Nature. 2013;501:217–21.
51. International League Against Epilepsy Consortium on Complex Epilepsies. Electronic address e-auea. Genetic determinants of

common epilepsies: a meta-analysis of genome-wide association studies. Lancet Neurol. 2014;13:893–903.
52. EuroEPINOMICS-RES Consortium. Epilepsy Phenome/Genome Project, Epi4K Consortium. De novo mutations in synaptic transmission genes including DNM1 cause epileptic encephalopathies. Am J Hum Genet. 2017;100:179.
53. Mailman MD, Feolo M, Jin Y, et al. The NCBI dbGaP database of genotypes and phenotypes. Nat Genet. 2007;39:1181–6.
54. KOMP Repository UD. KOMP repository knockout mouse project. [cited 2020 July 28]. Available at: https://www.komp.org/
55. Friedel RH, Seisenberger C, Kaloff C, Wurst W. EUCOMM--the European conditional mouse mutagenesis program. Brief Funct Genomic Proteomic. 2007;6:180-185. https://doi.org/10.1093/bfgp/elm022
56. NorCOMM. North American Conditional Mouse Mutagenesis Project. [cited 2020 July 28]. Available at: http://www.norcomm2.org/
57. Medicine Texas A&M Institute for Genomic Medicine. The world's largest collection of C57 ES cells and mice. [cited 2020 July 28]. Available at: https://www.tigm.org/
58. National Institutes of Health. Knockout mouse phenotyping. [cited 2020 July 28]. Available at: https://commonfund.nih.gov/komp2
59. Harte-Hargrove LC, French JA, Pitkanen A, Galanopoulou AS, Whittemore V, Scharfman HE. Common data elements for preclinical epilepsy research: standards for data collection and reporting. A TASK3 report of the AES/ILAE Translational Task Force of the ILAE. Epilepsia. 2017;58(Suppl 4):78–86.
60. Galanopoulou AS, French JA, O'Brien T, Simonato M. Harmonization in preclinical epilepsy research: a joint AES/ILAE translational initiative. Epilepsia. 2017;58(Suppl 4):7–9.
61. Hernan AE, Schevon CA, Worrell GA, et al. Methodological standards and functional correlates of depth in vivo electrophysiological recordings in control rodents. A TASK1-WG3 report of the AES/ILAE Translational Task Force of the ILAE. Epilepsia. 2017;58(Suppl 4):28–39.
62. Kadam SD, D'Ambrosio R, Duveau V, et al. Methodological standards and interpretation of video-electroencephalography in adult control rodents. A TASK1-WG1 report of the AES/ILAE Translational Task Force of the ILAE. Epilepsia. 2017;58(Suppl 4):10–27.
63. Moyer JT, Gnatkovsky V, Ono T, et al. Standards for data acquisition and software-based analysis of in vivo electroencephalography recordings from animals. A TASK1-WG5 report of the AES/ILAE Translational Task Force of the ILAE. Epilepsia. 2017;58(Suppl 4):53–67.
64. Raimondo JV, Heinemann U, de Curtis M, et al. Methodological standards for in vitro models of epilepsy and epileptic seizures. A TASK1-WG4 report of the AES/ILAE Translational Task Force of the ILAE. Epilepsia. 2017;58(Suppl 4):40–52.
65. Simonato M, Iyengar S, Brooks-Kayal A, et al. Identification and characterization of outcome measures reported in animal models of epilepsy: protocol for a systematic review of the literature—A TASK2 report of the AES/ILAE Translational Task Force of the ILAE. Epilepsia. 2017;58(Suppl 4):68–77.
66. Ben-Menachem E. Epilepsy in 2015: the year of collaborations for big data. Lancet Neurol. 2016;15:6–7.
67. Boshuisen K, Arzimanoglou A, Cross JH, et al. Timing of antiepileptic drug withdrawal and long-term seizure outcome after paediatric epilepsy surgery (TimeToStop): a retrospective observational study. Lancet Neurol. 2012;11:784–91.
68. Vale CL, Rydzewska LHM, Rovers MM, et al. Uptake of systematic reviews and meta-analyses based on individual participant data in clinical practice guidelines: descriptive study. BMJ. 2015;350:h1088.
69. Nevitt SJ, Marson AG, Davie B, Reynolds S, Williams L, Smith CT. Exploring changes over time and characteristics associated with data retrieval across individual participant data meta-analyses: systematic review. BMJ. 2017;357:j1390.
70. Nevitt SJ, Sudell M, Weston J, Tudur Smith C, Marson AG. Antiepileptic drug monotherapy for epilepsy: a network meta-analysis of individual participant data. Cochrane Database Syst Rev. 2017;12:CD011412.
71. Fallah A, Guyatt GH, Snead OC, et al. Predictors of seizure outcomes in children with tuberous sclerosis complex and intractable epilepsy undergoing resective epilepsy surgery: an individual participant data meta-analysis. PLoS One. 2013;8:e53565.
72. Offringa M, Bossuyt PMM, Lubsen J, et al. Risk factors for seizure recurrence in children with febrile seizures: a pooled analysis of individual patient data from five studies. J Pediatr. 1994;124:574–84.
73. Hampel KG, Thijs RD, Elger CE, Surges R. Recurrence risk of ictal asystole in epilepsy. Neurology. 2017;89:785–91.
74. Lamberink HJ, Otte WM, Geerts AT, et al. Individualised prediction model of seizure recurrence and long-term outcomes after withdrawal of antiepileptic drugs in seizure-free patients: a systematic review and individual participant data meta-analysis. Lancet Neurol. 2017;16:523–31.
75. Hemingway H, Asselbergs FW, Danesh J, et al. Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. Eur Heart J. 2018;39:1481–95.
76. Weber GM, Mandl KD, Kohane IS. Finding the missing link for big biomedical data. JAMA. 2014;311:2479–80.
77. Denaxas SC, Morley KI. Big biomedical data and cardiovascular disease research: opportunities and challenges. Eur Heart J Qual Care Clin Outcomes. 2015;1:9–16.
78. Denaxas SC, George J, Herrett E, et al. Data resource profile: cardiovascular disease research using linked bespoke studies and electronic health records (CALIBER). Int J Epidemiol. 2012;41:1625–38.
79. Pathak J, Kho AN, Denny JC. Electronic health records-driven phenotyping: challenges, recent advances, and perspectives. J Am Med Inform Assoc. 2013;20:e206–e211.
80. Morley KI, Wallace J, Denaxas SC, et al. Defining disease phenotypes using national linked electronic health records: a case study of atrial fibrillation. PLoS One. 2014;9:e110900.
81. Denaxas S, Direk K, Gonzalez-Izquierdo A, et al. Methods for enhancing the reproducibility of biomedical research findings using electronic health records. BioData Min. 2017;10:31.
82. Levin MA, Wanderer JP, Ehrenfeld JM. Data, big data, and metadata in anesthesiology. Anesth Analg. 2015;121:1661–7.
83. Mooney SJ, Westreich DJ, El-Sayed AM. Commentary: Epidemiology in the era of big data. Epidemiology. 2015;26:390–4.
84. Holmes LB, Wyszynski DF, Lieberman E. The AED (antiepileptic drug) pregnancy registry: a 6-year experience. Arch Neurol. 2004;61:673–8.
85. Shorvon SD, Goodridge DM. Longitudinal cohort studies of the prognosis of epilepsy: contribution of the National General Practice Study of Epilepsy and other studies. Brain. 2013;136:3497–510.

86. The Lancet Neurology. EURAP signals a new era in epilepsy research. Lancet Neurol. 2011;10:591.
87. Reid AY, St Germaine-Smith C, Liu M, et al. Development and validation of a case definition for epilepsy for use with administrative health data. Epilepsy Res. 2012;102:173–9.
88. Fonferko-Shadrach B, Lacey AS, White CP, et al. Validating epilepsy diagnoses in routinely collected data. Seizure. 2017;52:195–8.
89. Meeraus WH, Petersen I, Chin RF, Knott F, Gilbert R. Childhood epilepsy recorded in primary care in the UK. Arch Dis Child. 2013;98:195–202.
90. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the general practice research database: a systematic review. Br J Clin Pharmacol. 2010;69:4–14.
91. Shivade C, Raghavan P, Fosler-Lussier E, et al. A review of approaches to identifying patient phenotype cohorts using electronic health records. J Am Med Inform Assoc. 2014;21:221–30.
92. Faught E. Antiepileptic drug trials: the view from the clinic. Epileptic Disord. 2012;14:114–23.
93. Kim H, Thurman DJ, Durgin T, Faught E, Helmers S. Estimating epilepsy incidence and prevalence in the US pediatric population using nationwide health insurance claims data. J Child Neurol. 2016;31:743–9.
94. Gaitatzis A, Carroll K, Majeed A, Sander JW. The epidemiology of the comorbidity of epilepsy in the general population. Epilepsia. 2004;45:1613–22.
95. Hesdorffer DC, Ishihara L, Mynepalli L, Webb DJ, Weil J, Hauser WA. Epilepsy, suicidality, and psychiatric disorders: a bidirectional association. Ann Neurol. 2012;72:184–91.
96. Josephson CB, Lowerison M, Vallerand I, et al. Association of depression and treated depression with epilepsy and seizure outcomes: a multicohort analysis. JAMA Neurol. 2017;74:533–9.
97. Su CC, Chi MH, Lin SH, Yang YK. Bidirectional association between autism spectrum disorder and epilepsy in child and adolescent patients: a population-based cohort study. Eur Child Adolesc Psychiatry. 2016;25:979–87.
98. Sundelin HEK, Larsson H, Lichtenstein P, et al. Autism and epilepsy: a population-based nationwide cohort study. Neurology. 2016;87:192–7.
99. Arana A, Wentworth CE, Ayuso-Mateos JL, Arellano FM. Suicide-related events in patients treated with antiepileptic drugs. N Engl J Med. 2010;363:542–51.
100. Pugh MJV, Hesdorffer D, Wang C-P, et al. Temporal trends in new exposure to antiepileptic drug monotherapy and suicide-related behavior. Neurology. 2013;81:1900–6.
101. Josephson CB, Gonzalez-Izquierdo A, Denaxas S, et al. Serotonin reuptake inhibitors and mortality in epilepsy: a linked primary-care cohort study. Epilepsia. 2017;58:2002–9.
102. Ridsdale L, Charlton J, Ashworth M, Richardson MP, Gulliford MC. Epilepsy mortality and risk factors for death in epilepsy: a population-based study. Br J Gen Pract. 2011;61:e271–e278.
103. Kaiboriboon K, Schiltz NK, Bakaki PM, Lhatoo SD, Koroukian SM. Premature mortality in poor health and low income adults with epilepsy. Epilepsia. 2014;55:1781–8.
104. Devinsky O, Dilley C, Ozery-Flato M, et al. Changing the approach to treatment choice in epilepsy using big data. Epilepsy Behav. 2016;56:32–7.
105. Grinspan ZM, Shapiro JS, Abramson EL, Hooker G, Kaushal R, Kern LM. Predicting frequent ED use by people with epilepsy with health information exchange data. Neurology. 2015;85(12):1031–8.
106. Thurman DJ, Kobau R, Luo YH, Helmers SL, Zack MM. Healthcare access among adults with epilepsy: the U.S. National Health Interview Survey, 2010 and 2013. Epilepsy Behav. 2016;55:184–8.
107. Aghaei-Lasboo A, Fisher RS. Methods for measuring seizure frequency and severity. Neurol Clin. 2016;34:383–94.
108. Bidwell J, Khuwatsamrit T, Askew B, Ehrenberg JA, Helmers S. Seizure reporting technologies for epilepsy treatment: a review of clinical information needs and supporting technologies. Seizure. 2015;32:109–17.
109. Osorio I, Schachter S. Extracerebral detection of seizures: a new era in epileptology? Epilepsy Behav. 2011;22(Suppl 1):S82–7.
110. van Andel J, Thijs RD, de Weerd A, Arends J, Leijten F. Non-EEG based ambulatory seizure detection designed for home use: what is available and how will it influence epilepsy care? Epilepsy Behav. 2016;57:82–9.
111. Van de Vel A, Verhaert K, Ceulemans B. Critical evaluation of four different seizure detection systems tested on one patient with focal and generalized tonic and clonic seizures. Epilepsy Behav. 2014;37:91–4.
112. Jory C, Shankar R, Coker D, McLean B, Hanna J, Newman C. Safe and sound? A systematic literature review of seizure detection methods for personal use. Seizure. 2016;36:4–15.
113. Krauss GL, Ryvlin P. Non-EEG seizure detection is here. Neurology. 2018;90:207–8.
114. Poh M-Z, Loddenkemper T, Reinsberger C, et al. Convulsive seizure detection using a wrist-worn electrodermal activity and accelerometry biosensor. Epilepsia. 2012;53:e93–e97.
115. Beniczky S, Polster T, Kjaer TW, Hjalgrim H. Detection of generalized tonic-clonic seizures by a wireless wrist accelerometer: a prospective, multicenter study. Epilepsia. 2013;54:e58–e61.
116. Narechania AP, Garic II, Sen-Gupta I, Macken MP, Gerard EE, Schuele SU. Assessment of a quasi-piezoelectric mattress monitor as a detection system for generalized convulsions. Epilepsy Behav. 2013;28:172–6.
117. Patterson AL, Mudigoudar B, Fulton S, et al. SmartWatch by smartmonitor: assessment of seizure detection efficacy for various seizure types in children, a large prospective single-center study. Pediatr Neurol. 2015;53:309–11.
118. Poppel KV, Fulton SP, McGregor A, Ellis M, Patters A, Wheless J. Prospective study of the Emfit movement monitor. J Child Neurol. 2013;28:1434–6.
119. Beniczky S, Conradsen I, Moldovan M, et al. Quantitative analysis of surface electromyography during epileptic and nonepileptic convulsive seizures. Epilepsia. 2014;55:1128–34.
120. Beniczky S, Conradsen I, Moldovan M, et al. Automated differentiation between epileptic and nonepileptic convulsive seizures. Ann Neurol. 2015;77:348–51.
121. Conradsen I, Moldovan M, Jennum P, Wolf P, Farina D, Beniczky S. Dynamics of muscle activation during tonic-clonic seizures. Epilepsy Res. 2013;104:84–93.
122. Sarkis RA, Thome-Souza S, Poh M-Z, et al. Autonomic changes following generalized tonic clonic seizures: an analysis of adult and pediatric patients with epilepsy. Epilepsy Res. 2015;115:113–8.
123. Cogan D, Nourani M, Harvey J, Nagaraddi V. Epileptic seizure detection using wristworn biosensors. Conf Proc IEEE Eng Med Biol Soc. 2015;2015:5086–9.
124. Kroll RR, Boyd JG, Maslove DM. Accuracy of a wrist-worn wearable device for monitoring heart rates in hospital

inpatients: a prospective observational study. J Med Internet Res. 2016;18:e253.
125. Eggleston KS, Olin BD, Fisher RS. Ictal tachycardia: the head-heart connection. Seizure. 2014;23:496–505.
126. Zijlmans M, Flanagan D, Gotman J. Heart rate changes and ECG abnormalities during epileptic seizures: prevalence and definition of an objective clinical sign. Epilepsia. 2002;43:847–54.
127. Conradsen I, Beniczky S, Wolf P, Henriksen J, Sams T, Sorensen HB. Seizure onset detection based on a uni- or multi-modal intelligent seizure acquisition (UISA/MISA) system. Conf Proc IEEE Eng Med Biol Soc. 2010;2010:3269–72.
128. Conradsen I, Beniczky S, Wolf P, Kjaer TW, Sams T, Sorensen HB. Automatic multi-modal intelligent seizure acquisition (MISA) system for detection of motor seizures from electromyographic data and motion data. Comput Methods Programs Biomed. 2012;107:97–110.
129. Conradsen I, Beniczky S, Wolf P, Terney D, Sams T, Sorensen HB. Multi-modal intelligent seizure acquisition (MISA) system—a new approach towards seizure detection based on full body motion measures. Conf Proc IEEE Eng Med Biol Soc. 2009;2009:2591–5.
130. Van de Vel A, Cuppens K, Bonroy B, et al. Non-EEG seizure detection systems and potential SUDEP prevention: state of the art: review and update. Seizure. 2016;41:141–53.
131. Van de Vel A, Cuppens K, Bonroy B, et al. Non-EEG seizure-detection systems and potential SUDEP prevention: state of the art. Seizure. 2013;22:345–55.

How to cite this article: Lhatoo SD, Bernasconi N, Blumcke I, 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;61:1869–1883. https://doi.org/10.1111/epi.16633