1) Abstract of the study

The purpose of this study is to create a database of clinical electroencephalogram signals using both archival data amassed over the past decade as well as new recordings, generated by the Neurology Department at Temple University. The archive presently contains approximately 24,000 EEG recordings. Accompanying physician notes also indicate various details about each patient such as medical history, presenting complaints, and medications. The proposed study will organize, de-identify, and catalog these signals so they can be subsequently released to the research community in a non-proprietary file format. Researchers ranging from neuroscientists to engineers to machine learning scientists will use this data as a tool for understanding brain function. To our knowledge there has been no attempt ever made to create an EEG data bank of the size and scope we are proposing. Even though a particular neurologist may read thousands of EEGs over the course of a career, a computer can identify trends and variability in tens of thousands of recordings. We therefore expect that this will be an important contribution to the medical and scientific community.

2) Protocol Title

Development of Electroencephalogram Data Corpus

3) IRB Review History

N/A

4) Investigator

Iyad Obeid, PhD

Department of Electrical and Computer Engineering

215-204-9033 / iobeid@temple.edu

5) Objectives

The objective of this study is to create the world's largest database of electroencephalogram signals. Combined with appropriate identifiers such as patient demographics, medical history, and medications, this data will provide an unprecedented ability to study the variability and functionality of the human brain. At the conclusion of the study, the data will be made available to the international research community for study.

6) Background

EEG is a relatively mature clinical modality in which an array of electrodes (typically 21 but up to 64) is affixed to the scalp. These electrodes transduce electrical signals that originate from the cortical neurons situated underneath them. Owing to the spatial

distribution of the electrode across the scalp, EEG is said to have poor spatial resolution but excellent temporal resolution.

The use of computers and computer algorithms to study EEG signals is currently an active research area, especially with respect to epilepsy, traumatic brain injury, stroke detection, and pharmacology. However, despite the facts that (a) EEG is a common and relatively inexpensive clinical tool and (b) modern computing technology can reasonably analyze terabytes of data, researchers in this field have only worked with modest data sets. The largest publically available EEG data set is the CHB-MIT Scalp EEG Database published in 2000 with data from 22 epileptic pediatric patients covering a total of 916 hours. While substantial, modern computing technology is capable of tracking trends and variations in substantially larger datasets. In contrast, the proposed study will create a dataset spanning approximately 12,000 patients over 10,000-15,000 hours. This represents a ten-fold increase in raw data but over 500-fold increase in terms of number of patients (as compared to the CHB-MIT database). The diversity of patients in terms of age, ethnicity, socio-economic background, and medical history points to a broad range of variability in the data.

The advantage of using computers to analyze massive data sets is the prospect of identifying trends that are not readily obvious to the naked eye. Even a neurologist who reads thousands of EEGs over the course of a career may not notice subtle trends or variability patterns that a computer can detect. In our case, we expect that the database we create will be used by others to answer valuable questions such as how brain activity change in response to various medications and medical conditions, and whether these changes vary across demographic lines.

7) Setting of the Human Research

We are proposing to take a disparate collection of archival as well as newly generated clinical EEG data and create a single unified dataset for the research community. All the archival data we will use has been collected over the past ten years by clinicians at Temple University Hospital, Department of Neurology. New EEG recordings made in the Neurology Department will also be added to the databse.

The process of de-identifying, organizing, tagging, and converting the data to a non-proprietary file format will be performed on-site at Temple Neurology. In this manner, any sensitive data bearing personal health information will not be compromised. Once the data has been de-identified, some of the post-processing may take place

on Main Campus in Dr. Obeid's laboratory in the College of Engineering.

8) Resources Available to Conduct the Human Research

The only facilities required to generate the proposed dataset are standard computing tools such as personal computers and large backup hard drives. Software tools for data aggregation such as Matlab and C/C++ are already in possession by Temple University.

As described in (7), the proposed work will take place mostly at Temple Neurology as well as on main campus in the College of Engineering. We expect this work to require 6-9 months to complete.

The subjects in this study will be a combination of legacy patients whose records have been archived at Temple Hospital, as well as new (ongoing) patients receiving routine clinical care.

All personnel working with sensitive medical records will be required to comply with Temple's policy on personal health information. Specifically, they will pass the Web Based Research Compliance Courseware (CITI Training Program) and read through the three online primers on HIPAA regulations located at http://www.temple.edu/research/regaffairs/irb/irb_references.html

9) Prior Approvals

N/A

10) Study Design

a) Recruitment Methods

There will be no need to recruit subjects since this is a retroactive study of existing archival data.

b) Inclusion and Exclusion Criteria

The only criterion for inclusion is a complete EEG exam with accompanying medical information that can identify the age, gender, medical condition, and medications of the subject.

c) Local Number of Subjects

All archival data to be used will have been collected at Temple Neurology over the past ten years. New data will be added to the archive monthly as the records become available. Based on past performance, it is expected that approximately 3000 new EEGs will be made and added to the database each year.

d) Study-Wide Number of Subjects

Based on an initial estimate, we believe there will be EEGs from an initial archive of about 24,000 records plus approximately 3,000 more per year.

e) Study Timelines

Generating the initial dataset from the archival EEG data is estimate to take 6-9 months. New data will be added on a continuous basis as it becomes available.

f) Study Endpoints

The study will remain open so that new data can be added to the database as it becomes available.

g) Procedures Involved in the Human Research

As part of their standard clinical practice, the Neurology Department at Temple University Hospital maintains an archive of approximately 24,000 clinical EEGs that span roughly ten years. The EEGs are saved in a proprietary clinical format. In addition, the Department has clinician notes (saved in MS Word format) that give a brief synopsis of the medical condition of each patient including the general state of their health as well as any medications they are taking. For some patients, there is also structural brain data in the form of MRIs.

The goal of this protocol will be to create the world's largest database of EEG data by re-purposing Temple Neurology's massive archive. Each of the archived records will be stripped of the patient's name and the accompanying video will be deleted also. However other information such as date and time of recording, gender, age, ethnicity, relevant medical history, and medications will be retained since the archive loses its value without such important indicators. Examples of EEG EDF and physician report files that have been redacted in accordance with this protocol have been uploaded as addendums to this document. Each subject will be referred to by a unique "subject identification number" that is completely not related to the patent's medical records number or social security number. The investigators will use an "open" (i.e. non-proprietary) file format such as the European Data Format. Critical moments from the EEG will also be flagged for later study.

The process of creating this database will require a combination of manual and automated data processing. Whenever there is an opportunity to automate the de-identification and archival process, we will write a custom computer program to handle that. In other

cases, the data may be manipulated by hand. The work will be performed by graduate and undergraduate students under the supervision of Dr. Obeid.

Once the initial database (from archival data) is complete, new records will be added to the database monthly. It is anticipated that these records will be superior to archival data owing to improved electronic health record information for the subjects.

Once the database is complete, it will be made available for free to researchers on the World Wide Web with an annual or semiannual update of new records. Users will be asked to register their names and affiliations before being allowed to download the data. Users will be obliged to cite Temple University as the source of their data if they choose to publish any related findings.

Since it may be desirable to retrieve the identity of a particular database subject for further analysis, a “key” will be created that maps each unique subject identifier number to the subject’s name and/or Temple medical records ID number. This file will be kept under the sole possession of Dr. Jacobson with one hard copy and one soft copy burned to a CD.

In addition to making this data available to the global research community, the PIs will be performing analysis on this data as well, both internally at Temple University and in collaboration with external groups. When collaborating with third parties, Temple’s investigators will be expressly forbidden from sharing any privileged patient information.

h) Data and Specimen Banking

The de-identified data will be banked indefinitely as described in the previous section.

i) Data Management

The investigators expect to do the majority of the data preparation on-site at the Temple Neurology department. However once the data has been de-identified, it may be necessary to transport some of that data back to Dr. Obeid’s laboratory on Main Campus. Data will not be placed in the “cloud” but will rather be transported on a hard drive or other physical media. In this manner, the data will be secured until the entire archive is ready for dissemination.

j) Confidentiality

In order to ensure confidentiality, the data will be physically kept at Temple Neurology until they have been de-identified. At that time, the data will either remain at Temple Neurology or brought to Dr. Obeid’s laboratory on Main Campus using physical media. The

data will not “enter the cloud” until it has been completely de-identified and the entire database has been constructed.

Confidentiality will also be ensured by removing patient name and video from the records. Patients will be referred to in the database using an anonymous ID number that bears no resemblance to their Temple medical records number. Although a “key” file will be set up to map patients’ true identities to their anonymous ID number, this file will be in the sole ownership of Dr. Mercedes Jacobson (Temple Neurology). Any investigator wishing to access this key will require further IRB approval.

k) Provisions to Monitor the Data to Ensure the Safety of subjects

N/A

l) Withdrawal of Subjects

N/A

11) Risks to Subjects

As this is a study of de-identified data, there is no risk to subjects.

12) Potential Benefits to Subjects

Since this is a retroactive study of clinical data, there is no direct benefit to the subjects.

13) Provisions to Protect the Privacy Interests of Subjects

The patient data will be de-identified on-site at Temple Neurology. Although an “ID list” will be created that maps an anonymous patient ID number to their true identity, this list will be sealed and remain in the possession of Dr. Mercedes Jacobson at Temple Neurology.

14) Compensation for Research-Related Injury

N/A

15) Economic Burden to Subjects

None

16) Consent Process

Since this study is only considering de-identified data, it will not be necessary to acquire consent from the patients.

17) Process to Document Consent in Writing

N/A

18) Vulnerable Populations

While it is possible that the EEG archive we will be working with contains data from vulnerable populations, we will not be removing them from the study.

19) Drugs or Devices

N/A

20) Multi-Site Human Research

N/A

21) Sharing of Results with Subjects

N/A

ADDENDUM: Waiver or Alteration of the Consent Process

We believe this research qualifies for a Waiver under Criterion 1 of form 415C

- This research is NOT FDA-regulated
- This research does NOT involve non-viable neonates
- This research involves no more than minimal risk to the subjects: The primary purpose of this work is to de-identify a data archive. Once the data have been de-identified, there is no risk to the subjects since there will be no way of linking the data back to them. For contingency purposes, we will retain a “key” that links subjects’ true identities to their anonymous data identification number. That key will remain in the sole possession of Dr. Mercedes Jacobson at Temple University Hospital and will not be unsealed without first seeking an addendum to this IRB.
- The research could NOT be practicably carried out without the waiver or alteration. Since the study aims to include data from over 12,000 subjects, some of which was collected as far back as 10 years ago, it would be too burdensome to track down and seek consent from each of them.
- Whenever appropriate, the subjects will be provided with additional pertinent information after participation. This provision will not be applicable since subjects will not know they are included in this study.

ADDENDUM: HIPAA Waiver of Authorization

We believe this research qualifies for a HIPAA Waiver of Authorization on the grounds that:

- There is an adequate plan to protect the identifiers from improper use and disclosure. See Addendum above for detailed explanation.
- There is an adequate plan to destroy the identifiers at the earliest opportunity consistent with conduct of the research, unless there is a health or research justification for retaining the identifiers or such retention is otherwise required by law. Since the final dataset we create will be put in the public domain, there is no formal “end-date” for the research in question. Therefore we do not foresee any timeframe after which it will be appropriate to destroy the identifier key. However, the key will remain confidential. In the event that Dr. Jacobson leaves the university or otherwise does not wish to maintain the key, a replacement investigator will be identified and the IRB will be informed.
- There are adequate written assurances that the protected health information will not be reused or disclosed to any other person or entity, except as required by law, for authorized oversight of the research study, or for other research for which the use or disclosure of protected health information for which an authorization or opportunity to agree or object is not required by.45 CFR 164.512. As we have stated in this protocol, the patient identification will be kept confidential in the possession of a single investigator (Dr. Jacobson). Any access to the key will require an amendment to this protocol and approval from the IRB.
- The research could not practicably be conducted without the waiver or alteration. See Addendum above for detailed explanation.
- The research could not practicably be conducted without access to and use of the protected health information. Since the data, in its existing format, is identified with patient name, patient medical records number, and patient video, it will be impossible for us to create the desired anonymous data archive without first de-identifying the data. This will, by definition, necessitate access to protected health information. However, only a limited staff will be involved in that effort and they will be supervised by the PI and co-investigators listed on this protocol. The work will be done on site at Temple Neurology, so that private health information will never physically leave the hospital.