The University of San Francisco

USF Scholarship: a digital repository @ Gleeson Library | Geschke Center

Master's Theses

Theses, Dissertations, Capstones and Projects

Spring 5-19-2017

The Acquisition and Analysis of Electroencephalogram Data for the Classification of Benign Partial Epilepsy of Childhood with Centrotemporal Spikes

Jessica A. Scarborough
University of San Francisco, jess.scarborough12@gmail.com

Follow this and additional works at: https://repository.usfca.edu/thes

Part of the Diagnosis Commons, Longitudinal Data Analysis and Time Series Commons, Neurology Commons, Neurosciences Commons, Numerical Analysis and Scientific Computing Commons, and the Sleep Medicine Commons

UNIVERSITY OF SAN FRANCISCO

SCHOOL OF NURSING AND HEALTH PROFESSIONS

The Acquisition and Analysis of Electroencephalogram Data for the Classification of

Benign Partial Epilepsy of Childhood with Centrotemporal Spikes

A thesis submitted in partial satisfaction of the

requirements for the degree Master of Science in Health Informatics

by

Jessica Anne Scarborough

Thesis Supervisor:

William J. Bosl, Ph.D., Chair
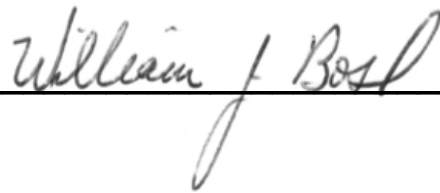
Director of Health Informatics

School of Nursing and Health Professions

University of San Francisco

May 2017

The thesis of Jessica Anne Scarborough is approved.

William J Bosl   5/18/2017

William J. Bosl,

Director of Heallth Informatics Program Director

Committee Chair

May 2017

The Acquisition and Analysis of Electroencephalogram Data for the Diagnosis of Benign Partial

Epilepsy of Childhood with Centrotemporal Spikes

ACKNOWLEDGEMENTS

The Acquisition and Analysis of Electroencephalogram Data for the Diagnosis of Benign Partial

Epilepsy of Childhood with Centrotemporal Spikes

by

Jessica Anne Scarborough

In this thesis, I will expand upon each step in the process of acquiring and analyzing

electroencephalogram (EEG) for the classification of benign childhood epilepsy with

centrotemporal spikes. Despite huge advancements in the field of health informatics—natural

language processing, machine learning, predictive modeling—there are significant barriers to the

access of clinical data. These barriers include information blocking, privacy policy concerns, and

a lack of stakeholder support. We will see that these roadblocks are all responsible for stunting

biomedical research in some way, including my own experiences in acquiring the data for the

second chapter of this thesis.

This second chapter expands upon just one possible advancement that can be achieved

when researchers attain clinical data (in this case, EEG data). BECTS is a type of epilepsy that

only displays epileptiform activity on night-time EEGs. We hypothesize that a brain affected by

BECTS is also developmentally different during the daytime, and based on this assumption, our

analysis aims to uncover these electrodynamic distinctions. After course-graining raw EEG

segments, we extracted sample entropy, recurrence rate, laminarity, and determinism using

recurrence quantitative analysis. Our results displayed two major findings. First, awake BECTS and control patients can be classified with no overlap using all of these features. Second, BECTS patients show differences in sleep state RQA values from centrotemporal and non-centrotemporal regions. We cannot confirm if these differences display epileptiform activity, however, because we do not have controls for sleep studies. With proper development and implementation, this research has the potential to become a clinical decision support tool and decrease the need for inconvenient sleep studies.

Table of Contents

I. Accessing Clinical Data for Research

A. Background

Despite a large push for health scientists to explore the world of big data (Khoury & Loannidis, 2014; Margolis et al., 2014), attaining useful health data remains a firm roadblock in much of health informatics research. The United States government has invested almost $30 billion into the development, regulation, and incentivizing of electronic health records (EHRs), yet the promises of an interoperable health information exchange (HIE) have not been realized (Marchibroda, 2014). Individual EHR vendors often act as data silos, and institutions are unable to effectively communicate with each other. Likewise, researchers often report unreasonable challenges when attempting to gather data in a usable manner from third-party EHRs, medical devices, and medical data collection software. In some cases, these third-party vendors may be attempting to block the transfer of information to make it difficult to change software, known as "locking in" users (ONC, 2015, pp. 13, 17-18). Ubiquitous and often unfounded fears of the HIPAA Privacy Rule can be heard in research laboratories and institutional review board (IRB) meetings in academic medical centers across the country. Many of these challenges can be overcome with the if stakeholders to health software development, government policy, and hospital administration support the needs of researchers. In this chapter, I will discuss why access to clinical data remains a significant barrier to biomedical research and recent policy and culture changes that have the potential to ease these burdens.

B. Barriers to Clinical Data Access

There is a scarcity of formal studies regarding the complications in attaining medical data for biomedical and public health research (ONC, 2015, p. 7). Despite this, there is significant anecdotal evidence from conference proceedings and online informatics discussions that points to significant limitations faced by researchers trying to obtain data. Using these complaints as a foundation, I procured empirical evidence and government reports regarding significant obstructions researchers find when attempting to utilize biomedical research data. Throughout this chapter, I will demonstrate three major barriers in accessing medical data for research:

1. Information Blocking

2. Privacy Law Concerns

3. Lack of Stakeholder Support

As defined by the ONC, information blocking occurs "when persons or entities knowingly and unreasonably interfere with exchange or use of electronic health information" (ONC, 2015, p. 11). Next, concerns related to the Health Insurance Portability and Accountability Act (HIPAA) are often misdirected (Herdman, Moses, States, & (U.S.), 2006), as protected health information is removed from most research data in a process known as de-identification. Despite this, these concerns play a significant role in stunting medical research (Dunlop, Graham, Leroy, Glanz, & Dunlop, 2007). The final barrier, a lack of stakeholder support, is demonstrated with anecdotal evidence and the need for government intervention to achieve "meaningful use" of electronic health records.

Information Blocking

Information blocking, a term that was first formally defined by the Office of the National Coordinator for Health Information Technology (ONC)(Adler-Milstein & Pfeifer, 2017), refers

to actions taken by individuals or entities who purposefully block or hinder the transfer of health information without reason (ONC, 2015, p. 11). Information blocking is an act that is difficult to define, because an individual or entity is only committing information blocking (as defined by the ONC) if they are aware that their actions are preventing the reasonable transfer of health information. Purposefully preventing information exchange to comply with privacy laws is a "reasonable" action; therefore, doing so would not be considered information blocking. Conversely, deliberately citing a privacy law (e.g. HIPAA's Privacy Rule) as an excuse for not sharing patient health information is considered an act of information blocking (ONC, 2015, p. 16). Figure 1 below visualizes these three requirements in order for an act to be considered information blocking.

Figure 1. Requirements to categorize an action as information blocking, as defined by the ONC. Taken from (ONC, 2015, p. 11)

There are many motivations for individuals or entities to block the exchange of health information—economic, technological, and practical—that are documented from anecdotal

evidence. Although health IT developers (software vendors) are the primary culprits of information blocking (ONC, 2015, p. 15), some providers may see an economic incentive to preventing the authorized transfer of medical records in order to impede patients from leaving their practice. Vendors may purposefully not pursue reasonable technological advances that increase interoperability in order to "lock-in" providers to one system (ONC, 2015, p. 13).

Information blocking has been discussed anecdotally since the advent of the Health Information Technology for Economic and Clinical Health (HITECH) Act and the subsequent push for interoperability in healthcare. These concerns were brought to a national stage when the ONC published an information blocking report in April 2015 (ONC, 2015). This report took aim at EHR vendors who are believed to partake in the majority of information blocking, outlined mechanisms used to unnecessarily impede the transfer of health data, and laid out steps for remedying the situation. This report, however, is not intended to act as the final investigation into information blocking (ONC, 2015, pp. 19-20, 29). There are approximately 60 unsolicited complaints made to the ONC regarding information blocking in 2014 alone (ONC, 2015, p. 15). These complaints, along with some empirical evidence, are at the foundation of the report to Congress. Even with this evidence, the ONC strongly recommends that additional research is performed to confirm or disprove anecdotal complaints and gather more information from the perspective of providers and vendors being accused of information blocking (ONC, 2015, pp. 19-20).

In this report and later publications, the ONC outlines various methods that vendors and providers use to block the appropriate transfer of patient information; many of these information blocking techniques affect research in addition to clinical care. For example, the ONC states that complaints have been made alleging that "developers may be engaging in information blocking

as a means of 'locking in' providers and consumers to rigid technologies and information sharing

networks that reinforce the market dominance of established players and prevent competition

from more innovative technologies and services" (ONC, 2015, pp. 17-18). When EHRs act as

data silos, clinical experiences suffer. Providers are unable to access family history, drug

allergies, or previous test results if the patient hasn't been seen at provider's institution yet.

Likewise, researchers may be unable to attain data, despite IRB approval for its extraction. The

report continues, "Some of these developers cite security concerns and business justifications for

these practices, while others provide no justification or, in some cases, appear to acknowledge a

strong preference not to exchange information using federally adopted standards and to instead

drive more users to exchange information using proprietary platforms and services." This blatant

preference for preventing data liquidity is damaging to current patients (who may receive sub-

optimal care when their medical records are not shared with authorized providers) and future

patients (who could benefit from ground-breaking research that requires a decreased burden for

HIE). The upcoming case study will demonstrate tangible consequences incurred when software

developers force proprietary formatting of health data. Additionally, the inaccurate citation of

HIPAA's Privacy Rule will be discussed in the next section of this chapter. Although

information blocking can occur in many forms, the outcomes are strikingly consistent to medical

researchers—sample sizes drop, the significance of results is diminished, and/or research

pursuits are abandoned altogether.

The wide-ranging reactions to this report demonstrate that information blocking remains

a contested issue. A letter of dissent, signed by the executives of Epic, Allscripts, McKesson, and

other prominent EHR vendors, was presented to the ONC shortly after the publication of their

report (Burchell et al., 2015). These vendors argued that what was described as "information

blocking" in the report was not intentional, but rather, a result of the high cost of developing and maintaining an interoperable interface. In this letter, the vendors maintain that there aren't enough users who want these features to justify their development. The ONC report, which was based significantly on anecdotal evidence, admits that further research must be done. Within the report, they state, "There is little quantitative data available with which to reliably identify and measure the extent of information blocking… In particular, ONC lacks methods and data to precisely determine why a provider is not exchanging when they should have the capability to do so." Due to its recent publication, there has been minimal follow-up research published regarding information blocking since this report to Congress. Yet, the ONC is not the only organization that believes EHR vendors are purposefully failing to deliver products that allow for data transfer. In a 2012 article titled *Escaping the EHR Trap—The Future of Health IT,* Dr. Kenneth Mandl and Dr. Isaac Kohane lament:

> We believe that EHR vendors propagate the myth that health IT is qualitatively different from industrial and consumer products in order to protect their prices and market share and block new entrants… This attitude has thwarted medicine's decades-long quest for an electronic information infrastructure capable of providing a dynamic and longitudinal view of the health care of individuals and populations. EHR companies have followed a business model whereby they control all data, rather than liberating the data for use in innovative applications in clinical care (Mandl & Kohane, 2012).

These authors, representing the Children's Hospital Informatics program, the Harvard-MIT Division of Health Sciences and Technology, and the Center for Biomedical Informatics at Harvard Medical School, note that despite general advancements in the field of data interoperability, EHR vendors have purposefully lagged behind these developments in order to

silo patient data. Although the ONC report is clearly based on anecdotal evidence, these complaints represent industry-wide concerns that cannot be dismissed.

In order to understand the causes of information blocking, EHR vendors must grant some transparency regarding their practices and contracts. The ONC expressed concern that attaining quantitative answers regarding the prevalence and mechanisms of information blocking (i.e. differences in charges for transferring data between institutions) is particularly challenging due to "gag clauses" often found in vendor agreements between providers and EHR companies (ONC, 2015, p. 16). Said gag clauses can prevent providers from discussing details of their contract, particularly cost. Given that intentionally prohibitive pricing is frequently referenced as a source of information blocking (ONC, 2015, p. 13), these restrictions limit the quantitative conclusions of any research on the matter (ONC, 2015, pp. 31-32).

Despite previously stated concerns regarding the shortage of empirical evidence related to information blocking, congress has moved quickly to improve standards for HIE. The 21st Century Cures Act, written and passed with bipartisan support, was signed into law by President Obama at the end of 2016. In addition to increasing funding to medical research substantially, this legislation aims to improve HIE interoperability by setting data exchange standards and barring acts of information blocking("21st Century Cures Act," 2016). Due to its recent enactment, the long-term effects of this law are unknown, but it is undoubtedly a step towards progress in achieving an interoperable healthcare system. Given that most EHR vendors appear to believe that their practices do not constitute information blocking, it is possible that more action will be necessary to create change in the private sector. This legislation does, however, set a tone for the expectations vendors will be held to for achieving continued Meaningful Use certification. If methods for detecting and discouraging information blocking are implemented in

addition to this redefined tone, the United States healthcare system will take a large step towards interoperability—leading to an improved standard of care and biomedical research.

Privacy Law Concerns

The introduction and implementation of the Health Insurance Portability and Accountability Act (HIPAA) was not met with enthusiasm from all providers. In theory, HIPAA was designed to improve patient privacy and electronic billing, yet many practitioners saw these rules as onerous, costly, and unconducive to quality care (Barieri, 2003; Bowers, 2001; Kumekawa, 2005). As a field, healthcare is notoriously resistant to change; still, over 20 years later, HIPAA has become an accepted industry standard, safeguarding the privacy of patients and their protected health information (PHI) (Solove, 2013). Despite this acceptance in the clinical world, concerns related to HIPAA (some reasonable and some unfounded) continue to hinder the progress of medical research (Dunlop et al., 2007; Gostin & Nass, 2009; Wei, 2015).

Apprehensions related to the inappropriate disclosure of PHI typically reference HIPAA's Privacy Rule. This rule sets standards for the protection of PHI within a covered entity. These covered entities include health plans, health clearinghouses, and healthcare providers. Per Susan McAndrew, the Deputy Director for Health Information Privacy in the Office for Civil Rights (OCR), data that have already been excised from a covered entity (e.g. de-identified and moved into a data warehouse) are no longer a concern of the Privacy Rule (Herdman et al., 2006, p. 8). Regulations are concerned with patient information that is being transferred between covered entities or between a covered and non-covered entity. In these circumstances, the Privacy Rule stipulates limitations for the use and disclosure of PHI.

With biomedical research in mind, the Privacy Rule creates several approaches for the transfer of identified medical records outside of one covered entity. These include the following: (Herdman et al., 2006, pp. 8-9)

1. Acquiring informed consent from each patient before recruitment into a study

2. Obtaining institutional review board (IRB) approval for the use of a limited data set (may contain broad geographical information and some dates)

3. (In very limited cases) Attaining IRB approval to use PHI without the patient's consent, stipulating that this PHI will be protected and remain unpublished

Although these procedures may delay research temporarily, it is not without good reason. This purpose is stated by Dr. Roberta Ness of the University of Pittsburg: "Researchers fundamentally believe in and are engaged in protecting confidentiality as much as is possible, because they fundamentally understand… that without the protection of confidentiality, there will be no trust in research and, therefore, [they] will be unable to conduct research." In addition to the ethical obligation to protect patient confidentiality, scientists also have benevolently selfish reasons for ensuring that PHI in their control remains secure. If the public's faith in the protection of PHI is diminished, less patients will agree to sharing their data, leading to reduced research opportunities and weaker results. With that being said, these sensible privacy requirements may inherently cause a delay to biomedical research. For example, obtaining informed consent, writing IRB proposals, and attending IRB meetings all require time and resources. Even given these acceptable burdens, the observed impact to biomedical research is disproportionate to the intended pauses set forth by the Privacy Rule (Ness, 2007).

Some inadvertent burdens that stem from the Privacy Rule are caused by confusion regarding what the IRB is capable of approving. One of their many roles, IRBs are responsible

for interpreting HIPAA regulations; as such, these review boards can waive requirements as they see fit (NIH, 2003). Nevertheless, according to Paul Feldman, Director of the Health Privacy Project, "IRBs believe they have no authority to approve alterations to or waivers of authorization for protocols not subject to the Common Rule. They do" (Herdman et al., 2006, p. 21). This statement was made at a forum regarding the effect of the HIPAA Privacy Rule on health research in 2006. It is now over 10 years later and there is little written about whether current IRBs are willing to waive HIPAA requirements when given the appropriate documentation of purpose and security. It is likely that a lack of discussion on this topic is a positive outcome—fewer researchers are experiencing this issue making further investigations unnecessary. Lessons learned from the confusion surrounding the initial roll-out of HIPAA regulations should be documented in order to improve the implementation of future health privacy policies.

De-identification, the process of removing PHI from medical records, is a common option for researchers wishing to attain patient data and avoid HIPAA constraints. There are two methods of de-identification, the "Expert Determination" method and the "Safe Harbor" method. The "Expert Determination" method, also known as the statistician method, requires that a person with appropriate field knowledge and statistical skills declares that the data is not individually identifiable ("Health Insurance Portability and Accountability Act of 1996," 1996). Next, the "Safe Harbor" method, doesn't require a statistician's approval. Instead, de-identification requires the removal of the 18 individual identifiers from medical records("Health Insurance Portability and Accountability Act of 1996," 1996). These identifiers can be found in Appendix A. The removal of all patient identifiers is the overwhelming method of choice,

despite the possibility of gaining valuable demographic information in a secure manner through the "Expert Determination Method."

The Privacy Rule stipulates that records de-identified with "expert determination" can be shared freely (Amatayukal, 2003; NIH, 2004); nonetheless, misguided concerns regarding the ability to share de-identified data are often accepted as valid. The statistician method of de-identifying data is not universally accepted by IRBs, because these boards are often unsure of how to regulate whether it is being done properly (Herdman et al., 2006, p. 22). A statistician's endorsement that data has been appropriately de-identified is only as infallible as the assumptions said statistician made during the evaluation process (Herdman et al., 2006, p. 17). For example, the statistician is only required to certify that the risk of re-identification is "very small," but per guidance published by the Department for Health and Human Services (HHS), there is no definition for what "very small" is. Given these uncertainties, IRB concern may be warranted.

In addition to good faith concerns, IRBs can be unsure of their liability for guaranteeing that de-identification is performed to adequate standards. By the very nature of IRB independence from a central agency, board requirements and decisions can differ widely between organizations (Herdman et al., 2006, pp. 25-28). While some variation is inevitable, not permitting a Privacy Rule-approved method of de-identification because of unfamiliarity with the technique may create unfair differences between institutions. Resolving these concerns will require addressing IRB reservations and unfamiliarity with the statistician method. First and foremost, detailed documentation of the de-identification process should be provided to the IRB, including assumptions made and steps taken to ensure there's a "very small" chance of re-identification. HHS guidance regarding the two methods of de-identification, including a lengthy

explanation of the "Expert Determination" technique, can be found online (OCR, 2017). And IRBs that are unfamiliar with their role in overseeing a statistician de-identification process should be provided with mechanisms to increase their understanding and awareness of this method.

The 21st Century Cures Act, touched upon in the "Information Blocking" section of this chapter, also works to improve medical research by lightening Privacy Rule restrictions for medical researchers. Before this update, PHI could be utilized by a covered entity for health care operations, including treatment, billing, and hospital procedures. This legislation now classifies research as part of "health care operations," giving researchers and business associates within covered entities autonomous access to medical records. Additionally, when authorization forms are required, the forms specify "one-time" consent, where the authorization to use said data is indefinite unless the patient later revokes authorization. This PHI is still protected by the HIPAA Privacy and Security Rules, involving significant protection standards and breach notification rules. These alterations to the Privacy Rule have the potential to make a large impact by giving researchers access to data. In order to realize this full impact, however, IRBs must be informed and educated on these changes.

HIPAA was not created with research as a priority. (Herdman et al., 2006, p. 13). This lack of foresight demonstrates a consequence of inconsistent stakeholder support in the pursuit of improving health research, which will be discussed further in the next section of this chapter. For good reason, obtaining medical records with PHI does require extra effort in obtaining patient consent or IRB approval. Yet, without the appropriate training in how to navigate these regulations, research will be stymied.

Lack of Stakeholder Support

Clinical research has stakeholders in various sectors, ranging from clinicians to policymakers to medical institutions. In a clinical setting, the ability to acquire clinical data begins with data collection and storage mechanisms. The advent of electronic health records centered around improving patient care with increased safety, efficacy of treatments, and ease of billing. These are all crucial goals; yet, for the purposes of this discussion, these objectives make it clear that research was not a priority. We can therefore postulate that this lack of forethought provides some explanation for the difficulties encountered in obtaining data from electronic medical records. Stakeholders in clinical data collection and storage mechanisms traditionally include patients, clinicians, IT personnel, data warehouse engineers, hospital administrators, and corporate employees. Medical centers may choose to prioritize research as an institutional objective, adding additional stakeholders related to research and education. Simply put, there are fundamental differences in how various stakeholder prioritize institutional objectives. When choosing between electronic health record vendors and investment into backend design, corporate employees and administrators may focus on institutional reputation and budget-saving measures. When expressing their opinions regarding the use of technology in a clinical setting, clinicians and patients may believe improved workflow and diagnostic rates are fundamental to success. Research principle investigators, on the other hand, may prioritize a well-defined data warehouse that allows for the institution's researchers to access de-identified data with ease. A primary step to improving access to medical data in research must be coordinating stakeholder support of these goals.

When included in the planning stages, these research stakeholders can influence fundamental government regulations, institutional mission statements, and organization cultural

attitudes related to the secure transfer of medical records. To do so, however, all stakeholders must be invited to provide their expertise and opinions. According to Marcy Wilder, Esq., the lead lawyer working on HIPAA's Privacy Rule, their team was not able to expand upon research regulations during the writing process, because there simply was a lack of expertise in the matter (Herdman et al., 2006, p. 14). As discussed previously, some fundamental concerns related to the Privacy Rule may have been avoided if clinical research stakeholders played a role in writing said policy. And despite significant alterations to the Privacy Rule, which simplify the transfer of PHI to covered entity researchers, the academic field will need to be educated on these changes in order for them to be implemented fully.

The Centers for Medicare & Medicaid (CMS) has created more recent policy incentivizing electronic health records (EHRs) that adhere to Meaningful Use standards. These standards are an example of creating policies that work to appropriately prioritize research goals in addition to clinical health. Meaningful Use regulations incentivize eligible providers and institutions to utilize EHRs that meet or exceed expectations of utilizing electronic health records in a "meaningful" way. In addition to increasing patient safety (e.g. requiring EHRs to perform drug-drug or drug-allergy interaction checks in Stage 1), this program requires participants to eventually utilize EHRs that promote public health (e.g. standards for reporting to public health repositories) (CMS, 2010). Additionally, in the beginning of 2017, CMS created a centralized repository for this reporting (CMS, 2017). Agencies that are able to accept public health and/or clinical data may sign up to be included, allowing clinical data sources to report to multiple agencies connected through this centralized repository. Here, we can see a path towards the interoperability required for achieving some of the true potential that EHRs hold.

Finally, the importance of economic incentive cannot be understated. There is little speculation regarding why vendors and institutions may want to prevent the easy transfer of medical information; easily transferred data can lead to easily changed systems. To some degree, this economic incentive needs to be respected, because it is precisely why developers work to create innovative products; it is why large academic medical centers exist. However, the EHR software industry has been made possible by Meaningful Use incentives that have a clear intention of promoting a health care system that benefits from easy, secure information exchange. Software developers, therefore, must recognize when benefits to consumers greatly outweigh economic incentives. Likewise, economic interests could and should be leveraged as an advantage to creating interoperable systems. From the perspective of institutions, academic medical centers that invest in software that easily transfers health data can produce research with more impressive sample sizes, more impressive results. Good research encourages additional funding; additional funding improves hospital and medical school rank, attracting even better talent to pursue even better research. Vendors, too, can benefit from this cycle. Institutions that wish to leverage data-driven research will invest in EHR systems and other health IT software (i.e. EKG analysis software) that makes the transfer and de-identification of medical records simple. Vendors may be concerned about making their software interoperable, but the consequence of not doing so could be much worse—isolating themselves from a market that is demanding interoperability.

C.   *Case Study: Acquiring Clinical Data from Boston Children's Hospital's Child Neurology Department*

Background

Since the advent of health informatics, Boston Children's Hospital (BCH) has been a leader in striving for innovative techniques that improve health care with the use of information technology (IT). Their Computational Health Informatics Program (CHIP) was awarded "Best Health Care Organization" by Health 2.0. In an award that took into consideration the last 10 years of health technology innovation, Boston Children's Hospital's global reputation was clearly demonstrated with this recognition. Furthermore, research division has created an Innovation and Digital Health Accelerator, which hopes to vet and foster technologies that advance the field of digital health. With a focus in remote care, clinical decision support, and innovation platforms (i.e. Fast Healthcare Interoperability Resources, also known as FHIR), Boston Children's Hospital is a clear forerunner in the health informatics field. This forward-thinking institution stands out with their investment in informatics, clearly recognizing the role that data-driven research will play in the future of medicine.

Research pursued by CHIP includes Health Data Fusion, SMART Health IT, HealthMap, and Apache cTakes. Each of these endeavors requires significant amounts of health data (sometimes millions of patient records) and the infrastructure necessary to obtain, process, and store large amounts of data. Of note, SMART Health IT is a platform comprised of open standards and open source software that allows external innovators and researchers to design health applications that can be used across the healthcare field. For example, "Cardiac Risk" is an app that uses an interactive interface to show a patient's risk of heart attack or stroke within the next 10 years. Once these tools are developed, applications may be added to the SMART

App Gallery at no charge. These applications can be "bolted on" to various health IT platforms. The very foundation of SMART Health IT relies on developers who are willing to share their standards, software, and applications with clinical and academic communities around the world. It is this culture of open innovation that feeds the research at Boston Children's Hospital. Yet, despite a strong concentration in big data research, this case study will demonstrate that BCH researchers are not without their own plights when attempting to access clinical data.

My time spent at Boston Children's Hospital enlightened me to the realities of accessing clinical records for biomedical research. To expand upon the research discussed in the following chapter, I traveled to BCH with the goal of extracting de-identified electroencephalogram (EEG) records from the Neurology Department's electronic medical record system. We worked in conjunction with Dr. Tobias Loddenkemper, an Associate Professor of Neurology at Harvard Medical School, with over 250 publications. At BCH, he is the Director of Clinical Epilepsy Research within the Neurology Department. As seen in many departments at Boston Children's Hospital, both excellence in clinical care and pioneering research are Neurology Department stakeholder priorities. During my time, I found that despite being a world leader in informatics and data-driven research, this clinical department still faces difficulties with fragmented, difficult to extract data.

Results

I traveled to Boston Children's Hospital with the goal of learning how to extract de-identified EEG records to be included in on-going epilepsy research with Dr. William Bosl. Our research required de-identified EEGs stored in the European Data Format (EDF), a standard format for the storage and transfer of time series data (e.g. EEG data) (Kemp, Varri, Rosa, Nielson, & Gade, 1993). We were specifically aiming to extract EEGs from developmentally

normal patients and patients who are diagnosed with Benign Epilepsy with Centro-Temporal Spikes.

Appropriate precautions were taken to adhere to the HIPAA Privacy Rule, but none of these proved onerous. For one, I completed HIPAA training modules to ensure my understanding and awareness of the policies researchers are expected to follow. Because my machine was not issued by Boston Children's Hospital, virtual private network (VPN) software was installed. Within this VPN, I could access clinical data in a secure, HIPAA-friendly manner. Although these steps required additional time and resources, the value of keeping clinical data secure and private is worth these minimal burdens.

With the goal of learning how to independently extract de-identified EEG records, I met with clinicians, researchers, and IT experts within the department of Neurology. In these meetings, I learned that in the Department of Neurology, EEG data is not stored within or attached to the EHR system that contains the majority of each patient's clinical data. This translates to manually searching the EEG collection software for each individual patient. Additionally, the EHR software has no querying capabilities. To a clinician, this is likely of little to no consequence. There's little need for them to see "All patients billed with ICD code X" or "What percentage of patients are taking medication Y?" Yet, these capabilities may be fundamental to a researcher's workflow. Without the ability to search for patients who were diagnosed with BECTS, Dr. Bosl and I relied on a list of patients provided by Dr. Loddenkemper. With this list, we could search through the EEG collection software to obtain the appropriate EEGs.

After selecting a patient of interest, a user selects the data they wish to extract (i.e. time segments of a chosen EEG recording), and they begin the export process. The EEG collection software exports data in the following two outputs:

1. De-identified EEG data in proprietary formatting

2. Identifiable EEG data in EDF formatting

Unfortunately, neither of these options works, as we needed de-identified data in a research-friendly (EDF) format. To achieve this desired output, researchers devised a workaround. First, they exported using the first option, de-identified data in the proprietary format. This file was then re-uploaded to the EEG collection software and re-exported in EDF using the second option. This left us with an EDF-formatted file with no PHI. This new procedure is relatively simple in theory, but in practice it changes a 5-minute export process to a 20-minute export process per patient. To achieve a 40-person sample-size, the export task went from requiring less than 3.5 hours to over 26.5 hours. In short, an assignment that could be accomplished in one morning now requires at least three full days' worth of work.

Plainly, this workaround is not ideal, and the researchers in this department have been in contact with the software developers for the EEG collection program. A continuing request for an option to download de-identified data into EDF format has been communicated to the developers for over a year with no response. Anecdotally, researchers have reported that other bugs and complaints have been addressed in software updates, but this request has been largely ignored.

Discussion

This experience at Boston Children's Hospital demonstrates many of the barriers researchers face when attempting to access appropriate clinical data. Although extra steps were taken to adhere to the HIPAA Privacy Rule, they were not unreasonable. Conversely, despite being part of an institution that actively strides to promote research utilizing patient data, software shortcomings made the process decidedly difficult.

An issue that effects both clinicians and researchers, it is disappointing to see clinical data such as EEG recordings not connected to a patient's electronic medical record. Having data linked appropriately is an unobtrusively helpful feature to any medical team. Linked data allows clinicians to perform fundamental tasks, such as viewing both patient history and testing data (e.g. radiology report, blood test results, etc.) in the same application. In the Neurology Department at BCH, if a researcher comes across an anomaly in an EEG recording, there is no clinical data attached to the file to help elucidate this finding. Instead of an integrated platform, where the EEG results can be viewed within a patient's EHR, researchers and clinicians are faced with two isolated systems.

The demand to develop interoperable health systems is likely just breaching the market, and finding vendors that are willing to integrate with each other is difficult. SMART Health IT and SMART on FHIR hope to solve this issue by opening the market to software developers who are motivated to create health technology that prioritizes interoperability over data silos with "locked-in" customers.

The additional barriers I came across, namely the inability to query the EHR and a challenging data export procedure, are prime examples of the consequences that stem from a lack of stakeholder support. The ability to query an EHR requires a significant investment in

development and resources, and discussing the logistics of creating such software could be a topic for another thesis. As most EHR software makes querying prohibitively difficult, organizations that wish to perform these tasks choose to export their data into a second, more analytics-friendly database (Mandl & Kohane, 2012). Still, an integrated querying feature would be a powerful tool in EHR software. If more research stakeholders were included in the planning stages of software development, these benefits could be articulated and the query functionality may be prioritized.

Another instance of absent stakeholder support, EEG collection software developers are not prioritizing the production of the specified format export feature. Of note, stakeholder support of research needs is likely minimal within a clinical software company. It is likely that more important features are taking precedence, but without the company's feedback it is impossible to be sure. There are always going to be "new and improved" features on the horizons for health IT development, and there is no shortage of wants from clinicians, researchers, and security specialists. Without a seat at the table, however, the wants and needs of researchers will be continually overlooked.

D.   Conclusion: Overcoming Barriers to Clinical Data Access

As future informaticists, nearly every student in University of San Francisco's Master of Science in Health Informatics (MSHI) program will face the persistent issue of encountering enticing data that is difficult to access, de-identify, or reformat. As such, it is imperative to understand these data acquisition challenges in order to achieve sought-after goals—the development of learning systems, clinical decision support tools, and natural language processing software.

Government policy has been an incredible driver in the advancement of the U.S. health care system, promoting HIE and interoperability, but some health policy did lead to unintended consequences. In both the case of HIPAA and Meaningful Use, the government stepped into the private health care sector in order to foster a secure and productive healthcare industry. Despite these common goals, the clear difference in consequences to biomedical research demonstrates that biomedical research was given much more consideration during the development of Meaningful Use. The HIPAA Privacy Rule was not intended to stymie research; that was made clear by the open-minded discussions and policy changes that followed the significant feedback submitted by medical researchers after deployment. Additionally, researchers are not advocating for weaker protection of PHI, especially when concerns are still being raised about the risks of re-identification (Benitez & Malin, 2010). Yet, it is now clear that creating policy that affects research without including research stakeholders in the process led to unexpected results—confusion on IRB responsibility, exceedingly cautious restrictions to access of patient data, and information blocking through inapplicable claims of Privacy Rule restrictions.

Changes in policy (i.e. The 21st Century Cures Act) and culture surrounding the transfer of health information to research professionals within a covered entity will require time and resources. Many organizations have decided to stop relying on third-party software vendors to prioritize the needs of research stakeholders; instead, they have sought in-house solutions. For example, academic medical centers in Cleveland, Ohio, have organized a research-friendly data warehouse that includes de-identified patient data from Cleveland Clinic, University Hospitals, and Metro Health Medical Center. All students and employees at any of these institutions are covered under the same IRB approval, allowing them to access this de-identified data without submitting a proposal. This effort is led by a research team at Case Western Reserve University's

Institute for Computational Biology. BCH's SMART Health IT is another demonstration of the changing culture surrounding health care interoperability. With developers striving to create open-API software that is accessible across platforms, SMART Health IT has created a path towards interoperability. There are plenty of viable means to achieving interoperability, and choosing just one is wholly unnecessary. The most important step in accomplishing the goal of effortless HIE is continued motivation and stakeholder support; without it, incredible technology (such as the SMART "Cardiac Risk" platform) will remain isolated from health care clinicians and researchers.

II. Benign Partial Epilepsy of Childhood with Centrotemporal Spikes

A. *Background*

Epileptic encephalopathies are a relatively new category for seizures that present with developmental delays or even regression. These conditions are correspondingly associated with neurocognitive and behavioral dysfunction that may persist long after seizures cease, usually in early adolescence (Engel, 2001). While most research has focused on early warning signs of seizure onset, another significant challenge is to monitor brain development. In doing so, it may be possible to detect functional characteristics—biomarkers—that indicate the brain has entered a dynamical state in which seizures are likely to occur and to predict the developmental trajectory.

Benign Childhood Epilepsy with Centro-Temporal Spikes

Benign childhood epilepsy with centro-temporal spikes (BECTS), also known as Rolandic Epilepsy, is the focus of this research. With a typical onset age of 7-10, BECTS usually remits before the age of 18 (Callenbach et al., 2010). BECTS seizures typically occur during sleep and can be characterized as simple partial seizures that lead to motor, hemifacial, and rapid spasms (Miziara & Manreza, 2002). Despite its name, this "benign" epilepsy is associated with developmental delays. Rarely, BECTS has been shown evolve into more severe epilepsy types (e.g. continuous spike-and-wave during slow-wave sleep syndrome, Landau-Kleffner syndrome, intractable seizures, etc.) leading to even more severe neurological impairment (Callenbach et al., 2010; Joost Nicolai et al., 2007). Despite the developmental defects observed in children

diagnosed with BECTS (Callenbach et al., 2010), it is often left untreated due to comparable long-term effects from taking anti-epileptic medication (Kwan & Brodie, 2001; Shields & III, 2009).

Techniques for Complex Systems Analysis

An automated method for quantifying epileptiform discharges in children has been developed by a co-investigator using a wavelet transform (Chavakula et al., 2013). Automated detection of continuous epileptiform activity is important because it can be present in the absence of overt seizures, yet has serious consequences. The complex systems methods developed in this project may be considered a further step in the development of automatic methods for detecting continuous spikes and waves during slow sleep (CSWS) and assessing severity. In another study by the co-investigator, spike counts were found to be relatively unaffected by confounding factors such as timing of epilepsy onset, clinical seizure type and frequency, and medication treatment and dose (Azhar, et al, in preparation). Furthermore, recently published research (W. J. Bosl, Tager-Flusberg, & Nelson, 2011) demonstrated that MSE could be used to distinguish infants at high risk for autism spectrum disorder (ASD) from those at low risk (on the basis of sibling history). A number of studies are beginning to show that patterns in nonlinear EEG features can be used as biomarkers for many neuropsychiatric disorders (Stam, 2005).

B.   Methods

Invariant measures convey information about the dynamics of the system from which they are derived. The challenge for clinical neuroscience is to determine which invariant measures, if any, are most relevant to detection and monitoring of specific conditions of clinical interest. One approach is to compute as many invariant measures as possible from populations of

32

patients with a condition of interest, and from controls, then use statistical learning algorithms to find the combinations of features, or feature patterns, that are most highly associated with the condition of interest. This data-driven approach may not be sufficient for a complete scientific understanding of the relationship between brain dynamics and behavioral, cognitive, or affective conditions, but may be sufficient for discovering useful clinical biomarkers. Data-driven discovery may also point in the direction of the most likely neural correlates of relevant behavioral constructs or cognitive phenotypes (W. Bosl, 2017).

The primary goal of the methods and results presented in this paper are clinical. We seek clinical decision support methods that will eventually provide the practicing physician, psychologist, or neurodiagnostic technologist with tools to enable risk assessment for BECTS during a child's routine neurological examination. The ultimate goal is avoiding unnecessary and expensive overnight EEG monitoring when possible.

Study Population

Epilepsy patients and a control group were seen at a tertiary epilepsy center in the Division of Epilepsy and Clinical Neurophysiology, Department of Neurology, Boston Children's Hospital (BCH). Patients with benign focal epilepsy of childhood were selected retrospectively. BECTS was confirmed by clinical history and EEG findings by an experienced neurologist. Patients were consecutively selected from the epilepsy outpatient clinic at Children's Hospital Boston, and were included if they were between 4 and 16 years old, presented with unilateral nocturnal motor seizures, and exhibited uni-or bilateral central-temporal sleep activated sharp waves with a frontal dipole on interictal EEG. Eleven age-

matched controls with normal EEG were also selected from the same clinical database. In total, 19 BECTS patients and 11 controls were used for this study.

Data Collection:

EEG data was sampled at 200 Hz for all Epilepsy Center subjects. Based on the Nyquist criterion, this implies that frequencies up to 100 Hz are included. From the BECTS cases, an experienced neurophysiology resident not directly involved in this study used visual review to select 30-second samples containing no spikes or evidence of epileptiform activity from awake EEG data. Similarly, 30-second segments were selected from awake subjects in the control group after visual review. A single segment of equal length was collected on 19 channels located according to the standard 10-20 system. EEG samples that were taken during sleep sessions with the BECTS patients. These segments were classified into sleep stages 2 and 3 by an experienced polysomnographic technologist and reviewed by the resident neurophysiologist. All **awake** EEG samples collected that were chosen, both BECTS cases and controls, appeared normal to the neurophysiologist on visual analysis. The sleep segments (BECTS cases only) were likely to contain nocturnal epileptiform spikes, but were not reviewed further for this study. No other filtering was performed on the EEG signals.

Signal processing

Multiscale entropy and recurrence quantitative analysis (RQA) values were computed from EEG signals using two steps. First, the coarse-graining described by (Costa et al., 2005) uses successive averaging of a time series to create new coarse-grained time series. For a

window size t, t = 1, 2, … j, the jth coarse-grain series, y^t_j, is computed by averaging non-overlapping windows, as shown in the figure below:
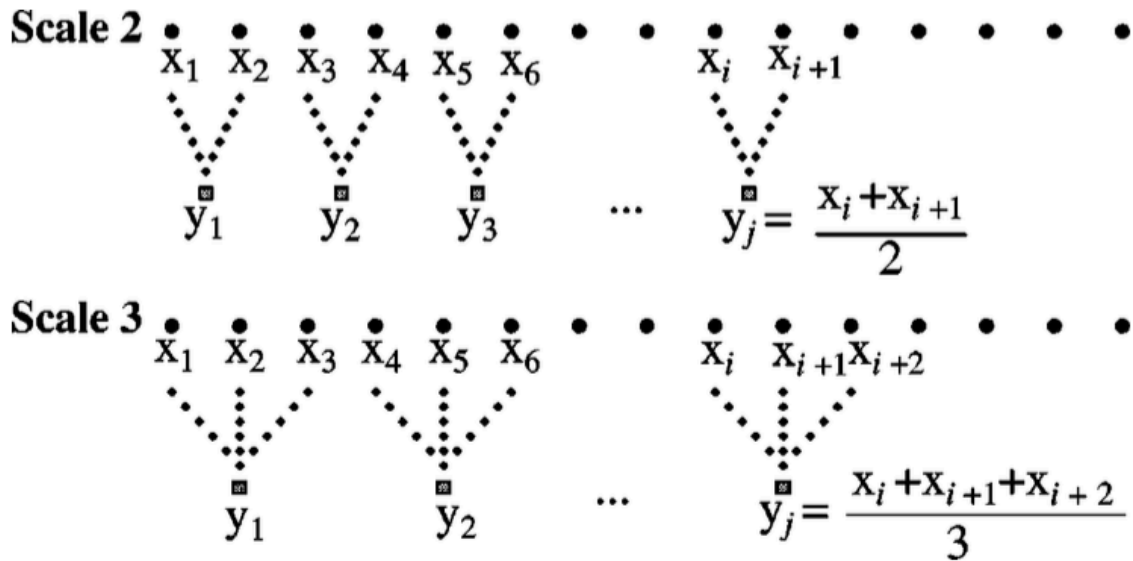

Figure 2. Schematic illustration of the coarse-graining procedure. Adapted from (Costa, Goldberger, & Peng, 2005)

Multiscale entropy (MSE) was computed using the modified sample entropy defined in (Xie, He, & Liu, 2008), which has been used in other studies for complexity analysis of physiological time series. Details about computing MSE are discussed in (Bosl et al. 2011). Multiscale entropy is a useful nonlinear method for analyze biological signals and distinguishing healthy from pathological states (Catarino, Churches, Baron-Cohen, Andrade, & Ring, 2011; Costa et al., 2005; Norris, Stein, & Morris, 2008; Takahashi et al., 2010; Vandendriessche et al., 2014).

Another approach to computing nonlinear time series properties is recurrence quantitative analysis (RQA). RQA is an empirical approach to analyzing time series data and is in principle capable of characterizing all of the essential dynamics of a complex system and is useful for

analyzing "real-world, noisy, high dimensional data" (Webber & Zbilut, 2005). It has proven to be a powerful tool already in physics, geophysics, engineering and biology (Komalapriya et al., 2008; Norbert Marwan, Romano, Thiel, & Kurths, 2007). Applications to neurology and neuroscience are in the early stages. In principle, RQA is capable of detecting significant state changes in a dynamical system (Norbert Marwan et al., 2007; Schinkel, Marwan, & Kurths, 2009; Webber & Marwan, 2015), which suggests that it may be appropriate for detecting developmental changes in brain function that are associated with chronic neurological and mental dysfunction.

RQA values were computed for all of the scales derived from each EEG channel using publicly available software (N. Marwan, 2012; Norbert Marwan et al., 2007). For input to the algorithms, the embedding dimension (m) was 10, the time delay was 2, and the threshold for the recurrence plot (tau) was 30. For detailed discussions of how these values may be determined, see (Chen, Zou, & Zhang, 2013; Niknazar, Mousavi, Vahdat, & Sayyah, 2013; Webber & Marwan, 2015).

Multiscale sample entropy was also computed and included in this set of EEG signal features and is denoted by SampE as in (W. J. Bosl et al., 2011). Sample entropy and the entropy derived from RQA, denoted "L_entr", are expected to be similar quantitative measures, but derived from different algorithms. Thus, multiscale curves were computed for eight features or values and each EEG sensor time series. Table 1 lists the nonlinear values computed for this study and provides a brief description of their meaning in a physical context.

RQA Variable	Symbol	Description
Recurrence rate	RR	The probability that a system state recurs in a finite time. RR has been found useful for detecting evoked response potentials (ERPs) using single trials (Schinkel et al., 2009).
Determinism	DET	DET comes from repeating patterns in the system and is an indication of its predictability. Regular, deterministic signals, such as sine waves, will have higher DET values, while uncorrelated time series, such as chaotic processes and random numbers, will cause low DET.
Laminarity	LAM	Laminarity represents the frequency of occurrence of laminar states in the system without describing the length of these laminar phases. More frequent appearance of laminar states may relate to more frequent "seeds" for synchronized dynamics (Hirata & Aihara, 2011), which may be related to epileptiform spiking on an EEG trace.
Max line length	L_max	Lmax is related to the largest Lyapunov exponent of a chaotic signal, which is a dynamic complexity measure that describes the divergence of trajectories starting at nearby initial states, (Gomez & Hornero, 2010). Lower values are typically associated with pathological conditions (Goldberger, 1997; Peng, Costa, & Goldberger, 2009).
Entropy derived from diagonal lines	L_entr	This measure of entropy is derived from the diagonal lines of the recurrence plot. It is related, but not identical to, other measures of entropy, such as the sample entropy used in previous studies (W. J. Bosl et al., 2011)
Trapping time	TT	Trapping time is an estimate of the time that a system will remain in a given state - "trapped" state. Thus, lower TT values may be an indication of more frequent transitions between dynamical states and less system stability.

Table 1. Recurrence Quantitative Analysis variables and their interpretation.

C. Results

Significant group differences were found for several RQA features, including recurrence rate, determinism, and laminarity, as well as sample entropy. The awake BECTS patients can be distinguished from controls with perfect accuracy in our sample population.

The scalp plots in the figures show multiscale curves for SampE, RR, DET, and LAM. 95% confidence intervals are shaded around the curves, demonstrating the relatively small variance.
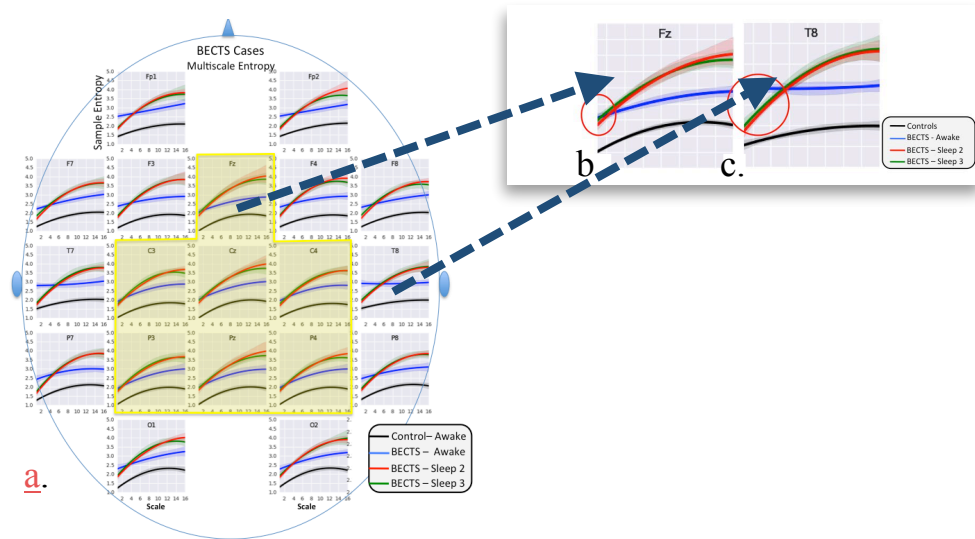

Figure 3. a. Multiscale curve with centrotemporal region highlighted; b. Multiscale curves typical of centrotemporal sensors; c. Multiscale entropy curves typical of non-central region

Determinism (DET) and Laminarity (LAM) have been found in previous studies to be lower closer to the epileptic zone in patients with focal epilepsies (Ngamga et al., 2016)
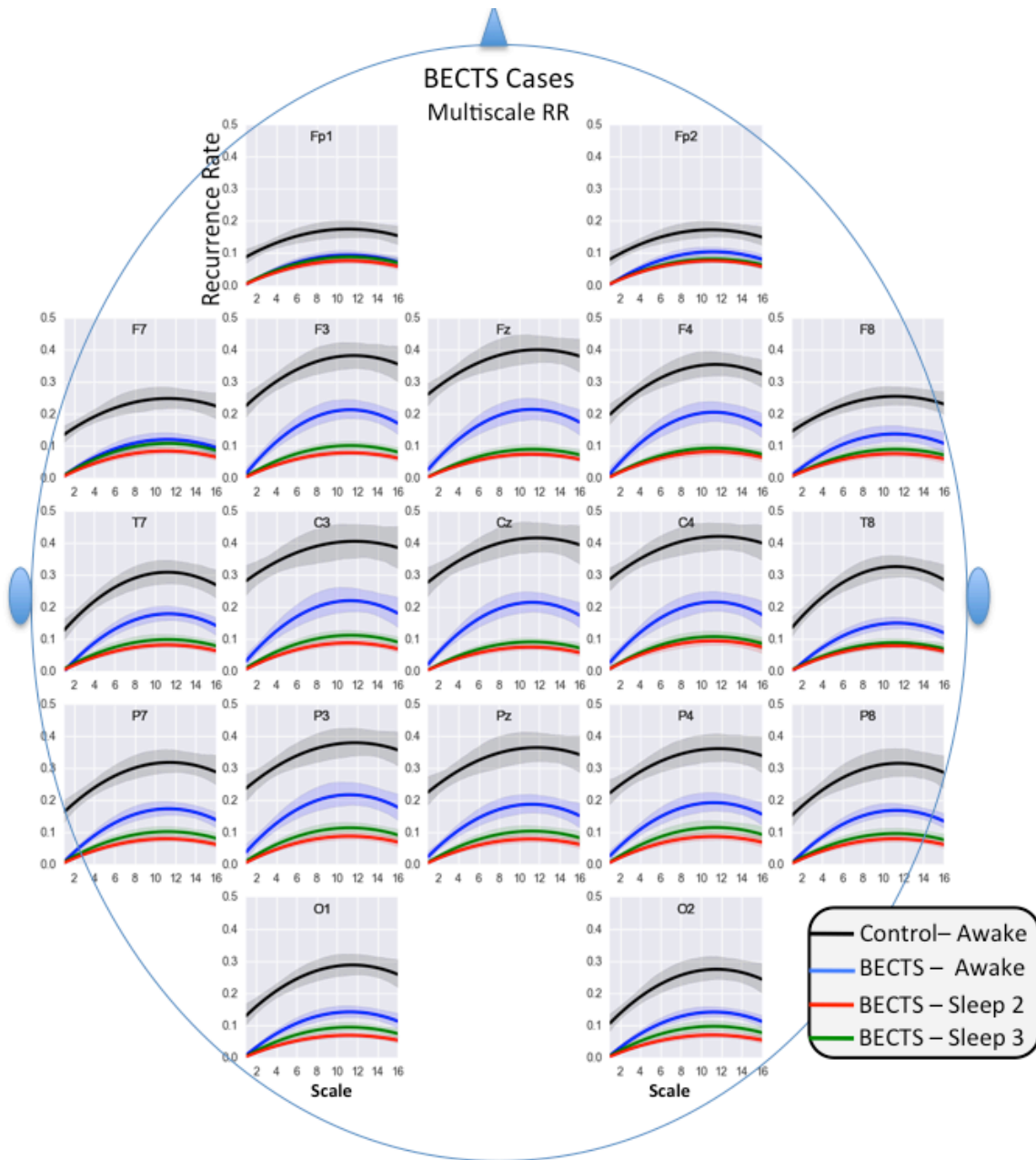
Figure 4. Multiscale RR curves for awake and asleep cases.

The separation in BECTS patients between awake multiscale RR curves and the sleep-2 and sleep-3 curves in the centro-temporal region (Fz, C3, Cz, C4, P3, Pz, P4) is greater than away from this region.
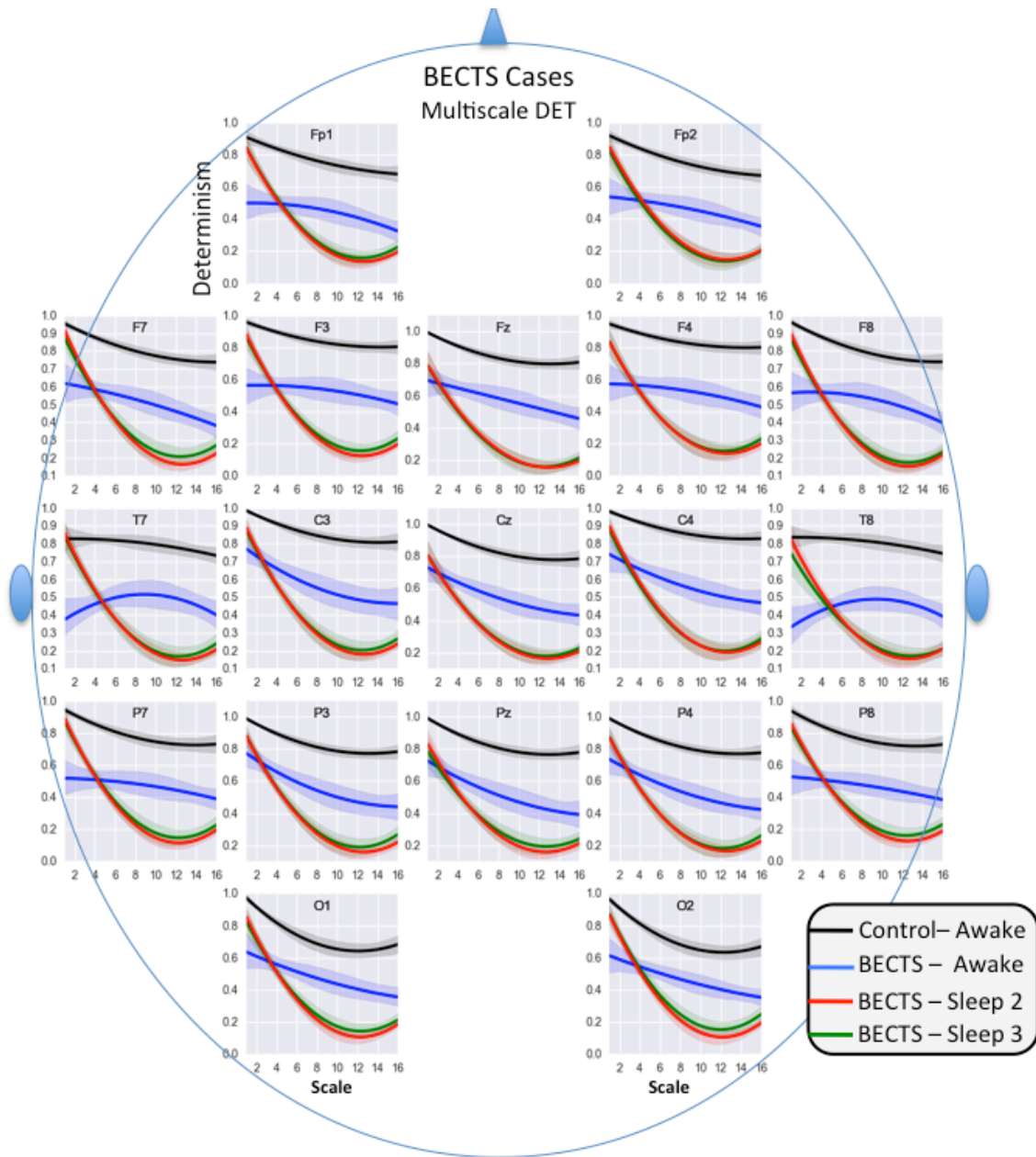
Figure 5. Multiscale DET curves for awake and asleep cases.

Lower scale multiscale DET curves show differences in awake BECTS patients between the centrotemporal and non-centrotemporal regions.
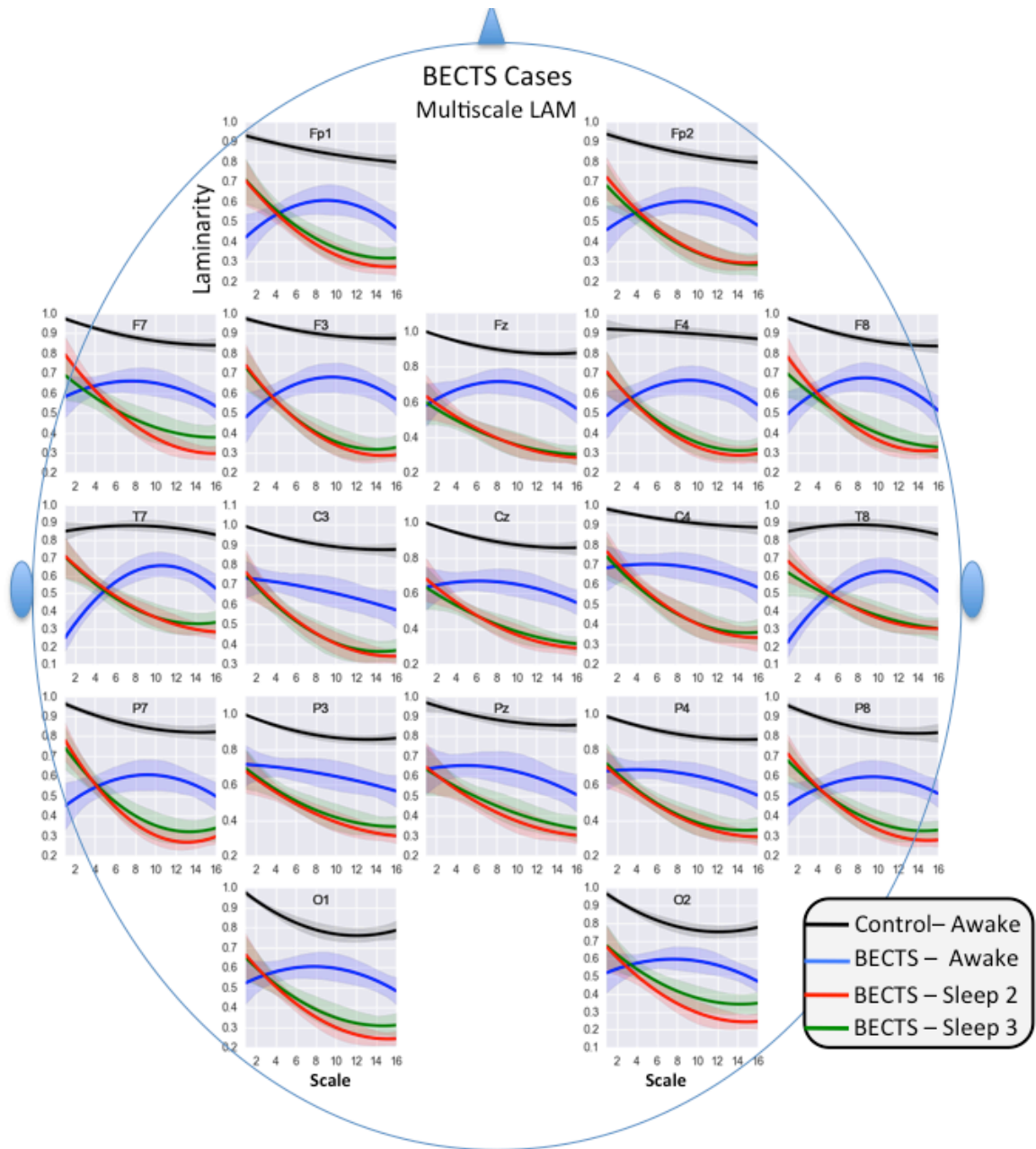
Figure 6. Multiscale LAM curves for awake and asleep cases.

Similarly, multiscale LAM curves reveal differences in lower scales within the centrotemporal region from other regions.

Taken together, multiscale entropy, recurrence rate, determinism, and laminarity appear to be potential biomarkers for BECTS, and may also be useful as research methods for understanding the role of sleep in BECTS nighttime seizures.

D. Discussion

Multiscale analysis, which gives insight into frequency dependence, provides more discriminatory information than quantitative EEG analysis nonlinear analysis of original EEG signals alone.

Our results suggest that changes in the nonlinear values (entropy, RR, DET, and LAM) from awake to sleep 2 and 3 are different in the centrotemporal region from other regions. This raises the question: is this because of epileptiform activity in this region, or are sleep-potentiated dynamical changes in this region of the brain promoting epileptic activity in this region. This study would be much stronger if sleep stage 2 and 3 EEG measurements were available for the control subjects for comparison. It cannot be determined if the differences between awake and sleep multiscale curves for BECTS patients are significantly different from differences that would be observed in controls in equivalent awake or sleep stages. Although the analysis to answer this question would be relatively straightforward, all-night EEG studies are not commonly done on patients who do not have a suspected sleep disorder or sleep-potentiated epilepsy such as BECTS. Hence, this data may be difficult to obtain. Given the challenges faced when trying to access clinical data, as discussed in chapter 1, this was an insurmountable problem for this project.

Although it is clear that significant differences were found between awake BECTS subjects and controls, further analysis is required to determine what neurophysiological differences in the BECTS patients is causing the differences. Nevertheless, our analysis demonstrates that functional brain differences may be present in BECTS patients even when visual review of the awake EEG does not reveal any abnormalities.

E.   *Conclusion*

Modern classification or statistical learning methods present a challenge for the clinician who wants clinical decision support methods that improve patient outcomes through better detection or monitoring of treatment progression. Some data analytics methods find patterns in data that are correlated with outcomes, as we have shown in our results, yet are somewhat opaque. It is difficult to interpret the features and patterns that are predictive in a scientific or biomedical conceptualization. This may be secondary to the clinician, but the research scientist needs to understand the causes that are producing observed results. Simpler models, such as linear regression, make the relationship between the outcome variable and the predictors relatively easy to interpret. However, these simple models are unable to discover more complex relationships between predictor variables. More complex modeling methods, such as support vector machines, decision trees, random forests, and even nearest neighbor methods are more difficult to interpret (James, Witten, Hastie, & Tibshirani, 2013). They often give better predictive results, which may be the primary clinical goal and thus more desirable than immediate scientific understanding.

With the successful implementation of our results into a clinical setting, we can reduce the need of overnight EEG monitoring. This is helpful for both patients and families, who can be highly inconvenienced by these long stays at a clinic. The successful creation of an algorithm,

however, is simply the first step in an implementation process. In order to be implemented as clinical decision support, we will need clinical trials with much larger patient populations, sleep EEGs for control subjects, and the development of software to analyze EEG data for real-time feedback for physicians.

III. References

Adler-Milstein, J., & Pfeifer, E. (2017). Information Blocking: Is It Occurring and What Policy Strategies Can Address It? *The Milbank Quarterly, 95*(1), 117-135.

Amatayukal, M. (2003). Another Layer of Regulations: Research Under HIPAA. *Journal of AHIMA, 74*(1), 16A-D.

Barieri, R. L. (2003). HIPAA: The good, the bad, and the ugly. *OBG Management, 15*(7), 8-11.

Benitez, K., & Malin, B. (2010). Evaluating re-identification risks with respect to the HIPAA privacy rule. *Journal of the American Medical Informatics Association, 17*(2), 169-177. doi:https://doi.org/10.1136/jamia.2009.000026

Bosl, W. (2017). EEG-Derived Endophenotypes In S. L. Santangelo & V. Jagaroo (Eds.), *Cognitive and Behavioral Phenotypes: New Trends for Research, Classification, and Diagnosis in Neuropsychology and Psychiatry*. New York: Springer.

Bosl, W. J., Tager-Flusberg, H., & Nelson, C. A. (2011). EEG Complexity as a Biomarker for Autism Spectrum Disorder. *BMC Medicine, 9*(18).

Bowers, D. (2001). The Health Insurance Portability and Accountability Act: is it really all that bad? *Baylor University Medical Center Proceedings, 14*(4), 347-348.

Burchell, L., Corley, S., Chapman, P., Loomis, R., Marshall, M., Reeves, R., . . . TerMaat, S. (2015). [EHRA Information Blocking Response Letter].

Callenbach, P. M. C., Bouma, P. A. D., Geerts, A. T., Arts, W. F. M., Stroink, H., Peeters, E. A. J., . . . Brouwer, O. F. (2010). Long term outcome of benign childhood epilepsy with centrotemporal spikes:

Dutch Study of Epilepsy in Childhood. *Seizure, 19*, 501-506.

Catarino, A., Churches, O., Baron-Cohen, S., Andrade, A., & Ring, H. (2011). Atypical EEG complexity in autism spectrum conditions: a multiscale entropy analysis. *Clin Neurophysiol, 122*(12), 2375-2383. doi:10.1016/j.clinph.2011.05.004

Chavakula, V., Fernández, I. S., Peters, J. M., Popli, G., Bosl, W., Rakhade, S., . . . Loddenkemper, T. (2013). Automated quantification of spikes. *Epilepsy and Behavior, 26*(2), 143-152. doi:https://doi.org/10.1016/j.yebeh.2012.11.048

Chen, L., Zou, J., & Zhang, J. (2013). *Dynamic feature extraction of epileptic EEG using recurrence quantification analysis*. Paper presented at the Proceedings of the 10th World Congress on Intelligent Control and Automation (WCICA2012).

CMS (Producer). (2010). Meaningful Use Stage 1 Requirements Overview. [Presentation] Retrieved from https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/downloads/MU_Stage1_ReqOverview.pdf

CMS. (2017, January 24). Centralized Repository.   Retrieved from https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/CentralizedRepository-.html

Costa, M., Goldberger, A. L., & Peng, C. K. (2005). Multiscale entropy analysis of biological signals. *Phys Rev E Stat Nonlin Soft Matter Phys, 71*(2 Pt 1), 021906.

Dunlop, A., Graham, T., Leroy, Z., Glanz, K., & Dunlop, B. (2007). The Impact of HIPAA Authorization on Willingness to Participate

in Clinical Research. *Annals of Epidemiology, 17*(11), 899-905. doi:http://doi.org/10.1016/j.annepidem.2007.05.006

Engel, J., Jr. (2001). A proposed diagnostic scheme for people with epileptic seizures and with epilepsy: report of the ILAE Task Force on Classification and Terminology. *Epilepsia, 42*(6), 796-803. doi:epi10401 [pii]

Goldberger, A. L. (1997). Fractal variability versus pathologic periodicity: complexity loss and stereotypy in disease. *Perspect Biol Med, 40*(4), 543-561.

Gomez, C., & Hornero, R. (2010). Entropy and Complexity Analyses in Alzheimer's Disease: An MEG Study. *Open Biomed Eng J, 4*, 223-235. doi:<u>10.2174/1874120701004010223</u> <u>TOBEJ-4-223 [pii]</u>

Gostin, L. O., & Nass, S. (2009). Reforming the HIPAA Privacy Rule

Safeguarding Privacy and Promoting Research. *JAMA, 301*(13), 1373-1375. doi:10.1001/jama.2009.424

H.R.34 - 21st Century Cures Act, Pub. L. No. 114-255 § 3010(a)(3) (2016 December 13, 2016).

H.R. 3103, Pub. L. No. 104-191 § 164 (1996).

Herdman, R., Moses, H. L., States, U., & (U.S.), N. C. P. F. (2006). *Effect of the HIPAA Privacy Rule on Health Research : Proceedings of a Workshop Presented to the National Cancer Policy Forum*, Washington, D.C.

Hirata, Y., & Aihara, K. (2011). Statistical Tests for Serial Dependence and Laminarity on Recurrence Plots. *International Journal of Bifurcation and Chaos*, 1077-1084.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. New York: Springer.

Joost Nicolai, Inge Van Der Linden, Johan B. A. M. Arends, Saskia G. M. Van Mil, Jacobiene W. Weber, Johan S. H. Vles, & Aldenkamp, A. P. (2007). EEG Characteristics Related to Educational Impairments in Children with Benign Childhood Epilepsy with Centrotemporal Spikes. *Epilepsia, 48*(11), 2093-2100. doi:10.1111/j.1528-1167.2007.01203.x

Kemp, B., Varri, A., Rosa, A. C., Nielson, K. D., & Gade, J. (1993). A simple format for exchange of digitized polygraphic recordings. *Electroencephalography and Clinical Neurophysiology, 82*(5), 391-393. doi:10.1016/0013-4694(92)90009-7

Khoury, M. J., & Loannidis, J. P. A. (2014). Big data meets public health. *Science, 346*(6213), 1054-1055. doi:10.1126/science.aaa2709

Komalapriya, C., Thiel, M., Romano, M. C., Marwan, N., Schwarz, U., & Kurths, J. (2008). Reconstruction of a system's dynamics from short trajectories. *Phys Rev E Stat Nonlin Soft Matter Phys, 78*(6 Pt 2), 066217.

Kumekawa, J. (2005). Overview: HIPAA: How Our Health Care World Has Changed. *The Online Journal of Issues in Nursing, 10*(2).

Kwan, P., & Brodie, M. J. (2001). Neuropsychological effects of epilepsy and antiepileptic drugs. *The Lancet, 357*(9251), 216-222. doi:<u>http://dx.doi.org/10.1016/S0140-6736(00)03600-X</u>

Mandl, K. D., & Kohane, I. S. (2012). Escaping the EHR Trap -- The Future of Health IT. *New England Journal of Medicine, 322*, 2240-2242. doi:10.1056/NEJMp1203102

Marchibroda, J. (2014). *"Health Policy Brief: Interoperability,"* Retrieved from <u>http://www.healthaffairs.org/healthpolicybriefs/brief.php?brief_id=122</u>

Margolis, R., Derr, L., Dunn, M., Huerta, M., Larkin, J., Sheehan, J., . . . Green, E. D. (2014). The National Institutes of Health's Big Data to Knowledge (BD2K) initiative: capitalizing on biomedical big data. *Journal of the American Medical Informatics Association, 21*(6), 957-958. doi:<u>https://doi.org/10.1136/amiajnl-2014-002974</u>

Marwan, N. (2012). http://www.recurrence-plot.tk/.

Marwan, N., Romano, M. C., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports, 438*, 237-329.

Miziara, C. S. M. G., & Manreza, M. L. G. (2002). Benign Focal Epilepsy of Childhood with Centrotemporal Spikes (BECTS). *Arq Neuropsiquiatr, 60*(2-B), 390-394.

Ness, R. B. (2007). Influence of the HIPAA Privacy Rule on Health Research. *JAMA, 298*(18), 2164-2170. doi:10.1001/jama.298.18.2164

Ngamga, E. J., Bialonski, S., Marwan, N., Kurths, J., Geier, C., & Lehnertz, K. (2016). Evaluation of selected recurrence measures in discriminating pre-ictal and inter-ictal periods from epileptic EEG data. *Physics Letters A, 380*(16), 1419-1425. doi:10.1016/J.PHYSLETA.2016.02.024

NIH. (2003). *Institutional Review Boards and the HIPAA Privacy Rule*. Retrieved from https://privacyruleandresearch.nih.gov/irbandprivacyrule.asp.

NIH. (2004). *Clinical Research and the HIPAA Privacy Rule*. Retrieved from https://privacyruleandresearch.nih.gov/clin_research.asp.

Niknazar, M., Mousavi, S. R., Vahdat, B. V., & Sayyah, M. (2013). A new framework based on recurrence quantification analysis for epileptic seizure detection. *IEEE J Biomed Health Inform, 17*(3), 572-578.

Norris, P. R., Stein, P. K., & Morris, J. A., Jr. (2008). Reduced heart rate multiscale entropy predicts death in critical illness: a study of physiologic complexity in 285 trauma patients. *J Crit Care, 23*(3), 399-405.

OCR. (2017). Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule.

ONC. (2015). *2015 Report to Congress on Health Information Blocking*. Retrieved from Washington, DC: https://www.healthit.gov/sites/default/files/reports/info_blocking_040915.pdf

Other requirements relating to uses and disclosures of protected health information., https://www.law.cornell.edu/cfr/text/45/164.514, § 164.514 Stat. (2000 28 December 2000).

Peng, C.-K., Costa, M., & Goldberger, A. L. (2009). Adaptive Data Analysis of Complex Fluctuations in Physiologic Time Series. *Advances in Adaptive Data Analysis, 1*(1), 61-70.

Schinkel, S., Marwan, N., & Kurths, J. (2009). Brain signal analysis based on recurrences. *J Physiol Paris, 103*(6), 315-323. doi:S0928-4257(09)00029-1 [pii] 10.1016/j.jphysparis.2009.05.007

Shields, D., & III, O. C. S. (2009). Benign epilepsy with centrotemporal spikes. *Epilepsia, 50*(Suppl. 8), 10-15. doi:10.1111/j.1528-1167.2009.02229.x

Solove, D. J. (2013). HIPAA Turns 10: Analyzing the Past, Present and Future Impact. *Journal of AHIMA, 84*(4), 22-28.

Stam, C. J. (2005). Nonlinear dynamical analysis of EEG and MEG: review of an emerging field. *Clin Neurophysiol, 116*(10), 2266-2301.

Takahashi, T., Cho, R. Y., Mizuno, T., Kikuchi, M., Murata, T., Takahashi, K., & Wada, Y. (2010). Antipsychotics reverse abnormal EEG complexity in drug-naive schizophrenia: a multiscale entropy analysis. *Neuroimage, 51*(1), 173-182. doi:S1053-8119(10)00157-6 [pii]

10.1016/j.neuroimage.2010.02.009

Vandendriessche, B., Peperstraete, H., Rogge, E., Cauwels, P., Hoste, E., Stiedl, O., . . . Cauwels, A. (2014). A multiscale entropy-based tool for scoring severity of systemic inflammation. *Crit Care Med, 42*(8), e560-569. doi:10.1097/CCM.0000000000000299

Webber, C. L., & Marwan, N. (Eds.). (2015). *Recurrence Quantification Analysis*. New York: Springer.

Webber, C. L., & Zbilut, J. P. (2005). Recurrence quantitative analysis of nonlinear dynamical systems. In M. A. Riley & G. Van Orden (Eds.), *Tutorials in contemporary nonlinear methods for the behavioral sciences*: National Science Foundation.

Wei, W. (2015). The Effects of HIPAA's Privacy Rule on Medical Research. *Faculty Curated Undergradate Works, Paper 22*.

Xie, H.-B., He, W.-X., & Liu, H. (2008). Measuring time series regularity using nonlinear similarity-based sample entropy. *Physics Letters A, 372*(48), 7140-7146. doi:doi:10.1016/j.physleta.2008.10.049

IV. Appendix

A. 18 PHI Identifiers.

Adapted from ("Other requirements relating to uses and disclosures of protected health

information.," 2000)

1. Names
2. Geographic subdivisions greater than state
   a. The initial three digits of a zip code may be included if following conditions are met:
      1. There are more than 20,000 people within the zip codes starting with those three digits
      2. The initial three digits of regions that do not have 20,000 people within them are changed to 000
3. All date elements except for year in patients under 89 years old, and all age elements (including year) for patients over 89 years old
4. Telephone numbers
5. Fax numbers
6. Electronic mail addresses
7. Social security numbers
8. Medical record numbers
9. Health plan beneficiary numbers
10. Account numbers
11. Certificate/license numbers
12. Vehicle identifiers, serial numbers, and license plate numbers
13. Device identifiers and serial numbers
14. Web universal resource locators (URLs)
15. Internet protocol (IP) address numbers
16. Biometric identifiers including voice and finger prints
17. Full face photographs or photographs containing identifying features
18. Any other unique identifier (except a code created for the original covered entity to be capable of re-identification)