NHS
Health Research Authority

This protocol has regard for the HRA guidance and order of content

FULL/LONG TITLE OF THE STUDY

Retrospective Analysis of Resting-State EEG in the Diagnosis of Epilepsy to validate a computational biomarker for seizure susceptibility

SHORT STUDY TITLE / ACRONYM

Computational Decision Support in Epilepsy using retrospective EEG

PROTOCOL VERSION NUMBER AND DATE

Version 10 01 October 2019

RESEARCH REFERENCE NUMBERS

IRAS Number:	260729
SPONSORS Number:	CFT/103939
FUNDERS Number:	103939

Computational Decision Support in
Epilepsy using retrospective EEG

SIGNATURE PAGE

The undersigned confirm that the following protocol has been agreed and accepted and that the Chief Investigator agrees to conduct the study in compliance with the approved protocol and will adhere to the principles outlined in the Declaration of Helsinki, the Sponsor's SOPs, and other regulatory requirement.

I agree to ensure that the confidential information contained in this document will not be used for any other purpose other than the evaluation or conduct of the investigation without the prior written consent of the Sponsor

I also confirm that I will make the findings of the study publicly available through publication or other dissemination tools without any unnecessary delay and that an honest accurate and transparent account of the study will be given; and that any discrepancies from the study as planned in this protocol will be explained.

For and on behalf of the Study Sponsor:

Signature:
.....

Date:
...../...../.....

Name (please print):

Sharon Hudson.....

Position:

Research Manager.....

Chief Investigator:

Signature:
.....

Date:
...../...../.....

Name: (please print):

Dr. Rohit Shankar.....

LIST of CONTENTS

KEY STUDY CONTACTS

Chief Investigator	Dr. Rohit Shankar Email - rohit.shankar@nhs.net Telephone – 01872 221553
Study Co-ordinator	Sharon Hudson Email - Sharon.hudson2@nhs.net Telephone – 01209 204021
Sponsor	Cornwall Partnership NHS Foundation Trust Sponsor Contact and Representative: Mrs Sharon Hudson Kernow Building Wilson Way Pool TR15 3QE Email - Sharon.hudson2@nhs.net Telephone – 01209 204021
Joint-sponsor(s)/co-sponsor(s)	N/A
Funder(s)	Innovate UK Neuronostics Ltd.
Key Protocol Contributors	Dr. Wessel Woldman Email – w.woldman@neuronostics.com Telephone – 01392 575727 Mr Ross Parkman Email – r.parkman@neuronostics.com Telephone – 01392 575720
Committees	N/A

STUDY SUMMARY

Study Title	Retrospective Analysis of Resting-State EEG in the Diagnosis of Epilepsy to validate a computational bio-marker for seizure susceptibility
Internal ref. no. (or short title)	Computational Decision Support in Epilepsy using retrospective EEG
Study Design	Quantitative (retrospective) Research
Study Participants	Subjects that underwent an EEG as part of suspected epilepsy; they either ended up with a diagnosis of epilepsy or a differential diagnosis (syncope, non-epileptic seizures (PNES))
Planned Size of Sample (if applicable)	825
Follow up duration (if applicable)	N/A
Planned Study Period	01/11/2019-31/12/2020
Research Question/Aim(s)	1. Validate a set of computational biomarkers for seizure susceptibility on a large cohort of study participants that were diagnosed with epilepsy and controls that ended up with another diagnosis. This means we will examine if the methodology works and has the potential to contribute to lower misdiagnosis-rates. 2. Examine whether the method has the potential to reduce waiting times and the number of clinical appointments before final diagnosis.

FUNDING AND SUPPORT IN KIND

FUNDER(S) (Names and contact details of ALL organisations providing funding and/or support in kind for this study)	**FINANCIAL AND NON FINANCIALSUPPORT GIVEN**
Innovate UK	Funding: BioEP: Proof of Concept to Neuronostics and the University of Exeter (via grant 103939, Health & Life Sciences - Round 2)
Neuronostics Ltd	Commercial Funding

ROLE OF STUDY SPONSOR AND FUNDER

The Cornwall Partnership NHS Foundation Trust will act as sponsor for this study, assuming overall responsibility of the work carried out in this project. The study sponsor will ensure that the research team has access to resources and support to deliver the research as proposed and that responsibilities for management, monitoring and reporting of the research are in place prior to the study commencing. The sponsor will ensure that there is agreement on recording, reporting and reviewing significant developments as the research proceeds and approve any modifications to design, obtaining requisite regulatory authority. The sponsor will assume responsibility for operating the management and monitoring systems of the research. Prior to the study commencing the sponsor will be satisfied that:

- The research will respect the dignity, rights, safety and well-being of participants and the relationship with healthcare professionals.
- Where appropriate the research has been reviewed and approved by an NHS Research Ethics Committee and/or the Health Research Authority Approval Programme.
- The Chief Investigator, and other key researchers have the requisite expertise and have access needed to conduct the research successfully.
- The arrangements and resources proposed for the research will allow the collection of high quality, accurate data and the systems and resources will allow appropriate data analysis and data protection.
- Organisations and individuals involved in the research agree the division of responsibilities between them.
- Arrangements are in place for the sponsor and other stakeholder organisations to be alerted to significant developments during the study, whether in relation to the safety of individuals or scientific direction.
- There are arrangements for the conclusion of the study including appropriate plans for the dissemination of findings.

The sponsor plays no role in the design of this study, and will have no role in data analysis or interpretation, or writing up of findings of the study.

Role of Funder(s)

Innovate UK

The role of this funder is to provide grant funding to enable the study to be conducted.

Neuronostics Ltd.

The role if this funder is to provide financial support to the study either as the recipient of Innovate UK grant funds or through commercial fundraising activities. Staff employed by the company will perform the following roles: data storage, curation, archiving and analysis, manuscript writing and dissemination of results.

ROLES AND RESPONSIBILITIES OF STUDY MANAGEMENT COMMITEES/GROUPS & INDIVIDUALS

Members of a local chapter of Epilepsy Action (South-West, lead by Mr. Simon Privett) were engaged during several Coffee & Chat meetings in order to receive feedback on our research study to ensure the potential outcomes align with the priorities of people with lived experience. Furthermore, Mr. Privett was consulted and reviewed the IRAS & HRA drafts: he provided feedback on the underlying rationale, the clarity of the document, and the data management plan.

PROTOCOL CONTRIBUTORS

Protocol developed by Dr Rohit Shankar (CI), Sharon Hudson (SC), Ross Parkman (Data Controller), Dr Wessel Woldman (Data Analyst), Professor John Terry (Project Mentor), and reviewed by Professor Rod Taylor (independent statistical expert).

Study design by Dr Wessel Woldman and Professor John Terry.

Data acquisition by the Participating Sites.

Data collection by the Sponsor through Dr Rohit Shankar and Sharon Hudson (SC).

Data analysis by Dr Wessel Woldman and Dr Leandro Junges.

Data curation and secure storage by Ross Parkman.

Interpretation of results by Dr Rohit Shankar, Dr Wessel Woldman, Dr Leandro Junges and Professor John Terry.

Manuscript writing and dissemination of results by Dr Rohit Shankar, Dr Wessel Woldman, Dr Leandro Junges and Professor John Terry.

The Sponsor (Cornwall Partnership NHS Foundation Trust) have no direct involvement or final decision on study design and data analysis.

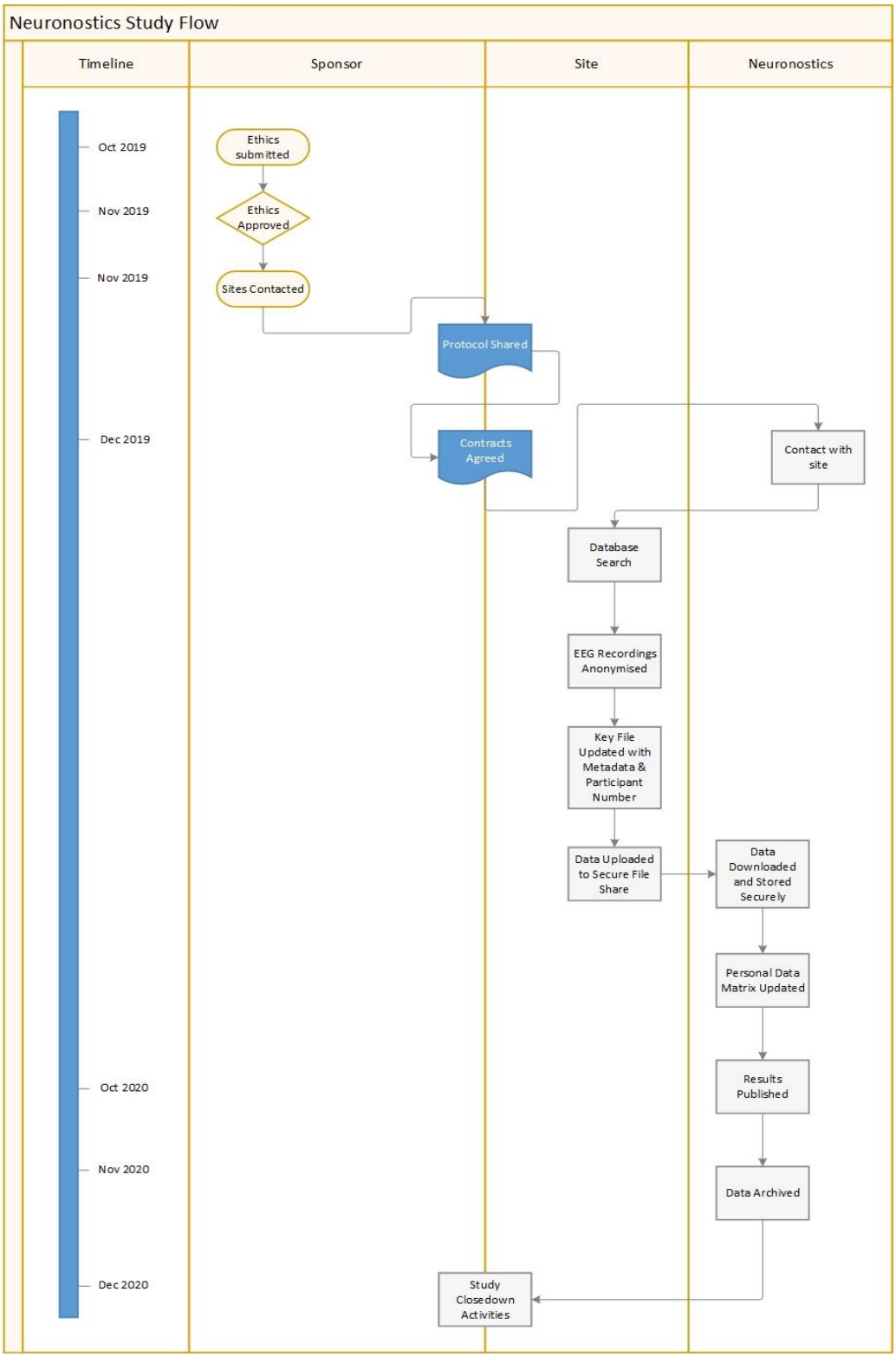
The public involvement group (Epilepsy Action chapter of the South-West lead by Mr Simon Privett) have been consulted to ensure that the aims and outputs of the study align a worthwhile topic for people with lived experience, and that there are no significant risks associated with the study.

A group of clinical neurophysiologists and neurologists (Taunton, Exeter, Cornwall, Bristol) were consulted to ensure the aim of this study addressed a worthwhile topic of public concern, and that there are no significant risks associated with the study. Their opinion was sought on anonymisation and pseudonymisation of the data.

KEY WORDS:

Epilepsy; seizure; EEG; diagnostic delay; misdiagnosis; biomarkers.

STUDY FLOW CHART



Computational Decision Support in
Epilepsy using retrospective EEG

STUDY PROTOCOL

Retrospective Analysis of Resting-State EEG in the Diagnosis of Epilepsy to validate a computational biomarker for seizure susceptibility

1 BACKGROUND

The accurate diagnosis and prognosis of epilepsy represents a significant unmet medical need: ~0.8% global population (including 600k in UK) have epilepsy. Due to the unpredictable nature of seizures, epilepsy is difficult to diagnose and treat. In the UK 125k people/y are referred to first seizure clinics with suspected epilepsy of which 40k receive a confirmed diagnosis of epilepsy. At present - in the absence of observable epileptiform abnormalities (such as spikes or seizures observed in an EEG) or a lesion in an MRI scan) - there are no clinically robust markers of epilepsy. Estimates suggest that in the developed world over 2M people annually attend a first-seizure clinic suspected of having epilepsy. Of these between one third and one half will have a first inconclusive EEG (Smith, 2005). Consequently, there is an average delay of a year to confirm a diagnosis of epilepsy (Joint Epilepsy Council Report, 2011).

Mathematical models provide a powerful and useful tool with which to identify and understand biological mechanisms that may lead to the risk of having seizures as well as how they generate, propagate and terminate (Wendling, 2005). Mathematical models that combine experimental and clinical detail at diverse scales have revealed the importance of many microscopic and macroscopic mechanisms in the generation of seizure-like activity, ranging from genetic and molecular mechanisms to changes in the excitability of neural populations leading to the generation of pathological oscillations (for review see Woldman & Terry (2015); Soltesz & Staley (2008)). Due to the increased availability of data recordings (EEG, MRI, MEG, CT, PET), there has been a significant increase in research studies that aim to identify novel biomarkers from these recordings with potential clinical value, using various different techniques (e.g. time-series analysis, computational modelling, machine learning).

By combining mathematical and computational techniques, we have identified properties in the resting-state EEG (eyes closed, relaxed) of people with epilepsy that differ from those of controls as well as their first-degree relatives (Chowdhury et al., 2014). Developing these approaches and applying them to routine recordings from individuals with epilepsy against a control cohort (Schmidt et al., 2016) revealed levels of diagnostic accuracy similar to current general (i.e. non-specialist) neurology practices (60% sensitivity, 87% specificity, N=68). Crucially, our method correctly classified several subjects using their first EEG, whereas clinical diagnosis was confirmed only after prolonged telemetric recordings over many months.

Since our methods and analysis depend on short segments of resting-state EEG only, its accuracy and efficacy do not rely on capturing epileptiform abnormalities, in contrast to the current use of EEG in diagnosing epilepsy. Since many EEGs return negative, clinicians are often faced with the problem of deciding on whether to opt for longer recordings of EEG or ambulatory or video EEG, which is currently the final method in the diagnostic cascade. This is time-consuming, expensive and relies on the availability and expertise of trained EEG-readers. By optimally interrogating short segments of background activity with mathematical and computational analysis, our methods, in the short term, provide additional evidence that could guide clinicians in future diagnostic steps.

We will conduct a retrospective study consisting of a large cohort (N = 825) of EEG recordings from people with epilepsy and representative controls (people that came into the clinic with suspected seizures / epilepsy, but ended up with a differential diagnosis) from centres running epilepsy clinics. We will derive a set of computational biomarkers from relevant EEG segments free of epileptiform discharges; therefore, this method could be effective even if the EEG is negative in the traditional sense. We will use the first EEG present in clinic available for each participant, and compare the performance of our method against the established diagnostic pathway, allowing us to examine whether our methodology could reduce diagnostic delay and decrease misdiagnosis rates (see

Schmidt et al. (2016); Woldman et al. (2019)). Retrospective data collection would take place at national centres through NHS Trusts.

This protocol outlines an investigation into the accuracy and efficacy of a set of computational biomarkers for epilepsy on a large cohort of representative recordings, in order to investigate their potential in providing additional decision support in the process of diagnosing people with suspected epilepsy.

2 RATIONALE

Neurological conditions are becoming an increasing healthcare burden globally. The financial costs are significant, estimated to exceed €145B in 2010 in Europe (Gustavsson et al., 2010) A significant proportion of these costs are due to misdiagnosis and delayed diagnosis. For example, in the UK alone, the cost of misdiagnosis of epilepsy is estimated at £268M annually (Joint Epilepsy Council Report, 2011). The personal burden of these conditions is also significant. An estimated 154M Disability Adjusted Life Years (6.3% of global disability burden) in 2015 were attributable to neurological conditions (Feigin et al., 2017).

Clinical diagnosis of epilepsy relies on a detailed case history, and involves typically at least one electroencephalogram (EEG) monitoring. However, conventional EEG has low sensitivity and specificity (Smith, 2005), and people may have non-epileptic seizures or mannerisms similar to the symptoms typically observed during seizures, but that are not caused by the synchronous neuronal activity underlying epileptic seizures. Epilepsy is challenging to diagnose and misdiagnosis of epilepsy is a serious issue, associated with several negative personal and socio-economic consequences for a subject, as well as significant costs at the national and global level. Over half a million people annually experience diagnostic delays in the developed world, with misdiagnosis-rates estimated in the range 20-40% (Smith et al. (1999); Chowdhury et al. (2008)). In numerous conversations with people with lived experience it became apparent that diagnostic delay was experienced in a significant number of cases and was a strongly negative experience. Furthermore, in a significant number of these cases the people with lived experience voiced frustration with getting seemingly contradictory indications after tests (such as EEGs) were taken from different consultants, with unclear explanations and outcomes.

Taken together, these statistics illustrate the necessity for research into objective, accurate and robust biomarkers that can aid diagnosis of epilepsy and seizures, in order to provide more accurate diagnosis faster, which would lead to better health and economic outcomes.

Since the methods we apply use background EEG only, they have the potential to be applicable and effective for all recordings, and crucially, does not depend on the observation of interictal discharges or seizures. However, given the fact that a misdiagnosis comes with significant consequences it is crucial to examine the overall efficacy and robustness of our methods on a large data-set of representative cases. By establishing the sensitivity and specificity in a retrospective case-control study we will be able to examine the performance of the computational markers: 1) whether it accurately differentiates representative controls (differential diagnosis, such as syncope) from people with epilepsy and 2) whether it differentiates people with focal epilepsy from people with generalised epilepsy and 3) whether it correctly classify people with epilepsy on what is currently regarded as a “negative” recording (i.e. no epileptiform discharges or spikes). Taken together, these algorithms could then provide objective decision support for clinicians during the diagnostic pathway.

3 THEORETICAL FRAMEWORK

The performance of our biomarkers has been demonstrated in retrospective analyses of EEG recordings from smaller cohorts. In a test cohort of 98 adult subjects with confirmed epilepsies (both generalised and focal epilepsies), our methodology achieved 60% sensitivity and 87% specificity, when it was used as a standalone diagnostic test (e.g. with no clinical case history or clinical analysis

of EEG recordings). Following the BMJ guidelines for diagnostics trials (Guidelines: Best Practice), this ability to differentiate between the “grossly affected” (e.g. with epilepsy) from the “clearly well” (e.g. without epilepsy) is equivalent to a Phase I clinical trial. In several cases our method classified someone with epilepsy using a segment of data from an EEG that would be considered negative under the current clinical pathway. These results compare favourably with levels of sensitivity and specificity achieved clinically following the complete pathway to diagnosis, involving potentially multiple follow-ups. Smith estimates these to be 25-56% sensitivity and 78-98% specificity (Smith, 2005) for EEG.

This methodology and analysis have been described and published in peer-reviewed journals (Schmidt et al., 2014 PLoS Computational Biology https://doi.org/10.1371/journal.pcbi.1003947 ; Chowdhury et al., 2014 PLoS One https://doi.org/10.1371/journal.pone.0110136 ; Schmidt et al., 2016 Epilepsia https://doi.org/10.1111/epi.13481), and the most recent work on differentiating focal and generalized epilepsies is currently under review and available on a pre-print server (Woldman et al., 2019 http://biorxiv.org/cgi/content/short/576785v1).

The next necessary step to examine the full potential of the methodology in providing decision support in epilepsy by reducing diagnostic delay and cases of misdiagnosis, is to test it on a new independent data-set and against a set of respresentative controls (Phase II). The representative control cohort would consist of EEG recordings from people that came into clinic with suspected seizures or epilepsy but ended up with a differential diagnosis such as syncope or psychogenic non-epileptic seizures (PNES). To the best of our knowledge, there is no other work done previously, which attempts to reduce diagnostic delay and misdiagnosis rates during the diagnosis of epilepsy using resting-state EEG segments.

4 RESEARCH QUESTION/AIM(S)

4.1 Objectives

The primary aim is to validate a set of computational biomarkers as potential decision support in epilepsy on a large cohort of study participants that were diagnosed with epilepsy and controls that ended up with another diagnosis (such as syncope or non-epileptic seizures). The goal is to examine if the methodology works robustly on this large cohort, and can theoretically contribute to the reduction of misdiagnosis rates.

The secondary aim is to examine whether the computational biomarkers could contribute to reducing the waiting time and the number of clinical appointments needed before a final diagnosis is made.

4.2 Outcome

We will calculate a set of statistical measures to examine the performance of our methodology including sensitivity, specificity and accuracy. We will further consider standard ROC-analysis and compute the area under the curve (AUC). We will report full statistical procedures and associated code, and register the study with ClinicalTrials.gov.

5 STUDY DESIGN and METHODS of DATA COLLECTION AND DATA ANALYIS

This study is a retrospective study – in order to increase the certainty of the eventual diagnosis that was made for each study participant (e.g be as certain as possible that the controls do not have epilepsy etcetera) – an EEG expert or (clinical) neurophysiologist will be able to go into the hospital’s or center’s database, and identify and anonymise the EEG recordings. We will then retrieve the anonymised EEG recordings for each participating center using a harddisk or a safe and secure electronic pathway. Furthermore, for each center we will obtain an electronic information-sheet containing relevant meta-data for the study participants (age, gender, diagnostic status at time of

EEG, medication at the time) and unique but anonymous study participant IDs. The data will be stored on our database management system, within which the computational and mathematical analysis will be carried out.

This retrospective study involves data analysis and mathematical modelling of the EEG recordings collected, and statistical analysis of the method's performance on the data-set.

Our method works by taking the following steps: to each EEG recording, we apply an algorithm that automatically detects relevant segments to our analysis (free of artefacts). Each segment of relevant resting-state EEG will then be transformed into a subject-specific "network structure". This is done by calculating the correlations between the activity of different EEG-channels, obtaining a network structure that summarizes the way information flows between the regions in the network. This network structure is then combined with a mathematical model. The mathematical model is an abstract representation that allows us to get two types of behaviour: 'healthy' resting-state activity and 'pathological' seizure activity. By combining the individually derived network structure with the mathematical model, we simulate a computer-generated EEG, which serves as a proxy for the original segment derived from the study participant. We then examine this computer-generated EEG by calculating two biomarkers:

1. A global marker that quantifies how easy it is for the entire network to make the transition to seizure activity in the model
    a. The underlying rationale for this marker is that for the networks derived from the representative controls, it is much harder to generate seizure state in the model in comparison to the subjects with epilepsy
2. A local marker that quantifies whether there are particular regions in the network that are particular prone to generating or participating in seizure activity in the model.
    a. The underlying rationale for this marker is that for the networks derived from the cohort with focal epilepsies, there will be particular regions within the network more prone to generate and participate in seizures, when compared to the generalised epilepsy cohort.

In order to quantify the performance of our method, we will split up the database in a training set (75% of all participants) and a test set (remaining 25 % of all participants). During the training phase, we will optimize our methodology actively using the subject's diagnostic status, that is, whether the subject had a confirmed diagnosis of epilepsy (focal or generalized) or ended up with a differential diagnosis (e.g. syncope, psychogenic seizures). When the methodology has been trained, it will be validated on the test set: this allows us to quantify the accuracy of our method on an independently set that was held out during the training phase. Using the method developed during the training phase, we will classify each participant from the independent test set and check this against their actual diagnostic status to see whether the algorithm correctly identified the subject as having epilepsy (true positive) or as not having epilepsy (true negative). We will quantify the number of cases in which the method has made an incorrect assessment on the participants in this test set:

- false positive: participant identified as having epilepsy by the test method, but did not have epilepsy in actuality.
- false negative: participant identified as not having epilepsy by the test method, but did have epilepsy in actuality.

The collected data will be analysed and modelled using the following programming languages: Matlab (mathematical modelling, numerical analysis, and data-analysis), R (statistical analysis) and Python (machine learning, visualisation).

It is important to stress that there will be some variability of the number of EEGs available for individual study participants: this will depend on the actual diagnostic and clinical pathway that the subject went through at the time. This would enable us to test our research aims (accuracy and diagnostic delay).

Data retrieval will commence once study approval has been granted in 2019. Pseudonymised data will be transferred to the chief investigator on a rolling basis through a password protected and encrypted file sharing portal where it will be stored on an encrypted workstation. The EEG recordings will be pseudonymised by associating the patient identifier with a coded identifier whilst still with the source clinic. The code will be stored in the research file and will also be kept at the source clinic. At the end of the study, the patient identifier will be deleted so that fully anonymised data only will be stored beyond the end of the study.

Once there a significant portion of recordings is collected data-analysis, we will be able to commence the data analysis and mathematical modelling during a training phase. The training phase will finish by analysing and refining our method on 75% of the entire dataset. We expect data acquisition to be finalised by the end of Q1 of 2020, at which point we will perform a full analysis on the entire data-set (in particular the test set). Mid 2020 the results will be written and submitted to peer reviewed journals, and we aim to submit the final report to the HRA by the end of 2020.

6 STUDY SETTING

Data will be collected across multiple sites within the NHS. The current site list has been generated using a combination of existing local contacts with clinicians and using the local NIHR coordinator to distribute 'Expressions of Interest' forms across the NIHR/CRN networks.

At each site the local direct care team within the Neurology clinics will be performing the participant identification and database searches required due to existing expertise with the computer systems involved and past patient knowledge.

7 SAMPLE AND RECRUITMENT

7.1 Eligibility Criteria

7.1.1 Inclusion criteria

Subject was suspected of having had a seizure or epilepsy (fits, faints or funny turns), and as part of the diagnostic process one or more EEGs was recorded

The subject ended up with a confirmed diagnosis of epilepsy or of the differential diagnosis such as syncope, or psychogenic seizures (diagnosis must have been at least 1 year ago, and not changed since)

For each subject identified we would like to have all the available EEG files within the centre, with the following metadata:

Primary meta-data (crucial):

- Age at the subject at time of each available EEG
- Treatment status at the time of each available EEG (including drug-load)
- Gender of the individual
- Ethnicity of the individual
- Confirmed diagnosis: details on the exact diagnosis made (syndrome and or condition)

Secondary meta-data (optional):

- Aim of each available EEG at the time

- Information on whether any other conditions are present such as Alzheimer's disease, schizophrenia, Intellectual Disability
- If available: information on when the diagnosis was made
- If available: interpretation of each available EEG

Specifics for the EEG recordings:

- Montage (10-20 preferred)
- Number of channels (minimum 19 channels)
- Referencing method (common average preferred)
- Format of the file (EDF preferred)
- Consistent channel labels for all EEGs provided from each centre
- Information concerning the time of day during the recording
- Information on the sampling frequency
- Faulty channels (not more than 2 preferred, all should be indicated though)
- Pre-processing details (information as to whether any filters were used, for example)

7.1.2 Exclusion criteria

- Subject was not suspected of having had a seizure or epilepsy
- Unavailable information concerning the final diagnosis of the subject (epilepsy or other)
- Incomplete or unreliable meta-data, such as the age, gender and treatment-status at the time of the EEG recording (primary meta-data)
- Recordings which do not comply with inclusion criteria.

7.2 Sampling

7.2.1 Size of sample

In our previous work we have found effect sizes ranging from 0.8 (Chowdhury et al, 2014) to 1.3 (Schmidt et al, 2016) in separating groups of people with epilepsy from controls. We conservatively assume a reduction in observed effect size to 0.6 (95% power with alpha of 0.01, standard deviation: 1.4, see Donner et al., (1981)). Consequently, we estimate that we require the following group sizes for the training set (75%):

- Focal epilepsies: 220
- Generalised epilepsies: 220
- Representative controls: 110

And another 172 cases for the test set (25% of the training set), making a total of **N = 688**. For the cohort of representative controls, cases will be aimed to match typical differential diagnosis rates: 66% (vasovagal) syncope, 23% psychogenic non-epileptic seizures, 11% cardiogenic and other causes. Using this sample size will allow us to test the ability of our method to differentiate the representative controls from the epilepsy cohort, and then within the epilepsy cohort the focal epilepsies from the generalized epilepsies (this is why the group size in the focal and generalized epilepsies is twice as big as the representative control cohort).

In practice, it is unlikely that it is always possible to obtain a long enough segment (~ 20 seconds) of artefact free EEG, especially in pediatric cases. Furthermore, there may be faulty EEG channels, channel noise, or missing information. By observing an experienced neurologist and clinical

neurophysiologist in retrieving suitable EEG recordings retrospectively from electronic health records at St Thomas Hospital London, we estimate that conservatively adding 20% would guarantee the required sample size will be achieved: **N = 825.**

7.2.2 Sampling technique

Since this is a retrospective study, subjects can be selected at the NHS site by the data handler and included in one of the three following groups: focal epilepsy, generalised epilepsy, representative controls (syncope, or any other differential diagnosis).

There are two groups that study participants could be allocated to: the training set (75% of all participants) and the test set (25% of all participants). In order to get unbiased results and balanced data-sets, we will use a random allocation software package (R) to select 25% from the representative control cohort, the generalised epilepsies cohort and the focal epilepsies cohort.

The people carrying out the method optimisation using the training set will not have access at any point to the test-set, which will be strictly held out until the method is optimised and will then be used on the independent test-set.

We will calculate several, standard measures to quantify the performance of our methodology: the sensitivity, specificity and the accuracy. These methods are appropriate because the study is retrospective in nature, and therefore, we know with certainty whether a subject has epilepsy or not.

We will compute the area under the curve (AUC) using standard ROC-analysis. In the case of statistical testing, we will robustly test whether all the assumptions of a specific statistical test are being met and publish the study protocol beforehand that includes details on dealing with multiple comparisons (e.g. Bonferonni correction).

With respect to the risk of missing data, our previous analysis has shown that our analysis is robust up to 2-3 missing EEG channels, and in calculating the desired sample size we have accounted for the fact that some recordings will be very noisy and not allow us to identify segments appropriate for our analysis. We will further assess the effect of potential confounders such as age, gender and ethnicity using standard statistical approaches.

7.3 Recruitment

7.3.1 Sample identification

The identification of potential participants will be performed by members of the direct care team at each site using the digital records contained in their existing EEG capture/analysis software. The exact process for being able to identify a suitable record will vary between software platforms.

As the full recordings are required to be submitted following identification of a suitable patient, the task will require working knowledge of the EEG capture software and basic computer skills.

7.3.2 Consent

Through our engagement with the public engagement groups and charity (Epilepsy Action) we have demonstrated that our study is in the public interest and is aimed at improving patient care. We have consulted with the HRA CAG and have been advised that there should not be a breach of section 251

of the National Health Service Act 2006 and its current Regulations, the Health Service (Control of Patient Information) Regulations 2002.

The NHS Act 2006 and the Regulations should enable the common law duty of confidentiality to be temporarily lifted so that confidential patient information can be transferred without the discloser being in breach of the common law duty of confidentiality. Our requirements are for pseudonymized data to be transferred rather than anonymized requiring us to ensure patient confidentiality.

8 ETHICAL AND REGULATORY CONSIDERATIONS

8.1 Assessment and management of risk

As the study is retrospective in nature there is no requirement to interact with participants.

8.2 Research Ethics Committee (REC) and other Regulatory review & reports

We will not seek REC review as the study is limited to use of previously collected, non-identifiable data.

Regulatory Review & Compliance

Before enrolment of participants, the Chief Investigator/Principal Investigator will ensure that appropriate approvals from supplying organisations are in place through the sponsor.

For any amendment to the study, the Chief Investigator, in agreement with the sponsor, will submit information to the appropriate body in order for them to issue approval for the amendment. The Chief Investigator will work with sites (R&D departments at NHS sites as well as the study delivery team) so they can put the necessary arrangements in place to implement the amendment to confirm their support for the study as amended.

Amendments

As the study does not require NHS REC review or NHS management approval, amendments will be handled in line with the sponsors and site management organisational polices.

8.3 Peer review

The Protocol and Research Study has been reviewed by an expert statistician (Professor Gordon Taylor, Research Design Services South West), 3 separate anonymous Reviewers instructed by Innovate UK, and neurologist Dr Brendan McLean (Royal Cornwall Hospitals).

8.4 Patient & Public Involvement

Members of a local chapter of Epilepsy Action (South-West, lead by Mr. Simon Privett) were engaged with during several of their monthly Coffee & Chat meetings. We consulted them in order to receive feedback on the research stud, to ensure the potential outcomes align with the priorities of people with lived experience. Mr. Privett was consulted and reviewed the HRA Protocol: he provided feedback on the underlying rationale, the clarity of the document, and the data management plan.

8.5 Protocol compliance

If accidental protocol deviations were to occur, they should be rectified as soon as possible and reported to the Chief Investigator and sponsor for documentation.

Continued and intentional deviations from the protocol and non-compliance are not acceptable and could potentially be classified as a serious breach. These occurrences will be reported to the Chief Investigator and sponsor and action will be taken to terminate the deviation. Close monitoring will be carried out by the Chief Investigator and if continuation of deviation is observed, the perpetrator may be dismissed from involvement in the study.

8.6 Data protection and patient confidentiality

All investigators and study site staff must comply with the requirements of the GDP and DPA 2018 with regards to the collection, storage, processing and disclosure of personal information and will uphold the Act's core principles.

All personal data shall be pseudonymised by the direct care team prior to supply to the data controller through the use of a unique participant code appended with a sequential number for each EEG recording submitted for that participant. The files will then be uploaded using a secure file sharing portal which will restrict access using password protected folders for each site. The location of this data will differ for EEG files themselves and the linking code file which could be used to identify the individual.

The EEG files submitted to the data controller will be stored in an Azure Files share which is encrypted using 256-bit AES encryption at rest and transit. It is only accessible through the use of a secure Windows 10 workstation based in Azure also which is password protected and requires the use of a certificate-based VPN connection to prevent any public access to the data. The file share has a network security group preventing access from anywhere other than the research machines and only staff directly related to the research can login to the machines.

The linking file will not be shared with the data analyst in order to preserve physical separation of data sets to reduce the impact of a potential data breach. This linking file will reside within the originating trust's own IT environment.

All data will remain within the Azure 'North Europe' region which is physically located in Dublin, Ireland.

Only approved research staff will be able to access the data beyond essential administrative accounts for backup purposes etc.

8.7 Indemnity

Arrangements have been made through Cornwall Foundation Trust for insurance and/or indemnity to meet the potential legal liability of the sponsor for harm to participants arising from the management/conduct/design of the research.

NHS indemnity scheme will apply for insurance and/or indemnity to meet the potential legal liability of the investigator arising from harm to participants in the conduct of the research.

There are no arrangements in place for payment of compensation in the event of harm to the research participants where no legal liability arises.

8.8 Access to the final study dataset

Dr Rohit Shankar, Dr Wessel Woldman, Dr Leandro Junges and Professor John Terry will have access to the full dataset.

Other investigators may be allowed to access the data-set if a formal request outlining their plans is approved by the CI, PI and the data analysis team.

Data acquired as part of this study may be used in future research by the authors listed in this protocol.

9 DISSEMINIATION POLICY

9.1 Dissemination policy -

The data arising from the study is owned by the participating NHS sites.

On completion of the study, the data will be analysed and a Final Study Report will be send to REC within 12 months. Deferral of publication of the report may be sought in case any potential commercial sensitivity of the work arises. The Final Study Report will be accessed via the HRA Research Summaries website.

In case the study generates new intellectual property (IP), the rights to this will be governed as detailed in the Memorandum of Understanding between the Sponsor and Neuronostics.

A manuscript will be written by the CI and the data analysis team and reviewed by the PI and project mentors, for submission to a peer-reviewed journal, where the full study protocol and statistical analysis plan will be published. The Innovate UK grant number 103939 will be acknowledged in any publications arising from the study. The results of the study will be made available online on ClinicalTrials.gov.

9.2 Authorship eligibility guidelines and any intended use of professional writers

Dr Rohit Shankar – Consultant neuropsychiatrist at Cornwall Partnership NHS Foundation Trust, Associate Professor at the University of Exeter – CI, project management, ethical review submission, data interpretation, writing manuscript.

Sharon Hudson – Consultant at Cornwall Partnership NHS Foundation Trust – SC, project management, ethical review submission, writing manuscript.

Dr Wessel Woldman – Postdoctoral Fellow at the University of Birmingham and Scientific Director of Neuronostics – project management, study design, ethical review submission, data analysis, data interpretation, writing manuscript.

Dr Leandro Junges – Postdoctoral Fellow at the University of Birmingham and R&D at Neuronostics – data analysis, data interpretation, writing manuscript.

Professor John Terry – Professor at the University of Birmingham and Founder of Neuronostics – project management, study design, data interpretation, writing manuscript.

10 REFERENCES

Chowdhury et al. (2008). Misdiagnosis in epilepsy: a review and recognition of diagnostic uncertainty. European journal of neurology, 15(10), 1034-1042.

Chowdhury, Woldman et al. (2014) Revealing a brain network endophenotype in families with idiopathic generalised epilepsy, PLoS One, volume 9, no. 10, DOI:10.1371/journal.pone.0110136.

Donner et al. (1981). Randomization by cluster: sample size requirements and analysis. American journal of epidemiology, 114(6), 906-914.

Feigin et al. (2017). Global, regional, and national burden of neurological disorders during 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet Neurology, 877-897.

Gustavsson et al. (2011). Cost of disorders of the brain in Europe 2010. European neuropsychopharmacology, 718-779.

Joint Epilepsy Council of the UK and Ireland. (2011). Epilepsy prevalence, incidence and other statistics. Access: https://www.epilepsyscotland.org.uk/wp-content/uploads/2019/05 /Joint_Epilepsy_Council_Prevalence_and_Incidence_September_11_3.pdf

Soltesz and Staley.(eds): Computational Neuroscience in Epilepsy. AcademicPress, San Diego (2008)

Schmidt , Woldman et al. (2016). A computational biomarker of idiopathic generalized epilepsy from resting-state EEG, Epilepsia 7(10):e200-e204, 10.1111/epi.13481.

Smith et al. (1999). The misdiagnosis of epilepsy and the management of refractory epilepsy in a specialist clinic. Q J Med. 92: 15–23.

Smith. (2005). EEG in the diagnosis, classification, and management of patients with epilepsy. J Neurol Neurosurg Psychiatry 2005; 76(Suppl II): ii2-ii7.

Terry et al. (2012). Seizure generation: the role of nodes and networks. Epilepsia, 53(9), e166-e169.

Wendling. (2005). Neurocomputational models in the study of epileptic phenomena. Journal of Clinical Neurophysiology 22, 285–287.

Woldman, Terry. (2015) Multilevel Computational Modelling in Epilepsy: Classical Studies and Recent Advances, Bhattacharya BS, Chowdhury F (eds), Validating Neuro- Computational Models of Neurological and Psychiatric Disorders, New York, Springer, 161-188.

Woldman et al. (2019) Dynamic network properties of the interictal brain determine whether seizures appear focal or generalised. BioRxiv.

11. APPENDICES

11.1 Appendix 1- Required documentation

Statement of activities and schedule of events

CV of Dr Rohit Shankar

11.2 Appendix 2 – Schedule of Procedures (Example)

Since the study is retrospective, there is no need for engaging with the study subjects.

13.3 Appendix 3 – Amendment History

Amendment No.	Protocol version no.	Date issued	Author(s) of changes	Details of changes made