



Published in final edited form as:

J Clin Neurophysiol. 2015 June ; 32(3): 235–239. doi:10.1097/WNP.000000000000159.

Collaborating and sharing data in Epilepsy Research

JB Wagenaar, PhD¹, GA Worrell, MD PhD², Z Ives, PhD³, M Dümpelmann, PhD⁴, B Litt, MD^{1,5,*}, and A Schulze-Bonhage, MD PhD^{4,*}

¹Neurology, University of Pennsylvania

²Neurology, Mayo Clinic

³Computer and Information Science, University of Pennsylvania

⁴Epileptology, University Hospital Freiburg

⁵Bioengineering, University of Pennsylvania

Abstract

Technological advances are dramatically advancing translational research in Epilepsy. Neurophysiology, imaging and meta-data are now recorded digitally in most centers, enabling quantitative analysis. Basic and translational research opportunities to use these data are exploding, but academic and funding cultures are preventing this potential from being realized. Research on epileptogenic networks, anti-epileptic devices and biomarkers could progress rapidly, if collaborative efforts to digest this “big neuro data” could be organized. Higher temporal and spatial resolution data are driving the need for novel multi-dimensional visualization and analysis tools. Crowd-sourced science, the same that drives innovation in computer science, could easily be mobilized for these tasks, were it not for competition for funding, attribution and lack of standard data formats and platforms. As these efforts mature, there is a great opportunity to advance Epilepsy research through data sharing, and increase collaboration between within the international research community.

Keywords

Epilepsy; Data-sharing; Cloud-computing; Data Repositories; EEG

Background

Technological advances in the past 20 years have dramatically advanced translational research in epilepsy. Electroencephalography (EEG), imaging and other data are now recorded digitally in most centers, which allows for quantitative measurements and analysis. The explosion of research using these data across many clinical and basic disciplines is

Corresponding Author: Joost B. Wagenaar: Phone: 215-746-4850, joostw@seas.upenn.edu.

*Senior authors contributed equally to this manuscript.

^{1,9}Wagenaar, and Litt: 301 Hayden Hall, University of Pennsylvania, 3320 Smith Walk, Philadelphia, PA 19104

³Worrell : Mayo Clinic, Department of Neurology, Mayo Systems Electrophysiology Lab, Rochester, MN 55905

³Ives : Computer and Information Science Dept., University of Pennsylvania, 3330 Walnut Street, Philadelphia, PA 19104

⁴Dümpelmann and Schulze-Bonhage : Epilepsiezentrum, Universitätsklinikum Freiburg, Neurozentrum, Breisacher Str. 64, D-79106 Freiburg, Germany

impressive, but only a fraction of its potential is being realized. In particular, quantitative electrophysiology research and its source data from patients largely remain at acquiring institutions. This not only limits the broad applicability of the research, but also the ability to validate results. In addition, there are some unique challenges and obstacles to sharing human data, such as integrating different file-formats, deidentifying protected health information (PHI), and adhering to government regulations regarding these datasets. This is even more complex if video recordings are required. As a result, the pace of scientific progress is slowed. There is broad appreciation of this problem in translational neuroscience, as major grant agencies now require data sharing. Unfortunately this is rarely done with human and animal electrophysiology, especially EEG data, because of a lack of a suitable venue in which to share it, and the lack of enforced sharing of raw data after investigators publish results.

In order to compare experimental methods in a robust fashion, they must be tested on the same data under similar conditions, and results need to be validated and reproducible. (Ioannidis 2005; Frei et al. 2010; Ince et al. 2012) Such standards are extremely difficult to maintain in neurophysiology research: data are stored locally and protocols vary widely. Other fields provide successful examples of data sharing: genetics research is uploaded to GenBank (Benson et al. 2014) by many groups, and functional MRI imaging (Mennes et al. 2013), computational modeling (Hines et al.), and electrophysiology (Moody et al. 2001) have also developed public databases to share computer code and data. The recent Epilepsy Phenome-Genome Project demonstrates that a large multicenter study can collaborate with multimodal data via a web-based portal. (Nesbitt et al. 2013)

The fMRI Data Center is a pioneer in data sharing, and provides important experience regarding the academic utility. The platform's architects also note some important challenges to multi-center data sharing: maintaining patient privacy, releasing data prior to follow up studies (Van Horn and Gazzaniga 2013), and the reluctance of many towards accepting a new culture of open data sharing. (Mennes et al. 2013) Other concerns identified by the fMRI community certainly apply to epilepsy: issues surrounding storage space and access, uncertain procedures for crediting data acquisition, and the possibility of having another group refute your results with your own data, either with correct or perhaps flawed methods. This latter issue is often quoted by investigators reluctant to share data, significant overhead to research, as productive scientists are diverted from new work to fend off attacks from naïve or less skilled investigators. From these experiences, it is clear that these sorts of platforms for data sharing advance scientific discovery; but also require considerable effort to establish a rigorous and "open science ethos," to really have impact (Mennes et al. 2013).

Multi-scale EEG that includes microwire recordings providing single neuron activity and macroscopic local field potentials poses unique problems in data sharing because of file large sizes (over 1 TB/ day in some systems). In addition, varying vendor formats and subject privacy are further barriers. Data must be reliable and correctly annotated, which is problematic given disparities in clinical interpretation (Lehnertz and Litt 2005; Benbadis et al. 2009; Osorio et al. 2011). EEG analysis often involves complex mathematics which, when combined with expansive data, results in large computational overhead and monetary

cost. Benchmarking experiments (i.e. testing algorithms on novel data) also require a central body to curate “gold standard” training data and withhold testing data.

Sharing both human and animal research data is a crucial standard in research and is drawing increased attention from major research and funding agencies in worldwide, such as the NIH, European Union agencies, DARPA, National Science Foundation and private enterprises, such as the Allen Institute for Brain Science. Adherence to this standard is weak in EEG research. Journals do not require publication of data or raw computer code, and there is little infrastructure to share them. Clinical centers that robustly acquire data, particularly in day-to-day practice, are often distant from facilities with expertise to analyze it. Thus, while technological advances have dramatically increased the richness and promise of neurophysiologic research, lack of standards, shared data and algorithms limits the impact of this research on healthcare and translational neuroscience. The impact is not just theoretical. Important questions like the best approaches for localizing brain networks in disease, the most effective surgical techniques, imaging protocols and methods for targeting devices remain active open questions, with huge implications for clinical care. The lack of data sharing, standards and collaboration across centers slows progress in addressing these questions.

The need for collaborative efforts, validation and gold standards

The need for a central EEG database was championed by the seizure prediction community (Lehnertz and Litt 2005), who developed rigorous tools to compare the results of different algorithms on similar datasets. (Snyder et al. 2008; Schulze-Bonhage et al. 2011) Through this collaborative network and shared data approach seizure prediction performance was shown to be highly dependent on adequate EEG data (Mormann et al. 2005; Schulze-Bonhage et al. 2011; Stacey et al. 2011). Early studies had been severely hampered by small data sets from small numbers of patients (5–21), often containing incomplete sets of EEG electrodes, small numbers of seizures (<89) and limited structured metadata.

Over the past decade, two complementary groups have developed more rigorous research platforms to address these challenges. The first was established in Freiburg, Germany (<http://epilepsy-database.eu>) (Ihle et al. 2012) and led to an ambitious multinational European EEG database. (Ihle et al. 2012; Klatt et al. 2012) This database contains scalp and intracranial EEG data from 275 patients and their clinical metadata, which can be downloaded, analyzed locally (Schelter et al. 2010; Park et al. 2011), and allows different research groups to compare algorithm performance on the same dataset. (2007) The second effort, <http://ieeg.org>, is a comprehensive platform to share data, display, annotation and analysis tools, and algorithm. In addition to data sharing the platform is designed to keep track of experiment and data provenance, in order to validate research. This platform currently hosts over 1,000 human and animal data sets, images and like Epilepsiae, a substantial international user base. Comprehensive databases such as these two efforts are critical to standardize EEG and other data analyses, and to avoid bias. (Ihle et al. 2012)

Using collaborative data sharing platforms, performance estimations of algorithms can be validated at a new level. This method has particular power and utility when data are

assembled in a collaborative way, to fully represent the spectrum of a particular field or application. In epilepsy this approach allows investigators to not only test seizure prediction algorithms on a considerable diversity of patients, seizure patterns, seizure onset areas and etiologies (Alexandre Teixeira et al. 2014; Alvarado-Rojas et al. 2014), but also to train algorithms within a given patient without risk for over-fitting by separating training and test datasets. Furthermore, detection algorithms can be evaluated on the spectrum of EEG onset patterns found in focal epilepsy to analyze best features to capture these patterns (Meier et al. 2008), to apply learning algorithms to improve detection and to implement classifiers which serve best to integrate the information of different features within a group of patterns or in a patient-individual manner. Additional metadata on patient characteristics such as etiology of the disorder and antiepileptic drug levels can be relevant for patient stratification and for analyzing the effect of unstable dynamics related to drug effects on the performance of time series analyses. Thus, aspects like analyses like the occurrence of circadian patterns or the effects of medications on seizure clustering can be assessed.

The Epilepsae project and the ieeg.org efforts have provided a significant push to advance the collaborative Epilepsy research community. Both efforts seek to provide the community with a high quality resource for Epilepsy research, albeit using a slightly different approach. The next two sections highlight each of these efforts.

Epilepsiae

As part of a EU-funded project (FP7-project “EPILEPSIAE”, Evolving Platform for Improving the Living Expectations of Patients Suffering from Ictal Events, Grant 211713, www.epilepsiae.eu), a new database with so far unprecedented size and quality was created 2009–2012 aiming at the integration of both raw EEG data and derived EEG features from 275 patients undergoing long-term video EEG recordings from three European Centers, the University Hospital of Coimbra, the Salpêtrière Hospital Paris and the Epilepsy Center at the University Hospital of Freiburg. Criteria for data sets to be included were based on the need for continuous long-term recordings including both, sufficiently long periods of interictal EEG data and high numbers of seizures separated by at least hours.

The EU database structure is a relational database with different tables for raw data, time references and other metadata, including imaging, electrodes used with their positions in the standardized MNI space, structured annotations for seizure events including early and late propagation patterns, and other data like daily antiepileptic drug dosages (Ihle et al. 2012). The database allows applying predefined queries mostly serving to identify patient subgroups according to the type of recording, localization of seizure onset or ictal onset EEG pattern; for more complex views on the database and data selections, freely programmed SQL queries can be entered. Both types of enquiries can be applied via a specifically programmed web client.

The EU database contains continuous long-term recordings from 275 patients with focal epilepsy (Klatt et al. 2012). EEG data are stored as binary files allowing the export into many commercially available software used for review and annotation. Raw data are continuous with the possible exception of brief intermissions e.g. related to the performance

of imaging procedures during the monitoring period. The mean duration of recordings is 165 h, the mean number of seizures per patient is 9.8 (recordings with higher numbers of seizures were preferentially selected at the participating centers). Beyond raw EEG data, more than 20 linear and non-linear EEG-derived features, which had been used for seizure prediction, were calculated at the centers and integrated into the database, saving time for applications at other sites.

Beyond raw and derived data files, a large set of metadata are integrated to the database. In a subset of patients with intracranial EEG recordings, subclinical events were additionally marked by expert reviewers as their presence may pose particular problems to the assessment of valid seizure predictions (Feldwisch-Drentrup et al. 2011; Klatt et al. 2012). Beyond annotations, metadata contain information on patient history, seizure types, etiology, imaging findings, intracranial electrode positions and antiepileptic drug regimens. The quantity of the database content exceeds former databases by more than an order of magnitude and contains some 2,700 seizures from patients from childhood to late adulthood, and a total recording time of more than 40,000 hours.

ieeg.org

The International Epilepsy Electrophysiology Portal (<http://ieeg.org>) is a NINDS-funded cloud based collaborative platform for electrophysiology for Epilepsy research [14]. This platform was initially developed to share large intracranial epilepsy electrophysiology datasets but has evolved to include preclinical data and additional data-modalities (i.e. imaging, documents etc.). This platform hosts over 1,200 datasets of continuous scalp and intracranial EEG from both animal models of Epilepsy and patients, and has over 500 users from more than 30 countries. It provides a foundation for developing Big Data tools for data storage, wrangling and analysis. Key collaborators on this proposal are sharing their data on this platform and are committed to collaborative scientific efforts.

The platform can automatically transcode from multiple EEG standards and import/index various meta-data formats. This minimizes the burden on the user to standardize the datasets before uploading them to the platform. The list of supported file formats for automated import continues to expand as converters are written. Since code for the Portal is open source, users with new formats can request or sponsor new import software to be written for their project. Unsupported file-formats can be archived and made available on the platform for direct downloads. This ‘import-pipeline’ is currently actively being developed and will feature additional capabilities within the near future.

The ieeg.org platform runs on the commercial cloud (Amazon Web Services, using Amazon’s S3 and RDS) and on a local intranet (using Tomcat, NFS and MySQL). It provides APIs through which code may be run “near” the data or remotely; and may directly fetch, process, annotate, and store data. The APIs are flexible and cross-platform: core services are written in Java but the portal also provides a sophisticated MATLAB “toolbox” that provides a more familiar environment to many users. It also provides Web Service (REST) APIs, so local or remote code in virtually any language can be used. Data is transferred in a highly compressed format and can be cached on the client side in Mongo

DB, for fast retrieval of previously downloaded data. These ideas of projection, preprocessing, and caching can be easily extended to other data types. The platform is browser-based and includes online tools for data and image visualization, a space to share analysis algorithms, and other collaborative tools.

Developing this cloud-based platform has highlighted some of the challenges in integrating Big Data for the neurosciences, for example: (1) To standardize time-series data, all incoming datasets are converted to an open-standard lossless compressed format. However, not all provided data is easily accessible outside the manufacturer's proprietary software and the heterogeneous nature of new data prevents easy integration with standardized database schema, especially for animal datasets. (2) Portal use-cases vary significantly depending on the background and level of expertise of the portal user (i.e. clinicians, clinician scientists, computational data analysts and undergraduate students). (3) Data providers indicate that they need custom control of data access privileges. This allows researchers to selectively share data with a small group of researchers instead of making the data public.

The overall goal of ieeeg.org is to provide a cloud based research platform that provides useful tools for individual research laboratories as well as an integrative way to share, and find data from a large cohort of research projects.

Funding, models and the need to sustain these resources

A particular challenge to data sharing is to secure enough resources to sustain the effort beyond the initial funding period. In contrast to hypothesis driven research projects, the added value of data sharing platforms only materializes if the platform will be available for the foreseeable future. Buy-in from platform users also depends on the prospects of continuity of these platforms, especially when the platform is used to host data from multiple sources and aims to provide open access to the research community. This has been a major challenge for all efforts in this domain.

In case of Epilepsiae, access is now restricted to scientific groups that financially contribute to the maintenance of the database, following EU regulations. Even if the charge for access is less than 0.3% of the costs of establishing the database, and presently groups from many countries as well as companies use the data for their research, it has turned out that the use of the database is lower than of former freely available data sources of by far lower quality.

Instead of requiring a flat fee to access a data-resource, it is also possible to implement a 'pay-as-you-go' type infrastructure. In this case, the user only pays for the accessed resources (i.e. downloaded datasets, computational time). This requires tracking usage and more sophisticated billing infrastructure but can provide a more flexible approach to keeping the services available.

However, the key requirement to a successful data sharing ethos, and infrastructure is the commitment of research funding sources such as the NIH, NSF, DARPA, the EU and others. The NIH requirement to provide a data sharing plan in most funding proposals needs to be supported by funding the resources that could support such data sharing plans. Sharing data can only be useful to the research community when the data is made available in way that

the research community at large can find the data, the tools to manipulate it, and interpret the data adequately. This requires the data-sharing resource to provide the data in an accessible way (eg. standardized access), with full documentation and preferably a way to provide interaction between the data providers and data users.

The Future of these efforts in epilepsy

Curating, annotating and sharing high quality scalp and intracranial EEG recordings from patients with epilepsy has tremendous potential to advance and accelerate research in this field. Running published algorithms *and code* over gold standard data, like those published in Epilepsiae or ieeg.org will generate irrefutable performance results that can be compared between efforts. There are, however, limits to collecting just more clinical recordings, however. Even if the number and duration of recordings, and the number of seizures included in them, surpasses what is already available in shared databases, recordings obtained from patients admitted for presurgical evaluation or differential diagnosis at dedicated epilepsy centers are mostly limited to a period of 10–14 days. (Stacey et al. 2011) During this period, changes in the antiepileptic drug regimen are frequently present, and acute signal alterations related to the implantation of intracranial electrodes may confound some measurements. A study of Cook et al. (2013) has pointed to the fact that optimal performance of seizure prediction methods may only be achieved several months after implantation of a recording device, suggesting that instabilities in brain dynamics during the periods included in presently available databases may impair the optimization and lead to an underestimation of the performance of seizure prediction algorithms applied.

It is important to realize that research data is only useful if the dataset can be interpreted adequately. This means that sharing raw data without context, meta-data, data-acquisition methods etc. has limited value. Therefore, data sharing should constitute more than a simple data repository, especially if the hosted data is multi-modal, and originates from multiple institutions and includes clinical, and preclinical data. Common data elements, standardized nomenclature, and inventive methods to integrate, navigate, and search these data, will be very important to optimally using collections of shared data from an ever-growing array of implantable sensors, as technology and its translation into the clinical arena expand.

One of the greatest barriers to data sharing is the disincentive to doing so provided by competitive academic systems around the world. The H-index is widely used as a measure of productivity and impact of a researcher and is based on the number of publications by this researcher, and the number of references to these publications. Academic promotion procedures at universities use the ‘H-index’, and the amount of grant funding, but do not currently include a mechanism for rewarding data sharing and collaboration. To the contrary, they actively discourage sharing data when this might help a competitor publish earlier, report a new finding, or perhaps correct an erroneous conclusion.

We propose a mechanism for giving investigators credit for catalyzing research, in the form of an ‘index’ that tracks the importance of shared data through how much it is used and the quality of publications to which it contributes. It provides insight in the quality and significance of the data recorded by a particular scientist and complements the commonly

used H-index, which is designed to track the number and quality of publications. One possible implementation of the ‘S-index’ might increase the score of an investigator whenever a study is published that utilizes that dataset and the owners do not appear as first or last author on the publication, eliminating self-citation. Data providers might be cited in a specific section of papers in which they are used, rather than appear in the author list. S-index credit might be given to junior investigators who actually collect the data more than lab directors. We realize that introducing such a novel idea is non-trivial, will take time to reach acceptance, and need some refining, but we also feel that there is a need for such an index to promote data sharing and create a more collaborative scientific mindset. The nature of the online data-repositories can provide a vital element in providing such index, as it tracks data-usage and provides ways to reference used data in publications. The index will have meaning if it is given ‘teeth’, such as being used as a criterion for promotion or grant funding, as universities, journals and funding agencies assess specific investigators. Initiating such an index would only require crediting a particular data set on a shared resource, such as the IEEG-Portal or the European Epilepsy Database.

Research on epileptogenic networks and anti-epileptic devices will benefit greatly from a more organized collaborative effort to digest the massive amounts of data that are available in various research centers around the world. With increased interest in higher resolution data collected at higher bandwidths, the need for collaboration on data visualization and analysis tools will only increase. Cloud based, shared resources, such as ieeg.org provide the means to share, and validate algorithms, standardize research approaches and provide data access to the larger research community. They facilitate crowd-sourced science, for example in the search for biomarkers. Such an example is currently under way at the time of this writing, through the American Epilepsy Society Seizure Detection and Prediction Challenge <https://www.kaggle.com/c/seizure-detection>, hosted on Kaggle.com and co-sponsored by the NIH, AES, Epilepsy Foundation, the University of Pennsylvania, Mayo Clinic and <http://ieeg.org>. It is up to the community to seize this moment and start harnessing the potential of cloud computing, massive parallel processing and interdisciplinary collaborations in epilepsy and electrophysiology research. In addition to these lofty goals, more concrete advances in clinical care might move forward, as data from presurgical patients could be made available on line to guide their care, and provide a clear pathway to consensus on interventions with surgery and devices, as opposed to the regional approaches undertaken at a host of individual academic epilepsy centers.

Over the coming years we expect that more and more labs will begin to leverage the power of cloud storage and computing for managing and analyzing laboratory data. This will accelerate as costs from large vendors such as Amazon, IBM, and Google eventually drop below the costs incurred by having individual servers, data archives and dedicated personnel to run them. However, the use of globally accessible cloud-resources requires additional efforts to adhere to the various international regulatory frameworks for storing clinical, and pre-clinical data (i.e. HIPAA, and the European Data Protection Directive). As cloud based products increasingly gain ground in clinical and research efforts, it will be of vital importance to shape/refine policy around these topics at University and governing entities to minimize hurdles while protecting patient information. Balancing privacy issues with the

potential for big data to help the general public is likely to be a central point of discussion as data sharing accelerates research.

Unquestionably, the change in resource utilization provides an unprecedented opportunity to accelerate the pace of collaborative science through advances in cloud-based technologies. This is a vital time for leaders in our field to incentive researchers to collaborate, share, and consequently get much more out of our increasingly precious research dollars. The efforts that began with the meticulous attention to quality, detail and expert review, best exemplified in the Epilepsiae database now need to be combined with the cloud-based platform approach to analysis, sharing, and validation of research provided by ieeg.org. Together these efforts have tremendous potential to accelerate discovery in epilepsy care and research.

Acknowledgement

The IEEG.org project is funded by the NINDS/NIH 1U24NS063930-04 and NIH 5P20NS080181-03. The EPILEPSIAE platform was funded through the EU FP7 Project EPILEPSIAE (Grant 211713).

References

- Alexandre Teixeira C, Direito B, Bandarabadi M, Le Van Quyen M, Valderrama M, Schelter B, et al. Epileptic seizure predictors based on computational intelligence techniques: a comparative study with 278 patients. *Comput Methods Programs Biomed.* 2014 May; 114(3):324–336. [PubMed: 24657096]
- Alvarado-Rojas C, Valderrama M, Fouad-Ahmed A, Feldwisch-Drentrup H, Ihle M, Teixeira CA, et al. Slow modulations of high-frequency activity (40–140-Hz) discriminate preictal changes in human focal epilepsy. *Sci Rep.* 2014; 4:4545. [PubMed: 24686330]
- Benbadis SR, LaFrance WC, Papandonatos GD, Korabathina K, Lin K, Kraemer HC. Interrater reliability of EEG-video monitoring. *Neurology.* 2009 Sep 15; 73(11):843–846. [PubMed: 19752450]
- Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res.* 2014; 42:D32–D37. [PubMed: 24217914]
- Feldwisch-Drentrup H, Ihle M, Quyen ML Van, Teixeira C, Dourado A, Timmer J, et al. Anticipating the unobserved: prediction of subclinical seizures. *Epilepsy Behav.* 2011 Dec; 22(Suppl 1):S119–S126. [PubMed: 22078512]
- Frei MG, Zaveri HP, Arthurs S, Bergey GK, Jouny CC, Lehnertz K, et al. Controversies in epilepsy: debates held during the Fourth International Workshop on Seizure Prediction. *Epilepsy Behav.* 2010 Sep; 19(1):4–16. [PubMed: 20708976]
- Hines ML, Morse T, Migliore M, Carnevale NT, Shepherd GM. Model DB: A Database to Support Computational Neuroscience. *J Comput Neurosci.* 2014 Jul-Aug; 17(1):7–11. [PubMed: 15218350]
- Van Horn JD, Gazzaniga MS. Why share data? Lessons learned from the fMRIDC. *Neuroimage.* 2013:677–682. [PubMed: 23160115]
- Ihle M, Feldwisch-Drentrup H, Teixeira CA, Witon A, Schelter B, Timmer J, et al. EPILEPSIAE - a European epilepsy database. *Comput Methods Programs Biomed.* 2012 Jun; 106(3):127–138. [PubMed: 20863589]
- Ince DC, Hatton L, Graham-Cumming J. The case for open computer programs. *Nature.* 2012; 482(7386):485–488. [PubMed: 22358837]
- Ioannidis JPA. Why most published research findings are false. *PLoS Med.* 2005 Aug.2(8):e124. [PubMed: 16060722]
- Klatt J, Feldwisch-Drentrup H, Ihle M, Navarro V, Neufang M, Teixeira C, et al. The EPILEPSIAE database: an extensive electroencephalography database of epilepsy patients. *Epilepsia.* 2012 Sep; 53(9):1669–1676. [PubMed: 22738131]

- Lehnertz K, Litt B. The First International Collaborative Workshop on Seizure Prediction: summary and data description. *Clin Neurophysiol.* 2005 Mar; 116(3):493–505. [PubMed: 15721063]
- Meier R, Dittrich H, Schulze-Bonhage A, Aertsen A. Detecting epileptic seizures in long-term human EEG: a new approach to automatic online and real-time detection and classification of polymorphic seizure patterns. *J Clin Neurophysiol.* 2008 Jun; 25(3):119–131. [PubMed: 18469727]
- Mennes M, Biswal BB, Castellanos FX, Milham MP. Making data sharing work: the FCP/INDI experience. *Neuroimage.* 2013 Nov 15;82:683–691. [PubMed: 23123682]
- Moody GB, Mark RG, Goldberger AL. PhysioNet: a Web-based resource for the study of physiologic signals. *IEEE Eng Med Biol Mag.* 2001 May-Jun;20(3):70–75. [PubMed: 11446213]
- Mormann F, Kreuz T, Rieke C, Andrzejak RG, Kraskov A, David P, et al. On the predictability of epileptic seizures. *Clin Neurophysiol.* 2005 Mar; 116(3):569–587. [PubMed: 15721071]
- Nesbitt G, McKenna K, Mays V, Carpenter A, Miller K, Williams M. The Epilepsy Phenome/Genome Project (EPGP) informatics platform. *Int J Med Inform.* 2013 Apr; 82(4):248–259. [PubMed: 22579394]
- Osorio I, Lyubushin A, Sornette D. Toward a probabilistic definition of seizures. *Epilepsy Behav.* 2011 Dec; 22(Suppl 1):S18–S28. [PubMed: 22078513]
- Park Y, Luo L, Parhi KK, Netoff T. Seizure prediction with spectral power of EEG using cost-sensitive support vector machines. *Epilepsia.* 2011 Oct; 52(10):1761–1770. [PubMed: 21692794]
- Schelter B, Feldwisch-Drentrup H, Timmer J, Gotman J, Schulze-Bonhage A. A common strategy and database to compare the performance of seizure prediction algorithms. *Epilepsy Behav.* 2010 Feb; 17(2):154–156. [PubMed: 20044314]
- 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 Dec; 22(Suppl 1):S88–S93. [PubMed: 22078525]
- Snyder DE, Echaz J, Grimes DB, Litt B. The statistics of a practical seizure warning system. *J Neural Eng.* 2008 Dec; 5(4):392–401. [PubMed: 18827312]
- Stacey W, Le Van Quyen M, Mormann F, Schulze-Bonhage A. What is the present-day EEG evidence for a preictal state? *Epilepsy Res.* 2011 Dec; 97(3):243–251. [PubMed: 21885253]
- Human and automated detection of high-frequency oscillations in clinical intracranial EEG recordings. *Clin Neurophysiol.* 2007 May; 118(5):1134–1143. [PubMed: 17382583]