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

TITLE OF THE STUDY
Identifying new biomarkers of Parkinson's from routine brain imaging

PROTOCOL VERSION NUMBER AND DATE

1.0

11/06/2021

RESEARCH REFERENCE NUMBERS

IRAS Number:	280243
SPONSOR Number:	2557
FUNDER Number:	EP/T518153/1 (ESPRC)

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 publically 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):

.....

Position:
.....

Chief Investigator:

Signature: Date: 12/05/21

Name: (please print):

DR STEPHEN MULLIN

LIST of CONTENTS

KEY STUDY CONTACTS

Chief Investigator	Dr Stephen Mullin, University of Plymouth, 01752764487, stephen.mullin@plymouth.ac.uk
Study Co-ordinator	Megan Courtman, megan.courtman@postgrad.plymouth.ac.uk
Sponsor	Mrs Sarah Jones, University of Plymouth, plymouth.sponsor@plymouth.ac.uk
Joint-sponsor(s)/co-sponsor(s)	N/A
Funder(s)	Engineering and Physical Sciences Research Council
Co-Investigators	Prof. Emmanuel Ifeachor, e.ifeachor@plymouth.ac.uk Dr. Adam Streeter, adam.streeter@plymouth.ac.uk Dr. Mark Thurston, mark.thurston@nhs.net Dr. Lucy McGavin, lucy.mcgavin@nhs.net Dr. Camille Carroll, camille.carroll@plymouth.ac.uk Dr Jacob Day, jacob.day@plymouth.ac.uk Dr Lingfen Sun, l.Sun@plymouth.ac.uk Megan Courtman, megan.courtman@postgrad.plymouth.ac.uk
Committees	Dr Stephen Mullin, trial management group chair stephen.mullin@plymouth.ac.uk, 01752764487,

STUDY SUMMARY

Study Title	Identifying new biomarkers of Parkinson's from routine brain imaging
Study Design	Case control, cross-sectional study, database analysis, feasibility/pilot study
Study Participants	Case group: A clinical diagnosis of Parkinson's Control group: No clinical diagnosis of Parkinson's
Planned Size of Sample (if applicable)	20000
Follow up duration (if applicable)	N/A
Planned Study Period	01/07/2021 – 01/07/2024
Research Question/Aim(s)	To assess whether identifiable brain changes, seen on MRI/CT/NM brain scans, are viable biomarkers of Parkinson's disease onset and/or progression.

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 FINANCIAL SUPPORT GIVEN**
Engineering and Physical Sciences Research Council	EPSRC DTP and School PhD scholarship (MC)
National Institute for Health Research	NIHR Academic Clinical Lectureship (SM) NIHR Academic Clinical Fellowship (JD) NIHR Research Associate funding (MT)

ROLE OF STUDY SPONSOR AND FUNDER

The University of Plymouth has agreed to act as sponsor for the study. They have provided advice on governance aspects of the study. They will assist with dissemination of results and will oversee the study as well as provide indemnity insurance for it.

The Engineering and Physical Sciences Research Council has provided a PhD bursary to support the project.

ROLES AND RESPONSIBILITIES OF STUDY MANAGEMENT COMMITTEES/GROUPS & INDIVIDUALS

Study Steering Groups

The study was designed with input from people with Parkinson's. Explicit feedback was gained on the use of routine anonymised imaging without patient consent and this feedback led to and was incorporated into the study design.

A trial management group will oversee the study for its duration. It meets on a six monthly basis and will include lay members, People with Parkinson's (PwP) as well as researchers not directly involved in the project.

PROTOCOL CONTRIBUTORS

Dr. Stephen Mullin is a clinical lecturer in neurology with interests in PD genetics, prodromal PD and scaled data. He has undertaken research on clinical and PET imaging biomarkers of prodromal PD in genetically at risk subjects, and runs an internet based cohort study (>1500 subjects recruited: www.rapsodistudy.com). He led the first clinical trial of a genetically targeted therapy in PD. He will lead the project and oversee governance/clinical aspects. (stephen.mullin@plymouth.ac.uk)

Prof. Emmanuel Ifeachor is Professor of intelligent electronics systems. He has extensive experience of data analysis, AI and machine learning (ML) in healthcare. Previous projects include EXPERT DataCare and INFANT, which brought two medical products to market and the EU BIOPATTERN which he co-ordinated. He has used AI and ML to diagnose dementia from primary care data and for biomarker discovery at the point-of-care (EPSRC and EU-funded). He will oversee design/building of the AI pipeline. (e.ifeachor@plymouth.ac.uk)

Dr Lingfen Sun is Associate Professor of multimedia communications and networks. She has a background in the use of machine learning in image recognition and analysis and will provide input into these aspects of the project.

Dr. Adam Streeter is a medical statistician at the Peninsula School of Medicine and the University of Munster. He brings methodological expertise from his research interest in causal inference using routinely-collected and large data. He has experience of data analyses in the PD field. He will lead the data analysis/matching strategy. (adam.streeter@plymouth.ac.uk)

Dr. Mark Thurston is a consultant radiologist at UHPNT with an interest in the use of AI and neural networks. His primary research interest is the use of AI to improve and optimise care delivery and patient safety. He has developed infrastructure for the automated download and anonymisation of routine care images at UHPNT. Together with Dr McGavin, he will oversee construction of the scan compilation pipeline.

Dr. Lucy McGavin Consultant neuroradiologist at UHPNT. She has interests in functional neuroimaging in PD, and clinical implementation of advanced imaging techniques. Together with Dr Thurston she will oversee construction of the scan compilation pipeline. (lucy.mcgavin@nhs.net)

Dr. Camille Carroll Associate professor of neurology and CRN National Specialty Lead for Neurodegeneration. She leads a multi-centre clinical trial of a disease modifying therapy in PD and has an interest in wearable sensors for early disease detection. As Plymouth Parkinson's service clinical lead, she and her team will define the PD cases, referring to existing databases held at UHPNT. She will advise on clinical/governance aspects. (camille.carroll@plymouth.ac.uk)

Dr Jacob Day Neurology registrar at Derriford hospital and academic clinical fellow with an interest in artificial intelligence and genomics. He is a member of the Parkinson's disease team. Dr Day will be involved in coordinating governance and logistical aspects of the project and designing the pipeline for download and anonymisation of images. He will act as the data controller for the project. (Jacob.day@plymouth.ac.uk)

Megan Courtman is a PhD candidate who will be working on the project and will have primary responsibility for the data analysis and compilation. She has an MSc in Data Science (distinction). (megan.courtman@postgrad.plymouth.ac.uk)

KEY WORDS: Parkinson's disease, brain imaging, biomarker discovery, artificial intelligence, computer vision

STUDY FLOW CHART

Key WP: Work package, T: Task, M: Milestone

Phase I

WP1: Patient/Public Involvement and governance
T1.1 PwP/lay member participation in design (SM, CC)
T1.2 Secure ethics (SM, CC)
T1.3 oversight by patient/public representatives (SM, CC)
M1.1 public/PwP involvement in design (achieved) M1.2 secure ethics (before start)

WP2: Data collection and preliminary analysis (UHPNT)
T2.1 Identify PD subjects (Routine care team)
T2.2 Link dataset, generate controls, anonymise (Data controller)
T2.3 Define analysis strategy, explore preliminary dataset (AS, MC, MT, SM, EI)
M2.1 PD case list (Month 3); M2.2 Link dataset (Month 5); M2.3 Analysis strategy (Month 11)

WP3: AI pipeline and data analysis (UHPNT)
T3.1 Define AI pipeline for image analysis (MC, SM, EI, MT)
T3.2 Image analysis (MC, SM, EI, MT)
T3.3 Dataset analysis (MC, AS, SM, EI, MT)
T3.4 Evaluation of performance (MC, SM, CC, EI, MT)
M3.1 AI pipeline (Month 5); M3.2 Image analysis (Month 13); M3.3 Data analysis (Month 13)

WP4. AI Pipeline (multi-site)
T4.1 Define pipeline for compiling scans (MT, MC, SM, CC, EI)
T4.2 Finalise AI pipeline for image analysis (MC, MT, AS, SM, EI)
M4.1 scan pipeline (Month 11); M4.2 AI pipeline (Month 11)

WP5. Project Management (multi-site)
T5.1 NHS site approvals (SM and CC)
M5.1 site approvals (M9);

Phase II

WP6 Data Collection and linkage (other sites).
T6.1 Identify PD subjects and supply to principal data controller (Routine care teams at peripheral sites)
T6.2 Compilation of pseudoanonymised master PD subject list + export (principal data controller)
T6.2 Link dataset, generate controls, pseudoanonymise, send to principal data controller (local data controllers)
M5.1 site approvals (M9);

WP7 Data analysis (all sites)
T3.2 Image analysis. (MC, SM, EI, MT)
T3.3 Dataset Analysis (MC, AS, SM, EI, MT)
T3.4 Evaluation of performance (MC, SM, CC, EI, MT)

WP8 Disseminate results (all sites)
T8.1 Publish research in peer reviewed journals (entire team)
T8.2 Present data at national and international conferences (MC, SM, CC, EI)
T8.3 Disseminate findings to people with Parkinson's locally and nationally (SM, CC, EI, MC)

Deliverables
D1 Ethical approval (before start)
D2 Combined UHPNT database (M12)
D3 Scan pipeline (M12)
D4 AI pipeline (M12)
D5 UHPNT Results (M12)
D6 Combined all site database (M12)
D5 all results (M12)

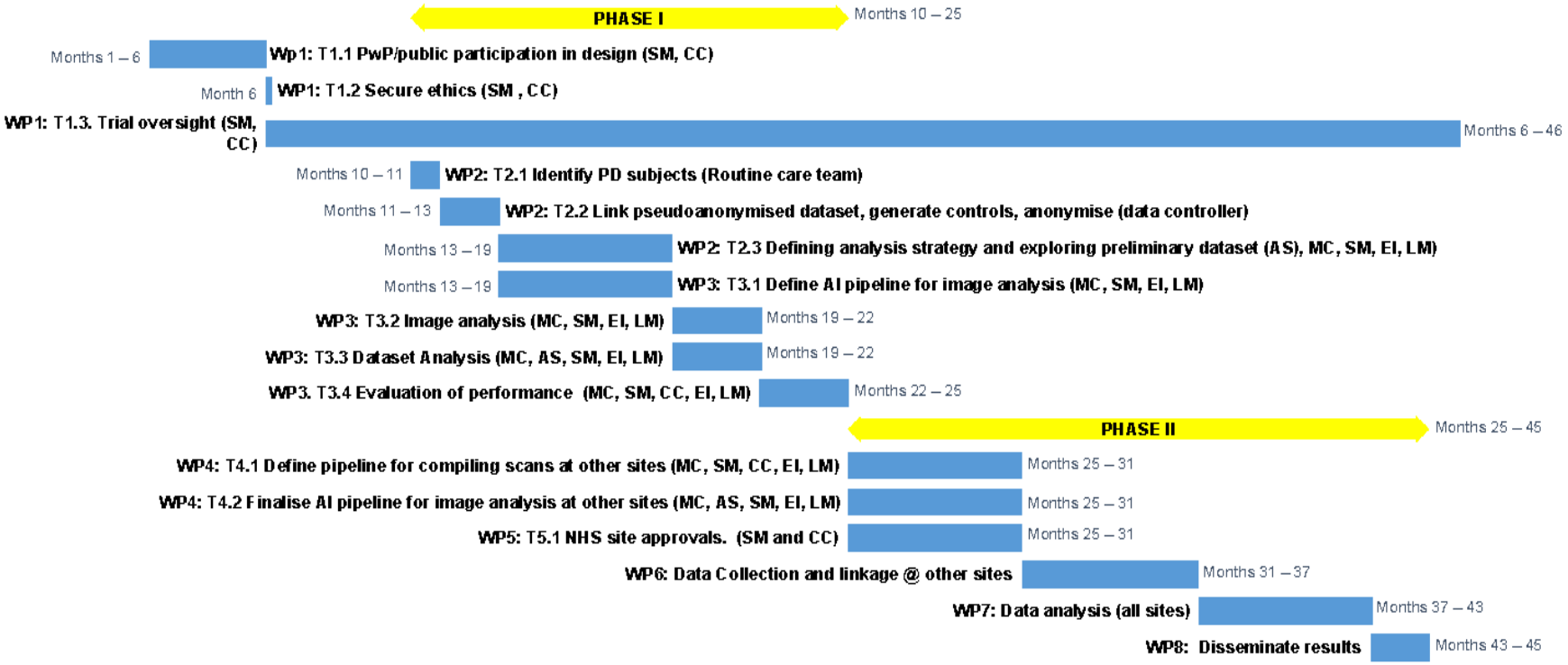

Figure 1 Workflow schematic

STUDY PROTOCOL

1 BACKGROUND

1.1 Parkinson's disease

Parkinson's disease (PD) is a neurodegenerative condition, which commonly presents in the fifth to sixth decade. It is caused primarily by the death of dopaminergic neurons within the basal ganglia, although other neurotransmitters and a variety of other brain structures are affected (Mullin & Schapira, 2015). The diagnosis is based primarily upon the presence of an asymmetrical resting tremor, slowness of movement (bradykinesia), rigidity and a clinical improvement following administration of dopaminergic therapy.

In addition to these 'motor' symptoms, a variety of 'non motor' symptoms are also features of Parkinson's disease. It is known that many of these, such as loss of smell (hyposmia) and Rapid eye movement (REM) sleep behaviour disorder, precede the onset of motor symptoms by a decade or more. This period is commonly referred to as the 'prodromal' phase of Parkinson's disease (Rana, Ahmed, Chaudry, & Vasan, 2015).

Most studies suggest that 30-50% of dopaminergic neurons within the basal ganglia must die before the manifestation of these 'motor' symptoms appear (Fearnley & Lees, 1991; Greffard et al., 2006; Ma, Röyttä, Rinne, Collan, & Rinne, 1997), although some data suggest this is an overestimate (Kordower et a, 2013). These findings suggest that pathological changes contributing to the onset of Parkinson's disease are present during the prodromal phase of Parkinson's disease.

1.2 Disease modification during the prodromal phase of PD

In recent years, many putative neuroprotective compounds have been put forward for use in Parkinson's disease (V. L. Dawson & Dawson, 2019). A number have yielded promising results in phase II studies (Athauda et al., 2017; Mullin et al., 2020). The prodromal phase of Parkinson's disease is seen as the ideal point to administer these drugs. This is because the death of the dopaminergic neurons which lead to the motor symptoms of PD could be prevented, hence avoiding the most debilitating features of PD. This heightened focus on the prodromal phase of PD led to the devising by the Movement Disorders Society of diagnostic criteria for prodromal PD (Berg et al., 2015).

2 RATIONALE

2.1 Current and potential imaging biomarkers of PD

The only robust diagnostic imaging test for Parkinson's disease is a dopamine active transporter (DAT) scan. This measures DAT levels within the basal ganglia, using single-photon emission computed tomography (SPECT) (Isaacson et al., 2017). This is a cumbersome and expensive test requiring prior oral administration of a radioactive tracer. This makes DAT scanning an unsuitable tool for a population level screening strategy.

There is evidence of brain changes in the prodromal phase of PD. DAT scans carried out in subjects with hyposmia, REM sleep behaviour disorder and LRRK2 mutations carriers have demonstrated abnormal dopamine uptake compared to controls (Artzi et al., 2017; Stokholm et al., 2017). Positron emission tomography (PET) has shown activation of glia in relevant brain regions amongst those with REM sleep behaviour disorder and LRRK2 carriers (Gersel Stokholm et al., 2020; Stokholm et al., 2017). Ultrasound of the substantia nigra has shown hyperechogenicity in those with REM sleep behaviour disorder and amongst LRRK2 mutations carriers (Li et al., 2017; Pullman et al., 2018).

MRI changes in a variety of brain regions have been detected in those with PD. Compared to age matched controls, cortical atrophy has been described in the frontal lobe, right hippocampus, left anterior cingulate and superior temporal gyri. Caudate, striatal and substantia nigra volume may be reduced in PD cases. High resolution MRI has described nigral appearances suggestive of brain iron accumulation in PD cases compared to controls. (Saeed et al., 2017)

Basal ganglia calcification is a common incidental finding on computer tomography (CT) head imaging, presenting in around 16-18% of consecutive routine CT scans. (M. F. de Oliveira, Silva, & de Oliveira, 2013) It is the principal finding in Fahr's disease, an autosomal dominant genetic disease in which parkinsonism is a common finding (Mufaddel & Al-Hassani, 2014). A cross sectional study identified an increased prevalence of basal ganglia calcification in newly diagnosed and established PD cases compared to controls (Vermersch, Leys, Pruvo, Clarisse, & Petit, 1992). Additionally, there is evidence that basal ganglia calcification may be a predictor of cognitive impairment and psychosis (Ostling, Andreasson, & Skoog, 2003). To date there has been no prospective evaluation of basal ganglia calcification as a PD biomarker in advance of motor PD symptoms.

These findings provide plausible evidence that early brain changes which are detectable in routine imaging modalities may be present in the prodromal phase of PD.

3 THEORETICAL FRAMEWORK

3.1 Difficulties associated with identifying imaging biomarkers amongst those with prodromal PD.

The main difficulty in identifying prodromal PD imaging changes is finding appropriate subjects for such studies, as in the vast majority of cases, prodromal PD cases are unaware of the significance of their symptoms. Moreover, prospective follow up of these subjects is time consuming and inefficient, as, with the exception of LRRK2 mutations carriers and those diagnosed with REM sleep behaviour disorder, the positive predictive value of these prodromal markers is low. This means there is a lack of high quality prospective data available to evaluate the specificity and sensitivity of these potential biomarkers.

3.2 The advantages of routine imaging data in biomarker discovery

CT brain scans are a common, routine investigation, available 24 hours a day at most hospitals. At University Hospitals Plymouth NHS Trust (UHPNT), for instance, 14,000 CT head scans are performed every year, predominantly the highest risk age group for PD (50+ years). They are cheap, quick, simple to perform and are the first line test for most neurological presentations, including head injury, stroke and confusion. NICE guidelines recommend a CT scan in anyone with acute confusion lasting more than 2 hours (NICE, 2019). Our experience is that PD patients have (often multiple) CT scans prior to their formal diagnosis, probably on account of their predisposition to falls and memory problems.

MRI brain scans are increasingly being utilised in both primary and secondary care. At University Hospitals Plymouth NHS Trust (UHPNT), 8700 MRI brain scans were performed last year. Commonly they are ordered following the onset of memory problems, a known prodromal feature of PD.

The use of routinely collected data allows analysis of past scans carried out prior to the later development of PD. This allows assembly of large prospective cohorts, including in some cases assessment of subjects at multiple time points. This permits accurate charting of the natural history of the disease and drastically increases the power of these analyses.

3.3 The advantages of an artificial intelligence approach

Whereas traditional approaches to defining and assessing imaging biomarkers rely on the generation of an initial hypothesis, artificial intelligence methodologies approach datasets in an agnostic manner. This allows generation of hypotheses distinct from conventional disease paradigms. In many cases this allows the detection of patterns which would not be visible to the naked eye. It also allows the analysis of a higher volume of scans than would be possible with traditional approaches.

4 RESEARCH QUESTION/AIM(S)

4.1 Objectives

1) To identify novel biomarkers of future PD in CT and MRI brains scan collected prior to the onset of PD using an artificial intelligence approach
2) To define whether prospective changes in these novel biomarkers correlate with the onset or clinical progression of PD
3) To define if novel biomarkers are present prior to the onset of motor PD symptoms

4.2 Outcome

The presence of novel putative biomarkers of future PD development as defined by a deep-learning method

5 STUDY DESIGN and METHODS of DATA COLLECTION AND DATA ANALYSIS

5.1 Overview

We intend to use historical CT/MRI/nuclear medicine brain scans to identify novel imaging biomarkers of prodromal PD. The primary data source for the study will be MRI and CT brain scans, whilst nuclear medicine imaging brain imaging (DAT scans) will be used to validate models produced and provide a functional outcome measure of brain dopamine uptake. We shall utilise a an artificial intelligence approach to compare scans of PD cases with matched controls in order to identify these imaging biomarkers.

A list of participants with a diagnosis will be compiled. This list, together with relevant clinical data, will be linked with historical CT/MRI/nuclear medicine scans carried out over the preceding years.

A control group of matched non-PD scans will also be compiled. Whilst the exact details of this process will be determined on the basis of validation experiments when the dataset is available, broadly we shall employ a strategy to gather control images which are matched by biologically relevant (e.g age and sex) and technically relevant (e.g the type of scanner used) variables. The final list of matching variables may be further refined after training, testing and validation.

In the case of both case and control scans, we shall use a filtering approach to remove scans which are likely to contain confounders. An example of such a confounder would be the presence of traumatic brain bleed in Parkinson’s patients. Brain bleeds in themselves would not predispose to Parkinson’s but may be more common in those with PD, because they are more prone to falls. In this case, the filtering strategy may include only outpatient scans (lower chance of acute pathology) and scans with reports that contain words likely to be associated with potentially confounding pathologies (e.g cancer, hydrocephalus, MDT) will be removed.

We will test the filtering strategy in manually inspected validation datasets, to clean the dataset. Depending on the validity of this approach, we may employ a optical character recognition (which will recognise text embedded within pixels which may contain identifiable information), to achieve this more efficiently. Although inevitably a degree of ‘noise’ will exist post filtering, we anticipate that this will be compensated for by the concurrent increase in power resulting from the enlarged sample size.

The dataset will be anonymised and a bespoke ML pipeline will be used to identify imaging features which may be indicative of prodromal PD. This initial stage will be carried out at University Hospital Plymouth NHS Trust (UHPNT), the Royal Cornwall Hospital NHS Trust (RCHNT) and Cornwall Partnership NHS Trust (CPNT). If successful, findings will be validated in a larger sample of scans compiled from hospitals regionally and nationally.

5.2 Compilation of PD case list(s)

A comprehensive list of PD cases will initially be compiled by routine care teams at UHPNT and other sites. It will consist of cases drawn from

1) Pre-existing clinical databases
2) Hospital electronic records (e.g. discharge summaries, patient letters)

Sites will/have been selected on the basis of the existence of such databases.

5.3 Data linkage

In order to comply with data protection regulations and to ensure anonymity of subjects, we have devised a bespoke mechanism to link clinical records with scans. A list of PD cases list will be compiled by the routine clinical care team from existing databases and records.

This list will be pseudoanonymised based on NHS number, year of birth, sex, outward code of post code. A member of the routine clinical care team will act as the data controller and will link the database with all available CT/MRI/nuclear medicine brain images using these identifiers. This information will be deleted prior to anonymisation.

NHS numbers from the regional Parkinson's disease registry will be used to query the Radiology Information System database (CRIS) to identify unique scan identifiers for relevant radiology studies that are suitable for inclusion in the study. The validity of linkage will be tested identifiers other than NHS numbers prior to anonymisation.

The unique scan identifiers will then be used to download the image data using a direct DICOM connection to the PACS system. In addition, meta data held on CRIS, will be extracted, downloaded and compiled. The downloaded image data will be saved in separate directories for cases and controls. They will also assemble a sample of MRI/CT scans from individuals without a PD diagnosis.

The data controller will hold an NHS contract and will be trained in the principles of GCP and the General Data Protection Regulation. This training will be undertaken via the NIHR and University of Plymouth e-learning portal respectively.

Data linkage will occur strictly on NHS servers. The linked database will be password protected. The data controller will act as an independent guardian of the dataset and will not be part of the team performing the analysis.

The linked dataset will be held by the data controller for the duration of the study to allow appending/re-export of the dataset and resolution of data queries. It will be password protected and stored on an NHS server.

This dataset will then be anonymised. The patient numbers and unique scan identifiers will be replaced by a one-way cryptographic grade non-reversible hash value of the original number. The date of the scan will be stored as an offset to reduce risk of identification from the time the scan took place. Technical metadata (e.g DICOM tags), which will aid analysis of the images, but does not contain identifiers, will also be included in the anonymised dataset.

Automated testing will be run on the anonymised scans to prove successful patient information removal. A manual audit of a representative sample of scan data encompassing a range of different scanning units will be performed by the data controller before the dataset is deemed suitable to be passed to the University of Plymouth for off-site analysis.

The combined database will be anonymised and made available to researchers for analysis. The anonymised dataset will retain information necessary to the analysis (free text of indication for scan, free text of report of scan, offset date, age, sex and scanner where images were taken).

With reference to data from sites other than, the UHPNT data controller will become the ‘primary data controller’, with ‘secondary data controller roles’ being created at each of the other NHS sites. Together they will constitute the ‘data control team’. A backup primary data controller will also be nominated at UHPNT.

Lists of pseudo anonymised PD cases from other sites will be forwarded to the primary data controller from secondary data controllers. The primary data controller will compile a pseudo anonymised list of all PD cases from all sites. The rationale for this approach is that we are aware that patients within the region may have had scans at other multiple hospitals and we wish to capture these scans within the database. Because we have the ability to directly upload scan images to UHPNT from other sites via the PACS system, the secondary data controller will have no role in selecting and anonymising imaging data, which will exclusively carried out by the primary data controller.

5.4 Transfer of data

UPHNT

In the case of UHPNT, pseudo anonymised CT/MRI/nuclear medicine images and PD case lists will be stored on the servers of UHPNT.

Other sites

Pseudo anonymised PD case lists and CT/MRI/nuclear medicine images will be stored exclusively on NHS servers. Transfer of PD of case lists and CT/MRI/nuclear medicine images between other NHS sites and UHPNT will be carried out via secure electronic transfer. In the case of the patient list this will be from nhs.net to nhs.net email address with an encrypted password protected file. Scan images will be sent/pulled via the regional PACS system.

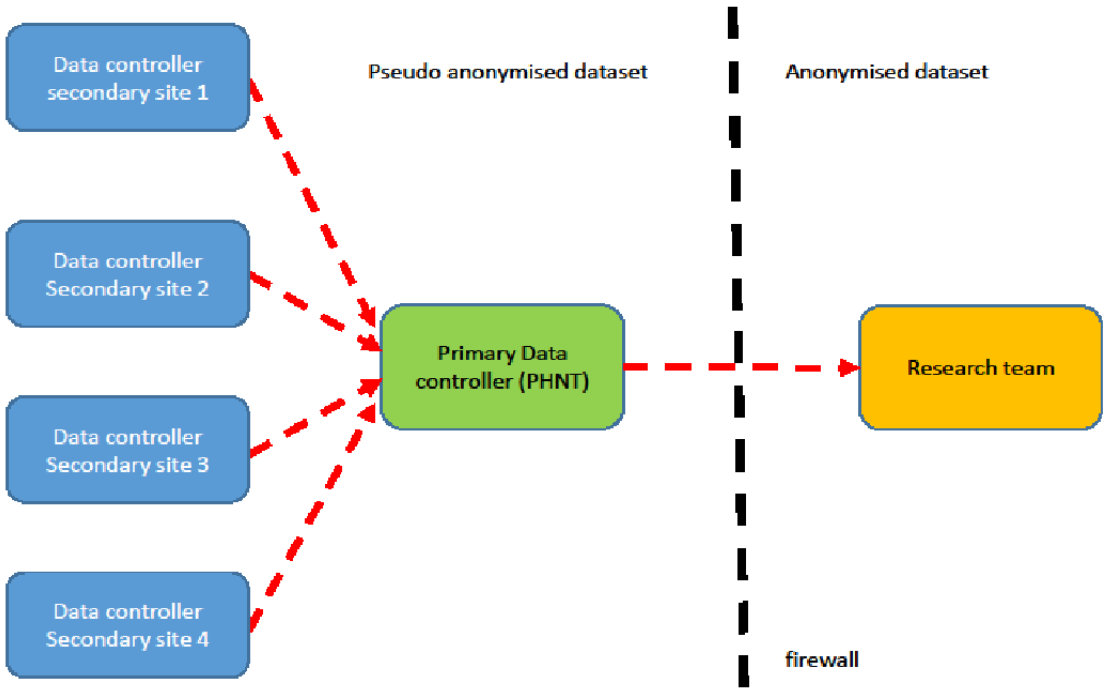

Figure 2 Schematic of study data flow

Export of anonymised dataset

Anonymised CT/MRI/nuclear medicine images and relevant clinical data for analysis will be transferred to a university server either via direct upload or using an external hard drive, depending on capabilities and governance arrangements.

5.5 Automated data extraction pipeline

To minimise the access of staff who are not clinically qualified, a pipeline will be developed to automate the process of matching, extracting and anonymising data from the pseudo anonymised datasets. A flow diagram of this process is provided in figure 3. Development of this pipeline will require input from university research staff holding an honorary NHS contract. Researchers will be provided with an example dummy datasets, compiled from randomly selected NHS numbers. Clinical details of patients within this dummy dataset will not be made available to researchers and it will be used only to test and validate the pipeline within NHS servers before being destroyed. Researchers will not have access to the pseudo anonymised dataset containing identifiers of Parkinson's patients. The code to extract the final dataset will be run by a member of the routine clinical care team who is not part of the research team.

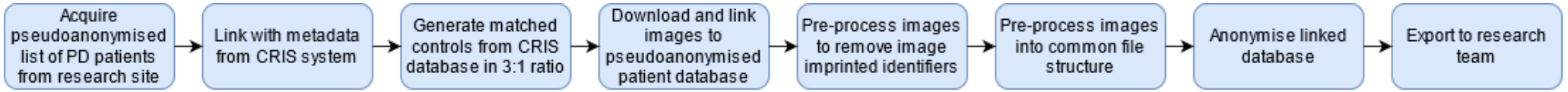

Figure 3 Schematic of data acquisition

5.6 AI Pipeline

Using CT/MRI/nuclear medicine scans as inputs, our tool will determine an index from the changes detected. In the basic mode it defines a region of interest and extracts features from it. A trained ML model maps the features into an index which classifies patients. In the advanced mode, a trained deep learning model determines the index. A database of labelled CT/MR/nuclear medicine scans will train the model to learn the complex relationships between features, PD development and the index. Other patient data may provide additional inputs. Machine learning methods, such as support vector machines and deep convolutional neural networks will enhance the models.

5.7 Analysis plan

Candidate biomarkers will be identified using methods which extrapolate areas of interest. For 2D slices, mapping techniques such as Gradient-weighted Class Activation Mapping (Selvaraju et al., 2020) will be used to highlight parts of the image which activate the class. For 2D slices and 3D volumes, Deep SHapley Additive exPlanations (SHAP) will be used to compute the contribution of each pixel to the prediction (van der Velden et al., 2020). Maps of these SHAP-values will visualise features which positively or negatively influence the classification.

Incorrectly classified examples will be reviewed using such visualisations to outline systematic errors and identify avenues for model optimisation.

In addition to the training set, a separate hold-back test set will be kept aside for analysing the performance of the final model. Exposure to the test set will be limited to after training and cross-validation has taken place.

The algorithm for identifying prodromal imaging biomarkers in the machine-learning stage will be developed and validated using a cross-validation strategy. External validation of the performance of the model will be undertaken using a representative, 'unseen', test dataset to assess its generalizability to other cohorts.

The performance of the model will be evaluated using a variety of metrics (e.g. sensitivity, specificity and area under the ROC curve). Calibration statistics and associated statistical significance for model predictions will be calculated and decision curve analysis carried out to assess its clinical utility.

We will also validate the performance of the model with stakeholders. We will involve clinical staff and PWP to develop use cases and clinical scenarios to test how the model would work in practice and to inform further development and its clinical use.

6 STUDY SETTING

Data will be generated at UHPNT and other NHS trusts. Data from multiple sites will be compiled and anonymised at UHPNT. An anonymised version will be exported for analysis at the University of Plymouth.

7 SAMPLE AND RECRUITMENT

7.1 Eligibility Criteria

7.1.1 Inclusion criteria

Case group: A clinical diagnosis of Parkinson's

Control group: No clinical diagnosis of Parkinson's

7.1.2 Exclusion criteria

The presence of abnormalities in CT/MRI/NM brain scans likely to confound AI pipeline

7.2 Sampling

7.2.1 Size of sample

5000 in case group + 15000 in control group

This is an exploratory study, and we have do not have sufficient information about the parameters to inform a proper sample size calculation, less so the final form of the model, on which to base our sample size calculation and which will only be realised through machine learning.

Preliminary work has been conducted using the publicly available Parkinson's Progression Markers Initiative dataset. Using a deep learning approach, a model has been developed to classify PD and control scans with an accuracy of 80%. The model was trained using one MRI scan per subject from 251 subjects. A greater sample size is likely to reduce the risk of overfitting, thereby allowing the development of a more accurate and generalisable model.

Accurate power calculations are problematic in the context of ML, as the variables contributing to the model's ability to differentiate cases and controls are numerous, not easily identifiable and often interacting. The automated nature of the data collection pipeline means that a large number of images can be acquired with relatively little increase in cost or time. A larger dataset also has the advantage that it decreases further the very unlikely possibility of inadvertent deanonymisation. Hence our sample size estimate reflects a justifiable and achievable aim based on population based estimates of PD prevalence within the region.

7.2.2 Sampling technique

PD cases will be identified from routine clinical databases, hospital electronic records and GP records. Data held on existing research and clinical databases related to the clinical features and progression of these subjects will be appended to these records. CT/MRI/NM scans will be matched using pseudoanonymised identifiers (NHS number, year of birth, outward code of postcode) by a member of the routine clinical care team who is not part of the research team. A number of matched control CT/MRI/NM scans will also be obtained as a control group. Once matched, the whole dataset will be anonymised for analysis, with clinically relevant details (such as scanner used, date of scan, indication for scan and report of scan) included within the dataset.

7.3 Recruitment

7.3.1 Sample identification

Parkinson's cases will be identified by a member of the routine clinical care team (who is not part of the research team) using clinical and research databases, electronic and paper based hospital and GP records. These data will be compiled by the secondary data controller at the respective site. Controls will selected on the basis of having a CT/MRI/NM scan which meets sample matching criteria, which may include but are not restricted to age, sex, scanner and hospital in which images were acquired.

7.3.2 Consent

The project will not seek explicit consent for the use of patient data. This is necessary, as it would be impractical to gain ethical approval from every subject with PD and would introduce bias into the analysis.

As such, we will take an extremely cautious approach to data handling and will ensure that only fully anonymised data is available to the researchers for analysis. Patient identifiers will be made available to the routine care team only. Pseudo anonymised datasets will only be available to those holding an NHS contract who are trained and familiar with the principles of Good Clinical Practice (GCP) and the

Data Protection Act 2018. Data within the anonymised dataset will be selected such that it is impossible to inadvertently de-anonymise subjects. We will gain explicit ethical approval for this approach.

8 ETHICAL AND REGULATORY CONSIDERATIONS

8.1 Assessment and management of risk

1) Ensuring anonymity of subjects

As this study will not be explicitly collecting consent from participants, the major challenge is accurately linking imaging and clinical data, then appropriately anonymising the combined dataset. The initial list of Parkinson's cases will be compiled by the routine clinical care team. Once compiled it will be pseudoanonymised (NHS number, year of birth, outward code of the postcode). A member of the routine clinical care team who is not part of the research team will carry out the data linkage of pseudoanonymised datasets. They will retain the pseudoanonymised dataset for the duration of the project to allow resolution of linkage queries or adding additional data as it becomes available. The linked deanonymised dataset will be deleted upon completion of the project. Development of the pipeline for data linkage and anonymisation will require input from university research staff. This will involve writing code for example dummy data. Research staff will not have access to the pseudoanonymised data and will not extract the final dataset.

2) Avoiding inadvertent deanonymisation of subjects

Identifiers which could inadvertently deanonymise subjects will be avoided (e.g. year of birth rather than DOB, outward code rather than the full postcode). Access to the dataset will be restricted only to the research team and will not be distributed outside of it.

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

Before the start of the study, a favourable opinion will be sought from a REC for the study protocol. A formal discussion has been held between the local authority and the CAG which has confirmed that this study does not need to be put before the CAG as local processes have been followed.

Substantial amendments that require review by NHS REC will not be implemented until that review is in place and other mechanisms are in place to implement at site. All correspondence with the REC will be retained.

The Chief Investigator will submit an annual progress report (APR) to the REC within 30 days of the anniversary date on which the favourable opinion was given, and will notify the REC of the end of the study. If the study is ended prematurely, the Chief Investigator will notify the REC, including the reasons for the premature termination. Within one year after the end of the study, the Chief Investigator will submit a final report with the results, including any publications/abstracts, to the REC.

The study will be registered on clinicaltrials.org once REC approval has been granted.

Regulatory Review & Compliance

Before any site can enrol patients into the study, the Chief Investigator will ensure that appropriate approvals from participating organisations are in place. Specific arrangements on how to gain approval from participating organisations are in place and comply with the relevant guidance.

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

Once it has been established that a protocol amendment is required, the Chief Investigator should assess whether they consider an amendment is substantial or not. This must be agreed with the R&D department and sponsor.

Substantial amendments must be reported using the amendment notification form to the R&D department, the main ethics committee who originally reviewed the study and on IRAS.

Non-substantial amendments do not need to be notified to the ethics committee. However details should be kept in the study file and the R&D department should be informed of the amendment.

The application should be accompanied by a covering letter which indicates the reason for qualification as a substantial amendment. Notifications should include either an extract of the modified documents showing previous and new wording where applicable or the new version of the modified documents (with the changes highlighted), identified with updated version number and date.

8.3 Peer review

Internal peer review has been carried out by Prof. Ray Jones, Professor of health informatics, University of Plymouth. External peer review was performed by Dr David Breen, Senior lecturer in Neurology, University of Edinburgh.

8.4 Patient & Public Involvement

Oversight will be provided by a trial management committee which will meet on at least a six monthly basis. The committee will include patient and lay representation and will be chaired by Dr Mullin.

8.5 Protocol compliance

Accidental protocol deviations can happen at any time. They must be adequately documented on the relevant forms and reported to the Chief Investigator and Sponsor immediately.

Deviations from the protocol which are found to frequently recur are not acceptable, will require immediate action and could potentially be classified as a serious breach.

8.6 Data protection and patient confidentiality

In order to comply with data protection regulations and to ensure anonymity of subjects, we have devised a bespoke mechanism to link clinical records with scans. A list of PD cases list will compiled by the routine clinical care team from existing databases and records. In the case of many NHS organisations, these databases already exist. If necessary we may enrich existing databases using diagnostic clinic codes obtained from discharge summaries, or other data mining strategies, which may provide additional PD cases. This will be subject to appropriate quality control measures to ensure the accuracy of these classifications..

This list will be pseudoanonymised (NHS number, year of birth, sex, first four letters of post code). A member of the routine clinical care team (known as the data controller) will link the database with all available CT/MRI/NM brain images using these identifiers. They will also assemble a sample of MRI/CT/NM scans from individuals without a PD diagnosis. These images, which will be matched to the case group, will act as control scans.

The data controller will hold an NHS contract and will be trained in the principles of Good Clinical Practice (GCP) and the Data Protection Act 2018. Data linkage will occur strictly on NHS servers. The linked database will be password protected. The data controller will act as an independent guardian of the dataset and will not be part of the research group.

This combined database will be anonymised and made available to researchers from Prof. Ifeachor or Dr Mullin’s respective research groups for analysis. The anonymised dataset will retain information necessary to the analysis (free text of indication for scan, free text of report of scan, date which scan

took place, year of birth, sex, scanner where images were taken). The linked dataset will be held by the data controller for the duration of the study to allow appending/re-export of the dataset and resolution of data queries. It will be password protected and stored on an NHS server.

Development of the pipeline for matching and anonymisation will require input from university research staff holding an honorary NHS contract. Researchers will be provided with example PACS and CRIS dummy datasets to allow the writing of computer code for data extraction. This will be carried out within NHS premises, using NHS computers. Members of the research team will not have access to the pseudoanonymised data and will have no role in the extraction of the final dataset. At no time will identifiable data be held on non NHS computers.

8.7 Indemnity

The University has in force a Public Liability Policy and the activities here are included within that coverage. https://www.plymouth.ac.uk/about-us/university-structure/service-areas/procurement/insurance-certificates

8.8 Access to the final study dataset

Full patient identifiers will only be available to the routine clinical care team. Pseudoanonymised datasets will only be accessible by the nominated member of the routine care team responsible for data linkage. Anonymised datasets only will be available to researchers to undertake analysis on.

9 DISSEMINATION POLICY

9.1 Dissemination policy

On completion of the study, the data will be analysed and tabulated and a Final Study Report prepared. As formal consent is not given and the data is anonymised, we do not have permission or the means to return results to the patients involved. We will however disseminate the results widely and make them available in an accessible format to the public. This will include Parkinson's disease patient groups in the areas where the data was collected.

9.2 Authorship eligibility guidelines and any intended use of professional writers

The final study report will reflect time spent on the work. The first author will have be the researcher who has carried out the majority of the study analysis and data collation. We anticipate this will be Megan Courtman. The Senior author will be the researcher who envisaged the design of the project and has done the majority of the supervision. We anticipate this will be Dr Mullin. Other authors will be included if they have made contributions to the following aspects:

1) Study design
2) Securing governance and ethical approvals
3) Data collation
4) Data analysis
5) Data interpretation
6) Manuscript drafting

10 REFERENCES

Artzi, M., Even-Sapir, E., Lerman Shacham, H., Thaler, A., Urterger, A. O., Bressman, S., et al. (2017). DaT-SPECT assessment depicts dopamine depletion among asymptomatic G2019S LRRK2 mutation carriers. *PloS One*, *12*(4), e0175424. http://doi.org/10.1371/journal.pone.0175424

Athauda, D., Maclagan, K., Skene, S. S., Bajwa-Joseph, M., Letchford, D., Chowdhury, K., et al. (2017). Exenatide once weekly versus placebo in Parkinson's disease: a randomised, double-blind, placebo-controlled trial. *The Lancet*, *390*(10103), 1664–1675. http://doi.org/10.1016/S0140-6736(17)31585-4

Berg, D., Postuma, R. B., Adler, C. H., Bloem, B. R., Chan, P., Dubois, B., et al. (2015). MDS research criteria for prodromal Parkinson's disease. *Movement Disorders*, *30*(12), 1600–1611. http://doi.org/10.1002/mds.26431

Dawson, V. L., & Dawson, T. M. (2019). Promising disease-modifying therapies for Parkinson's disease. *Science Translational Medicine*, *11*(520), eaba1659.

de Oliveira, M. F., Silva, E. B. E., & de Oliveira, J. R. M. (2013). Prevalence of brain calcifications in a Brazilian cohort: A retrospective study in radiology services. *Dementia & Neuropsychologia*, *7*(2), 210–215. http://doi.org/10.1590/S1980-57642013DN70200012

Fearnley, J. M., & Lees, A. J. (1991). Ageing and Parkinson's disease: substantia nigra regional selectivity. *Brain : a Journal of Neurology*, *114 (Pt 5)*, 2283–2301. http://doi.org/10.1093/brain/114.5.2283

Gersel Stokholm, M., Garrido, A., Tolosa, E., Serradell, M., Iranzo, A., Østergaard, K., et al. (2020). Imaging dopamine function and microglia in asymptomatic LRRK2 mutation carriers. *Journal of Neurology*.

Greffard, S., Verny, M., Bonnet, A.-M., Beinis, J.-Y., Gallinari, C., Meaume, S., et al. (2006). Motor score of the Unified Parkinson Disease Rating Scale as a good predictor of Lewy body-associated neuronal loss in the substantia nigra. *Archives of Neurology*, *63*(4), 584–588. http://doi.org/10.1001/archneur.63.4.584

Isaacson, S. H., Fisher, S., Gupta, F., Hermanowicz, N., Kremens, D. E., Lew, M. F., et al. (2017). Clinical utility of DaTscan™ imaging in the evaluation of patients with parkinsonism: a US perspective. *Expert Review of Neurotherapeutics*, *17*(3), 219–225. http://doi.org/10.1080/14737175.2017.1256205

Kordower, J. Olanow, W., Dodiya, H., Chu, Y, Beach, T., Adler, C,. Halliday, G., Bartun., R. Disease duration and the integrity of the nigrostriatal system in Parkinson's disease, *Brain*, Volume 136, Issue 8, August 2013, Pages 2419–2431, https://doi.org/10.1093/brain/awt192

Li, X., Xue, S., Jia, S., Zhou, Z., Qiao, Y., Hou, C., et al. (2017). Transcranial sonography in idiopathic REM sleep behavior disorder and multiple system atrophy. *Psychiatry and Clinical Neurosciences*, *71*(4), 238–246. http://doi.org/10.1111/pcn.12483

Ma, S. Y., Röyttä, M., Rinne, J. O., Collan, Y., & Rinne, U. K. (1997). Correlation between neuromorphometry in the substantia nigra and clinical features in Parkinson's disease using disector counts. *Journal of the Neurological Sciences*, *151*(1), 83–87. http://doi.org/10.1016/s0022-510x(97)00100-7

Mufaddel, A. A., & Al-Hassani, G. A. (2014). Familial idiopathic basal ganglia calcification (Fahr`s disease). *Neurosciences (Riyadh, Saudi Arabia)*, *19*(3), 171–177.

Mullin, S., & Schapira, A. H. V. (2015). Pathogenic mechanisms of neurodegeneration in Parkinson disease. *Neurologic Clinics of NA*, *33*(1), 1–17. http://doi.org/10.1016/j.ncl.2014.09.010

Mullin, S., Smith, L., Lee, K., D'Souza, G., Woodgate, P., Elflein, J., et al. (2020). Ambroxol for the Treatment of Patients With Parkinson Disease With and Without Glucocerebrosidase Gene Mutations: A Nonrandomized, Noncontrolled Trial. *JAMA Neurology*, *77*(4), 427–434. http://doi.org/10.1001/jamaneurol.2019.4611

NICE. (2019). *Head injury: assessment and early management*. *nice.org.uk* (pp. 1–59).

Ostling, S., Andreasson, L. A., & Skoog, I. (2003). Basal ganglia calcification and psychotic symptoms in the very old. *International Journal of Geriatric Psychiatry*, *18*(11), 983–987. http://doi.org/10.1002/gps.997

Pullman, M., Ortega, R., Glickman, A., Deik, A., Raymond, D., Marder, K., et al. (2018). Increased substantia nigra echogenicity in LRRK2 family members without mutations. *Movement Disorders*, *33*(9), 1504–1505. http://doi.org/10.1002/mds.27443

Rana, A. Q., Ahmed, U. S., Chaudry, Z. M., & Vasan, S. (2015). Parkinson's disease: a review of non-motor symptoms. *Expert Review of Neurotherapeutics*, *15*(5), 549–562. http://doi.org/10.1586/14737175.2015.1038244

Saeed, U., Compagnone, J., Aviv, R. I., Strafella, A. P., Black, S. E., Lang, A. E., & Masellis, M. (2017). Imaging biomarkers in Parkinson's disease and Parkinsonian syndromes: current and emerging concepts. *Translational Neurodegeneration*, *6*(1), 8.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision*, *128*(2), 336–359.

Stokholm, M. G., Iranzo, A., Østergaard, K., Serradell, M., Otto, M., Svendsen, K. B., et al. (2017). Assessment of neuroinflammation in patients with idiopathic rapid-eye-movement sleep behaviour disorder: a case-control study. *Lancet Neurology*, *16*(10), 789–796. http://doi.org/10.1016/S1474-4422(17)30173-4

van der Velden, B. H. M., Janse, M. H. A., Ragusi, M. A. A., Loo, C. E., & Gilhuijs, K. G. A. (2020). Volumetric breast density estimation on MRI using explainable deep learning regression. *Scientific Reports*, *10*(1), 18095.

Vermersch, P., Leys, D., Pruvo, J. P., Clarisse, J., & Petit, H. (1992). Parkinson's disease and basal ganglia calcifications: prevalence and clinico-radiological correlations. *Clinical Neurology and Neurosurgery*, *94*(3), 213–217. http://doi.org/10.1016/0303-8467(92)90091-g

11. APPENDICES

11.1 Appendix 1- Required documentation

1) CVs of research team

11.2 Appendix 2 – Schedule of Procedures (Example)

Not applicable

13.3 Appendix 3 – Amendment History

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

List details of all protocol amendments here whenever a new version of the protocol is produced. Protocol amendments must be submitted to the Sponsor for approval prior to submission to the REC.