Objective Assessment of Upper Limb Motor Symptoms in Parkinson's Disease Using Body-worn Sensors

James Michael Fisher

A thesis submitted to the University of Newcastle upon Tyne

for the degree of Doctor of Medicine (August 2014)

Northumbria Healthcare NHS Foundation Trust

Institute for Ageing and Health

Abstract

Background

There is a need for an objective method of symptom assessment in Parkinson's disease (PD) to enable better treatment decisions and to aid evaluation of new treatments. Current assessment methods; patient-completed symptom diaries and clinical rating scales, have limitations. Accelerometers (sensors capable of capturing data on human movement) and analysis using artificial neural networks (ANNs) have shown potential as a method of motor symptom evaluation in PD. It is unknown whether symptom monitoring with body-worn sensors is acceptable to PD patients due to a lack of previous research.

Methods

34 participants with PD wore bilateral wrist-worn accelerometers for 4 hours in a research facility (phase 1) and then for 7 days in their homes (phase 2) whilst also completing symptom diaries. An ANN designed to predict a patient's motor status, was developed and trained based on accelerometer data during phase 2. ANN performance was evaluated (leave-one-out approach) against patient-completed symptom diaries during phase 2, and against clinician rating of disease state during phase 1 observations. Participants' views regarding the sensors were obtained via a Likert-style questionnaire completed after each phase. Differences in responses between phases were assessed for using the Wilcoxon rank-sum test.

Results

ANN-derived values of the proportion of time in each disease state (phase 2), showed strong, significant correlations with values derived from patient-completed symptom diaries. ANN disease state recognition during phase 1 was sub-optimal. High concordance with sensors was seen. Prolonged wearing of the sensors did not adversely affect participants' opinions on the wearability of the sensors, when compared to their responses following phase 1

Conclusions

- Accelerometers and ANNs produced results comparable to those of symptom diaries.

- Our findings suggest that long-term monitoring with wrist-worn sensors is acceptable to PD patients.

Dedication

For my parents

"You'll never walk alone"

Acknowledgements

Statement of work undertaken

The design of the study protocol was collaborative with input from my supervisors, Professors Richard Walker and Lynn Rochester, and from Nils Hammerla. I coordinated patient recruitment to the study, with invaluable assistance from the Northumbria Healthcare NHS Foundation Trust Parkinson's Disease Department. I coordinated all patient visits to the research facility during which I performed all clinical assessments. Subsequently I undertook all data entry of any clinical patient data, including home diaries.

Downloading, storage and management of the accelerometer data was undertaken by Nils Hammerla. Construction and development of the artificial neural network algorithm was undertaken by Nils Hammerla as a component of his PhD thesis in Computing Science. Calculation of the correlation between sensor and diary output for each participant was performed by Nils Hammerla. All subsequent analysis was my own work. All other statistical analysis was undertaken by me, with invaluable advice and guidance from Keith Gray. Assistance with analysis of qualitative data was provided by Kate Greenwell.

Table of Contents

List of Tables

List of Figures

Abbreviations

[^{18}F]-DOPA PET = Fluorine-18-L-dihydroxyphenylalanine positron emission tomography

AD = Alzheimer's Disease

AIMS = Abnormal Involuntary Movement Scale

CARU = Clinical Ageing Research Unit

CF = Centre Frequency

CGI = Clinical global impression

CI = Confidence Intervals

COMTi = Catechol-o-methyltransferase Inhibitor

DaT = Dopamine Transporter

DaTSCAN^{TM} = Dopamine Transporter Scan (GE Healthcare)

DBS = Deep Brain Stimulation

ET = Essential Tremor

EMG = Electromyography

EPDA = European Parkinson's Disease Association

FFT = Fast Fourier transform

FRE = Flesch Reading Ease formula

g = Standard gravity of Earth (9.81 m/s^2)

GP = General Practitioner

Hz = Hertz

IBM-SPSS = International Business Machines Corporation Statistical Product and Service Solutions

INDET = Indeterminate (disease phenotype)

LED = Light Emitting Diode

MAOBi = Monoamine Oxidase B Inhibitor

MB = Megabyte

MCIC = Minimal clinically important change

MCRD = Minimal clinically relevant difference

MDS-UPDRS = Movement Disorders Society-sponsored Unified Parkinson's Disease Rating Scale

MEMS = Micro-electromechanical sensor

MMSE = Mini Mental State Examination

MoCA = Montreal Cognitive Assessment

MPTP = 1-methyl-4-phenyl-1, 2, 3, 6-tetrahydropyridine

MSA = Multiple system atrophy

NHS = National Health Service

NICE = National Institute for Health and Clinical Excellence

NMS = Non-motor Symptoms

NS = Non-significant

PD = Parkinson's Disease

PDNS = Parkinson's Disease Nurse Specialist

PDQ-39 = The 39 item Parkinson's Disease Questionnaire

PDD = Parkinson's Disease associated Dementia

PDSS-2 = Parkinson's Disease Sleep Scale

PDYS-26 = Parkinson's Disease Dyskinesia Scale

PIGD = Postural Instability/Gait Difficulty

PSD = Power Spectral Density

PSP = Progressive Supranuclear Palsy

QOL = Quality of Life

RCP = Royal College of Physicians

RBD = REM Sleep Behaviour Disorder

REM = Rapid Eye Movement

SF = Significant Figure

SPECT = Single-photon emission computed tomography

SVM = Support Vector Machine

SWEDD = Scans without evidence for dopaminergic deficit

TD = Tremor Dominant

UK = United Kingdom

UDysRS = The Unified Dyskinesia Rating Scale

UPDRS = Unified Parkinson's Disease Rating Scale

USB = Universal Serial Bus

VMT = Voluntary Movement Threshold

Chapter 1. Introduction

1.I   Overview of Thesis

Parkinson's Disease (PD) is a neurodegenerative disorder classically characterised by a number of motor symptoms, including tremor and bradykinesia. Periods of time where motor symptoms are adequately controlled are often referred to by patients and clinicians as 'on' time. Conversely, 'off' time describes periods where motor symptoms are prominent and are sub-optimally controlled. As PD advances variability in motor symptoms and fluctuations between these disease 'states' is seen. Furthermore, long-term administration of anti-parkinsonian medications is associated with the development of additional, involuntary movements known as dyskinesia.

In routine clinical practice, the assessment of PD motor symptoms is undertaken via history and examination. History taking is dependent on a patient's recollection of their motor symptoms and thus may be confounded by recall bias. Recall bias may be exacerbated by both the long period of time between clinic attendances and the fluctuant nature of motor symptoms in PD. Further complicating matters is cognitive impairment. Cognitive impairment is a common occurrence in PD, particularly with advancing disease, and may further impair the ability to obtain accurate information on motor symptoms via history taking. Clinical examination enables a clinician to gain an impression of a patient's motor symptoms at the point in time at which the assessment occurs. This evaluation may however be influenced by the mere presence of the clinician, since the act of being observed can impact on the motor symptoms a patient exhibits. Whether the findings from a single-moment-in-time examination can be accurately extrapolated to inform decisions about motor state over more prolonged periods is unclear.

Clinical rating scales are scoring systems that aim to quantify and objectify the assessment of Parkinsonian symptoms. The most widely used example, the Movement Disorders Society-sponsored Unified Parkinson's Disease Rating Scale (MDS-UPDRS),

has undergone extensive clinimetric evaluation. The MDS-UPDRS captures a variety of data on motor symptoms including patient/carer-reported information, clinician-led interviews and a formal motor examination. However, one of the major limitations of this assessment method is the time burden associated with its completion. Consequently, this tool is predominantly used in a research context, and is infrequently applied in routine clinical practice.

Patient-completed symptom diaries enable patients to keep a record of their motor symptoms during a monitoring period at home. Patients are typically asked to record their predominant disease state at regular intervals, often hourly or half-hourly. Patient-completed symptom diaries are used both in clinical practice and in clinical research. In clinical practice diaries may be used to enable a clinician to gain an appreciation of the temporal dynamics of a patient's symptoms; that is their fluctuation and evolution over the course of a day. Such information can then be used to assist with decisions regarding medications and their timings. In clinical research, diaries may be used to evaluate the duration of time a person experiences each disease state during a given period. This approach is often employed in drug trials to evaluate whether an experimental anti-parkinsonian agent results in a decrease in 'off' time. Diaries are however flawed. It is recognised that patient concordance with diaries is highly variable and that 'wilful fabrication' of entries (to avoid giving the impression of poor concordance) may compromise the validity of diaries. Furthermore, the inherent subjectivity of the disease state nomenclature in PD and failure of patients to recognise dyskinesia may also impair their validity.

Currently anti-parkinsonian medications provide only symptomatic benefit; there is no definitive evidence that any current therapy delays or halts disease progression. The search for an agent that is 'neuro-protective' (or 'disease modifying') is therefore currently a key research area in PD. It has been suggested however, that current trial outcome measures such as those cited above, may be insufficiently sensitive to measure PD-related disability optimally (Evans and Barker, 2011). Evans and Barker suggest research should explore novel methods of assessment in PD, including body-

2

worn sensors. In summary, the current assessment methods of motor symptoms in PD are flawed. There is therefore, a great need for an objective assessment method in PD.

In this thesis a literature review is presented summarising previous research on objective assessment methods of upper limb motor symptoms in PD. The primary focus of the literature review will be the use of accelerometers; sensors capable of detecting acceleration. Their application to the objective assessment of upper limb PD motor symptoms will be explored.

Much of the existing work with accelerometers worn on the upper-limbs has relied on analysis involving a single variable, or small numbers of variables. Whilst simplistic approaches such as these are attractive, it will be demonstrated in this thesis that such approaches produce suboptimal results, particularly when assessment is undertaken in the presence of volitional movement. Previous work in this field has predominantly undertaken data collection in controlled, laboratory-based settings with carefully selected participants performing scripted tasks. This approach is unlikely to yield data that provides a valid representation of 'real-life' data for a given patient. Furthermore, by virtue of the environment, such data collection is often only for short periods and may be insufficient to capture the motor fluctuations seen with PD. When undertaken, application of such models to unsupervised, home-based activities has been shown to yield disappointing results.

In this thesis I will describe the application of artificial neural networks to this problem. Artificial neural networks are complex computer algorithms capable of considering huge numbers of variables and critically, their relationships with one another. Application of such methods to accelerometer data collected from PD patients has yielded encouraging results, even in the presence of volitional movement and even when knowledge about patient activity has been removed.

In this thesis the use of wrist-worn accelerometers to capture data from PD patients during a seven-day, home-monitoring period is described. Capturing accelerometer data in the home environment produces two distinct advantages to data capture in more closely controlled and monitored environments. Firstly, the data recorded is likely to reflect the wearer's activities of daily living more closely than laboratory-derived data. Secondly, unsupervised data capture at home enables huge amounts of continuous data to be collected. Using the data captured during the home monitoring period an artificial neural network will be constructed - large volumes of data are critical to the development of such a method.

The artificial neural network developed and described in this thesis, will provide a 'decision' with regards the wearer's disease state for a given period of time. This thesis will present evidence for the validity of this assessment; firstly through comparison against the current gold standard for home monitoring, patient-completed symptom diaries, and secondly through comparison against clinical assessment in a controlled environment.

In summary, the aims of this work are to:

1) Establish whether wrist-worn accelerometers (and analysis with artificial neural networks) reproduce patient-completed symptom diaries during prolonged periods of unobserved home monitoring when prior knowledge of disease status is, for a given patient, removed?

2) Establish whether wrist-worn accelerometers (and analysis with artificial neural networks) reproduce clinician assessment of PD patients' disease state during periods of observation in a clinical environment, when prior knowledge of disease status is, for a given patient, removed?

One area of study that has arguably been neglected in previous research is consideration of the patient experience of wearing body-worn sensors. As stated previously, data capture in the home environment is likely to more closely represent the wearer's activities of daily living compared to laboratory-derived data. This claim is

however based on the assumption that the wearing of the sensor does not influence the wearer's usual pattern of activity. A detailed evaluation of the acceptability of the wrist-worn devices to patients will therefore be undertaken.

The third and final aim of this work is therefore to:

> 3) Establish whether prolonged monitoring using bilateral wrist-worn accelerometers is acceptable to patients?

1.II Parkinson's Disease

1.II.i Clinical Features

Parkinson's disease (PD) is a progressive neurodegenerative condition that was first described by James Parkinson in 1817 (Parkinson, 2002). James Parkinson's "Essay on the Shaking palsy" detailed the motor manifestations of the condition, namely bradykinesia, tremor and rigidity, and the typical, initial pattern of asymmetrical upper limb involvement (Kempster et al., 2007) which persists throughout the course of the disease (Lee et al., 1995). Expression of motor features in PD patients is variable, with the condition displaying heterogeneity in both its presentation and clinical course. Analysis of the DATATOP cohort of PD patients (Jankovic et al., 1990) provided evidence for the existence of two distinct motor phenotypes; 'tremor-dominant' and 'postural instability and gait difficulty' (PIGD). Subsequently it has been demonstrated that disease progression occurs more slowly in those with tremor predominance (Foltynie et al., 2002, Lewis et al., 2005) and that those exhibiting the PIGD phenotype were more likely to develop PD associated dementia (PDD) (Burn et al., 2003).

James Parkinson's original essay makes reference to a series of non-motor symptoms (NMS) experienced by his patients, yet it is only in recent years that their high prevalence (Martinez-Martin et al., 2007, Khoo et al., 2013) and their impact on patients' quality of life (Martinez-Martin et al., 2011) has been widely recognised. The NMS seen in PD encompass a constellation of symptoms (Chaudhuri et al., 2005): neuropsychiatric: including dementia, depression, psychosis, impulse control disorders, anxiety and apathy (Weintraub and Burn, 2011); autonomic dysfunction, resulting in

bladder disturbances, sexual dysfunction and orthostatic hypotension; constipation; and sleep disorders, including rapid eye movement (REM) behaviour disorder (RBD).

1.II.ii   Diagnosis

Despite almost 200 years passing since James Parkinson's initial description of the condition, PD remains a mainly clinical diagnosis based on the demonstration of 'parkinsonism'. Demonstration of 'parkinsonism' requires evidence of bradykinesia (slowness of initiating voluntary movement) and some, but not necessarily all, of the following clinical features: tremor, typically of a coarse nature and occurring at rest; rigidity of muscular tone and postural instability. These clinical features have been summarised by the United Kingdom (UK) Brain Bank Diagnostic Criteria for Parkinson's disease (Gibb and Lees, 1988) which is shown in Figure 1 below. These diagnostic criteria also include features that would exclude a diagnosis of PD and features that, if evident, provide further support for the diagnosis of PD.

Figure 1: UK Brain Bank Diagnostic Criteria for PD (adapted from Gibb and Lees, 1988)

Despite the existence of these diagnostic criteria, it is well recognised that there is an error rate associated with the diagnosis of PD. One study that examined the brains of 100 patients who had in life been diagnosed with PD, demonstrated absence of the

pathological hallmark (nigral Lewy bodies) in 24% of cases (Hughes et al., 1992). Later work, examining an unselected primary care population of patients suspected to have PD, demonstrated a diagnostic error rate of 15% (Schrag et al., 2002). Diagnostic accuracy in specialist movement disorder clinics has been demonstrated to be higher (Hughes et al., 2002). Diagnostic error can be attributed to a number of factors. Firstly, there is currently no definitive diagnostic test for PD; no serological or radiological investigations have been shown to have sufficient efficacy to be incorporated into routine clinical practice. Secondly, differentiating PD from other conditions can be challenging. Tremulous disorders such as essential tremor (ET) (Schrag et al., 2000) and dystonic tremor (Jedynak et al., 1991) may be incorrectly labelled as PD. Furthermore, the so-called "Parkinson's plus" disorders; progressive supranuclear palsy (PSP) (Litvan et al., 1996) and multiple system atrophy (MSA) (Quinn, 1989, Wenning and Geser, 2004) introduce further diagnostic challenges, and may be mistaken for PD even by a specialist movement disorders service (Hughes et al., 2002).

In recent years, huge advances have been made in the application of neuroimaging to movement disorders such as PD (Stoessl et al., 2011). The use of single-photon emission computed tomography (SPECT) with the membrane dopamine transporter (DAT) ligand [^{123}I]FP-CIT has enabled visualisation of decreased synaptic dopamine reuptake in PD patients (Booij et al., 1997). [^{123}I]FP-CIT SPECT scanning (DaTSCAN^{TM}) has been shown to be a sensitive and specific test for differentiation of PD from ET (Benamer et al., 2000) and its use for this purpose has subsequently been supported in the National Institute for Health and Clinical Excellence (NICE) guidelines for PD(2006). Despite the utility of DaTSCAN^{TM}, identification of intact dopamine pathways on functional imaging in patients who clinically appear to have PD, has been a relatively consistent finding, seen in around 4-15% (Stoessl, 2010, Bajaj, 2010). This finding has been termed "Scans without evidence for dopaminergic deficit" (SWEDD) and should prompt clinical re-evaluation of the patient.

As understanding of the NMS seen in PD has developed, evidence has accumulated supporting the existence of a 'prodromal' or 'pre-motor' stage to PD. In this stage of the disease neuro-degeneration has begun, but the classical motor symptoms of PD

are not yet evident (Postuma et al., 2012). Various markers of prodromal PD have been suggested, with evidence to support olfactory dysfunction (Ross et al., 2006, Ross et al., 2008), REM sleep behaviour disorder (RBD) (Postuma et al., 2009) and constipation (Abbott et al., 2001).

1.II.iii Epidemiology and Aetiology

PD is the 2nd most commonly occurring neurodegenerative disorder, after Alzheimer's Disease (AD), and its incidence increases with age (Nussbaum and Ellis, 2003). In a 2004 study, undertaken in North East England (Porter et al., 2006), the age-adjusted prevalence of PD was estimated to be 139 per 100,000. The prevalence of PD in developing countries has been estimated, in a rural Tanzanian study, to be lower (40 per 100,000) (Dotchin et al., 2008), but this is likely to be an under-estimate due to challenges associated with identification of those with early stage PD. The UK's population is ageing (2012a) and it is therefore expected that the prevalence of PD will increase. Recent research has estimated that there will be approximately 162,000 UK citizens with PD in 2020 (2012d), a 28% increase from 2009. The effects of the ageing population are likely to be even more dramatic in developing countries such as Tanzania; projections for 2025 forecast a 184% increase in the number of PD cases from 2005 (Dotchin et al., 2012).

The majority of cases of PD occur sporadically but a small number are inherited. Previous twin studies have suggested that genetic factors do not play an important role in causing 'typical' PD, but that they may do so in patients diagnosed with PD before the age of 50 (Grosset et al., 2009). Recent advances in genetic sequencing technology have however enabled identification of specific genetic loci in which variability of expression influenced the risk of developing the sporadic, 'typical' form of PD. It is suspected that further development in genetic technology over the coming years will identify more genes in which variability influences the aetiology of sporadic PD (Gasser et al., 2011).

A number of specific genetic mutations have been identified as causative for familial PD; these include both autosomal dominant and autosomal recessive forms. PARK1, a

8

mutation in the gene encoding for α-synuclein (a protein that aggregates within Lewy bodies (see 1.II.iv)), is inherited in an autosomal dominant pattern and results in early-onset of parkinsonian symptoms. PARK2 mutation is inherited in an autosomal recessive pattern, and refers to a gene that normally encodes parkin, a protein with a role in degradation of intra-cellular proteins (Morris, 2005). PARK2 mutation results in early or juvenile onset of parkinsonian symptoms, but often with a symmetrical presentation and with prominent dyskinesia. As described, the clinical characteristics of inherited forms of PD differ from those of 'typical' PD, but identification of these genetic mutations has helped with ongoing attempts to understand the pathogenesis of idiopathic PD.

A variety of possible environmental factors in the aetiology of PD have been investigated. There is strong epidemiological evidence supporting a potential protective effect of both caffeine and cigarette smoking (Ascherio et al., 2001, Hernán et al., 2002). Again, the exact biological mechanisms underpinning these associations are unclear. Caffeine does however function as a non-selective adenosine antagonist and antagonism of the adenosine 2A receptor (A_{2A}) can increase output from striatonigral neurones, producing anti-parkinsonian effects (Jenner et al., 2009). Phase 2 trials of preladenant; an experimental drug that antagonises A_{2A} receptors, suggest it may have clinical benefit in PD patients with motor fluctuations (Hauser et al., 2011b). Evidence around heavy metals, dietary factors and rural living is inconclusive, but pesticide exposure does seem to increase the likelihood of developing PD (Lai et al., 2002). The exact mechanism linking pesticide exposure to the development of PD is unclear. It is however known that some pesticides have similar chemical structures to 1-methyl-4-phenyl-1, 2, 3, 6-tetrahydropyridine (MPTP), a street drug contaminant known to induce a Parkinsonian syndrome in both human and non-human primates (Tetrud and Langston, 1989). It is important to note that MPTP-induced parkinsonism has a number of key differences when compared to idiopathic PD that limit its role in aetiological research; absence of Lewy bodies, lack of progression and restriction of pathology to the substantia nigra (Schapira and Jenner, 2011).

1.II.iv Pathogenesis

It is widely acknowledged that the pathological hallmark of PD is loss of dopaminergic neurons in the substantia nigra, associated with the presence of insoluble α-synuclein aggregates within cytoplasmic inclusion bodies; so-called Lewy bodies (Halliday et al., 2011). At present it remains unclear whether Lewy pathology is itself pathogenic or whether it represents a potentially beneficial compensatory survival mechanism employed by the neuron (Dickson et al., 2009). A number of possible explanations for cellular dysfunction and death have been suggested including mitochondrial dysfunction, oxidative stress, altered protein handling and inflammatory change as a consequence of reactive microgliosis (Schapira and Jenner, 2011). It is now recognised that Lewy pathology is not confined solely to the substantia nigra. The presence of these pathologic findings has been demonstrated in a variety of areas including the medulla oblongata, the olfactory system and even in colonic mucosa prior to the onset of motor symptoms (Shannon et al., 2012). Seminal work by Braak et al. (Braak et al., 2003) has proposed a staging sequence for PD, showing step-wise progression of Lewy pathology throughout areas of the nervous system that, for younger, more 'typical' PD patients, seems to correlate with the clinical features and natural history of the disease (Halliday et al., 2008). The mechanism by which PD pathology propagates is unclear, but one hypothesis is that α-synuclein may spread from cell to cell in a similar manner to that seen in prion diseases (Brundin et al., 2008).

1.II.v Pharmacological management and natural history

Replacement of dopamine via oral administration of levodopa, a dopamine precursor, has been the cornerstone of anti-parkinsonian treatment since the late 1960's and remains in widespread use to this day. It has been clearly demonstrated that levodopa, in a dose-response pattern, reduces severity of symptoms in PD (2004). As levodopa use became widespread, it became clear that a number of motor complications were associated with its prolonged use. Patients were noted to exhibit fluctuations in their motor state, experiencing 'on' periods, where motor symptoms were well controlled, and 'off' periods, where motor symptoms were more disabling (Marsden and Parkes,

1976). For some, this pattern of worsening motor control could be linked to medication timings, reflecting falling levels of plasma levodopa (so-called 'wearing off' effect). However, many patients were seen to exhibit more rapid, unpredictable switches between disease states that did not correlate with medication delivery. Furthermore, patients were noted to develop additional, levodopa-induced involuntary movements, typically referred to as dyskinesia. The term dyskinesia describes involuntary movements regardless of their aetiology (Hoff et al., 1999), but in the context of PD, takes one of two forms. 'Peak-dose dyskinesia' typically occurs during the period of maximal relief of parkinsonian symptoms and is predominantly choreiform in nature. 'End of dose dyskinesia' is seen less frequently and may take the form of either dystonic or ballistic movements (Marconi et al., 1994). In general dyskinesias are present at rest and may worsen when mental or motor tasks are undertaken (Durif et al., 1999). Rascol et al. (Rascol et al., 2000) demonstrated that for a cohort of early PD patients treated with levodopa, 45% were found to exhibit dyskinesia after five years of treatment. It is important to note that the mean daily dose of levodopa administered to these patients (753mg) was considerably higher than that typically used in current practice and this may have contributed to the high incidence of dyskinesia seen.

Figure 2 demonstrates how advancing disease and prolonged administration of levodopa impacts on motor symptoms, making control of motor symptoms increasingly challenging.

Figure 2: Schematic graphs demonstrating fluctuation in plasma levodopa over time in PD patients of varying disease duration: A, early; B, advancing; C, advanced [levodopa administration represented by the arrows on the x-axis] (adapted from (Obeso et al., 2000))
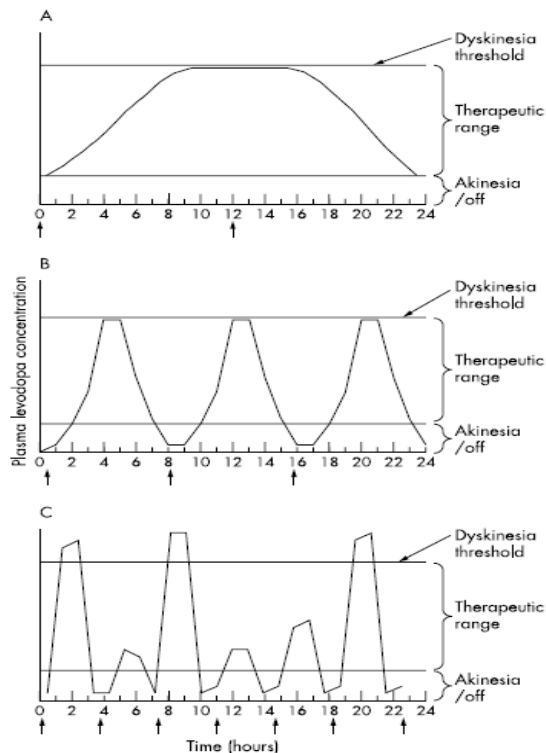
The pathophysiology underpinning levodopa induced motor fluctuations and dyskinesia remains unclear, although discontinuous delivery of dopamine to the brain, resulting in intermittent, rather than continuous, physiological stimulation, has been suggested as a possible causative factor (Rascol et al., 2011). Attempts to deliver levodopa continuously, in a manner more akin to normal physiology, have been explored with the use of intra-duodenal infusion of levodopa via a portable pump and produced reductions in both "off" periods and time spent dyskinetic (Antonini et al., 2007).

A number of medications have been shown to produce symptomatic benefit in PD but no conclusive evidence exists to support the existence of a 'disease-modifying' treatment; that is an agent that slows the natural rate of progression of the clinical course of PD (Olanow et al., 2011). The number of potential new PD medications currently under investigation is somewhat limited (Duncan et al., 2013). In the future, improved recognition of subjects who are at risk of developing PD may enable targeted

interventional trials investigating for disease modifying or preventative effects in an at risk cohort.

Chapter 2.   Literature Review

2.I   Methods of Assessment in Parkinson's Disease

There exist a number of different assessment methods in PD and an overview of the available methods is presented in Table 1 below.

Table 1: Methods of Assessment in Parkinson's Disease

Methods of Assessment in Parkinson's Disease		
Clinical Practice	**Other Objective Methods**	
History Taking	Archimedes Spirals	Force Transducers
Physical Examination	Handwriting	Gyroscopes
Clinical Rating Scales (current gold standard for clinical assessment)	Optoelectronics	Electromagnetic Sensors
	LASER	Computerised Assessments
Patient-completed Symptom Diaries (current gold standard for home monitoring)	Electromyography	Telemedicine
	Video Analysis	**Accelerometers**

Some of these assessment methods form part of routine clinical practice and these will be examined below (2.I.i). Various authors have attempted to develop new, objective methods of assessment and these will be considered in 2.II. Accelerometers, and their potential use as an objective assessment method in PD, will be discussed in 2.III and form the focus of this thesis.

2.I.i   Clinical Assessment

The main tenets of clinical assessment are history and examination. These will be addressed individually, but firstly consideration will be given to the typical model of care delivery for PD patients in the UK. The European Parkinson's Disease Association (EPDA) Charter (1997) recommended that all patients with PD should have the right to consult a doctor with a special interest in PD. Classically the role of a doctor with a special interest in PD has been fulfilled by a neurologist. There is however acknowledgement that a variety of other health professional groups can provide ongoing assessment and care of PD patients including, geriatricians, general practitioners (GPs), Parkinson's Disease Specialist Nurses (PDNS) and physiotherapists

(Kale and Menken, 2004). The NICE guidelines for PD (2006) for example, suggest that both geriatricians and physicians may be adequately qualified to make the initial diagnosis and to provide ongoing assessment. Given the constellation of symptoms seen in PD, the condition has been referred to as a 'geriatric syndrome' (Lauretani et al., 2012) and may be well suited to the comprehensive geriatric assessment employed by geriatricians (Powell, 2008). GPs have a role throughout the clinical course of PD that appears to expand as patients become progressively more dependant and move into care facilities, with patients then more likely to receive a GP visit than to attend a specialist clinic (Hindle et al., 2007).

The NICE guidelines on PD recommended that patients with early, mild disease should be reviewed by an expert on a 6-12 monthly basis, but that patients on treatment, or those at later stages of the disease, may require more frequent review. No guidance on the duration of clinical review exists that is specific to PD, however the Royal College of Physicians (RCP) have published specialty specific guidance on recommended outpatient clinic duration (2011) that is summarised in Figure 3.

Figure 3: Royal College of Physicians Guidance on Outpatient Clinic Duration (2011)

Specialty		Nature of visit	
		New	Follow up
Neurology		30 mins	15 mins
Geriatric Medicine	"General Medical"	15-25 mins	10 mins
	"Complex Elderly"	45-60 mins	20 mins

As described above, the complexity of PD may mean that the recommendation for "complex elderly" may be most appropriate for those patients on medication or with advancing disease. In between out-patient clinic attendances, patients may be reviewed by PDNS either face-to-face or often via telephone contact (MacMahon, 1999).

2.I.ii   History

A detailed history forms the cornerstone of clinical evaluation and a thorough history alone can often provide sufficient evidence to support a preliminary diagnosis

(Peterson et al., 1992). It is important to acknowledge that for many patients with PD, the presenting complaint that first brings them to the attention of medical services may be non-motor in nature (O'Sullivan et al., 2008). The process of eliciting the symptoms the patient is, and has been experiencing, is central to the doctor-patient interaction. Accurate identification of symptoms in PD is of particular importance. Firstly, treatment guidelines suggest that instigation of anti-parkinsonian medication need not happen at the point of diagnosis (2010). Instead they recommend that the timing of starting treatment should be governed by the patient's individual circumstances and their symptoms, although consensus opinion does seem to be moving towards early treatment in PD (Grosset and Schapira, 2008). Hence accurate assessment of the patient's symptoms is a vital process informing this management decision. Once established on anti-parkinsonian medication, ongoing assessment of symptoms is key to establishing whether or not the treatment is effective. Furthermore, a lack of improvement of symptoms with levodopa therapy, or a lack of symptom progression, may suggest diagnostic error (Gibb and Lees, 1988).

As the disease progresses, recognition of the development of new symptoms is crucial. The onset of both dyskinesia and motor fluctuations represent important milestones in the natural history of the condition (Jankovic, 2005) and have a detrimental effect on patient quality of life (Péchevis et al., 2005), as well as disease management and health economic implications (Dodel et al., 2001). Clinical history taking does however have a number of limitations. Firstly, patients may not volunteer a particular symptom that they have experienced due to a number of reasons. It may be that a patient is simply unaware of the presence of the symptom, as has been demonstrated with dyskinesia (Vitale et al., 2001). Alternatively, the patient may be aware of a particular symptom but not of its association with PD, and therefore may not volunteer it to their PD clinician. This occurs more commonly with non-motor symptoms as opposed to motor symptoms (Chaudhuri et al., 2010). Asking patients to recollect their symptoms introduces the possibility of inaccurate recall. Evidence exists in a PD population that suggests patients are able to reliably recall when their symptoms first began, but are less consistent when asked to recall the nature of the symptoms they had initially

experienced (Richards et al., 1994). Inaccuracies in the patient history may be exacerbated by a number of factors. Firstly the variability of symptom expression and the degree of fluctuation that may be seen renders accurate recollection of symptom nature and timings more challenging, particularly with advancing disease. Patients may go several months between seeing their PD specialist; long intervals between appointments may have a detrimental effect on patients' ability to accurately recollect their symptoms. This may be further confounded by changes to National Health Service (NHS) structure and delivery that may result in a decrease in the duration of consultations (Waghorn and McKee, 1999). Furthermore, the high prevalence of cognitive dysfunction in PD may, as the condition advances, further impair the ability of patients to accurately recall their symptoms.

2.I.iii   Examination

Physical examination has been a central part of clinical assessment for many years and reached a "golden age" in the late nineteenth century, when it was believed to offer unprecedented diagnostic certainty (Mangione and Peitzman, 1996). Technological developments have resulted in the increased availability of specialised diagnostic equipment and tests, which some health professionals may be perceive to be more objective than clinical examination (Anderson et al., 2001). There is also evidence to suggest that the standards of junior doctors' examination skills have declined, perhaps due to decreased confidence in the role of clinical examination (Oliver et al., 2013). An example of the subjectivity inherent to clinical examination, is the demonstration of considerable inter-observer disagreement between neurologists' assessments of tendon reflexes (Stam and Crevel, 1990).

In the context of PD there are some further limitations of clinical examination. Firstly, the presence of an observer, i.e. the examining clinician, can influence the clinical signs exhibited by a patient. It has long been recognised that psychological stress can result in worsening of a patient's tremor (Marsden and Owen, 1967). The activity a patient is engaged with at the time of examination may also influence the observed clinical signs; dyskinesia, for example, may be more prominent when patients are engaged in motor

or mental tasks (Durif et al., 1999). Furthermore, the variable nature of PD means that the examination findings at a given time may not be truly representative of the patient's condition. Patients at different stages of the 'medication cycle' can have vastly different clinical examination findings. Motor fluctuations may only be present for a relatively short period and the symptoms may be described by a patient, but not witnessed by the clinician at the time of examination. The limitations of clinical examination are compounded further by the relatively short duration of the clinical assessment in comparison to the periods between evaluations.

2.I.iv   Clinical Rating Scales

The process of measurement has previously been described as "an essential component of scientific research" (Streiner and Norman, 2008). Laboratory based research allows the environment to be carefully controlled and manipulated; measurement in this setting is therefore often relatively straightforward. In clinical research, objective outcome measures such as life-span or binary outcomes such as 'dead' or 'alive', presence or absence of disease, simplify the measurement process. Measurement in clinical research has become more complex since new interventions may influence quality but not quantity of life. The challenge facing clinical researchers exploring the efficacy of such interventions is how to measure something previously thought to be un-measurable. Marsden et al. (Marsden and Schachter, 1981), in a 1981 review article, summarised the assessment methods for PD available at that time and highlighted a number of clinical rating scales that had been developed. A clinical rating scale is a formalised battery of assessments that may include data relating to historical information pertaining to a patient's symptoms, information regarding interference with activities of daily living and clinical examination findings. Marsden reported on the use of a simple method of staging PD (Hoehn and Yahr, 1967) and this scale is still employed in the field of PD clinical research to this day. Almost 50 years have passed since this scale was described and during this period of time the field of clinimetrics has emerged. Clinimetrics describes the study of indexes and rating scales that are used to describe or measure symptoms, physical signs or other clinical phenomena (Fava et al., 2012). Clinimetrics is concerned with the thorough scientific

evaluation of such measurement instruments and there exist three scientific properties identified as being integral to the quality of such tools: reliability, validity and responsiveness. Reliability considers whether an instrument measures in an accurate, consistent and reproducible manner (Hobart et al., 1996). Validity addresses whether or not an instrument measures what it purports to measure. Responsiveness considers whether an instrument is able to measure change within persons over time and whether it is able to detect minimal clinically important differences (Guyatt et al., 1987).

The Unified Parkinson's Disease Rating Scale (UPDRS) was originally developed in the 1980's (Fahn et al., 1987) and by the year 2000 had become the most commonly employed clinical rating scale in PD clinical trials (Mitchell et al., 2000). The advantages and disadvantages of clinical rating scales such as the UPDRS will be discussed, and the evolution of the UPDRS into its current iteration, the Movement Disorders Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS), will be summarised. The MDS-UPDRS consists of 50 questions sub-divided into four sections, with scores that can be considered separately or summed to produce a total. The four sections are: I, Non-motor experiences of daily living; II, Motor experiences of daily living; III, Motor examination and IV, Motor complications. Several of the part I questions and all of part II, are given to the patient in the form of a questionnaire that they (and a caregiver if applicable) complete without involvement of a clinician. Part IV and the remaining part I questions are completed via an interview conducted by the clinician. Part III, the motor examination, is performed by the clinician. Despite the structured format of the interview sections of the scale, the potential for influence due to recall bias remains. There exists only weak evidence to support the reliability of patient completion of historical sections of the UPDRS (Louis et al., 1996). The multiple sections of the MDS-UPDRS do allow data to be captured regarding a number of areas of PD; the latest iteration for example, contains a more detailed assessment of non-motor symptoms, expanded from the previous version, in light of greater appreciation of their impact on QOL (Martinez-Martin et al., 2011). One limitation, arising as a direct consequence of the structure of the scale, is the time burden associated with its

completion. The MDS-UPDRS is estimated to require 30 minutes of rater involvement (Goetz et al., 2007). Patient-completed questionnaires require additional time for the patient but require no specific supervision from the rater. This time burden therefore limits the use of the MDS-UPDRS in routine clinical encounters where the time allotted for an appointment may be less than that required to perform the MDS-UPDRS. Such scales are therefore predominantly used in clinical research.

The 50 questions that make up the MDS-UPDRS are all scored using the same marking structure. The rating clinician (or the patient completing the questionnaire) has five potential responses to select, presented on a 0 - 4 ordinal rating scale. Each numerical value has an 'anchor' term that is consistent across every question; these are 0: normal; 1: slight; 2: mild; 3: moderate and 4: severe. Every potential response has a short section of text describing the criteria for each. Such structural consistency is advantageous since it may assist the clinician completing the assessment, may contribute to convincing clinimetric evidence of the scale's reliability (Goetz et al., 2008d) and is a clear improvement on the previous iteration where variable scoring systems were employed. There is however a problem inherent to the scoring system employed. The ordinal, scalar nature of the scoring system results in the forced assumption that differences between adjacent categories are the same i.e. the difference between mild and moderate (2 and 3) is identical to the difference between moderate and severe (3 and 4); this is unlikely to be the case (Evans and Barker, 2011).

Further consideration must also be given to *who* completes the assessment of a patient. A tool that can only be administered by a senior movement disorder specialist is unlikely to be of use in a clinical environment. Such tools, if they are to be clinically useful, need to be to be deliverable by other members of the movement disorders service, such as more junior medical staff and by nurse specialists. To assist raters with differing levels of experience to accurately and reliably administer the scale, various training programmes have been developed. For the original UPDRS, videotape-based training was developed for both the motor examination section (Goetz et al., 1995) and ADL scale section (Goetz et al., 2003b). The latest iteration has similar video-based training available on the Movement Disorders Society's website. Consideration has

also been given to the effect of rater experience on scoring produced by such scales. Acceptable inter-rater agreement has been demonstrated for the UPDRS between movement disorder specialists and other professional groups such as nurse specialists and more junior medical staff (Post et al., 2005, Bennett et al., 1997). Professional status had no impact on likelihood of successful completion of a certification programme for the UPDRS that involved rating of patients from video-taped examples (Goetz and Stebbins, 2004).

The MDS-UPDRS has a number of other limitations when subjected to thorough clinimetric evaluation. The first relates to the numerical output of the scale itself. To enable useful application of the scale across different patient groups at different times, the percentage of patients achieving the worst and best scores should be minimal; respectively referred to as 'floor' and 'ceiling' effects (McHorney and Tarlov, 1995). A large floor effect has been demonstrated for the motor complications sections of both the original UPDRS (Martinez-Martin and Forjaz, 2006) and the MDS-UPDRS (Goetz et al., 2008d), limiting the scale's utility in those with mild motor complications. A major problem, limiting the utility of clinical rating scales such as the MDS-UPDRS, is the lack of a definitive definition of what constitutes a minimal clinically relevant difference (MCRD) (Rascol, 2006). MCRD, also sometimes termed minimal clinically important change (MCIC), is the smallest difference in scores between consecutive assessments that is associated with a meaningful difference in a patient's status. A number of authors have tried to define this value for the UPDRS (Schrag et al., 2006, Hauser et al., 2011a) but no such work has yet taken place doing so for the MDS-UPDRS. In both of these studies, serial UPDRS assessments were performed and improvement was assessed using a 7-point Likert scale known as the 'Clinical global impression' scale (CGI). The person completing the CGI was variable, with both patients and clinicians doing so. The use of CGI is an example of an 'anchor-based approach' to addressing the question of MCRD; the measure of interest (e.g. UPDRS) is compared to another measure deemed to have clinical relevance (Crosby et al., 2003). Such an approach has a number of limitations, the first of which is the inherent susceptibility of the anchor to recall bias. Secondly, whether impression of improvement should be rated by patient

or clinician is unclear. Lastly, comparison between the anchor and the scale of interest is undermined if the relationship between CGI and UPDRS is non-linear; furthermore the MRCD is unlikely to be fixed and may vary as the condition progresses.

Prior to the development of the UPDRS a number of different clinical rating scales for PD had been developed and employed in different specialist centres (MDS, 2003). A major advantage of having one established, routinely employed clinical rating scale is that it enables greater clarity of communication between clinicians and makes comparison between patients more straightforward. Such a scale can be employed globally in international multi-centre trials to allow comparison across geographically dispersed patient groups. It is however important to ensure that clinical rating scales are not biased towards the culture in which they were developed. The original UPDRS, for example, included questions regarding dressing (difficulty with buttons) and cutting food; yet buttons and eating utensils may not be a normal part of life for some cultures.

As described above, the motor complications section of the MDS-UPDRS is limited by a significant floor effect. Furthermore, the data captured by the MDS-UPDRS regarding duration of motor fluctuations is rather crude. Patients are first asked to quantify the number of waking hours for a typical day in the last week, and are then asked to quantify the typical number of hours spent in the off state and the number of hours spent dyskinetic. Once more the method of data collection employed in this section renders the data captured susceptible to recall bias. Data on dyskinesia derived from the MDS-UPDRS, and its utility as a tool for assessment of patients with dyskinesia, is therefore limited. A number of clinical rating scales have been designed to specifically examine dyskinesia but there is no one rating scale that has gained widespread acceptance in the manner of the MDS-UPDRS. A 2010 review article (Colosimo et al., 2010) examined the clinimetric evidence underpinning eight different dyskinesia scales, including the motor complications section of the original UPDRS. The article's conclusion was that only two of the scales were recommended for use in a PD population, the Abnormal Involuntary Movement Scale (AIMS) (Guy, 1976) and the Rush Dyskinesia Scale (Goetz et al., 1994). The two most recently developed scales, the

Parkinson's Disease Dyskinesia Scale (PDYS-26) (Katzenschlager et al., 2007) and the Unified Dyskinesia Rating Scale (UDysRS) (Goetz et al., 2008c) were acknowledged to have potential for future use but required further external validation work.

In recent years patient-reported QOL measures have been employed more widely in PD research. It has been suggested that outcome measures of disability, as defined by patients in terms of their activities of daily living, may provide greater appreciation of disease progression than rating scales, where the emphasis is predominantly on clinician-defined motor examination scores (Lang et al., 2013). The 39 item Parkinson's Disease Questionnaire (PDQ-39) (Peto et al., 1995) is a patient reported quality of life measures designed specifically for PD and this has recently been employed as the primary outcome measure in large, pragmatic drug studies such as PD-MED (Gray et al., 2012, Clarke et al., 2012).

2.I.v   Patient-completed symptom diaries

The MDS-UPDRS clinical rating scale captures crude data regarding duration of motor fluctuations in a typical day. This is based on patients' recollection of the duration of these symptoms over the past week. Patient-completed symptom diaries also aim to capture data describing the amount of time spent in different disease states but over an extended period of time. Such diaries are typically highly structured, with patients required to record their predominant disease state for a given time period at, or as near as possible to, the particular time period. An example of a patient symptom diary is shown in Figure 4 below.

Figure 4: An example of a patient-completed symptom diary (Adapted from Hauser et al. 2000)

time	asleep	OFF	ON without dyskinesia	ON with non-troublesome dyskinesia	ON with troublesome dyskinesia
6:00 AM					
:30					
7:00 AM					
:30					
8:00 AM					
:30					
9:00 AM					
:30					
10:00 AM					
:30					
11:00 AM					
:30					
12:00 PM					
:30					
1:00 PM					
:30					
2:00 PM					
:30					
3:00 PM					
:30					
4:00 PM					
:30					
5:00 PM					
:30					

There is no consensus on the optimum duration of diary recording in PD patients, or for that matter in other health conditions where symptom diaries are employed. A one week period of diary completion was shown to provide sufficient data to make representative judgements about symptoms over a one month period, in PD patients with motor fluctuations and dyskinesia (Reimer et al., 2004). Prolonged periods of diary recording may provide greater amounts of clinical information, but in doing so place greater burden on the person completing the diary. For example, in patients with continence problems, symptom diary completion rates were shown to be superior for three day diaries compared to seven day diaries (Tincello et al., 2007). This concept of 'diary fatigue'; worsening of completion with increasing duration of recording, is well recognised. Tincello et al. also demonstrated that data from days one to three of the seven day diary, was less complete that the three day diary. In light of this finding the authors suggested the concept of 'diary despair' i.e. the prospect of completing seven days of data collection may produce a negative impact of completion rates from the outset of the recording period. Ultimately a balance must be struck between gathering sufficient clinical information and maintaining adequate patient concordance.

Diaries used for evaluation of PD motor symptoms are typically 'fixed schedule' diaries; the required times of diary entries are pre-specified by the clinical or research team providing the diary. Consideration must therefore be given to the interval period between required diary entries. Too long an interval may limit appreciation of symptom evolution over time. Short interval spacing may provide greater detail regarding the temporal dynamics of symptoms but places greater burden on participants due to the need for frequent reporting (Bolger et al., 2003). Ultimately diaries represent a relatively crude method of capturing data since they, for a given time period, distil long periods into just one data point. If a patient has experienced a change in their disease state during a particular time period, this is unlikely to be appreciated from the diary data alone.

Early versions of patient symptom diaries for PD required respondents to differentiate between two different motor states, 'on' and 'off'. Clinical trials of medications frequently relied on diaries to evaluate changes in the amount of 'off' time; using this as a measure of treatment efficacy. It is important to acknowledge that reducing 'off' time may not provide overall benefit to the patient, if by doing so the amount of time spent dyskinetic is increased; as such patient completed symptom diaries should include assessment of dyskinesia (Hauser et al., 1997)

The assumption that time spent dyskinetic is undesirable for patients was examined via simultaneous completion of two home diaries; a 'symptom diary' including dyskinesia categories and a 'reference diary', where patients were required to describe time periods as being either 'good' or 'bad' time (Hauser et al., 2000). It was demonstrated that 'on time without dyskinesia' and 'off' time were predominantly described as 'good' and 'bad' time respectively. Dyskinesia was however equivocal, with 58% of respondents describing time spent dyskinetic as being 'good' and 42% describing it as 'bad'. Survey evidence also suggests that PD patients who experience dyskinesia, often prefer this to the 'off' state (Hung et al., 2010). Interestingly caregivers of PD patients show the opposite preference when asked which disease state they preferred their relative to be in (Khlebtovsky et al., 2012). Hauser and colleagues went on to show that dichotomising the dyskinesia category into

'troublesome' and 'non-troublesome' dyskinesia provided functional separation; 'troublesome dyskinesia' was predominantly considered by patients to be 'bad' time and 'non-troublesome dyskinesia' considered to be 'good' time (Hauser et al., 2000). Clarity of category title is of critical importance to the reliability of patient completed symptom diaries. The category 'partial off', for example, has been shown to have comparatively poor inter-rater agreement between patient and clinician diary entries (Reimer et al., 2004). Hauser et al. provided further validation of their commonly used diary (Figure 4) in work where 302 PD patients across 10 countries completed diaries on three consecutive days of two consecutive weeks (Hauser et al., 2004). Inclusion criteria for this research were vague; patients who "were judged capable of accurately completing diaries" were eligible for participation, but no further qualification of this statement is given. This work does however suggest that the Hauser diary demonstrates sufficient test-retest reliability to be employed in clinical trials over three day recording periods.

Patient-completed symptom diaries have a number of advantages when compared to other assessment methods employed in clinical practice. Firstly, the diaries are completed by patients in their home environment. Their pattern of activity and behaviour during the assessment period is likely to be representative of their usual daily life; this is unlikely to be the case with assessments performed in a clinical environment. Furthermore, during a period of diary monitoring a patient will not typically have contact with a clinician. Patient-completed symptom diaries do therefore decrease the element of clinical bias introduced by both a clinical setting and the presence of a clinician. Whilst assessment using the MDS-UPDRS gives numerical values for the amount of time spent in different disease states, it gives no information on the "temporal dynamics" of symptoms (Zanni, 2007), i.e. how a patient's symptoms evolve over the course of a day. Evaluation of a completed patient symptom diary may enable a clinician to appreciate symptom progression over the days where monitoring has occurred. For example, diaries may allow identification of times in the day where a patient is 'wearing off' or times where a patient is experiencing dyskinesia. This information ultimately enables more informed treatment decisions to be made, be

they a change of type, dose, frequency or timing of a medication. The NICE guidelines for PD clearly state the need to empower patients to participate in judgements and choice about their own care. Patient completed symptom diaries can be considered a form of patient centred communication and may help to empower a patient by increasing their participation with their care (Zanni, 2007). Lastly, patient completed symptom diaries have also become widely accepted as a useful source of information for clinical trials. In clinical research studies examining the efficacy of anti-parkinsonian medications, commonly employed outcome measures include: the change in 'off' time from baseline to endpoint, the change in 'on' time from baseline to endpoint and the change in time spent dyskinesia (Papapetropoulos et al., 2008).

Patient completed symptom diaries also have a number of limitations; these can be categorised as relating to the diarist or to the diary itself. Patient concordance with symptom diaries is critical to their usefulness as a method of assessment and a number of factors may influence this. One might argue that PD patients may be more concordant with diary completion compared to patients with different health problems. Certain personality traits are seen commonly in PD (Todes and Lees, 1985); industriousness, exactness, meticulousness – the so-called 'Parkinsonian personality', and these traits may predispose to more accurate diary completion. Completion of a symptom diary requires a patient with intact cognitive abilities. PD associated dementia (PDD) is a common long-term outcome in PD, affecting up to 80% of patients; furthermore, 25% of non-demented PD patients exhibit mild cognitive impairment (Weintraub and Burn, 2011). Cognitive dysfunction may result in impaired patient recollection of both the need to complete the diary and also the nature of the symptoms recently experienced. Further complicating matters is the fact that patients with cognitive impairment are more likely to have advanced disease and thus experience motor fluctuations and dyskinesia, which render assessment, and ultimately management, more challenging. Consequently the use of patient-completed symptom diaries in patients with PD and cognitive impairment is limited.

Patient concordance with diaries can be affected 'honest forgetfulness'; where patients simply fail to remember to complete the diary or perhaps do not have the

diary to hand at a particular time (Bolger et al., 2003). Since patient recollection of symptoms is not infallible, diary respondents unable to recall a particular detail, may rely on 'inference'; coming to a judgement based on the fragments of information they do recall (Bradburn et al., 1987). Inference can produce inaccurate, misleading responses. The more contemporaneously diary entries are made, the less the delay between experiencing and recalling symptoms, the less the need for patient inference and consequently, the lower the potential for recall error. It is also recognised that patients may deliberately fabricate completion of missing diary entries. Research into compliance with paper diaries, albeit in patients with chronic pain, used the innovative technique of photo-sensors within the diary binder, enabling detection of when the binder had been opened and closed (Stone et al., 2003, Stone et al., 2002). Stone and colleagues found that on 32% of the diary days the binder had not been opened, yet entries had been recorded on 97% of these days. Furthermore, only 11% of entries had been made within a 30 minute window around the target time. This failure to make timely entries, with later backfilling of missing data, was termed 'hoarding' and it is hypothesised that patients do so to give the appearance of good compliance. This work also demonstrated that hoarding can mask declining diary concordance as the monitoring period ensues, so called 'diary fatigue'. 'Reported compliance' (whether or not an entry was made for a given time period) and 'actual compliance' (whether or not an entry was made within a 30 minute window around the particular time period) were compared (Stone et al., 2003). It was evident that actual compliance with paper diaries declined over time, yet reported compliance did not. A significant limitation of paper diaries employed in clinical practice is therefore the inability to know exactly when the entry was made.

The ability of a patient to correctly identify their disease state is central to the validity of patient completed symptom diaries. It is widely recognised that patients may be unaware of symptoms such as dyskinesia (Vitale et al., 2001) which may have a detrimental effect on diary validity. Previous research has highlighted the inaccuracy of patients' recognition of disease states (Goetz et al., 1997). In this work patients undertook four hours of symptom diary completion, whilst being simultaneously

observed and rated with the same diary system by clinicians; 20 out of 32 patients were shown to have <80% agreement with clinician rating. Various efforts have been made to provide patient education on identification of different disease states, aiming to improve the accuracy of diary completion. A patient training video, that included visual examples of patients in different disease state, has been shown to increase agreement in disease state recognition between patients and clinicians, with the effect maintained at one month (Goetz et al., 1997). Further work by Goetz, with a different cohort of PD patients, showed that despite the use of the same training video, mean agreement between rater and patient for 'on' and 'off' disease states was only 64%, falling to 59% after four weeks (Goetz et al., 2008a). The patients in this work had PD of considerably shorter duration to those in whom the video was first tested, and experienced no troublesome dyskinesia with only mild motor fluctuations (defined as no change in Hoehn and Yahr stage between on and off states). The poor agreement seen may well reflect the difficulties inherent in patient identification of more subtle motor fluctuations. It is also important to consider that showing training videos of patients with advanced, fluctuant disease may have a negative psychological impact on study participants and as such they should be employed with caution (Papapetropoulos et al., 2008).

Typically patient-completed symptom diaries are given to a patient in paper form which, whilst relatively in-expensive, has a number of limitations. Firstly, patients with PD often report deterioration in the quality of their handwriting. Decreasing size and deteriorating form of handwriting by PD patients was first described by Charcot (Charcot and Sigerson, 1879) and is termed 'micrographia'. Micrographia may result in illegible diary entries, although a well designed, structured diary should minimise the possibility of this through the use of tables with pre-prepared time slots that simply require a tick or cross in the relevant box. When completing a paper diary a patient may enter multiple entries for the same period. It is impossible to differentiate from the diary alone whether this dual-entry reflects patient uncertainty regarding their disease state or whether the patient has experienced both disease states during the given period. Advances in technology have enabled symptom diaries to be produced in

electronic forms. These electronic diaries have a number of advantages when compared to the traditionally employed paper diaries. Firstly, an electronic device can be programmed to provide audible reminders when an entry is due and when entries are made, they can be date and time 'stamped', thus allowing identification of the exact time at which the entry was made (Burton et al., 2007). Furthermore electronic diary entries will always be legible and the device can be programmed to only permit one category to be selected. The prime motivation behind the development of electronic versions of the patient symptom diary has been attempts to improve patient concordance; results have however been mixed. Work by Nyholm et al. showed no significant difference in concordance between the paper and electronic diaries and in fact demonstrated unexpectedly high patient concordance with paper diaries (Nyholm et al., 2004). More recent work reported extremely high compliance (99.98%) with electronic symptom diaries (Lyons and Pahwa, 2007). It is important to note that these authors considered a data entry to be valid if it were entered within 24 hours of the period rated. Recollection of symptoms, 24 hours after having experienced them, may not be reliable. One potential limitation of employing electronic symptom diaries is that participants need to be competent in use of the technology. Despite a relatively young study cohort (mean age of 58 years), Lyon et al. reported that half of their participants required help from their carers when completing diary entries and a third of participants did in fact express a preference for paper diaries over electronic diaries (Lyons and Pahwa, 2007). 40% of those using the electronic diary in Nyholm et al.'s work experienced difficulties using it.

2.I.vi   Summary

The current assessment methods of the motor symptoms of PD are flawed. Clinical evaluation with history and examination, clinical rating scales and patient diaries are susceptible to bias, introduced inevitably due to their widespread reliance on patient recall; subjective, and in the case of the later two methods, often impractical in routine clinical care of a PD patient. Whilst greater appreciation of clinimetrics within the medical research community has seen an improvement in the scientific quality of

30

clinical rating scales, the clinical meaningfulness of such scales remains unclear. There is therefore great need for a more objective outcome measure in PD.

2.II   Objective Assessment Methods

A number of potential objective assessment methods of motor symptoms in PD have been explored previously (see summary in Table 1); these will be considered below.

2.II.i   Archimedes Spirals

Occasionally in clinical practice, as part of a motor examination, patients are asked to draw onto paper an Archimedes spiral. A clinician will observe the patient drawing the spiral and pass judgement on its smoothness and symmetry. The presence of a tremor in the writing limb may result in a more tortuous spiral. Examples of spiral drawings, normal and abnormal, are shown in Figure 5 below.

Figure 5: Examples of spiral drawings: (a) 'perfect' computer-generated spiral; (b) healthy control patient; (c) patient with PD [adapted from Liu et al., 2005, Elble et al., 1990 and Saunders-Pullman et al., 2008]
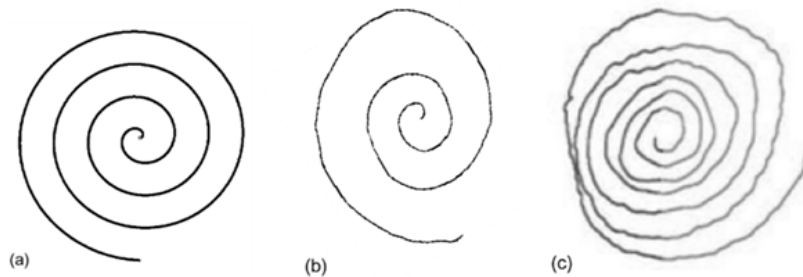
Ultimately clinician judgment of spiral drawings is highly subjective and open to bias due to a lack of blinding. Various attempts have therefore been made to objectify this assessment method, though early attempts were rather crude and in one case involved apparatus that required the participant's trunk to be strapped to a chair (Agostino et al., 1992). In this work, a double-jointed structure was attached at one end to a table and was splinted to the participants' hand at the other. This allowed free movement of the arm whilst potentiometers (voltage dividers) recorded

movement velocity in the shafts of the joints. Results from this work simply demonstrated that patients with PD were slower in their execution of movements compared to normal controls. More recent work took hand-drawn spirals, scanned them into a computer and used an automated computerised algorithm to quantify difference between the drawn spiral and the 'perfect' spiral (Kraus and Hoffmann, 2010). This method was validated against a spiral rating scale (Bain et al., 1993), where clinicians rated how tremulous a spiral appeared using a scoring system from 0 (normal) to a maximum of 10 (extremely tremulous). Congruence between computerised spiral analysis and clinician rating of the spirals was demonstrated, but it is important to note that the scale employed had not undergone thorough clinimetric evaluation. A critical flaw of this method is the retrospective analysis, which fails to capture data regarding the time taken to complete the drawing and hence the frequency of tremor exhibited is unknown.

In recent years, advances in technology have seen the development of digitising tablets. Digitising tablets consist of a writing surface beneath which lies a fine wire grid of evenly spaced horizontal and vertical wires (Elble et al., 1990). A pen, or stylus, that emits an electromagnetic field is used and when this is held in contact with the writing surface, a current is induced in the wires below the point of contact. Computational analysis of which wires are being induced enables the tablet to define the position of the stylus in terms of horizontal (x) and vertical (y) coordinates. Such devices are typically able to derive this positional data several hundred times per second meaning that both volitional and involuntary movements are readily captured. Analysing spiral drawing using digitising tables therefore enables both real-time capture of spiral drawings and more precise measurements to be made. A number of different research groups have employed this method, each relying on different algorithms to interpret the data obtained. No one method has been shown to have satisfactory scientific rigour to see this assessment method employed more widely.

The responsiveness of this assessment method to changes in motor function associated with medication administration has previously been demonstrated (Eichhorn et al., 1996). This work lacked evidence to support the validity of their

assessment method; no correlation was seen between their spiral-derived parameters and clinical assessment of motor function using the UPDRS. Some authors have presented weak evidence for the validity of this assessment method. Firstly, Saunders-Pullman and colleagues examined spiral analysis in patients with early PD (Saunders-Pullman et al., 2008). In this work multiple indices were generated mathematically to describe spiral drawings produced by the patients. Some of the indices produced were shown to correlate to a moderate degree with assessments of motor function using the UPDRS. The derivation of these indices is summarised in earlier work (Pullman, 1998); it is however important to note that this was based on a cohort of patients with tremors of varying aetiology and consequently subsequent application to a PD population may not be valid.

Later work using similar methodology and patients with early PD, also included age-matched controls (Stanley et al., 2010). In this work a mathematical formula was employed to describe the spiral and was then translated into a 'spiral severity score' (five-point scale (0-4)). Patients with purely unilateral parkinsonian symptoms, who therefore had a motor UPDRS score of 0 in their unaffected side, were found to produce spirals worse than those drawn by controls. The authors therefore conclude that this assessment method may be sufficiently sensitive to detect early motor abnormalities not discernible via clinical examination or clinical rating scales. A critical confounding factor in this work however is the failure to consider the handedness of the patients; whilst spirals were drawn repeatedly with both hands this factor was not addressed.

Digitising tablets have also been proposed as a potential assessment method for upper limb dyskinesia, since proximal and distal joints of the arm are typically involved and spiral drawing requires movement at both the wrist and the elbow (Liu et al., 2005). Liu and colleagues derived velocity signals via differentiation of the positional signals of the pen's movement in each direction. The frequency content of the overall signal obtained was derived using a Fast Fourier transform (FFT). Fourier's theorem states that a periodic signal of any complexity can be represented as a sum of sine and cosine waves of different frequencies. Fourier transforms are the mathematical tools that

compute the signal decomposition process (Riviere et al., 1997). FFT specifically refers to algorithms that fulfil this role but do so with relatively little computational 'cost', therefore allowing rapid data analysis. Fourier analysis allows visual representation of each frequency's contribution to the overall signal. This information can be displayed as a 'power spectrum', an example of which is shown in Figure 6 below.

Figure 6: A power spectrum of a patient's spiral drawing (adapted from Liu et al. 2005)
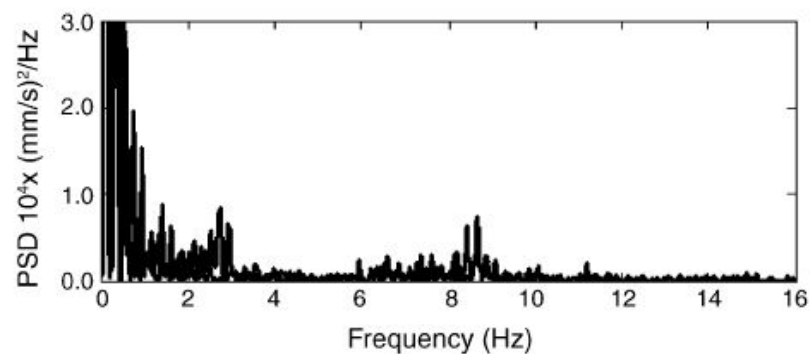
Liu and colleagues suggested that the power spectrum shown in Figure 6 contains three distinct components in the following frequency ranges; <1.0Hz, attributed to voluntary movement; 6-10Hz, attributed to action tremor and 1-5Hz, attributed to dyskinesia (Liu et al., 2005). Critically, participants in this study traced spirals as opposed to freely drawing them. The presence of a pre-existing shape to trace gives the participant on-going visual feedback during the task. This results in voluntary error correction movements throughout drawing; the frequency of which has previously been demonstrated to be 1.7-2.0 Hz (Liu et al., 2001). It is therefore likely that the prominence in the power spectrum between 1-5 Hz may be influenced by the presence of such movements and therefore have led to an overestimation of dyskinesia.

Other authors have employed spiral analysis in attempts to differentiate between PD patients and normal individuals (Aly et al., 2007) and also in tremulous patients with diagnostic uncertainty (Bajaj et al., 2011). Bajaj et al. compared parameters derived from spiral analysis to the results of DaTSCAN™ imaging and demonstrated that certain parameters differed significantly between patients with normal scans; so-called scans without evidence of dopaminergic deficit (SWEDDs) and abnormal scans; termed

tremulous PD. Certain spiral indices achieved sensitivity and specificity values close to those achieved by movement disorder specialists blinded to diagnosis, who were asked to review video-footage and attempt to differentiate between tremulous PD and SWEDDS patients.

Assessment using spiral analysis on digitising tablets has some clear advantages; the test itself is simple and quick to administer and provides a non-invasive evaluation. There are however a number of disadvantages. Spiral analysis provides assessment at one single moment in time and captures data in only two dimensions. Attempts have been made to infer movement in a third axis by examining the degree of pressure placed onto the screen by the writing implement, but this provides only a crude approximation. Furthermore, and crucially, no one method has high quality evidence supporting its validation as an assessment tool.

2.II.ii   Handwriting

Handwriting abnormalities are frequently reported by patients with PD, with complaints of decreasing size and deteriorating form of handwriting ('micrographia'). Contemporary forensic analysis of the handwriting of PD patients has revealed a host of abnormalities that may assist with differentiation between PD and healthy patients (Walton, 1997). Inspection of a brief handwriting sample, provided by a patient in the clinic, may form part of the clinical motor assessment. Such an assessment is inherently subjective and various authors have therefore explored methods to objectify it.

Manual measurement of the size of patients' handwriting has been explored (Wagle Shukla et al., 2012). This involved taking a series of measurements to estimate the area occupied on the page by a section of handwriting and comparing sections at the beginning and end of a writing task. Micrographia was arbitrarily defined as a decrease in area of ≥30% between first and last sections of writing, and was found to be present in 50% of the study cohort. The ability of subjective and objective assessments of handwriting to differentiate between PD and SWEDDS has been compared (Bajaj et al., 2012). The objective measures employed were manual measurements of sentence

length, sentence height and change in sentence height as the sentence progressed. This work demonstrated that subjective assessment of micrographia had 55% sensitivity and 85% specificity for differentiating PD from SWEDDS, an argument perhaps for the validity of simple, clinical evaluation. After derivation of cut-off values, objective measures were shown to only produce a slight improvement in sensitivity, but none in specificity, compared to subjective assessment. Given the time-consuming nature of performing such measurements manually it seems that this method adds little to this assessment.

Various authors have used digitising tablets to allow more detailed analysis of handwriting, since such technology can provide kinematic parameters in an automated fashion. Significant differences in kinematic parameters between patients and controls and between PD patients in the 'on' and 'off' states have been demonstrated (Tucha et al., 2006). Similar methodology has also been applied to PD patients with deep brain stimulation (DBS) in situ (Siebner et al., 1999). A significant difference in all kinematic measures of handwriting was observed between on and off stimulator settings. Selected kinematic measures were also shown to correlate significantly with the improvement in motor UPDRS scores for tremor, rigidity and bradykinesia in the writing limb. This finding provides evidence to support the use of objective assessment of handwriting as a surrogate measure of parkinsonian symptom burden.

Unlike some clinical assessment tasks, handwriting is a task that is performed frequently as part of everyday life and thus has relevance to the patient. It is important to acknowledge that capturing data on asymmetrical parkinsonian signs restricted to the non-writing hand introduces challenges; consideration must be given to the differences in handwriting between dominant and non-dominant hands. Furthermore such an assessment is only able to capture data at a given moment in time and is susceptible to influence from other additional involuntary movements and from other co-morbidities affecting the writing hand.

2.II.iii Optoelectronics

Optoelectronics is a field of scientific study that uses electronic devices capable of detecting light. Computer-assisted optoelectronic movement analysis has led to the development of a new field of study in laboratory quantitative analysis of movement. The first generation of optoelectronic technology involved attaching infra-red light emitting diodes (LEDs) onto various parts of the body. The infra-red light emitted by LEDs was detected by an series of cameras surrounding the wearer and fed into a computer capable of determining body position and movements in three dimensions (Steg et al., 1989). Early application of this technology to PD patients demonstrated that, when assessed in the off state, movement times during a given functional task were significantly greater than those in normal controls (Johnels et al., 1989). Current technology has replaced the LEDs with markers coated with reflective tape that instead reflect infra-red light emitted from cameras arranged around the gait laboratory (Ingvarsson et al., 1999). These cameras also detect the reflected infra-red light and discern body position in the same manner. This assessment method has been shown to enable high accuracy measurement of body position and detection of gait abnormalities in PD patients (Galna et al., 2012).

Application of this technology to the assessment of upper limb motor symptoms is less well established. Finger tapping movements, a measure of bradykinesia, have been analysed using an optoelectronic system with a sensor attached to the distal phalanx of the index finger (Agostino et al., 2003). For a variety of kinematic measures significant differences were seen between PD patients and age-matched controls; PD patients performed the task more slowly, with more pauses and exhibited smaller amplitude of movement. Such systems are costly, require a large testing environment and are complex to analyse. Their role is established in the assessment of gait but not in evaluation of upper limb movements. Consequently such systems are inappropriate for monitoring upper limb symptoms in PD patients in a home environment.

2.II.iv   LASER

Laser analogue sensors have been used to quantify both rest and action tremor (Beuter et al., 1994). The system employed in this work required a patient's finger to be placed at a fixed distance from the laser source and for a small piece of white card to be attached to the finger to provide a reflective surface for the laser. Some of the reflected laser light from the finger returns through a receiver lens and is projected onto a position sensitive device. Movement of the finger and subsequent changes in the angle of reflection allows optical triangulation of the finger. This system was shown to be able to detect presence of tremor, both resting and action, but was unable to provide reliable quantification. The same method was latterly shown to be sensitive enough to detect tremor in the unaffected upper limb of PD patients i.e. where clinical assessment had found no evidence of tremor (Beuter et al., 2005).

A major concern with the use of lasers is the potential for retinal damage if the eyes were exposed to the beam. Precautions can be taken, such as the use of an opaque screen placed around the subject's hand (Beuter et al., 1994) but the risks associated with this assessment method may preclude unsupervised home monitoring. The other major limitation of this assessment method is that systems such as these purely measure linear velocity in one plane, requiring the finger to remain perpendicular to the laser beam (Norman et al., 1999). It is questionable whether this assessment method can be translated to clinical assessment where consideration of movement in three dimensions is essential.

2.II.v   Electromyography

Electromyography (EMG) allows measurement of the electrical activity occurring within skeletal muscles. Voltages are recorded using either needle electrodes inserted through the skin directly into muscles, or via surface electrodes secured to the skin overlying muscles. Prolonged EMG recordings and spectral analysis techniques have been applied to the assessment of tremor in PD (Scholz et al., 1988, Bacher et al., 1989). In the three PD patients they studied they were able to demonstrate an improvement in tremor in one, following administration of additional medication, and

similarities in daily recordings in a patient on a stable treatment regimen. This provides limited evidence to support the reliability and responsiveness of this assessment method. The use of EMG to distinguish between controls and pathological tremor (PD and essential tremor, ET) has been demonstrated with high sensitivity and specificity seen (Spieker et al., 1995). This work also demonstrated that this method was reproducible over a three day recording period. Work by Boose et al. attempted to demonstrate criterion-related validity of EMG by examining correlation between tremor severity with assessment of tremor via both clinician and patient rating scales (Boose et al., 1994). Importantly, the nature of the clinician rating scale used was not described. No significant relationship between EMG recording and physician rating was detected, however a significant correlation between patient self-rating and EMG measures was seen. Later work by Spieker (Spieker et al., 1997, Spieker et al., 1998) used a larger cohort of patients and well-established clinical rating scales such as the UPDRS. This work did show a significant correlation between EMG-measured tremor occurrence and tremor severity as judged by UPDRS rest-tremor sub-score. This finding is offered by the authors as evidence of the criterion-related validity of EMG as an assessment method for tremor. It is important to note however that EMG-measured tremor *intensity* showed poor correlation with clinician rating of tremor. The UPDRS assessment of rest tremor used in this study grades tremor by amplitude not duration. The fact that clinician judgement of tremor amplitude correlated with duration of tremor but not tremor intensity is counter-intuitive. Clearly tremor amplitude and duration are not dependent variables and this disparity in demonstrated correlations calls into question the validity of this method. Tremor has also been detected on brief EMG recordings in PD patients in whom no tremor was discernible clinically (Lakie and Mutch, 1989); this finding was presented as evidence for the sensitivity of EMG for tremor.

Research with EMG has predominantly focussed on tremor, with little consideration to other motor symptoms. At present, insufficient evidence exists to support the role of EMG as an effective form of home monitoring of upper limb PD symptoms.

2.II.vi Video Analysis

Kupryjanow et al. described the use of a 'Virtual-Touchpad' (Figure 7) as a potential new method of objective assessment of upper limb symptoms in PD (Kupryjanow et al., 2010). This system consisted of a mobile computer interface capable of instructing patients to perform a series of upper limb tests. These tests were recorded by an attached web-cam. The recording obtained was analysed in real time at a rate of 15 frames per second and a hand detection algorithm, capable of analysing the contour of the hand, was able to classify the gesture visible at that given time (Figure 7; a-d).

Figure 7: 'Virtual-Touchpad' in use with recognisable gestures displayed (a-d) (adapted from Kupryjanow et al. 2010)
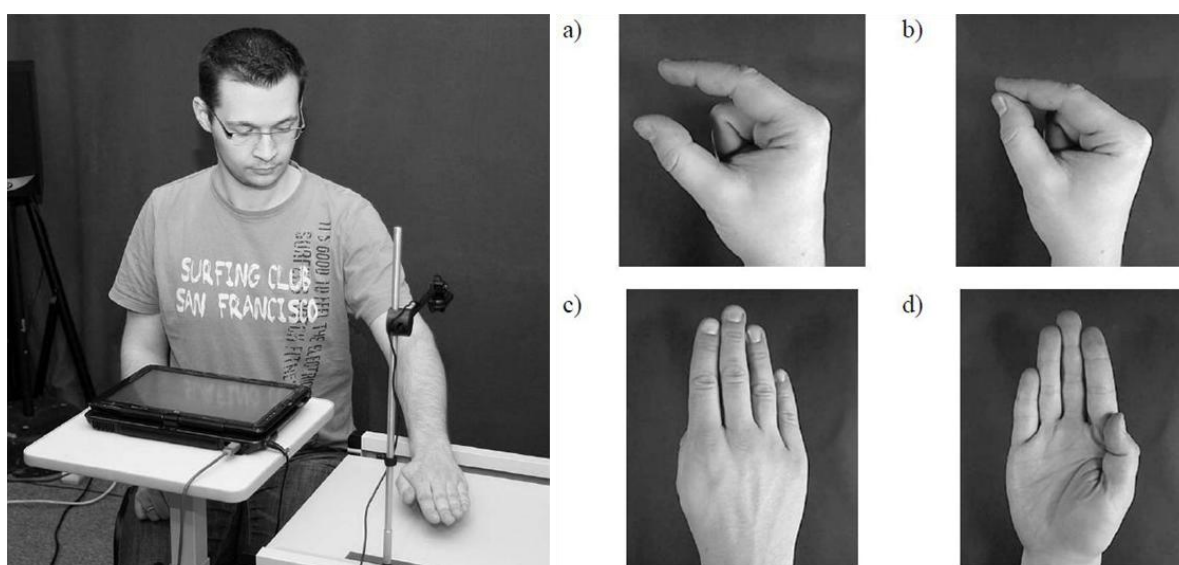
The four different hand configurations that were detectable allowed assessment of both finger-taps and pronation/supination movements. The lack of any attached equipment to the patient and the potential for assessment in a patient's own home make this a potentially attractive option. This work is however at an early stage of development and the authors are yet to publish work involving patients with PD.

Criss et al. also published early work exploring video-analysis as a potential objective measure of finger tapping (Criss and McNames, 2011). They filmed finger tapping movements but under carefully controlled conditions that required a particular

40

background and specific lighting arrangements. They applied what they termed a 'model-based statistical video processing algorithm' that decomposed the images derived from the video into a series of interconnecting polygons that represented the phalanges of the index finger. The angle between each, and its change over time, was thus derived and used to measure finger tapping speed. Again, this concept is in its infancy and is yet to be applied to patients with PD.

2.II.vii Force Transducers

A number of authors have explored the use of force transducers for the assessment of PD motor symptoms. A transducer is a device that converts energy from one form to another, in this case typically mechanical energy to electrical energy. One research group chose to adapt a handheld toy racing car controller for use for this purpose (Sauermann et al., 2005). Patients with PD were asked grip the device and to maximally depress and release a push-button with their thumb as fast as possible over a 30 second period. Movement of the button results in modulation of the voltage output and changes in this were detected using a computer. Moderate correlation between selected parameters derived from the force transducer and the UPDRS score for the hand pronation/supination task were demonstrated. No such correlation was seen when the finger tapping task was considered which, given the similarity of the UPDRS task to the device task, is surprising and calls into question the validity of the system as a tool for motor assessment. Another group chose to mount a pair of force transducers on the index finger and thumb (Brewer et al., 2009); squeezing of the thumb and index finger exerted force on the sensors which was converted to electrical energy and recorded by a computer. In this work the mean of the forces exerted was displayed graphically on a computer screen in real-time. Participants were asked to modulate the applied force to track a target waveform that appeared on the same computer screen for a period of three minutes. A series of parameters derived from the force transducers were identified and from these a regression model was constructed that predicted UPDRS score with moderate success.

Force transducers have also been used to evaluate upper limb dyskinesia, albeit in a small study containing only 8 dyskinetic patients (Caligiuri and Peterson, 1993). Participants exerted force through index finger flexion on a force transducer and were asked to try to maintain isoelectric force (an oscilloscope was used to provide visual feedback for participants). During this process FFT was used to obtain spectral amplitude values for frequencies <2Hz. The sum of these amplitude values was used as their measure of dyskinesia. A statistically significant difference in this value, between assessments performed before and after ingestion of levodopa, was demonstrated. There were no attempts to evaluate the validity of this measure and as such the authors' conclusion that this system is a valid measure of dyskinesia can be refuted.

2.II.viii Gyroscopes

A gyroscope is a device capable of measuring angular velocity; the rate of change of angular displacement of an object. Gyroscopes have been in existence for several hundred years, but recent advances in technology have enabled the development of micro-electro-mechanical system (MEMS) gyroscopes that are small enough in size to be contained on a computer microchip. A number of authors have explored the use of gyroscopes in the assessment of motor symptoms in PD, often in combination with accelerometers which will be considered in isolation later.

Gyroscopic-derived features have been shown to have significant correlation to clinical rating of dyskinesia severity (Burkhard et al., 1999). Critically the clinical rating scale employed to assess dyskinesia severity was not a recognised dyskinesia scoring scale and as such had not been subject to clinimetric validation. Furthermore participants were recorded for periods of only ten seconds and during this time were asked to refrain from making any volitional movements. Artificial neural networks (advanced analytical methods that will be discussed in detail later), have been applied to gyroscope-derived data to evaluate the severity of dyskinesia (Mera et al., 2012a). Whilst results demonstrated strong correlation with those of clinician rating of dyskinesia, this work was also limited by short duration data collection (20 seconds) and the requirement that participants refrained from volitional movements. Clearly

such constraints limit the applicability of such a method to more prolonged unobserved periods of 'real-life' data collection.

Bradykinesia has also been evaluated using gyroscopes, with significant differences in gyroscope-derived variables demonstrated between 'on' and 'off' clinician-defined disease states (Espay et al., 2011). In this work sensors (containing three accelerometers and three gyroscopes) were worn on the thumb and index finger; only gyroscopic-derived measures were used. Whilst a large cohort of 85 PD patients were included, any person exhibiting tremor with a severity of >1 ("slight") on the UPDRS was excluded; whether such results would have been replicated with tremulous patients in the cohort is questionable. A different research group used a forearm sensor, containing three gyroscopes, that was worn by participants whilst performing a series of scripted activities of daily living during a 45 minute period (Salarian et al., 2007). A strong, significant correlation was demonstrated between selected gyroscope-derived parameters and clinician UPDRS scores for both tremor and bradykinesia. Data was also captured whilst patients were moving freely for a period of between 3-5 hours. Analysis of these periods revealed a decline in correlation but, for selected parameters, a strong significant correlation with UPDRS remained. These findings provide encouragement that symptom assessment can be achieved even in the presence of volitional movements. Gyroscopes have shown promise as potential tools to evaluate motor symptoms in PD; their application is however limited by their high rate of power consumption, which may ultimately restrict their ability to perform prolonged monitoring.

2.II.ix   Electromagnetic sensors

Various authors have explored the use of electromagnetic sensors in the identification and quantification of motor symptoms. Such systems use electromagnetic transmitters and a number of sensors whose position, relative to the transmitter, is ascertained from fluctuations in the magnetic field produced.

This method has previously been shown to be able to detect a difference in tremor amplitude between on and off states (Rajaraman et al., 2000). In this work six patients

with prominent upper limb tremor wore a sensor mounted on the tip of the index finger; analysis was undertaken both before and after administration of a dopamine agonist. There was however no attempts to validate measures of tremor against clinical measures such as UPDRS tremor scores. Using a similar set-up, good correlation between physician estimation of tremor amplitude and sensor-derived measures was subsequently demonstrated (O'Suilleabhain and Dewey Jr, 2001). As reported by O'Suilleabhain, the accuracy of the system decreased markedly if the distance between transmitter and sensor exceeded 30 inches. Further limiting the application of this method is the problem of interference; any electrical device in the vicinity of the measuring apparatus will result in electromagnetic 'noise', as it will also produce an electromagnetic field that will interact with the sensors. The high prevalence of electrical devices in the modern home and the restricted range of measurement mean that this method is not currently a viable option for assessment of patients' symptoms in their own home.

Several authors have explored the use of magnetic devices to objectify the assessment of the finger tapping test, a section of the MDS-UPDRS used frequently in clinical practice to assess bradykinesia. The apparatus developed by one group involved the placement of copper wire coils on both the thumb and index finger of subjects (Kandori et al., 2004). An alternating current was passed through one coil resulting in an oscillating magnetic field around the coil. This magnetic field induced a voltage in the neighbouring coil, the magnitude of which is inversely proportional to the square of the distance between the coils. High frequency recording of this voltage allowed calculation of the distance between the coils at a given time, and hence the speed and acceleration of the finger taps to be derived by differentiation. This work demonstrated that the average amplitude, speed and acceleration of taps performed by PD patients decreased as Hoehn and Yahr staging increased. There were a number of flaws to this study however; a lack of appropriately age-matched controls and the absence of tremulous PD patients, since the presence of involuntary movement, such as tremor, is likely to interfere with assessment. Later work used the same apparatus but extracted more detailed features from the data to capture information on

variation within the signal; a significant difference between PD patients and normal, elderly subjects was demonstrated for all indices (Shima et al., 2008).

Objective assessment of finger tapping is inherently limited, since such movements have limited direct applicability to movements performed during activities of daily living. Whilst no formal evaluation of its acceptability to patients was undertaken, it seems unlikely that such apparatus would be tolerated by participants for prolonged period. Consequently, continuous long-term monitoring of symptoms using this method does not appear, at present, to be feasible

2.II.x   Computerised assessments

Various authors have used computerised finger tapping tests as an assessment tool in PD, postulating that rapid tapping movements of the fingers are analogous to the finger tapping test for bradykinesia used in the UPDRS.

A moderate, but statistically significant correlation between the rate of alternating tapping of the "S" and ";" keys on the keyboard of a standard personal computer and motor UPDRS scores has been demonstrated (Giovannoni et al., 1999) . This work failed to consider potential confounding factors such as age, cognitive function and practice effects. Other work sought to correlate finger tapping on a computerised drum machine with nigro-striatal integrity, as assessed with fluorine-18-L-dihydroxyphenylalanine positron emission tomography ([^{18}F]-DOPA PET) (Pal et al., 2001). Only one of the tapping tests yielded a statistically significant correlation and its size was only moderate. The use of an electronic piano keyboard and a computer-based musical instrument digital interface to quantify finger tapping movements has also been trialled (Bronte-Stewart et al., 2000). PD patients and age-matched controls were asked to perform a 'trill'; repetitive, alternating finger tapping on adjacent keys, for 60 seconds. This assessment method produced kinematic measures that showed significant differences when patients were examined in both 'on' and 'off' states. Later work, using similar methodology and equipment, demonstrated that features derived from the keyboard task correlated with both UPDRS motor score and the bradykinesia component of it (Tavares et al., 2005). They were also able to demonstrate

improvements in both UPDRS and kinematic measures when treatment; medications or deep brain stimulation (DBS), was instigated.

The Purdue pegboard test (Figure 8) is a well established test of manual dexterity that measures the total time taken to insert 25 pegs from a rack into a series of holes. Systems similar to this have been applied to PD patients with significant correlation evident between derived scores and UPDRS motor scores (Müller and Benz, 2002, Muller et al., 2000). The method was also able to detect improvement in motor symptoms after ingestion of levodopa (Müller and Benz, 2002). Comparison has been made between these two assessment methods and results suggest that performance at pegboard tasks correlate more closely with motor UPDRS scores than finger-tapping test performance (Muller et al., 2000). In addition, pegboard scores are more responsive to effects of treatment and to advancing disease severity when compared to tapping tests.

Figure 8: The Purdue Pegboard Test (Sawyer and Bennett, 2006)
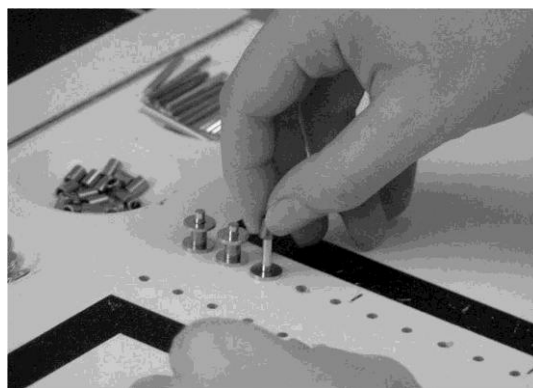
Figure 9 displays an example of an 'at home testing device', which incorporates a combination of tests into a test battery for use at home, including keys for finger tapping tests and an 8-peg pegboard (Goetz et al., 2009). In this work participants were required to complete the test battery weekly for 6 months; derived scores were compared with UPDRS assessments at baseline, 3 and 6 months. Patient satisfaction with this method of assessment was reportedly high and compliance with test completion was in excess of 90%. Whilst the study was underpowered to detect

change (since its primary aim was to evaluate feasibility), a statistically significant improvement in reaction times was seen as the study period progressed. This improvement may represent training effect and this finding brings into question whether such methods of assessment, when applied on a regular basis, can be a valid measure of symptoms over time.

Figure 9: At Home Testing Device (adapted from Goetz et al., 2009)

The use of cheap computer peripherals such as computer-game joysticks and steering wheels has also been explored in PD patients (Allen et al., 2007). Patients used the peripherals to move an on-screen cursor in response to a series of targets and scores were derived from their efficacy of target following. The study was limited by a small sample size but the method was able to distinguish between controls and PD patients. Distinguishing between sub-groups of PD patients with varying severity of bradykinesia, as measured by the UPDRS, was not possible. Cunningham et al. (Cunningham et al., 2011) developed a programme installed onto a notebook computer that aimed to determine the on/off status of the user. Patients used the computer at home for four days and completed tests twice daily; once whilst 'off' and once whilst 'on'. The tests involved moving and clicking a mouse between a series of targets whilst various timings were recorded. The authors claimed that the programme was able to reliably distinguish between disease states, but these claims must be considered in light of there being no external validation of patient identification of their disease state.

2.II.xi   Telemedicine

Telemedicine can be defined as the use of information and communications technology to provide healthcare services to individuals who are some distance from the healthcare provider (Roine et al., 2001). Video-conferencing technology enables clinicians to remotely assess patients, thus removing the need for the patient to travel to the clinic environment. A number of studies have demonstrated the feasibility of using such technology for the out-patient management of patients with PD (Hubble et al., 1993, Hoffmann et al., 2008, Dorsey et al., 2010, Cubo et al., 2012, Dorsey et al., 2013). Comparison of the assessment of motor UPDRS via telemedicine to in-person assessment has shown high levels of inter and intra-rater reliability (Dorsey et al., 2010, Hoffmann et al., 2008), although low levels of agreement have been observed for the assessment of dyskinesia using the Rush DRS (Cubo et al., 2012). It is important to note that the assessment of rigidity is precluded by such a method, since this cannot simply be observed and requires manipulation of a patient's limb by the examining clinician. Assessment of postural stability also requires a 'hands-on' approach and may in fact be unsafe to perform without a clinician being present. Two randomised controlled trials of telemedicine for PD have been performed (Dorsey et al., 2010, Dorsey et al., 2013). These showed conflicting results in terms of the effect of telemedicine on patient-rated quality of life and UPDRS; clearly the inability to blind in such studies is a limitation.

Such technology has the potential to improve the availability of specialist services for patients who find them difficult to access; those in nursing facilities (Biglan et al., 2009) or those who live great distances from their clinic (Dorsey et al., 2010). The costs of setting up such a system have declined markedly in recent years as technology has advanced and become more readily available; such systems may even prove to be more cost efficient (Dorsey et al., 2013). Crucially, many patients hold concerns regarding the confidentiality of telemedicine consultations (Chua et al., 2001). Furthermore, patients undergoing telemedicine consultations reported feeling more shy and nervous about speaking than patients who underwent face-to-face consultations. It is known that psychological stress can exacerbate the motor

symptoms experienced by patients (Marsden and Owen, 1967, Durif et al., 1999). Other work has suggested that patients preferred on-line assessments to office-based assessments, yet interestingly the converse was reported by clinicians (Cubo et al., 2012). This may reflect the fact that many doctors hold concerns regarding the confidentiality of telemedicine consultations and the legal ramifications of such interactions (Nayak et al., 2012). Despite these concerns, this method of assessment has huge potential and its use is likely to expand in the coming years.

2.II.xii Summary

A wide variety of different methods have been explored as potential objective assessment tools of upper limb motor symptoms in PD. Some methods have shown promise when used in a controlled setting on selected patients. Many of the methods described above however, are simply not appropriate for continuous home monitoring of patients; reasons for this include:

- the size, complexity or expense of the equipment required

- the need for clinical supervision during evaluation periods

- inability to evaluate symptoms for a prolonged period

- lack of convincing evidence to support the efficacy of the tool as an assessment method by which upper limb motor symptoms can be evaluated

The use of body-worn accelerometers has shown promise as a potential objective assessment method in PD. A recent review article (Maetzler et al., 2013) summarised the multitude of areas in which accelerometers have been applied in PD, including assessment of both motor and non-motor disabilities (see Table 2).

Table 2: Areas in which body-worn accelerometry has been applied in Parkinson's Disease research

Motor Disabilities	Non-Motor Disabilities
Gait and transfers	Sleep
Freezing of Gait	Autonomic Dysfunction
Balance	
Falls	
Tremor	
Bradykinesia	
Dyskinesia	

The motor disabilities highlighted in Table 2 will form the focus of this thesis. Whilst the use of body-worn accelerometers is well established in gait and balance assessment, description of this field lies out with the remit of this thesis. A review of the literature describing the application of accelerometry to the objective assessment of upper limb motor symptoms will now be presented.

2.III   Accelerometers

2.III.i   What is an accelerometer and how does it work?

An accelerometer is a device that measures acceleration. An accelerometer, when static, can measure the force imparted by earth's gravitational field and can derive from this the angle of tilt of the device relative to the earth. When an accelerometer is moved it is capable of determining the acceleration of the movement it is subject to. There are a variety of different types of accelerometer in use today (Wong et al., 2007). Piezo-electric accelerometers rely on the piezo-electric effect, a property exhibited by certain materials such as quartz. Such accelerometers contain a seismic mass, mounted such that acceleration of the sensor results in movement of the mass according to Newton's Second Law (Yang and Hsu, 2010). Displacement of the seismic mass within the accelerometer results in compression of the piezo-electric material. Application of a mechanical stress to the material leads to the flow of electrical charge within it, with a voltage proportional to the stress applied. From the magnitude of this voltage, the magnitude of acceleration can be calculated. Piezo-resistive accelerometers are similar, but conversely the application of a force results in an increase in the material's electrical resistance. Micro-electromechanical systems

(MEMS) accelerometers (Figure 10) are a newer form of accelerometer that have come to prominence in the last 30 years with the advent of nano-technology. Such devices are miniscule; often measuring less than a millimetre across. Despite their size these devices work on the same principle as other accelerometers; the motion of a seismic mass is detected when the accelerometer is moved by an external force. Most accelerometers are sensitive to acceleration in only one plane of movement; uni-axial accelerometry. By introducing an additional accelerometer, placed perpendicular to the original device, a bi-axial accelerometer can be created. Addition of a third accelerometer, perpendicular to the other two, allows assessment of acceleration in three dimensions; tri-axial accelerometry.

Figure 10: A typical MEMS accelerometer (2012b)
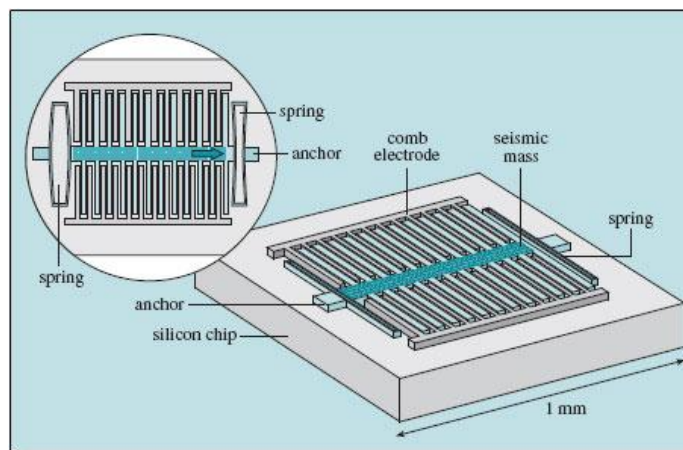
2.III.ii   Actigraphy and its limitations

Various authors have experimented with accelerometers sited at the belt (Saito et al., 2004), ankle (Busse et al., 2004, Skidmore et al., 2008) and triceps (Cereda et al., 2010) to measure the number of steps taken daily, using this as a proxy for general motor activity. This research has demonstrated that PD patients take fewer steps per day compared to controls (Busse et al., 2004, Saito et al., 2004) and that step count in a given period declines as the severity of PD increases (Skidmore et al., 2008). Other authors have used accelerometers on the thigh (Chastin et al., 2010) or combinations of sensors on both thighs and the sternum (White et al., 2006, White et al., 2007) to

51

identify the time spent in different postures e.g. sitting, lying, standing. Differentiation between upright posture and non-upright postures (i.e. sedentary behaviour) can also be used as a surrogate for general functional activity level. It has been demonstrated that bouts of sedentary behaviour lasted for longer periods in PD patients compared to controls, but that there was no difference in the overall amount of sedentary behaviour between groups (Chastin et al., 2010). Such assessments provide a crude global evaluation of motor function but give no information about specific symptoms and their severity.

With regard to the upper limbs, similar devices have been applied to the wrist and used to measure motor activity; this practice is commonly referred to as actigraphy. Such devices typically employ a uni-axial accelerometer which, upon detecting acceleration larger than a pre-set threshold, records this as an 'activity count'. Such devices typically record the number of activity counts in a short time period (typically 5-15 seconds). This data can then be longitudinally displayed over more prolonged periods. In healthy individuals, it has been shown that wearing the actigraph on either the dominant or non-dominant wrist results in no significant difference in recorded activity level (Van Hilten et al., 1993d). Clearly in PD populations, where motor signs exhibited in the limbs are typically asymmetrical, consideration must be given to which side of the body the actigraph is sited. Despite undertaking a thorough literature review it is not possible to quote a globally accepted definition of 'actigraphy', with different authors employing the term in different contexts. Thus for the purposes of this thesis, 'actigraphy' has been defined as the process of detecting acceleration values above a pre-determined threshold, to determine an 'activity count'.

The use of actigraphy has demonstrated that PD patients, when compared to controls, showed less motor activity when monitored over a one week period (van Hilten et al., 1993a) and that more severely affected patients appear to have a less variable profile of diurnal activity. Activity levels have been shown to decline with age for both PD and control groups (van Hilten et al., 1995). This work found no evidence of a relationship between disease duration and measures of activity. However, after dichotomising the PD group by disease severity (using an arbitrarily defined cut of UPDRS score of 18) the

authors demonstrated that those patients with a worse disease burden recorded significantly lower levels of activity. Some early work with actigraphy in PD (Van Hilten et al., 1993c, van Hilten et al., 1991) allowed generation of simple histograms that displayed activity counts for a patient over a prolonged monitoring period; an example of which is seen in Figure 11.

Figure 11: 24 hour activity profile for PD patients measured with wrist-worn actigraph (adapted from van Hilten et al. 1991) [NB: solid line represents mean activity of healthy controls]
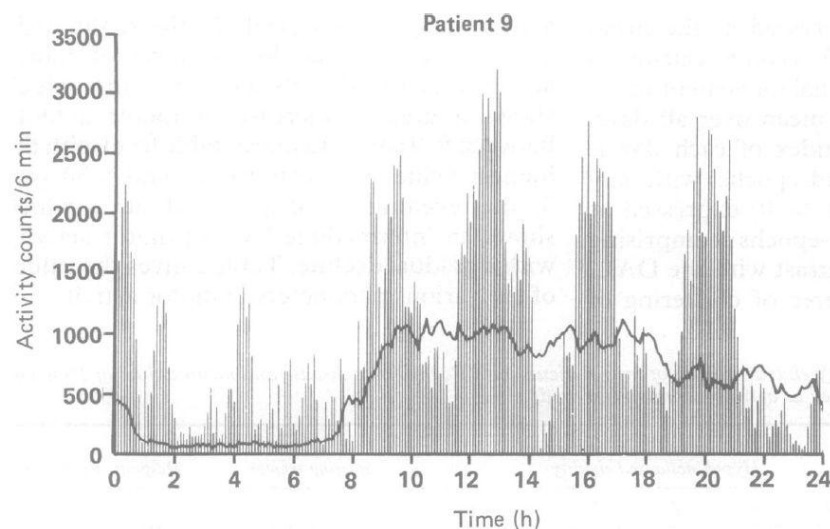
In this work the authors attributed peaks in the activity count to time with dyskinesia and the troughs to periods of hypokinesia. There was however poor correlation between activity count and clinical rating of the severity of dyskinesia as measured by AIMS. This finding suggests that the assumption that peaks correspond to periods with dyskinesia may not be valid. An alternative explanation of periods with high activity levels could be good treatment effect from medications and a resultant increase in mobility. In attempts to evaluate the responsiveness of wrist-worn actigraphy, one study monitored activity in a cohort of PD patients before and after instigation of an additional anti-parkinsonian drug (Katayama, 2001). Activity levels were shown to significantly increase in the second monitoring period and the authors attributed this to positive treatment effect. It could however be argued that the increases in activity count seen were due to an increase in dyskinesia, provoked by the new medication.

Dyskinetic patients were in fact excluded from this study, but other work has shown that dyskinetic patients, monitored with actigraphs, can exhibit activity levels in the range of 'normal' activity levels exhibited by controls (Garcia Ruiz and Bernados, 2008). Tremor also presents a problem, since rapid, repetitive involuntary movements may be mistaken for volitional movement. Many authors chose to try and exclude data in the frequency range of parkinsonian tremor by applying a 'band pass filter' to the derived signal. A band pass filter simply removes certain frequencies from the sample, aiming to reduce the influence of movements of that frequency on the sample obtained. A visual explanation of band-pass filtering is presented in Figure 12. Despite using this signal processing tool it is important to acknowledge that many authors simply chose to exclude both tremulous and dyskinetic patients (Van Hilten et al., 1993c, Katayama, 2001, Van Hilten et al., 1994, van Hilten et al., 1993a); consequently the applicability of this assessment technique to a unselected PD cohort is unvalidated.

Figure 12: Band-pass filtering: a) Initial power spectrum; b) Application of high and low-pass band filters; c) Resultant pass-band spectrum obtained
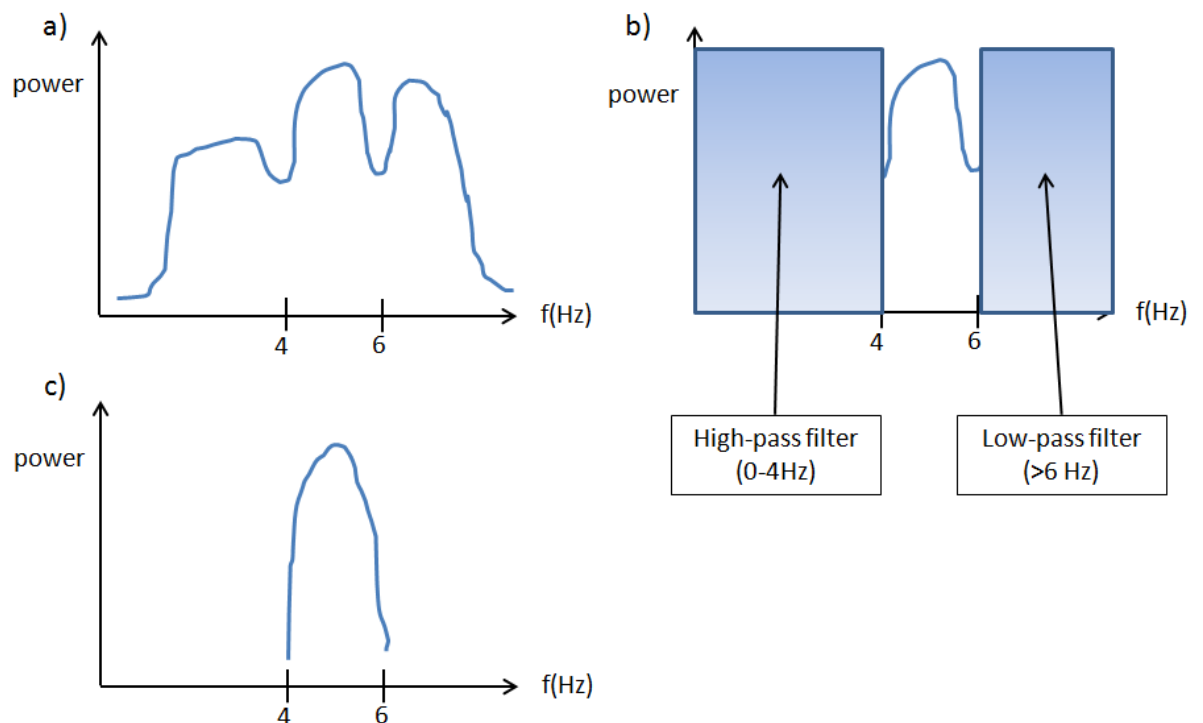
Various authors have attempted to demonstrate criterion validity of actigraphy by examining correlation between actigraphic-derived variables and the UPDRS, the accepted gold standard for the assessment of motor symptoms. Two research groups demonstrated a significant negative correlation between actigraphic-derived measures of activity and motor UPDRS scores, suggesting that more severe PD symptoms are associated with declining activity levels (Garcia Ruiz and Bernados, 2008, Van Hilten et al., 1994). Other work by Van Hilten, using an actigraph on the non-dominant wrist, demonstrated seemingly contradictory results (Van Hilten et al., 1993b). This work reported a *positive* correlation between all measures of activity and the total UPDRS score, suggesting that more severe motor symptoms are associated with increased activity. Their explanation of this finding, that worsening functional impairment of the dominant arm is associated with increased use of the non-dominant arm, seems unlikely. The presence of dyskinetic patients within the study sample may have contributed to this, since greater UPDRS scores may be associated with more prominent dyskinesia, itself producing increased activity counts. In addition, clinical scores of tremor severity correlated significantly with activity measures, suggesting that perhaps tremor may also have interfered with results, with involuntary movements leading to elevation in recorded activity. The band-pass filter used in this study (0.25 – 3.0Hz) will not have excluded parkinsonian tremor entirely.

Another limitation of actigraphy is capture of artefactual data caused by positional changes of the sensor within a gravitational field rather than by movement-induced accelerations. These accelerations are of low frequency and tend to produce a frequency peak at round 0.25Hz (Figure 13a). Application of a tri-axial accelerometry and calculation of a three-dimensional vector can reduce the contribution of gravity to the signal and remove this peak from a frequency spectra (Van Someren et al., 1996) (Figure 13b).

Figure 13: Average normalised power spectra for 15 healthy adults during a writing task: (a) spectrum produced by a uni-axial accelerometer; (b) for the same movements, the spectrum calculated from tri-axial accelerometry (adapted from Van Someren et al. 1997)
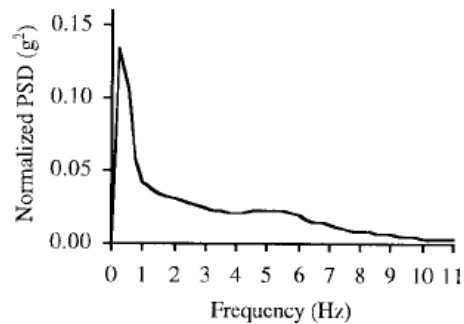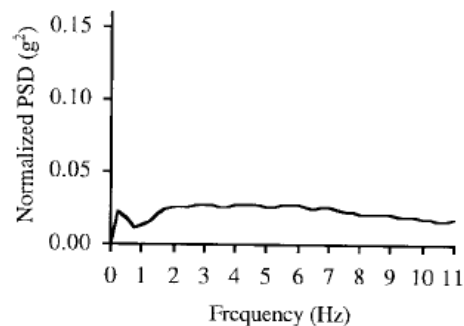
It has been recommended therefore, that for uni-axial devices such as actigraphs, the lower limit of the band-pass filter should be increased to 0.5HZ to minimise the influence of gravitational artefact on results (Van Someren, 1997).

Lastly, Lloret et al. (Lloret et al., 2010) compared actigraphy-derived activity levels during periods of on, off and dyskinesia. This was evaluated in a clinical environment, as part of a levodopa challenge test, and in patients' homes, as patients underwent a 72 hour monitoring period whilst completing symptom diaries. In both instances there was no significant difference demonstrated in activity level between on and off states. Activity levels were however significantly higher for periods of dyskinesia compared to on time.

In summary, actigraphy is a crude method of motor assessment that is capable of quantifying the amount of movement but lacks the ability to reliably differentiate

between movements of different types. The use of actigraphy in the objective measurement of PD is therefore inherently limited by this lack of specificity.

2.III.iii Assessment of tremor

In attempts to make accelerometry a more objective and specific method of assessment in PD, a variety of authors have examined its use in identifying individual Parkinsonian motor symptoms, as opposed to overall activity levels. The bulk of this research has focussed on assessment of tremor. Work in the 1960's (Cowell et al., 1965, Owen and Marsden, 1965) represents some of the earliest applications of accelerometry to this problem. This early work was limited by the technology available at the time; sensors were unable to store data and therefore transmitted it at the point of collection to a nearby receiver. Consequently the patient had to be kept within 30 feet of the receiver at all times and the set-up was also prone to interference from nearby electrical equipment. Limitations of the available technology, even up until the end of the 20th century, led to restrictions in both acceptability and the validity of accelerometry in tremor assessment. The accelerometers used have often been bulky (Pimlott et al., 1983) with the potential to interfere with the movements they were attempting to measure. For example, the apparatus used in one study involved two accelerometers mounted onto a wooden board strapped to the wearer's hand and a second wooden board attached to the opposite aspect of the hand to sandwich the limb within the recording device (Ghika et al., 1993). This system was described by the authors as "portable" but appeared extremely crude and was likely to be incompatible with prolonged wearing and continuous monitoring.

Additional equipment required for functioning of the sensors, such as a battery or a recording device, was often wired to the accelerometer providing further potential restriction of movement (Owen and Marsden, 1965, Cowell et al., 1965, Potvin et al., 1975, Jankovic and Frost, 1981, Pimlott et al., 1983, van Someren et al., 1993b, Smeja et al., 1999, Foerster and Smeja, 1999, Thielgen et al., 2004). Patients were often required to wear this heavy additional equipment on the belt. Such studies gave little or no consideration to patient concordance or patient feedback on their experiences

wearing the devices. One group reported that two of the eight participants in their study "objected" to wearing their device for 24 hours and hence only abbreviated recordings were possible (van Someren et al., 1993b). This work also highlights other potential problems with the technology; data were missing for 25% of participants as a result of device power failure. Advances in technology have lead to smaller components and the ability to store data on miniscule solid-state recording devices. This has enabled more prolonged monitoring periods without the need to carry peripheral devices; more recent work involved continuous data collection for a period of 22 days (Binder et al., 2009). Rapid recent technological development has enabled MEMS accelerometers to be mass produced and this is reflected by the ubiquity of such accelerometers in modern 'smart-phones'. Both the i-Phone[TM] (Lemoyne et al., 2010, Joundi et al., 2011) and Blackberry[TM] smart phones (Daneault et al., 2013) have, in proof of concept work, been used to evaluate tremor with moderate success.

In studies exploring the use of accelerometry to assess tremor there has been wide variation in the number of sensors used, the number of measurable axes in each, and the placement of the sensors about the body; Table 3 summarises, in chronological order, this information. It can be seen that more recent work has tended to employ tri-axial accelerometry and to use fewer sensors.

Table 3: Summary of the number, nature and site of accelerometers used in studies using accelerometry to assess tremor in PD

| Author (et al.) | Year | Accelerometers | | |
		Number used	Number of axes in sensor	Site(s) of placement
Owen	1965	1	Uni-axial	Wrist
Potvin	1975	1	Uni-axial	Dorsum of index finger
Jankovic	1981	1	Tri-axial	Hand, foot or chin
Pimlott	1983	2	Uni-axial	Both on the same hand (10cm apart)
Bain	1993	2	Uni-axial	Dorsum of each hand
Van Someren	1993	1	Uni-axial	Dominant wrist
Smeja	1999	4	Uni-axial	Dorsum of both hands, thigh of dominant leg
Foerster	1999	4	Uni-axial	Both hands, thigh of dominant leg, sternum
Hoff	2001	1	Tri-axial	Wrist (most affected side)
Zeuner	2003	2	Tri-axial	Dorsum of both hands
Caligiuri	2004	1	Tri-axial	Index finger
Thielgen	2004	6	Uni-axial	Both hands, right thigh, sternum (x3)
Van Someren	2006	1	Uni-axial	Wrist
Machowska	2007	1	Bi-axial	Dorsum of hand
Binder	2009	1	Uni-axial	Wrist
Rigas	2009	4	Tri-axial	Both forearms, both shins
Giuffrida	2009	1	Tri-axial	Middle finger
LeMoyne	2010	1	Tri-axial	Mounted in the dorsum of a glove
Costa	2010	1	Uni-axial	Index finger
Muthuraman	2011	1	Not specified	"Hand" (most affected side)
Joundi	2011	1	Tri-axial	Forearm or lower leg
Daneault	2013	1	Tri-axial	Held in the hand (smartphone)

The majority of research in this area has focussed on very short data recording periods, in the region of seconds to minutes (Owen and Marsden, 1965, Cowell et al., 1965, Potvin et al., 1975, Jankovic and Frost, 1981, Foerster and Smeja, 1999, Zeuner et al., 2003, Caligiuri and Tripp, 2004, Farkas et al., 2006, Machowska-Majchrzal et al., 2007, Lemoyne et al., 2010, Joundi et al., 2011, Muthuraman et al., 2011, Giuffrida et al., 2009). These recordings are typically undertaken in highly controlled situations, usually in a clinical facility, with participants typically replicating a series of movements analogous to those performed during clinical examination. Six authors have however described more prolonged periods of data collection. In three of these studies the prolonged period of data collection was preceded by a calibration phase, where short duration recordings were made to provide data to inform subsequent analysis of the

longer, monitoring periods. After calibration had occurred, prolonged monitoring was undertaken for 24 hours either in a rehabilitation centre (Thielgen et al., 2004, Smeja et al., 1999) or in the patient's own home (Hoff et al., 2001b). Other work did not include initial calibration and instead proceeded straight to prolonged recording for 24 hours (Pimlott et al., 1983, van Someren et al., 1993b), or in one case for 6 weeks (Binder et al., 2009).

A number of different analysis methods have been applied to the accelerometer data obtained when studying tremor in PD. In order to identify patterns within data, complex recordings must be summarised by a set of 'features'. A feature is a measure that describes a characteristic of the data at a given time or over a period of time. There are a wide variety of such features used in this research field, some of which are easy to comprehend from a clinical viewpoint, whilst others are more abstract. Simple examples include mean acceleration in a given period, the root mean square of acceleration or jerk (the derivative of acceleration with respect to time i.e. the rate of change of acceleration). Fourier analysis and FFT provide a mathematical way to decompose a signal into its constituent frequencies (Riviere et al., 1997). FFT is widely used in accelerometric analysis of tremor due to the inherent sinusoidal waveform of tremulous movements and allows generation a power spectrum (Figure 6). This analysis method enables calculation of further features that may be used for analysis such as peak frequency, median frequency, centre frequency (CF; the frequency that delineates median area below the power spectrum), frequency dispersion (CF±SD) and power spectral density (PSD; power per frequency band, typically measured in $(m/s^2)^2$/Hz) (Grimaldi and Manto, 2010). Once these features have been selected and the data itself collected there have been a variety of analytical approaches employed by different authors. Some simply plot power spectra for the data collected and describe their visual interpretation of the spectra (Joundi et al., 2011). Many authors (Machowska-Majchrzal et al., 2007, Zeuner et al., 2003, Costa et al., 2010, Jankovic and Frost, 1981, Lemoyne et al., 2010) chose to compare the means of particular features between two groups, arguing in some cases that demonstration of a statistically significant difference between the two means allows differentiation between the

groups. Machowska-Majchrzak et al. (Machowska-Majchrzal et al., 2007) for example, demonstrated significant differences between the means of various parameters (CF, CF±SD) between a group of controls and a group of patients with pathological tremor, including PD. It is important to stress that demonstrating a significant difference between means for a given feature does not imply that the method will be capable of discriminating between the two groups, when presented with new data to interpret. Muthuraman et al. (Muthuraman et al., 2011) acknowledged that most of their selected features showed overlap between PD and ET patients and were therefore of limited use in differentiating between the groups. They did however demonstrate, using a receiver operating characteristic (ROC) curve to set cut-off values, that 94% classification accuracy between conditions could be achieved when the feature variable mean peak power of harmonic peaks was employed.

Other authors constructed 'decision rules', used to classify accelerometer signal content into tremor containing or non-tremor containing. In some cases authors referred to the use of 'tremor detecting algorithms' but provided no details regarding how they developed these tools (Binder et al., 2009). A number of authors opted to categorised signal as tremor or non-tremor based on four rules. (Foerster and Smeja, 1999, Smeja et al., 1999, Foerster et al., 2002, Thielgen et al., 2004). To be classified as tremor the signal had to have a peak frequency occurring within a predefined frequency band (2-7Hz), have >50% of the total power within the peak frequency, have an acceleration greater than a predetermined amplitude and lastly, a duration greater than one second. Van Someren's group used similar decision rules based on amplitude threshold, frequency band and duration but sub-divided non-tremor into 'rest' or 'activity', based on whether or not the amplitude threshold was exceeded (van Someren et al., 1993b, Van Someren et al., 1998). Figure 14 provides a schematic overview of their decision rules being applied to a dummy signal.

Figure 14: Example signal and schematic overview of a tremor decision algorithm used to discern 'tremor', 'rest' and 'activity' (adapted from van Someren et al. 1993b)
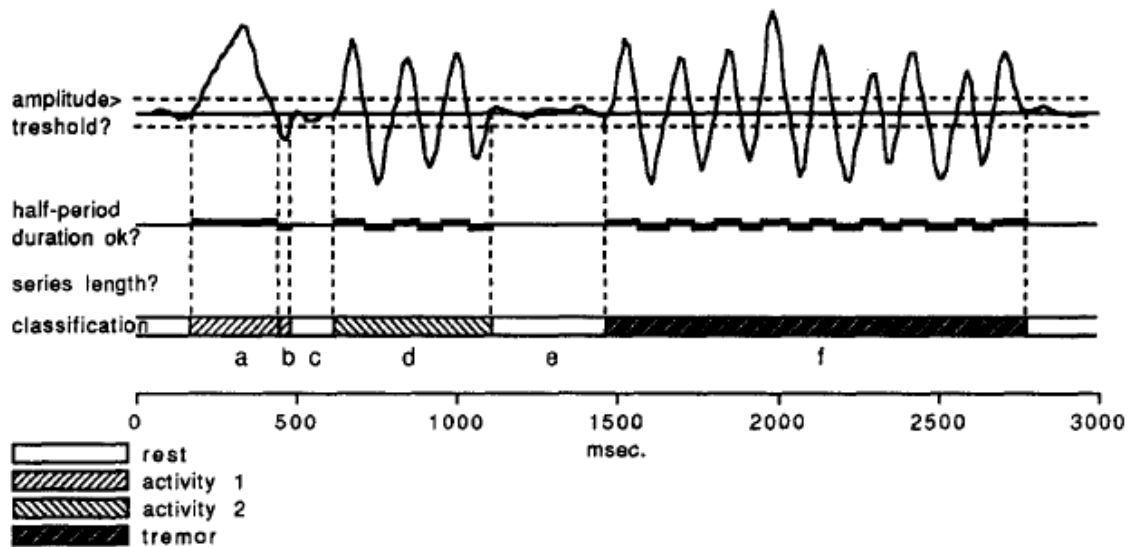
An important consideration when interpreting the methodologies used is the sampling rate employed, i.e. the number of samples per unit time taken from a continuous signal to make a discrete signal. This is typically expressed as Hertz (Hz); samples per second. Foerster et al. (Foerster and Smeja, 1999) and Smeja et al. (Smeja et al., 1999) used a relatively low sampling rate of 16Hz as a consequence of limitations in their storage capacity and battery power. The Nyquist frequency represents the minimal frequency at which sampling must be performed to ensure changes in the signal are appreciated (Figure 15b). To do this, a signal must be sampled at a frequency twice that of the highest frequency it exhibits; the Nyquist frequency. Sampling at rates below the Nyquist frequency results in failure to appreciate changes in the signal (aliasing) as demonstrated in Figure 15a.

Figure 15: The effect of sampling frequency on the signal obtained

a)

Signal of frequency, *f,* being sampled at frequency *f*

● represents sampling

—— represents interpretation of the signal

b)

Signal of frequency, *f,* being sampled at frequency 2*f*
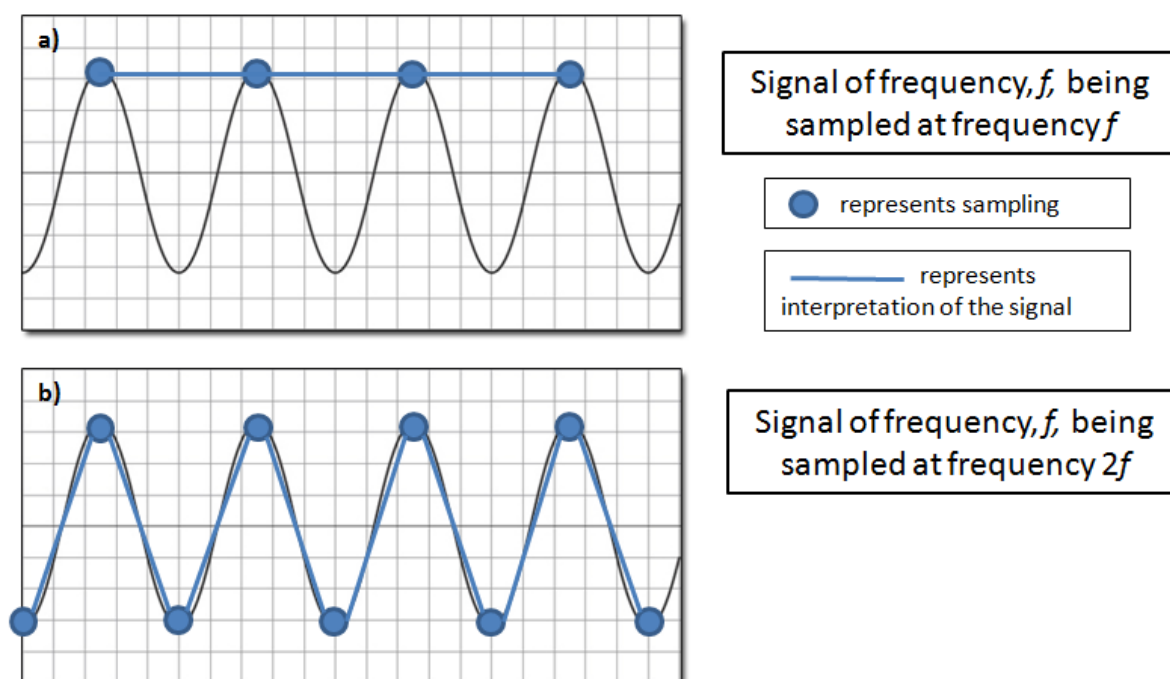
Whilst rest tremor tends to occur between 3-8Hz, it has been demonstrated that action tremor in PD can occur at higher frequencies (6-8Hz) (Liu et al., 1999). The Nyquist frequency is therefore 16Hz (2 x 8Hz); sampling at frequencies below this may result in aliasing.

A variety of other, more complex analysis methods have been applied to this problem and these include multiple linear regression models (Giuffrida et al., 2009), discriminant function analysis (Caligiuri and Tripp, 2004) and artificial neural networks (ANNs) (Rigas et al., 2009). ANN are computational methodologies that perform multifactorial analyses (Dayhoff and DeLeo, 2001). ANNs are said to be 'trained' by providing input data along with corresponding known output values (e.g. UPDRS tremor score). ANNs examine the closeness of fit with output variables and repetitively adjust the weight given to interconnections within the network to incrementally bring outputs closer to their desired values. ANNs require complex computer programming and as such the early research examining their application to medical problems was predominantly published in computer science forums, often with limited clinician involvement. This lack of clinician rigour is evident in the initial work of Rigas et al.

(Rigas et al., 2009) where, despite the use of advanced computational tools, accelerometer data was acquired from a number of junior doctors asked to simulate tremor whilst wearing sensors. The publishing of review articles on ANNs in mainstream medical journals (Dayhoff and DeLeo, 2001) has increased awareness about the topic amongst clinicians. This, coupled with increased collaborative working, has led to increased clinician involvement in studies applying ANNs to clinical problems. ANNs will be discussed in greater detail latterly (2.III.iv) since their use in the assessment of dyskinesia is more developed.

To address the question of whether accelerometry is an acceptable method of assessment of tremor in PD, the validity, responsiveness and reliability of this assessment method will be considered. Several authors have chosen to examine correlations between kinematic features and clinician assessments in attempts to demonstrate the criterion validity of this assessment method.

Early work reported extremely high correlation between accelerometric 'tremor score' and clinician scoring of tremor (Potvin et al., 1975), but relied on a non-validated clinical tremor scoring system developed specifically for use in this study. The widespread use of the clinimetrically validated UPRDS, and its most recent iteration, the MDS-UPDRS provides a comparable, reliable clinical rating scale against which new assessment methods can be compared. Giuffrida et al. (Giuffrida et al., 2009) replicated upper limb motor UPDRS assessment whilst participants wore a finger and wrist-worn accelerometer. This study employed a multiple linear regression model that correlated quantitative features of kinematic signals with clinician UPDRS scores for rest, postural and kinetic tremors. This model produced high correlations across all three assessments and the model's ability to generalise to new data was also proven, with high correlations also demonstrated using 'leave one out' analysis. There was however a bias towards less severe tremor in this study. No patients were deemed to exhibit severe tremor during any tremor task and as such the model's ability to discern severe tremor is therefore questionable.

Caligiuri et al. (Caligiuri and Tripp, 2004) also demonstrated significant correlation between accelerometer variables and tremor severity, as judged by clinician rated

UPDRS tremor scores. This study did however require participants to wear a sensor wired to a computer, and to perform clinical assessments in a highly controlled manner. Conversely, some authors have found no correlation between clinician rating of tremor (UPDRS) and tremor duration or amplitude derived using actigraphy (Binder et al., 2009). This failure to demonstrate correlation may reflect the uncontrolled nature of data collection in this study, with patients being monitored at home for a period of 6 weeks. This work did however demonstrate correlation of actigraphic measures with patient home diaries, where patient self-rated their tremor using UPDRS tremor scores, but the validity of patient self-rating was not established. These findings do act to reinforce the limitations of intermittent clinician assessment and suggest that continuous monitoring methods may reflect more accurately the variation in clinical symptoms exhibited by PD patients. More recent work has explored the use of artificial neural networks (ANN) to assess tremor, initially on simulated tremor patients and latterly on patients with PD (Rigas et al., 2009, Rigas et al., 2012). This work used a series of six accelerometers with participants performing a specified series of tasks. 'Leave one out' validation demonstrated that the ANN was capable of discerning between tremors of differing severities (as judged by UPDRS tremor score) with high reliability.

Hoff et al. (Hoff et al., 2001b) demonstrated reliable detection of tremor using their detection algorithm, reporting high sensitivity and specificity values when decisions on tremor status from accelerometer data were compared with clinical assessment of tremor. They then demonstrated moderate correlation between UPDRS tremor score and both duration and intensity of tremor. It is important to note that this study included data derived from 87 trials performed by 60 PD patients. Linear regression analysis was applied in this study but this method is reliant on data being independent; this is not the case when repeated measurements have been made on the same person (Sainani, 2010) and may have led to over-estimation of correlation. Other research groups have also examined the sensitivity and specificity of their tremor detection methods. Van Someren et al. (Van Someren et al., 1998) used decision rules to detect the presence of absence of tremor. This work relied upon a cohort of 16

control patients in whom the system was tested to examine for the false positive detection rate. The authors concluded that the specificity of their tremor detection algorithm was extremely high (0.94). It must be noted however that the 16 controls in this study were patients with Alzheimer's dementia (AD) and this is likely to represent an inappropriate control group. The assumption that activity levels of AD patients are similar to that of normal healthy people may be invalid. If healthy controls had been used instead, these people may have exhibited more volitional activity than AD patients and this may have resulted in an increase in the amount of time incorrectly labelled as tremor. More recent work by Van Someren et al. again used wrist-worn actigraphy in a highly controlled environment, but this time with more complex decision rules (Van Someren et al., 2006). Whilst tremor classification was highly accurate in their initial cohort of PD patients, the generalisability of this model to new, previously unseen data is questionable, since classification performance deteriorated when applied to a new cohort of patients.

Various authors have shown that accelerometry is able to detect change over time by demonstrating improvement in tremor parameters following a variety of interventions; beta-blockade (Owen and Marsden, 1965), rehabilitation (including optimisation of medications) (Thielgen et al., 2004), thalamotomy (van Someren et al., 1993b), DBS insertion (Papapetropoulos et al., 2008), cabergoline therapy (Binder et al., 2009) and relaxation therapies (Schlesinger et al., 2009). There is however limited evidence for test-retest reliability of this assessment method. Jankovic et al. (Jankovic and Frost, 1981) performed accelerometric recordings for a 30 second period whilst participants performed specified upper limb movements; each being performed 3 times. The authors describe the accelerometric features derived from the three assessments as being "consistent" with one another, but this conclusion appears to be purely based on comparison of the values derived as opposed to any statistical analysis. Smeja et al. (Smeja et al., 1999) undertook 24 hour monitoring with a number of sensors that also allowed posture to be detected. They compared tremor parameters obtained whilst in the sitting position before 12am and after 12am and demonstrated significant differences for tremor occurrence, frequency and amplitude between these time

66

periods. Whilst this does suggest test-retest reliability it is important to note that similar data is not presented for positions other than sitting, when the occurrence of increased volitional movement may make data interpretation less reliable.

Direct comparison between different authors' work is difficult, since different research groups use different sensors, sited in different places and analyse the data using different methods. Many systems have shown promise when participants are monitored whilst performing specified tasks in a highly controlled environment. However, the presence of volitional movement often renders the analytical methods employed less effective. Many of the systems described above therefore restrict patients to performing set clinical tests as opposed to free-living. Similarly, many of the studies highly select patients to avoid those with pronounced fluctuations that could render analysis less effective. In summary, no one system has convincing evidence of its efficacy as an assessment method for Parkinsonian tremor.

2.III.iv Assessment of dyskinesia

Application of traditional analysis approaches to the measurement of dyskinesia has produced limited success. Hoff and colleagues used a series of 8 uni-axial accelerometers attached in pairs to the upper leg, wrist, trunk and upper arm (of most affected side) to assess dyskinesia (Hoff et al., 2001a). Participants exhibiting dyskinesia were asked to perform a series of tasks during which clinicians rated the severity of the observed dyskinesia using AIMS and the DRS. Some of the tasks required the patient to be sat at rest, others required volitional movement. The accelerometric feature selected in this case was the mean amplitude of the dominant frequency seen between the frequency bands 1-4 and 4-8Hz, when the signal was subjected to frequency analysis. Strong, statistically significant correlations were seen between this feature and clinical ratings of severity of dyskinesia for tasks performed whilst patients were at rest (0.80-0.87). However, the strength of this correlation declined markedly for the tasks where volitional movements were present (0.58–0.77). A different group used a rather cumbersome tri-axial accelerometry unit that required taping to the patient's shoulder with adhesive tape to monitor dyskinetic patients for a

period of 20 minutes (Manson et al., 2000). Strong, significant correlations between clinical rating of dyskinesia severity (AIMS and Goetz scales) and the mean acceleration in the frequency band 1-3Hz were demonstrated for periods when participants were seated (0.95-0.99). The strength of this correlation declined markedly when patients were walking (0.72) and as such, periods when patients were walking were latterly excluded from analysis. Unlike Hoff's work however, correlations were maintained in the presence of volitional movement (0.88-0.97). It is however important to note that this analysis method showed very low specificity for mild dyskinesia, with overlap into range of mean accelerations exhibited by healthy control patients.

More recent work (Mera et al., 2013) developed a linear regression model based on features derived from wrist-worn sensors (accelerometers and gyroscopes) and clinical assessment of dyskinesia severity with mAIMS. Optimal performance was seen with a single feature (logarithm of the power spectrum area between 0.3-3.0 Hz), which showed moderate correlation (0.70) to clinical severity of extremity dyskinesia. Data capture was however only undertaken for short periods (20 seconds) in controlled, clinical environments with participants asked to remain stationary whilst assuming a particular body position e.g. arms outstretched. The ability of the system to discern dyskinesia from voluntary movements is therefore unknown.

Artificial neural networks: a possible solution?

 Much of the research presented so far has attempted to find a single feature, or a small number of features, that definitively differentiates between presence or absence of a symptom, or between differing severities of a given symptom. Selection of these features is often driven by medical intuition; for example, examining mean acceleration may be fruitful since it seems logical to suggest that patients with bradykinesia may exhibit slower acceleration compared to healthy controls. Whilst in some cases results have been encouraging, these have tended to require highly-selected patients in controlled, supervised environments. Furthermore, the introduction of volitional movement renders many of these assessment methods obsolete. When consideration is given to the sheer complexity of human movement, one might argue that it is highly unlikely that one feature will possess the ability to

differentiate between subtleties of movement. Furthermore, the assumption that a simple linear relationship exists between features and clinical rating scales may not be valid.

The ability to instead consider many different features that describe human movement and to apply differing levels of importance to each feature may enable more accurate classifications to be made. Artificial neural networks (ANNs) may be a methodology by which this could be achieved. ANNs are computational methodologies capable of performing multifactorial analyses (Dayhoff and DeLeo, 2001). The inspiration for these computational methodologies comes from biology, specifically from the neuron and the complex networks that neurones form to produce neurological systems. Artificial neurones form connections within an artificial neural network and each neuron is essentially a mathematical equation capable of performing a non-linear summing function. The synapses that connect neurones provide a weighting, meaning different input variables can be assigned more or less importance within the ANN. Early versions of ANNs consisted solely of one such layer but more recent iterations, the most well-known being the multilayered perceptron (MLP), have more than one.

An ANN requires data to be inputted and, once analysis of this data has occurred, will produce output data. The development process of an ANN begins with 'training' of the network. In its initial 'untrained' state, the weights on interconnections within the ANN are assigned at random. The ANN is 'trained' using a compendium of data which consists of input data, along with its corresponding known output data. The error between ANN output and the known output is calculated. The weighting assigned to the neuronal connections is then adjusted and the process is repeated. This process is repeated many times, with weightings repeatedly modified in an attempt to minimise the error between ANN output and known output. 'Training' of an ANN can therefore be summarised as serial presentation of data to a network with repeated adjustment of the weightings assigned to input variables and their interconnections until the ANN output is concordant with the known output.

ANNs are typically displayed graphically in layers. See the example in Figure 16 below which displays an ANN developed for use in a study that used ANNs to assess

dyskinesia (Keijsers et al., 2003a). Here the input layer represents a series of accelerometric variables derived from worn body sensors. The output layer in this case represents an integer on a dyskinesia severity scoring system. It is important to note a third layer, the 'hidden' layer, depicted between input and output layers. It is evident that each neuron is connected to every neuron in the subsequent layer. The hidden layer contains computational tools and weightings that are unseen to the user, hence the term 'hidden'. In this example, the ANN output of dyskinesia severity is compared to expert assessment i.e. clinician rating.

Figure 16: Schematic overview of a neural network approach to assessing the severity of dyskinesia (adapted from Keijsers et al. 2003a)
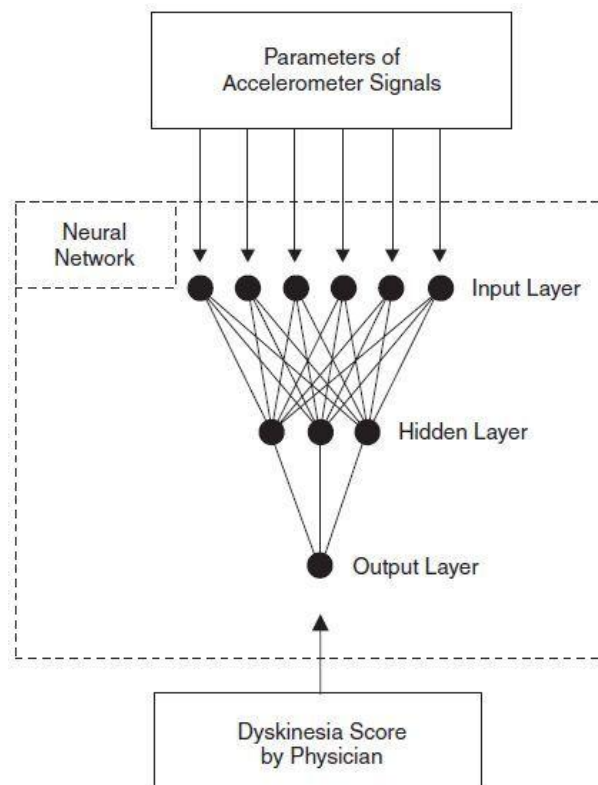
Work by Keijsers et al. (Keijsers et al., 2000) provides an opportunity to directly compare the results achieved by artificial neural networks to those produced by more traditional analysis approaches, since Keijsers used the same data set that was collected by Hoff et al. (Hoff et al., 2001a). The results from this work by Keijsers et al. are summarised below in Table 4. Firstly, in the training set, results demonstrate high

70

network performance, with strong, significant correlations seen between dyskinesia severity as judged by the ANN and by a clinician using a modified version of the AIMS (mAIMS). Correlations were higher in those tasks performed at rest and were comparable to those produced by the analysis method employed by Hoff et al. Correlations between sensor and clinician ratings can be seen to decline for the tasks where volitional movement was present, but it is evident that the ANN outperforms the method employed by Hoff and colleagues. It is important to acknowledge the large mean error values which are likely to result in misclassification of AIMS scores.

For neural networks to be useful and applicable to clinical medicine they must be able to demonstrate similar high performance when applied to other patients; this is typically termed 'generalisability'. Keijsers and colleagues assessed the ability of their neural network to generalise using the 'leave-one-out method'. Using this method the neural network is trained with data from all but one of the study patients, 15 in this case. The resulting neural network is then applied to the remaining patient, which presents the neural network with new, previously unseen data. This process can be repeated for each patient within the study cohort (therefore producing 16 separate neural networks). The network outcome for each previously unseen patient is then compared to the clinician assessment using mAIMS and the results averaged across the study cohort (Table 4). One can conclude that overall, the network performs less well when presented with previously unseen data; correlations are less marked, but remain significant. The decline in correlation strength is more marked when tasks requiring volitional movement are considered. It is important to note however that network performance on the volitional movement tasks still outperforms the results achieved by previous analytical approaches.

Table 4: Performance of the ANN developed by Keijsers et al. (Keijsers et al., 2000) in classifying dyskinesia severity compared to clinical evaluation

Task	Correlation between trained ANN output and actual mAIMS score		Correlation between predicted ANN output and actual mAIMS score (leave one out method)	
	Rs	ME	Rs	ME
Sitting	0.89*	0.37	0.76*	0.51
Counting	0.88*	0.46	0.74*	0.58
Spelling	0.94*	0.48	0.90*	0.57
Drinking	0.82*	0.51	0.60**	0.86
Putting on a coat	0.81*	0.33	0.62*	0.51
Buttoning	0.84*	0.32	0.75*	0.63
Walking	0.75*	0.49	0.67*	0.81

* p < 0.01
** p < 0.05
Rs: Spearman's correlation
ME: mean error

If such a method of assessment is to become a viable tool for home monitoring; a period when clinicians are oblivious to the activities being performed, the neural network employed would need to be able to accurately classify data without the need for prior knowledge of participants' activities during the monitoring period. Keijsers et al. attempted to assess whether their neural networks were capable of this by combining all of their data into one dataset, arguing that by doing so, all prior knowledge of the activity being performed was removed. Neural network classification performance for the combined data set is displayed below in Table 5. Significant, moderately high levels of correlation were seen within the training group. As the complexity of the neural network increased, with the addition of further more input units, the correlation also increased. As seen previously, application of the 'leave-one-out method' to test ANN performance with previously unseen data, results in declining levels of correlation.

Table 5: The influence of removal of prior knowledge about task and varying network architecture on ANN performance in classifying dyskinesia severity compared to clinical evaluation (adapted from Keijsers et al. 2000)

Number of input units in network	Correlation between ANN output and actual mAIMS score for all tasks	
	Rs	ME
Training		
16	0.76*	0.60
38	0.85*	0.48
Leave one out method		
16	0.64*	0.73
38	0.58*	0.66

* p < 0.01		
Rs: Spearman's correlation		
ME: mean error		

Keijsers et al. conclude that these findings demonstrate the ANN's ability to distinguish dyskinesia from voluntary movement and to assess its severity without any prior information about the tasks performed during the data collection period. It is tempting to infer from this conclusion that such a method will therefore be effective when translated to unsupervised home monitoring periods. It is important to acknowledge that the combined data set employed in this work does not truly remove all knowledge of the activities the patients were performing. Data were captured from patients whilst a set selection of activities was performed and even though the labels have been removed, such a data set cannot truly be considered as unsupervised data, since it still consists only of data derived during a known set of activities. Furthermore, some of the tasks performed by participants during data collection were tasks that would be infrequently performed during a typical day in real life (e.g. buttoning a garment, putting on a coat). The results are relatively impressive however, when you consider that the ANN's structure was relatively simple (consisting of only 16 input units) and was based on data recorded over only a very short period of time (one minute). More prolonged monitoring periods and more complex ANN architecture may produce more successful results, but it is important to highlight that more complex networks do not necessarily translate into better network performance when they are applied to new,

previously unseen data. It is evident from Table 5 that the network with the more complex architecture actually performs less well than the network with only 16 input units. This observation is likely explained by a process known as 'over-training'. Over-training can be thought of as the ANN simply 'memorising' the data. The model developed becomes too 'familiar' with the training data and thus loses the ability to generalise when presented with new, previously unseen data. This 'overtraining' effect may also be seen when too many training iterations are performed (i.e. readjustment of weights and examination of error between ANN output and known outputs). This is visually represented in Figure 17 below, where it can be seen that performing more training iterations results in decreased error in the training set. The same can be said for the testing set but only to a point (indicated by n_1 on the x-axis). When the number of training iterations exceeds n_1, the result is increasing error due to failure of the ANN to generalise to new, previously unseen data.

Figure 17: Training, and over-training, of a neural network (adapted from Dayhoff and DeLeo, 2001)
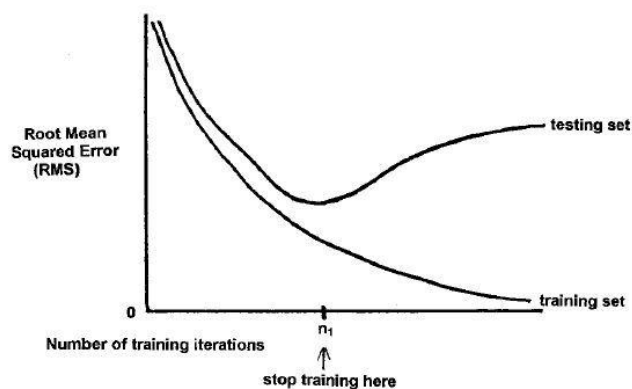
More recent work by Keijsers et al. (Keijsers et al., 2003b) again focussed on the application of ANNs to the assessment of dyskinesia. In this work participants were monitored with a series of tri-axial accelerometers on the sternum, the wrist with the most prominent dyskinesia and bilaterally on the thighs and shoulders. Participants were monitored for a period of 150 minutes in an occupational therapy department whilst performing a series of 35 functional daily life activities. Severity of dyskinesia was rated by clinicians using the modified AIMS and this formed the output for the

74

ANNs. ANNs were developed for each body part (arm, trunk, leg). Data were presented to the ANN in one minute epochs. Based on testing of the ANN developed for the arm (via the leave one out method), correlation of 0.80 was demonstrated between sensor prediction and clinician rating of dyskinesia severity. ANN performance, when presented with more prolonged periods of time (15 minutes), was also evaluated. When epochs of 15 minutes were considered, correct classification was seen in 93.7% of cases (where correct was defined as a difference between network output and clinician mAIMS score of <0.5). Furthermore, the correlation between sensor prediction and clinician rating improved to 0.88. However, like Keijser's previous work, it was evident that classification performance declined in the presence of volitional movement and that there were limited data representing more severe dyskinesia (no participant experienced dyskinesia rated as mAIMS = 4). Consequently there were problems differentiating mild dyskinesia from the absence of dyskinesia, particularly in the presence of volitional movement.

Tsipouras and colleagues have also developed ANNs to assess dyskinesia. Their 2011 paper (Tsipouras et al., 2011) used a series of six tri-axial accelerometers (with gyroscopes) to capture data whilst participants performed a limited series of simple activities in a laboratory environment. The ANN employed consisted of 10 input layers, one hidden layer and four output layers that corresponded to clinical ratings of dyskinesia using the UPDRS. Examining the wrist sensors specifically, it was evident the ANN was able to identify the absence of dyskinesia with high sensitivity and specificity. Performance for dyskinesia of varying severity was however poor with high rates of misclassification. The authors attributed this finding to difficulties inherent to the UPDRS scoring system. A more likely explanation is bias amongst the study cohort; the majority of data epochs represented periods with no 'disabling dyskinesia' (64.8%) and epochs with 'severely' and 'completely disabling' dyskinesia only accounted for 0.5% of epochs. Failure to expose the ANN to sufficient data describing more severe dyskinesia probably accounts for its poor performance. Later work, which employed similar methodologies, demonstrated the superiority of ANNs over other statistical approaches to analysis (Tsipouras et al., 2012). However, as per previous work, the

ANN developed performed well in the absence of dyskinesia but struggled to reliably categorise dyskinesia when present.

2.III.v   Assessment of bradykinesia

Attempts have also been made to objectify the assessment of bradykinesia using accelerometers. Early work involved both PD and control patients wearing wrist-worn accelerometers whilst performing button-pushing tests to assess tap rate and movement time (Dunnewold et al., 1997). This work demonstrated the superiority of considering multi-axis measures (i.e. vectors) of acceleration, as opposed to uni-axial measurement. There was however very limited correlation between accelerometer-derived results and bradykinesia clinical assessment scores derived from the UPDRS. The band-pass filters employed in this study are also questionable; filtering resulted in only signal between 1.0 and 3.2 Hz being analysed. The authors' justification for this bandwidth was twofold; firstly they stated that these filters would exclude tremor from the signal and secondly they reported that voluntary movements do not occur at a frequency greater than 3.3 Hz. There is however, evidence in this study that healthy controls exhibited voluntary movements at frequencies up to 5.3 Hz and thus acceleration values measured in this work may have been underestimated. Work by Van Hilten et al. assessed both hypokinesia, using activity monitoring, and bradykinesia, using tapping tests alone (van Hilten et al., 1998). This work produced no evidence of any relationship between measures of bradykinesia and hypokinesia. Furthermore, neither of the measures displayed any meaningful correlation with clinical assessments of disease status using UPDRS motor scores.

Later work by Dunnewold et al. employed similar methods to their previous work but incorporated a series of accelerometers and a body-position detection algorithm (Dunnewold et al., 1998). These additions enabled reliable detection of periods when the wearer was supine. The authors then excluded these periods of time from assessment of bradykinesia, since they reasoned that whilst supine, patients were unlikely to exhibit much voluntary movement. Despite this, in a similar manner to other such studies, this work did not demonstrate meaningful correlation between

accelerometer-derived measures of bradykinesia and clinical assessment using the UPDRS. Interestingly the authors attribute this lack of correlation to limitations of the UPDRS as opposed to limitations of their methods. They argue that the UPDRS it is subjective, provides only momentary assessment and is based on movements that have little in common with movements occurring in daily life. Hence, they conclude that a lack of correlation between their measures and the UPDRS was not surprising. Whilst Dunnewold et al.'s criticisms of the UPDRS are valid, a counter argument would suggest that the UPDRS, despite its limitations, is the current gold standard of motor assessment in PD and as such the need to demonstrate content validity lies with the newly proposed assessment method.

One group (Okuno et al., 2006) took a different approach to the assessment of bradykinesia, choosing to instrument the finger tapping test employed in the UPDRS with two finger mounted tri-axial accelerometers and metal touch sensors on the finger tips. Subjects and controls performed finger taps for 60 seconds wearing the apparatus whilst bradykinesia was rated by a clinician using the UPDRS. Descriptive statistics presented in this work suggest that as the UPDRS score for finger tapping increases, the standard deviation of finger tapping intervals increases and the mean velocity of movement decreases. No significance values are quoted in this work however and the methodology is described only briefly. Furthermore, the role of control patients is unclear; it is presumed that the controls provided the data for those people exhibiting a UPDRS score of zero but this is not explicitly stated.

2.III.vi Global assessments of motor function

Thus far, much of the accelerometer-based research presented has focussed on one motor symptom, often excluding patients who experienced other symptoms. Clearly this limits the potential applicability of monitoring systems to an unselected patient cohort, since patients with PD frequently exhibit combinations of motor symptoms that may vary during a typical day. Work where multiple symptoms have been evaluated in combination will now be discussed.

Firstly, Mera et al. aimed to measure both tremor and bradykinesia using a home-based monitoring system (Mera et al., 2012b). The system consisted of a finger-worn sensor with a wrist-worn command module and a tablet computer that guided patients through a series of motor tasks derived from the UPDRS motor section. The sensor employed contained three accelerometers and three gyroscopes. The emphasis in this research was not to measure symptoms, but to establish whether or not participants were able to correctly perform the tasks at home without clinical supervision. The authors calculated a 'voluntary movement threshold' (VMT) to help establish whether tasks were being performed correctly, since some tasks required voluntary movement (e.g. bradykinesia) and others required the participant to be at rest (e.g. rest tremor). The VMT was defined as the average peak in the power spectrum of the integral of gyroscopic angular velocity plus two standard deviations. A band pass filter was incorporated to exclude tremor frequencies. Rest tremor task data was rejected if the VMT was exceeded; bradykinesia related tasks were rejected if the VMT was *not* exceeded. Overall, 97% of task data was accepted suggesting that patients were able to perform the tests correctly. It is important to note that 5% of rest-tremor task assessments were rejected, perhaps simply due to voluntary movements or perhaps related to the presence of dyskinesia. Similar to the VMT employed by Mera et al., work by Zwartes et al. incorporated an 'activity classifier' into their analysis to discern whether the wearer was performing a static or a dynamic activity (Zwartjes et al., 2010). This work included six PD patients with DBS in the subthalamic nucleus; none of whom experienced dyskinesia. Participants wore a series of combined accelerometer and gyroscope sensors on the trunk, sternum and on the wrist, thigh and foot of the most affected side. A 'decision tree' was constructed which considered numerous factors including the relative orientation of the sensors to one another to determine posture. This allowed the authors to determine periods of walking, sitting, standing and lying and furthermore whether active arm movement (AAM) was present in each posture. This allowed assessment to be more focussed; rest tremor was only assessed when AAM was absent (and the converse for kinetic tremor) and bradykinesia was only assessed when AAM was present. The system identified rest tremor with high accuracy and sensor measures of rest tremor were shown to

correlate significantly with the UPDRS. Performance in identification of all other symptoms was less impressive with moderate, albeit significant, correlations to the UPRDS seen. The authors also tested the system's ability to discern between different stimulator settings (on, off and intermediate). Whilst a significant difference was shown in rest tremor between all stimulator states, no such differences were demonstrated for other symptoms. Critically there was no statistically significant difference in participants' UPDRS scores between the different stimulator settings; this may simply reflect the small sample size or perhaps the insufficient 'wash-out' time between adjustment of DBS setting and assessment. Recent work suggests this 'wash-out' time can be highly variable amongst patients and that researchers should opt for conservative 'wash-out' periods to avoid this problem (Cooper et al., 2013). The authors reach the questionable conclusion that the ability of their system to detect a change not discernible by the UPDRS, demonstrates its usefulness as a tool to optimise PD treatment.

Addressing the challenge of measuring multiple symptoms at once requires more complex analytical methods and ANNs have been employed by a number of authors. Roy et al. recently published work describing an ANN they developed to measure tremor and dyskinesia in PD patients (Roy et al., 2013). This ANN was trained on features derived from both wrist-worn accelerometry and surface EMG data recorded during a four hour period participants spent in a laboratory setting; a sensor was worn on each limb. The ANN ultimately provided, for every one second interval, a measure of severity (mild, moderate, severe) for tremor and dyskinesia. When the ANN was applied to a previously unseen test cohort, comparison of its outputs to clinician ratings showed identification of tremor and dyskinesia with high sensitivity and specificity (90.2% and 93.5%; 91.7% and 89.5% respectively). Deeper analysis of the architecture of the ANN demonstrated that the addition of features from sEMG resulted in large improvements in network performance. This study demonstrated high ANN performance in symptom recognition in a relatively uncontrolled environment. There was however, a lack of patients exhibiting more marked symptoms in this study; patients exhibiting severity scores of 3 and 4 were combined into one category,

79

termed "severe". The applicability of this ANN to more severely affected patients is therefore questionable.

More prolonged assessment of multiple symptoms was performed by Griffiths et al., in work that required participants with PD to wear a wrist-worn tri-axial accelerometer for a period of 10 days at home (Griffiths et al., 2012). The authors presented data suggesting that the presence and severity of both bradykinesia and dyskinesia can be reliably identified using their accelerometer-derived bradykinesia and dyskinesia 'scores'. Single patient data is also presented with the authors suggesting that temporal fluctuations of symptoms, including changes in light of modifications to medications, can be appreciated. Crucially the authors have no method by which they can demonstrate the validity of these conclusions, since no alternative method of symptom assessment was performed during home monitoring periods, such as patient completed symptom diaries. The authors do report moderate correlation between bradykinesia scores and clinic-based UPDRS scores (parts III) and also weak correlation between dyskinesia scores and UPDRS scores (part IV). The limitations of such clinical scoring systems however render such evidence weak. When discussing their methodology, the authors describe employing an 'expert systems' approach; a computer science phrase that simply describes a computer system that is designed to emulate the decision making process of a human expert. Deeper analysis of their methods raises concerns about how the algorithms employed on the home data were originally developed. The algorithm for dyskinesia recognition was developed using laboratory based assessments where patients with PD performed "naturalistic" tasks whilst concurrently being assessed by clinicians using the AIMS. Whilst significant levels of correlation were evident between the dyskinesia assessment score and the AIMS, only 12 patients contributed to this stage with each being recorded for only 12 minutes. Given that the authors divided the accelerometer data into two minute long epochs this resulted in only six data points for each participant, a very small amount of data on which to base a model. The genesis of the bradykinesia scoring algorithm can also be questioned. In this case, participants were asked to slide their forefinger between two dots on a piece of cardboard for 30 seconds; essentially a variant of the

key-tapping test. The authors demonstrated a moderate correlation between bradykinesia as measured by the finger slide test and their bradykinesia scoring algorithm. The applicability and relevance of such a task to activities occurring during unsupervised home-based monitoring is however highly questionable. Details as to the exact nature of the "algorithms" in this work are limited and this may reflect the fact that the assessment system employed in this study is a commercially available product.

Patel et al. employed a support vector machine (SVM) classifier to the assessment of tremor, bradykinesia and dyskinesia; a strategy of supervised machine learning that is similar to that employed by ANNs (Patel et al., 2009). This approach produced low error rates between clinician-rated severity scores of tremor, bradykinesia and dyskinesia to the output of the SVM (2.8%, 1.7% and 1.2% respectively). The SVM involved five different input features and these were derived from a series of eight body-worn accelerometers. Data were collected whilst participants performed a series of scripted, clinically-based motor tasks derived from the UPDRS. Authors chose to rely on certain tasks to examine for particular symptoms but the appropriateness of their selection is questionable. For example, data derived during the finger tapping task were used to assess for tremor but not for bradykinesia as would be the norm in clinical practice. From Patel's results it was also evident that several motor tasks were suitable for estimating the severity of more than just one symptom. Despite the limitations of this work this result does suggest that machine learning approaches may be able to discern symptoms irrespective of the task participants were performing at the time, a promising result for home-based monitoring.

Rather than examining specific motor symptoms, other authors chose to look at disease status. Hoff et al. assessed 15 PD patients who experienced motor fluctuations using a series of seven body-worn accelerometers worn at home for a period of 24 hours (Hoff et al., 2004). During the monitoring period participants completed a half-hourly symptom diary in which they recorded whether they were 'off', 'on' or experiencing dyskinesia. Measurement of body position, bradykinesia, hypokinesia and tremor were performed using the same analysis methods described previously by this research group (Dunnewold et al., 1997, Dunnewold et al., 1998, Hoff et al., 2001b).

These produced three different variables that were used separately to detect on and off states. Hoff suggested that inter-individual variability in activity levels precluded the use of a single cut-off point for each measure capable of differentiating between disease states. Instead post-data collection analysis using a receiver operating curve (ROC) was performed to produce cut-offs individual to each patient for each variable. Results from this work showed only limited sensitivity and specificity (ranging between 0.66-0.76 and 0.60-0.71 respectively) for the detection of on and off states. There were inadequate numbers of patients with dyskinesia to draw conclusions regarding its identification. The authors suggest that incorrect diary completion by patients may have contributed to the system's inadequate performance, but have no method by which they can externally validate such a claim. It is also important to highlight that the authors use of a 1.0–3.5 Hz bandpass filter, when examining for hypokinesia and bradykinesia, may have excluded many volitional movements that occur at frequencies <1.0 Hz. This paper marks an important milestone in the development of PD symptom monitoring systems. As the authors acknowledge, previous analytical approaches, whilst successful when examining symptoms in isolation, failed to produce similar success when applied to less carefully selected patients with multiple symptoms, monitored in an uncontrolled environment. The limited number of variables that such models can consider is inadequate when the complexity of human movement and the variability of symptoms in PD are considered. Analytical tools capable of considering a large number of variables and their interconnectedness may represent the next step in the evolution of home monitoring systems.

Keijsers et al. acknowledged these problems and sought to address them through the use of an ANN (Keijsers et al., 2006). In this study patients were monitored continuously with a series of six tri-axial accelerometers during a three hour period spent performing a series of ADLs in an occupational therapy department. All patients were known to exhibit motor fluctuations. If no such fluctuation occurred during their monitoring period participants were either given an additional levodopa dose to induce 'on' time or their dose was delayed to induce 'off' time. The method by which disease status was discerned is unclear in this study. The authors comment that only

patients who could "clearly identify their motor state" were included in the study but do not describe how this judgment was made. The authors also reported that "the experimenter used any sign of PD symptoms to determine whether a patient was in an off state". It is therefore unclear whether judgement on disease status was made by participant, clinician or both. Critically, dyskinesia was not considered as a distinct disease state; instead 'on' periods with and without dyskinesia were simply classified as 'on'. The authors attempted to justify this decision by stating that their previous work (Keijsers et al., 2003b) had already demonstrated the ability of ANNs to distinguish dyskinesia from voluntary movements and that several of the study participants in the latest study had contributed to this earlier work. Incorporation of dyskinesia into what is considered 'on' time will result in 'pollution' of the quality of the data representing 'on'. There is a possibility that discerning between 'off' and 'on' will be made easier if what is deemed 'on' time also encompasses dyskinesia. This methodological flaw was further compounded when it transpired that "about half" of the patients suffered dyskinesia in the 'on' state.

Two different analytic approaches were employed. The first approach examined four different accelerometer-derived variables. For each variable ROC curves were used to compute a threshold to allow differentiation between on and off. The second approach involved development of an ANN. The same four accelerometer-derived variables were used as input features for the ANN on a minute by minute basis; 24 features were used in total (four features derived from each of the six sensors). The network output related to disease status and was either 'on' or 'off'. Training of the network was undertaken using back propagation; a process of optimisation where attempts are made to minimise the error between network output and clinician rating of disease status and the relative feature importance to the model is adjusted. Results of the single variable analysis for wrist-worn sensors were comparable to those of Hoff et al. (Hoff et al., 2004), with sensitivity and specificity values of 0.71-0.74 and 0.78 for on-off detection demonstrated. An ANN with two hidden units and four input parameters produced the highest performance with sensitivity and specificity of 1.00 and 0.98 respectively. This was a relatively simple ANN and further analysis of the

feature weighting reveals that 2 variables alone, %PF$_4$ (trunk) and %PF$_4$ (leg), accounted for 96% and 2% of network performance respectively. %PF$_4$ represents the percentage of peak frequencies that occurred above 4Hz, where peak frequency is the frequency that occurs with the largest power on Fourier analysis. Patients in the on state, performing volitional movements, are likely to exhibit movements of frequencies <4Hz. Conversely, tremulous patients in the off state are likely to exhibit movements of frequencies >4Hz due to tremor. It is however unclear why 'non-tremulous' patients in the off state, would exhibit movements of frequencies >4Hz, and the authors' suggestions (postural tremor, physiological tremor, body surface 'micro-vibrations') are unconvincing.

Whilst the ANN performance quoted in this study appears impressive it is crucial to note that the authors present only training data. No attempt has been made to apply the ANN to new, previously unseen data. It is therefore impossible to conclude whether the model will be applicable to other patients with PD; the impressive results seen may simply reflect 'over-training' and thus the model may lack the ability to generalise when presented with new data. The authors' conclusion that the system can "operate successfully in unsupervised ambulatory conditions" is therefore highly questionable. The data in this work cannot truly be considered unsupervised, since data collection was based on recordings during a known set of activities and do not reflect truly unsupervised data. Furthermore, the claim that the system is able to operate in ambulatory conditions is highly contentious when you consider that periods where patients were walking for longer than three minutes were excluded from data analysis.

2.IV   Summary

The above review highlights the wide variety of methods by which accelerometers have been applied to the evaluation of motor symptoms in PD. This review has demonstrated that analysis involving a single variable, or small numbers of variables, proved insufficient when volitional movement is introduced. Furthermore collecting data in controlled, laboratory-based settings with highly selected participants

performing scripted tasks is unlikely to produce data representative of real-life, home-based activities. This data collection is also limited in terms of time, since more prolonged data capture is not feasible in such environments. It is perhaps no surprise that application of such models to unsupervised, home-based activities has been shown to yield disappointing results. This review has also introduced artificial neural networks; complex computer algorithms capable of considering huge numbers of features and critically, their relations with one another. Application of such methods to accelerometer data collected from PD patients has yielded encouraging results, even in the presence of volitional movement and even when knowledge about patient activity has been removed.

The aims of this work are therefore as follows:

1)   Can wrist-worn accelerometers (and analysis with artificial neural networks) reproduce patient-completed symptom diaries during prolonged periods of unobserved home monitoring when prior knowledge of disease status is, for a given patient, removed?

2)   Can wrist-worn accelerometers (and analysis with artificial neural networks) reproduce clinician assessment of PD patients' disease state during periods of observation in a clinical environment, when prior knowledge of disease status is, for a given patient, removed?

3)   Is prolonged monitoring using bilateral wrist-worn accelerometers acceptable to patients?

To achieve these aims wrist-worn accelerometers will be employed to capture data from PD patients during a seven-day, home-monitoring period. Capturing data in this environment produces two distinct advantages to the more traditional methods of data capture described above. Firstly, the data recorded will reflect the wearer's activities of daily living more closely than laboratory-derived data. This is based on the assumption that the wearing of the sensor does not influence the wearer's usual pattern of activity; a detailed evaluation of the acceptability of the wrist-worn devices to patients will therefore be undertaken. Secondly, unsupervised data capture at home enables huge amounts of continuous data to be captured. Based on the data captured

during the home monitoring period, an artificial neural network will be constructed; large volumes of data are critical to the development of such a method. This artificial neural network will be capable of analysing periods of accelerometer data and providing a 'decision' as to the disease state of the patient at that time. Attempts will be made to validate the artificial neural network's 'decision' against patient completed symptom diaries and against clinical evaluations made in a laboratory environment.

Chapter 3.   Methods

3.I   Recruitment

Study design was observational, with PD patients being recruited from the Northumbria Healthcare NHS Foundation Trust Movement Disorder Service between January and November 2012. Patients were approached in both medical and PD nurse-specialist clinics that took place across the region covered by the service. The service covers both urban and rural areas in the Northeast of England with age-adjusted prevalence estimates of PD in these areas previously estimated to be 139 cases (95% CI 116-162) per 100,000 (Porter et al., 2006) and 142 cases (95% CI 118 to 165) per 100,000 (Walker et al., 2010) respectively. The target for recruitment was 32 participants. Confidential demographic information was collected for all patients approached, thus allowing comparison between participants and those declining participation in the study.

Inclusion criteria for entry into the study were:

- Ability to give informed consent
- Aged 18 years or above
- Diagnosis of idiopathic PD as per United Kingdom (UK) Brain Bank Criteria (Gibb and Lees, 1988)
- Under the care of the Northumbria PD service
- Taking immediate release levodopa medication

Exclusion criteria for the study were:

- Inability to provide informed consent

- Presence of significant cognitive impairment; defined as mini-mental state examination (Folstein and Folstein, 1975) (MMSE) score of <24

- Immobility, i.e. Hoehn and Yahr (Hoehn and Yahr, 1967) stage 5

- Not taking immediate release levodopa medications

Patients who fulfilled all inclusion criteria were recruited for study participation. The study consisted of two phases. During phase one, participants attended Newcastle University's Clinical Ageing Research Unit (CARU). The final study phase, phase two, took place in participants' homes. During both phases of the study participants wore bilateral wrist-worn sensors. Figure 18 summarises the research protocol, with each phase subsequently discussed in greater detail.

Figure 18: Summary of Study Phases 1 and 2
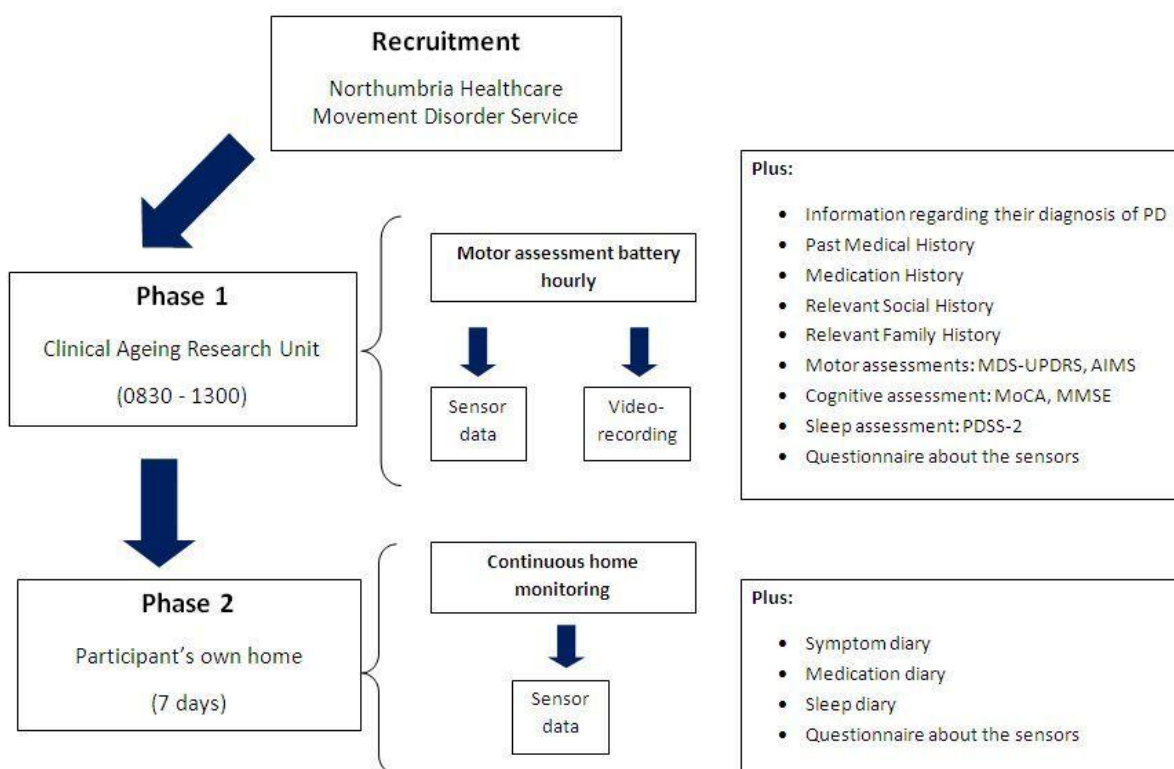
3.II Ethical Approval

The study proposal for this research project was submitted to the County Durham and Tees valley Research Ethics Committee on 13th September 2011. Final ethical approval, following minor amendments, was granted on 11th October 2011.

3.III Phase 1 Procedures (CARU)

Participants who consented to involvement with the study attended CARU and were encouraged to bring a companion (friend, relative or carer). Transport was arranged to

ensure that participants arrived at CARU at approximately 0830 hours. On arrival at CARU consent was reviewed and confirmed. Participants were allocated a randomly generated four letter identification code to allow subsequent identification of participants in a pseudo-anonymised manner. Participants were fitted with two sensors that attached to the wrists (Figure 19). Accelerometer data were recorded continuously for the duration of the participants' time at CARU. During this time subjects were free to move around the clinical environment without restrictions. Activities performed by participants were selected to include a variety of different postures and mobility states i.e. walking, standing, sitting etc. Participants also performed a number of activities during their time in CARU that were likely to be similar to activities of daily living performed in their own home; these included interacting with family members, reading, writing and eating.

Figure 19: Orientation of sensors on participants' wrists (adapted from Axivity AX3 User Guide(2013b))
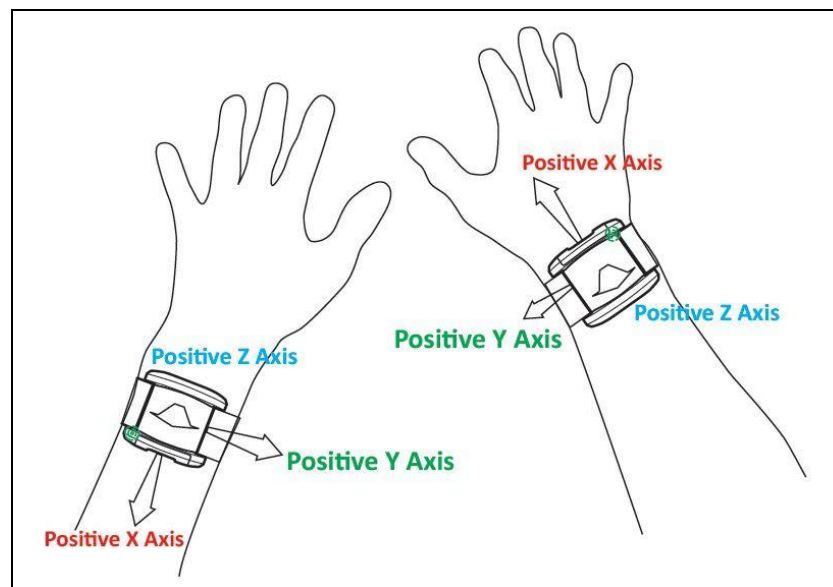
On the day of their attendance at CARU, participants were asked, where possible, to refrain from taking their first morning dose of immediate release levodopa. This enabled initial assessments at CARU to be undertaken with the patient typically in the 'off' disease state, a practice recognised as being important in enabling study of the

motor deficits due to the disease process of PD itself (Gordon and Reilmann, 1999). The initial assessment at CARU consisted of a battery of motor tests derived from two clinical rating scales, the MDS-UPDRS and the AIMS. Each individual motor test required the participant to perform a particular action during which the severity of clinical sign observed was rated against specified descriptors, thus providing an integer value representation of symptom severity (MDS-UPDRS: 0-4, AIMS: 1-5). The entire motor assessment battery is summarised in Table 6 below.

Table 6: Summary of CARU Motor Assessment Battery

Motor Feature	Test	Scale	Sub-section
Tremor	Rest Tremor	MDS-UPDRS	3.17
	Rest Tremor (with distraction)	MDS-UPDRS	(3.17)
	Postural Tremor	MDS-UPDRS	3.15
	Kinetic Tremor	MDS-UPDRS	3.16
Bradykinesia	Finger Taps	MDS-UPDRS	3.4
	Pronation-Supination	MDS-UPDRS	3.6
Dyskinesia	Upper Limbs	AIMS	5
	Lower Limbs	AIMS	6
	Trunk	AIMS	7
	Global Severity	AIMS	8

Key Outcome Measure:

- The rating clinician also passed judgement on participants' disease state at the time of each clinical assessment, selecting from one of the following categories: off, on or dyskinesia.

- Whilst participants were undertaking each motor assessment battery, a digital video-recording of the assessment was also taken. This enabled a second, blinded clinician (RW; an experienced movement disorder specialist) to also pass judgement on participants' disease state and to rate the assessments using the same clinical rating scores.

- Accelerometer-derived estimations of a participant's disease state at a given time would ultimately be compared against the clinician reported disease state recorded here.

After completion of the initial motor assessment battery participants were invited to take their immediate release levodopa medication. The motor assessment battery described in Table 6 was performed serially throughout each participant's time at CARU, typically 30 minutes after ingestion of the immediate release levodopa and at hourly intervals from the initial assessment. Serial assessments before and after ingestion of levodopa, allowed for appreciation of any potential variance in motor symptoms experienced by participants and enabled accelerometer data to be captured over the course of this cycle. As such, patients not on immediate release levodopa were excluded from this study. Those participants who were unwilling or unable to refrain from withholding their morning immediate release levodopa remained in the study.

Participants also underwent a full medical history and examination during phase one. The diagnosis of PD was confirmed using the UK Brain Bank Criteria. Medical history included evaluation of past medical history, family history, social history and review of participants' medications. To enable comparison between participants who were taking a variety of different dopaminergic agents with differing regimens, the medication history was used to calculate the levodopa equivalent dose (Tomlinson et al., 2010). During phase one all participants were also assessed using the complete MDS-UPDRS proforma which captured data regarding the motor and non-motor experiences of daily living, motor complications and included an entire motor examination. Dyskinesia was assessed using the AIMS, which has previously been recommended for use in a PD population (Colosimo et al., 2010) which critically appraised eight different dyskinesia rating scales. The AIMS evaluates the severity of the actual dyskinetic movements themselves, hence its selection for use in this work. The AIMS consists of 10 items, each rated by a clinician using a five point Likert scale from 1 to 5: absent, minimal, mild, moderate, and severe. Items 1 to 4 rate the presence and severity of abnormal movements in the face and mouth. Inclusion of assessment of orofacial dyskinesia reflects the fact that the AIMS was originally developed for the rating of tardive dyskinesia. Items 5 and 6 rate the presence and severity of abnormal movements in the limbs; item 7 does so for the trunk. Items 8 to

10 rate the global severity of the abnormal movements, the disability as a consequence of the abnormal movements and lastly, the patient's awareness of the abnormal movements. Cognition was assessed with both the MMSE and the Montreal Cognitive Assessment (MoCA) (Dalrymple-Alford et al., 2010). The MMSE is a widely used cognitive assessment tool that predominantly tests memory and language; a narrower range of cognitive domains than that which the MoCA addresses (Gill et al., 2008). The use of the MMSE has been questioned in a PD population, with the MoCA being shown to be more sensitive in detecting mild cognitive impairment (Nazem et al., 2009). Sleep was assessed using the Parkinson's Disease Sleep Scale (PDSS-2) (Trenkwalder et al., 2011), as during this research participants would wear the wrist-worn sensors during their sleep. Participants were categorised by disease phenotype (TD or PIGD) using a formula which categorises different MDS-UPDRS items by phenotype (Stebbins et al., 2013).

Once clinical assessments were completed the sensors were removed. The final assessment participants completed was a questionnaire designed to capture their thoughts and feelings about wearing the sensors. This consisted of nine statements (items). Participants were asked to display their level of agreement for each of the items using a five point Likert scale. There was additional space on the questionnaire for free text comments from participants.

3.IV   Phase 2 Procedures (Home)

During phase two of the study participants wore bilateral wrist sensors for a seven day period whilst in their own homes. The sensors were given to participants before leaving CARU, on completion of phase one. Sensors were fully charged to ensure a full seven day period of monitoring could be undertaken. Participants were asked to put on the sensors and commence the week long monitoring period at 0800 hours the following day. If the subsequent week was inconvenient for the participant, an alternative date was agreed and a fully charged set of sensors was provided at this time. During phase one all participants received a standardised briefing about phase two of the study. Participants received explicit guidance, along with accompanying

photographs to take away with them, explaining how to ensure that the sensors were placed on the correct wrist and in the correct orientation (coloured coded; see Figure 23). Participants were asked to wear the sensors continuously, including at night, and to go about their daily activities as normal. Participants were advised to discontinue wearing the sensors if they felt they had become particularly burdensome. Participants were aware that the sensors were waterproof and could be worn during washing or bathing but were informed that if they preferred, the sensors could be removed for short periods. Participants were advised to refer to the supplemental information provided (see appendices) to ensure the sensors were correctly re-sited after a period of non-wearing. Participants were informed that the sensors stored only data pertaining to acceleration, light and temperature, and reassured that all data was stored on the device itself, with none being transmitted elsewhere. Participants were reassured that if the sensors were damaged they would not be liable for any replacements. It was explained to participants that the flashing LED, clearly visible during phase one of the study, would be disabled for the home monitoring period in an attempt to make the device less conspicuous and to avoid any sleep interference during night-time.

Participants were asked to keep three daily diaries during phase two of the study; a symptom diary, a medication diary and a sleep diary. These diaries were printed in the form of a 'home monitoring booklet' (see chapter 12.III.i) to allow participants to carry the diary around as one document. Within the home monitoring booklet there were telephone numbers for the research team to ensure that participants had a point of contact in the event of any problems or difficulties with the sensors or diaries.

Key Outcome Measure:

- The symptom diary required participants to record, on an hourly basis, their predominant disease state. The options available to choose from were 'on', 'off', 'on with troublesome dyskinesia' and 'asleep'.

- Accelerometer-derived estimations of a participant's disease state at a given time would ultimately be compared against the patient reported disease state recorded here.

Written descriptions of what constituted each of the available disease states were included in the diary. The times at which responses were required were pre-printed to assist patient completion of the diary. During the standardised briefing performed in phase one, each different disease state was explained to participants. In an effort to decrease inaccurate diary completion, participants were made aware that accuracy of responses was more important than the number of responses (Bolger et al., 2003). The example completed symptom diary given to participants is shown in Figure 20 below.

Figure 20: The example symptom diary (AM only) provided for participants in the 'home monitoring booklet'

Time (AM) Starts at Midnight	ON	ON with troublesome dyskinesias	OFF	Asleep
12:00 am - 1:00 am				X
1:00 am - 2:00 am				X
2:00 am - 3:00 am				X
3:00 am - 4:00 am				X
4:00 am - 5:00 am				X
5:00 am - 6:00 am				X
6:00 am - 7:00 am				X
7:00 am - 8:00 am			X	
8:00 am - 9:00 am	X			
9:00 am - 10:00 am	X			
10:00 am - 11:00 am		X		
11:00 am - 12:00 pm			X	

When completing the symptom diary, participants were asked to avoid selecting two different disease states for one given time period; such an entry would be considered unclassifiable. Whilst the research team acknowledged that participants could exhibit a variety of disease states within a one hour period, participants were asked to pick the option that reflected their predominant disease state for that period. Participants were asked to complete the diary contemporaneously wherever possible, and to try to avoid

retrospective completion of the diary for long periods of time. If participants had not made a diary entry for many hours, rather than potentially incorrectly recalling their disease state, participants were asked to leave such an entry blank. Despite this advice, blank diary entries were recorded as unclassifiable for purposes of data analysis.

The sleep diary required participants to record the approximate time at which they went to bed, defined as the time at which they turned the lights out for bed, and the time they got out of bed in the morning. Participants were also asked to record any disturbances to their sleep overnight and the approximate times at which these occurred. The example of a completed sleep diary given to participants is shown in Figure 21 below.

Figure 21: The example sleep diary provided for participants in the 'home monitoring booklet'

At what time did you turn the lights off for bed?	11:15 PM
At what time did you get out of bed in the morning?	6:45 AM
Were there any disturbances to your sleep overnight? (Please try and include time if possible)	Got up to go to the toilet at around 1am. Couldn't get back to sleep for at least half an hour after this

The medication diary required participants to record their anti-parkinsonian medications for the week long period of home monitoring. For each of their anti-parkinsonian medications, participants were asked to record the name, dose, scheduled time of administration and the actual time of administration. The example completed medication diary that was given to participants is shown in Figure 22.

Figure 22: The example medication diary provided for participants in the 'home monitoring booklet'

Medication Name	Dose	Time Prescribed	Time taken (please write here)
Madopar	62.5mg	8:00 AM	7:45 AM
Madopar	62.5mg	12:00 PM	12:10 PM
Madopar	62.5mg	4:00 PM	4:00 PM

The home monitoring booklet contained seven days worth of blank diaries ready for completion by participants. Given the cognitive demands placed on participants by requiring them to complete three diaries accurately and contemporaneously, for the duration of the week long monitoring period, patients with a MMSE <24 were excluded from this study. The final page of the home monitoring booklet was another copy of the sensor questionnaire. Participants were specifically asked to ensure that this was the final task they completed, having removed the sensors after a week of wearing. If, for whatever reason, participants decided to discontinue the home monitoring period, they were asked to complete the questionnaire at this point. The same questionnaire was administered to participants to explore whether their responses differed between phases one and two. Participants were provided with pre-paid envelopes to return the diaries and sensors to the research team on completion of phase two. Envelopes were marked with each participant's unique, pseudo-anonymised four-letter code to ensure that the sensors could be matched to the correct participant. Return of the diaries and sensors marked the end of a participant's involvement in the study.

Figure 23: Axivity AX3 devices mounted in Velcro straps (colour coded to indicate the hand on which they were to be worn; red: right, blue: left)

3.V Accelerometer Protocol

3.V.i The accelerometer sensor

The device used was the Axivity AX-3 (Figure 24). The device contained a tri-axial micro-electromechanical (MEMS) accelerometer capable of sensing accelerations of up to 16g in three different axes, up to a frequency of 2kHz (2013a). The device also contained a real-time clock, a temperature sensor and an ambient light sensor. The device measured 6mm x 21.5mm x 31.5mm.

Figure 24: Axivity AX3 Device (adapted from Axivity AX3 Data Sheet(2013a))
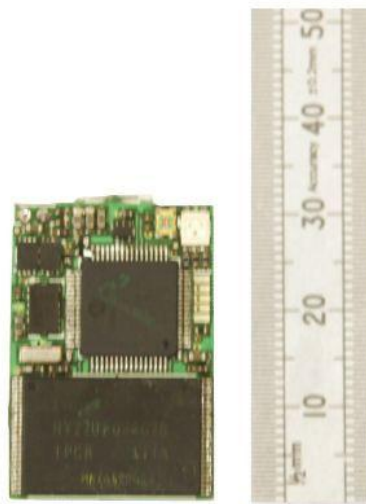
The device drew its power from a lithium polymer battery, which took approximately two hours to completely charge. The battery life of the device was primarily dictated by the sampling rate of the accelerometers. As demonstrated in Table 7, the greater the sampling rate the higher the rate of battery consumption. During this research all devices were set to sample at a rate of 100Hz, thus allowing a potential maximum of 14 days continuous recording. Such a sensor can also be equipped with a MEMS gyroscope, a device capable of measuring angular velocity. For the purposes of this research the decision was taken not to include gyroscopes in the sensor. Whilst gyroscopes can provide useful additional data there were concerns that their comparatively high power requirements would reduce battery life, such that the

sensor would be unable to perform prolonged monitoring periods without being recharged.

Table 7: Sample Rate and Battery Life (adapted from Axivity AX3 User Guide(2013b))

Sample Rate (Hz)	Battery Life (days)
400	5
200	9
100	14
50	21
25	34

The Axivity sensor was fully encapsulated within a hermetically sealed, fully waterproof polycarbonate casing. The casing increased the dimensions of the device to 39mm x 36mm x 12.5mm. The device was attached to the wearer's wrist using an adjustable Velcro strap. The combined weight of the sensor, casing and strap was 35 grams. The device, along with a description of each its component parts, is shown in greater detail in Figure 25 below.

Figure 25: Axivity AX3 Device Component Parts (adapted from Axivity AX3 User Guide(2013b))
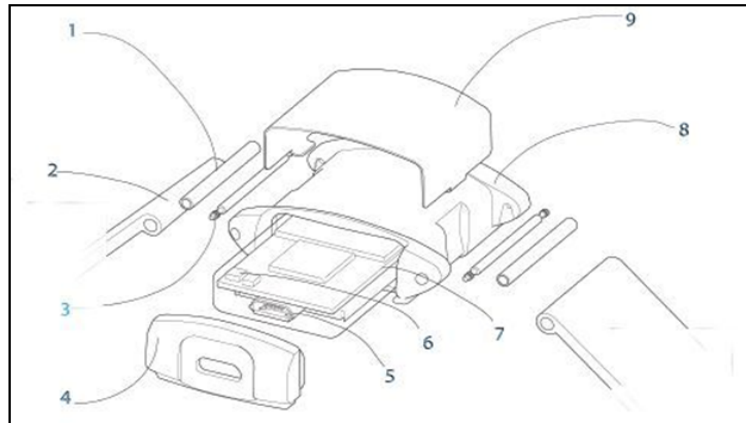
Item	Description	Note
1	Strap Bar	Used to reinforce the strap
2	Strap	Used to connect the device to the user's wrist
3	Strap Pin	Connects the strap to the casing. Made from stainless steel
4	Outer Casing lid	Made from food-safe, hypoallergenic Polycarbonate
5	USB Connector	Used for configuration, download and charging
6	LED	Used to display status to the user
7	Inner Puck	Contains all electronics in a waterproof package
8	Outer Casing	Made from food-safe, hypoallergenic Polycarbonate
9	Clip	Optional extra

Whilst the device was recording data, a green LED flashed on and off to indicate that data collection was ongoing. The three axes in which acceleration was measured are demonstrated graphically in Figure 26.

Figure 26: Orientation of Accelerometer Axes (adapted from Axivity AX3 User Guide(2013b))
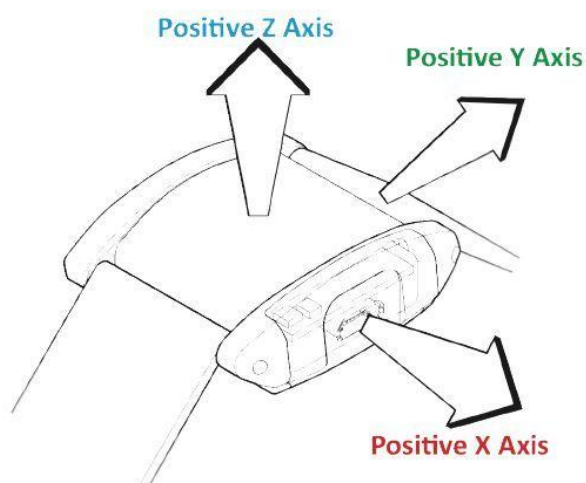
Data generated by the device was stored on a memory chip capable of storing a maximum 512 megabytes (MB) of data; adequate capacity to store data for the duration of the battery life. The device could be connected to a computer via a Universal Serial Bus (USB) cable to allow both reconfiguration of the device and downloading of the data collected. Downloading a complete data set of 512Mb took approximately six minutes. There were no buttons or switches on the device that required manipulation by the wearer.

Prior to initiating each motor assessment battery during phase one, markers were placed into the accelerometer data to indicate the start of the examination period. This was done by subjecting the sensors to a short impulse force produced by tapping firmly several times on the sensors. Firmly tapping the sensors resulted in a short period of unusually high acceleration that is easily discernible when the data is later analysed. An example of this process of data marking is depicted in Figure 27. Ensuring that this data marking occurred before each assessment meant that the period of

accelerometer data corresponding to the clinical assessment was readily identifiable on later inspection of the accelerometer data.

Figure 27: Data markers visible as large spikes in the tri-axial data stream (adapted from Axivity AX3 User Guide(2013b))
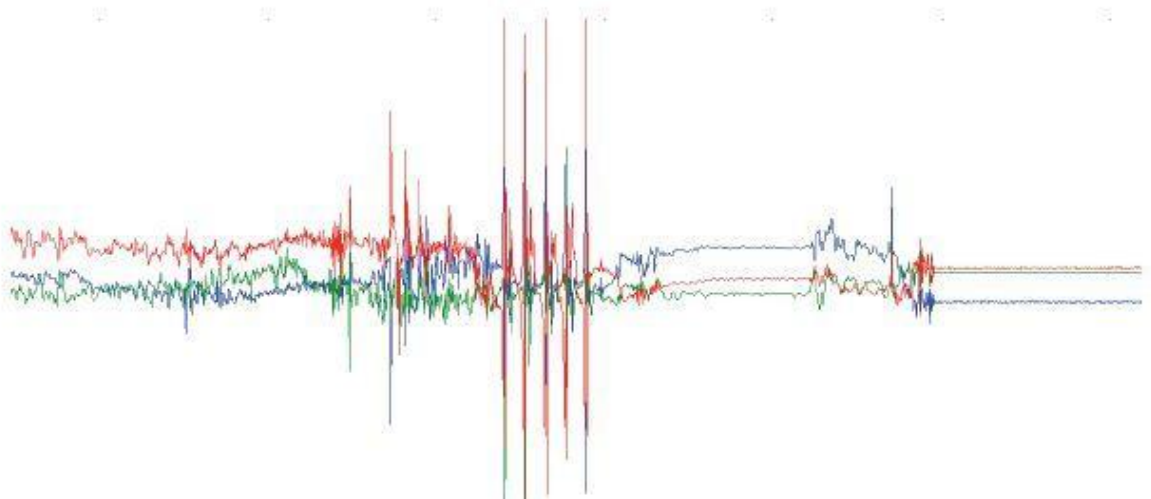
When the used sensors were received by the research team all data was downloaded via a USB connection. Once the data download was complete, the sensor's memory was erased, the battery recharged and the device could then be used for data collection with another participant. Prior to use by another participant both the sensor and strap were cleaned using an isopropyl alcohol-based wipe.

3.V.ii   Development of the Artificial Neural Network (ANN)

An ANN was developed from the data collected during the home monitoring period. The over-riding aim was to produce an ANN capable of examining previously unseen accelerometer data and providing a judgement as to the disease state of the wearer for that given time. The process of ANN development and training is summarised in the flow chart depicted in Figure 28.

The ANN was 'trained' using a compendium of data that consisted of input data, along with its corresponding known output data. Input data took the form of 70 different features extracted from the accelerometer data. Features were extracted for each one minute period of data collection during the home monitoring period. A 71st feature

was also included; this related to the patient's clinical phenotype as judged during the CARU assessments described above. The corresponding known output data, for each one minute period, was the known disease state for that moment in time. This output data was derived from the patient completed symptom diaries, where participants had recorded their predominant disease state on an hourly basis. Diary entries were assigned a numerical value (0, asleep; 1, off; 2, on; 3; dyskinesia).

In its initial 'untrained' state, the weighting of interconnections within the ANN was assigned at random and the error between ANN output and the known output calculated. The weighting assigned to the neuronal connections was then adjusted and the process repeated. This process, 'forward selection', was repeated many times, with weightings repeatedly modified in an attempt to minimise the error between ANN output and known output. 'Training' of an ANN can therefore be summarised as serial presentation of data to a network with repeated adjustment of the weightings assigned to input variables and their interconnections until the ANN output is concordant with the known output. The ANN employed in this work included two hidden layers and four outputs, each of which related to the following disease states; asleep, off, on and dyskinesia. Output for each disease state was provided in the form of a numerical value between 0 and 1, with the overriding rule that the sum of the four outputs must equal one. The ANN outputs can therefore be thought of as the likelihood of a given time period representing a given disease state.

Figure 28: Flow chart describing analysis process for phase I and phase II derived accelerometer data
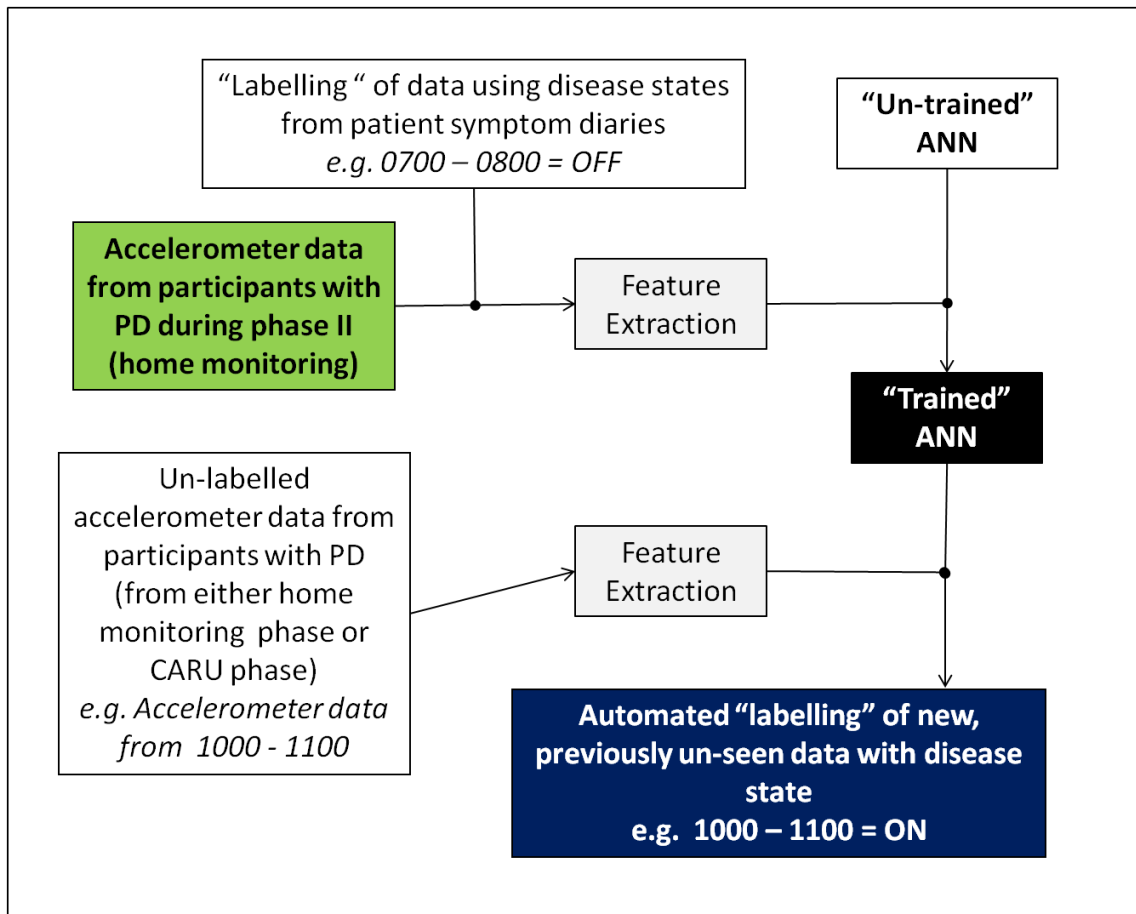
3.V.iii   From ANN outputs to disease state 'decision'.

Two differing approaches were undertaken to translate the four individual ANN outputs into a 'decision' about disease status. The first employed the 'expected value' equation. With this approach, the ANN provides output in terms of $p_{disease\ state}$, the probability of a given period of accelerometer data representing a particular disease state. The ANN provides a probability for each of the four possible disease states and thus the sum of $p_{asleep} + p_{off} + p_{on} + p_{dyskinesia}$ must equal 1. Each disease state was assigned a numerical value: asleep = 1, off = 2, on = 3, dyskinesia = 4. The vector product of the probability of each disease state was calculated to provide a prediction of disease state for a given time period:

$(p_{asleep} \times 1) + (p_{off} \times 2) + (p_{on} \times 3) + (p_{dyskinesia} \times 4)$ = disease state decision.

A decision regarding disease state is guaranteed by considering this value to one significant figure. An example of the 'expected value' equation applied to an example ANN output for a given minute is presented in Figure 29 below.

Figure 29: Example of disease state decision using the 'expected value' equation

Artificial Neural Network output for a one minute period (example)

e.g.

- $(p_{asleep} \times 1) + (p_{off} \times 2) + (p_{on} \times 3) + (p_{dyskinesia} \times 4)$ = disease state decision
- $(0.1 \times 1) + (0.2 \times 2) + (0.6 \times 3) + (0.1 \times 4)$ = disease state decision
- $0.1 + 0.4 + 1.8 + 0.4 = 2.7 = 3$ (1 s.f.) = **ON**
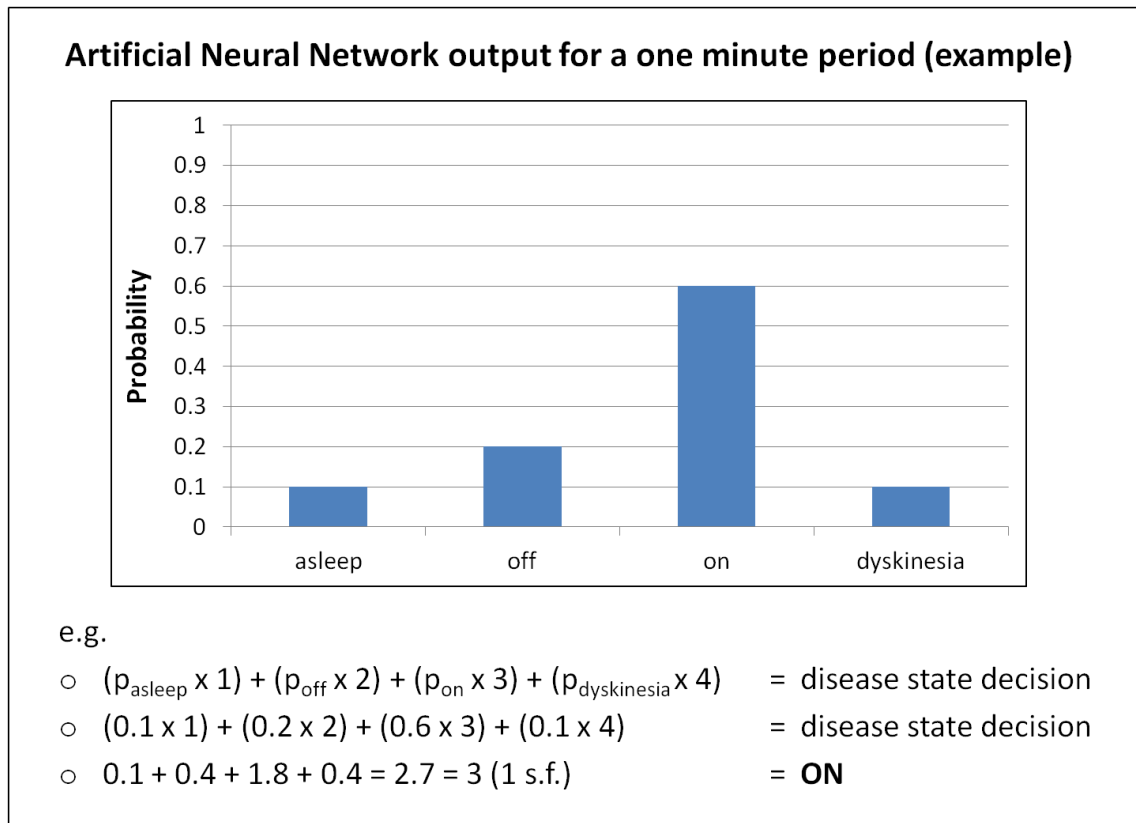
This method allows a decision regarding disease state to be made based on ANN output on a minute by minute basis. This method does however rely on a number of assumptions. Firstly, that the linear numerical system applied is valid, which itself is based on the assumption that the difference between disease states is linear. Secondly, this method requires the assumption that the network has equal sensitivity to each disease state. The expected value equation also treats every minute of data independently and gives no consideration to the minutes preceding. Clearly for a patient with PD their current disease state has a degree of influence on their disease

state in the forthcoming minute. A second analytical approach was therefore employed that instead focussed on a more clinically relevant period of time. During the home monitoring period of the study participants completed symptom diaries on a one hourly basis and hence the alternative analytical approach focused on one hour periods of time. The second approach aimed to extract multiple descriptors that summarised the network prediction across the hour period. Figure 30 below provides an overview of this method.

Within each one hour period the ANN provided a probabilistic output for each disease state, in the form depicted in Figure 29, at intervals of one minute. For each disease state, the minute by minute probabilistic output of the ANN was averaged to provide a value for the mean probability for a given disease state during the hour. The 60 one minute duration ANN outputs were also examined for the one minute period in which the highest probability for a particular disease state was seen. For the minute where the highest probability for a particular disease state was seen, probability values for all four disease outputs in that minute were taken. Figure 30 demonstrates how, for example, the one minute period with the highest probability for sleep was minute 1. Similarly, the one minute period where the highest probability for 'off' was minute 24. Ultimately, for each disease state, its maximum probability value in the hour (along with the corresponding other disease state probability values for that particular minute) were captured, providing a total of 16 values. These, added to the 4 mean values, provided a total of 20 values to describe the hour long period. From this 20 value representation of the hour a linear regression model was constructed to give a value equivalent to the diary value for this given hour, drawing on data from all participants.

Figure 30: Alternative analysis method applied to an hour of ANN outputs (NB not all one minute periods shown, only those with maximal p values for a given disease state)
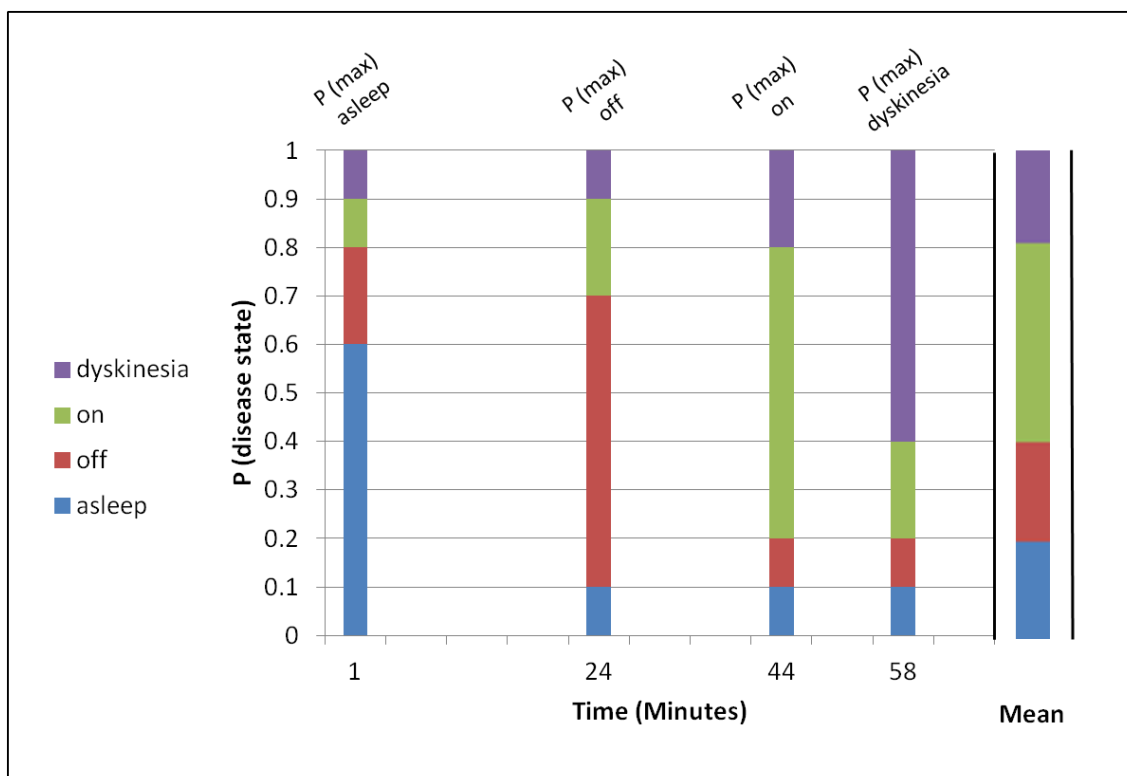
3.V.iv   Validation process

Assessment of ANN performance was undertaken using both data derived from home-based and CARU-based recordings; this validation process is summarised in the flow chart depicted in Figure 31.

Home Diaries

Subsequent testing of the ANN was performed using 'leave one out' methods. The ANN was trained using data from all but one participant. Data from the one remaining participant, derived during their home monitoring period, was then presented to the ANN as new, previously un-seen data. Employing the second analytical approach described in the methodology above, the ANN produced, for each hour long period, an estimation of the participant's disease state for that time period. The correlation between the ANN 'decision' on disease status and diary entries was examined using

Pearson's correlation coefficient. This process of 'leave one out' validation was repeated for all study participants.

CARU data

Video recordings of clinical assessments were also reviewed and rated in terms of patient disease status. All of the clinical assessments performed in CARU were time labelled, meaning that the accelerometer data recorded at the time of the assessment could also be identified. At the point of analysis the accelerometer data obtained during the clinical assessments was excluded. This step was taken because the activities performed as part of an abbreviated MDS-UPRDS are highly scripted and do not reflect activities that would realistically be performed by a person during everyday life. Instead, accelerometer data from the ten minute periods before and after the assessment was taken and divided into one minute epochs. The ANN was presented with data derived from each one minute epoch individually. For each one minute epoch the network provided a prediction of the likely disease state for that time. The output of the ANN was probabilistic, in the form of a number (0-1) for each state with the sum of these adding to one. For each one minute period the disease state that was assigned the highest value was deemed the ANN's 'decision' for that given time. Comparison between the ANN decision and clinician rating of disease status was then undertaken.

Figure 31: Flow chart describing process for validation of sensor/ANN data
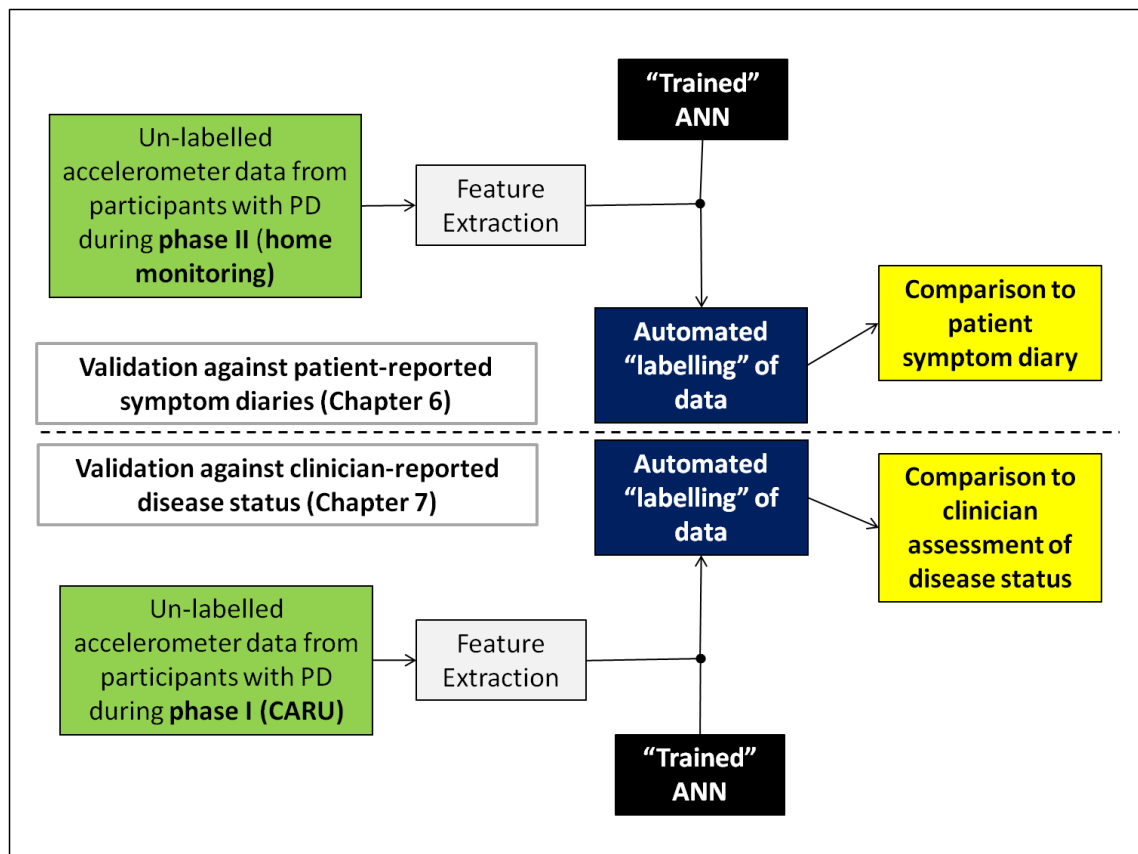
3.VI Statistics

Data were analysed using the International Business Machines Corporation Statistical Product and Service Solutions (IBM-SPSS) software package. Descriptive statistics, including means and standard deviations, were calculated. The distribution of data was examined for normality via construction of histograms. Normally distributed data were analysed with parametric tests. Non-normally distributed data were analysed using non-parametric tests, such as the Wilcoxon Ranked Sum. All reported p values are two-tailed for parametric tests. All significance values are quoted to three decimal places. Significance values of less than 0.001 are abbreviated to p <0.001. Those results not achieving significance are labelled 'NS'.

Chapter 4.   Study Participants

4.I   Recruited participants

Patients fulfilling inclusion criteria for the study were approached at clinics organised by Northumbria Healthcare NHS Foundation Trust's PD service. In total 34 participants were recruited for entry into this study. Basic demographic information for the 34 study participants is displayed in Table 8. Mean age was 68.9 years (SD=9.1). Mean disease duration was 9.8 years (SD=5.6). All patients were on anti-parkinsonian medication with a mean levodopa equivalent dose (LED) of 918.1 mg/24hours. 28 (82.4%) of the study cohort were also taking a dopamine agonist. 17 participants (50%) were on other adjunctive anti-parkinsonian agents, either monoamine oxidase B Inhibitors (MAOBi) and/or catechol-o-methyltransferase inhibitors (COMTi).7 participants were prescribed amantadine (20.6%). 2 participants (5.9%) were on an intra-jejunal levodopa infusion (Duodopa).

Clinical assessment in CARU revealed that 14 participants (41.2%) experienced dyskinesia; the mean AIMS score across the cohort was 15.7 (SD=9.0). Assessment of cognition with the MoCA and MMSE cognitive scoring tools revealed mean scores of 25.9 (SD=3.3) and 28.6 (SD=1.5) respectively. MDS-UPDRS scores by section were: I – 15.9 (SD=5.6); II – 20.1 (SD=9.7); III – 32.9 (SD=14.4) and IV – 6.9 (SD=3.6). Evaluation of sleep with the PDSS-2 revealed mean scores of 21.5 (SD=10.7).

Table 8: Demographics of study participants

Age (years)	68.9 (9.1)
Sex (% male)	67.6
PD Duration (years)	9.8 (5.6)
Total Daily Levodopa Dose (mg/day)	557.2 (293.8)
% on Dopamine Agonists	82.4
% on MAOBi	17.6
% on COMTi	35.3
% on amantadine	20.6
% on duodopa	5.9
Levodopa equivalent dose (mg/day)	918.1 (372.7)
Hoehn and Yahr Stage*	II (I-IV)
Experience dyskinesia (Yes:No)	14:20
Disease phenotype (TD:PIGD:INDET)	12:19:3
MOCA	25.9 (3.3)
MMSE	28.6 (1.5)
PDSS-2	21.5 (10.7)
MDS-UPDRS (Section I)	15.9 (5.6)
MDS-UPDRS (Section II)	20.1 (9.7)
MDS-UPDRS (Section III)	32.9 (14.4)
MDS-UPDRS (Section IV)	6.9 (3.6)
MDS-UPDRS (Total)	75.8 (22.9)
AIMS	15.7 (9.0)

Values expressed as mean (standard deviation) unless expressed otherwise

*median (range)

MAOBi = Monoamine Oxidase B Inhibitor; COMTi = Catechol-o-methyltransferase Inhibitor; TD = Tremor Dominant; PIGD = Postural Instability/Gait Difficulty; INDET = Indeterminate; MoCA = Montreal Cognitive Assessment; MMSE = Mini Mental State Examination; PDSS-2 = Parkinson's Disease Sleep Scale; MDS-UPDRS = Movement Disorders Society-sponsored Unified Parkinson's Disease Rating Scale; AIMS = Abnormal Involuntary Movement Scale

Patient recruitment in this study aimed to include a range of motor disability. The Hoehn and Yahr (H&Y) staging system, a measure of disability, was therefore incorporated into study exclusion criteria. Patients categorised as H&Y stage 5

109

(confined to bed or wheelchair unless aided) were deemed not eligible for study entry. The distribution of H&Y staging across study participants is displayed in Figure 32 below.

Figure 32: Hoehn and Yahr Staging of Study Participants
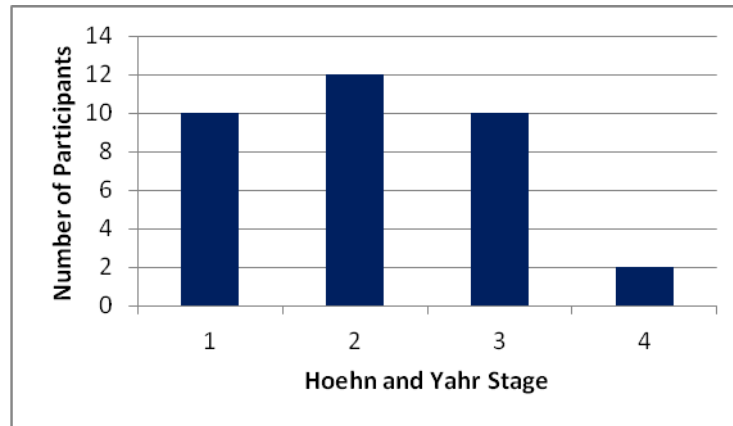
Study participants were also categorised by disease phenotype (Stebbins et al., 2013). Figure 33 displays the frequency of differing disease phenotypes amongst the study cohort.

Figure 33: Disease Phenotype of Study Participants
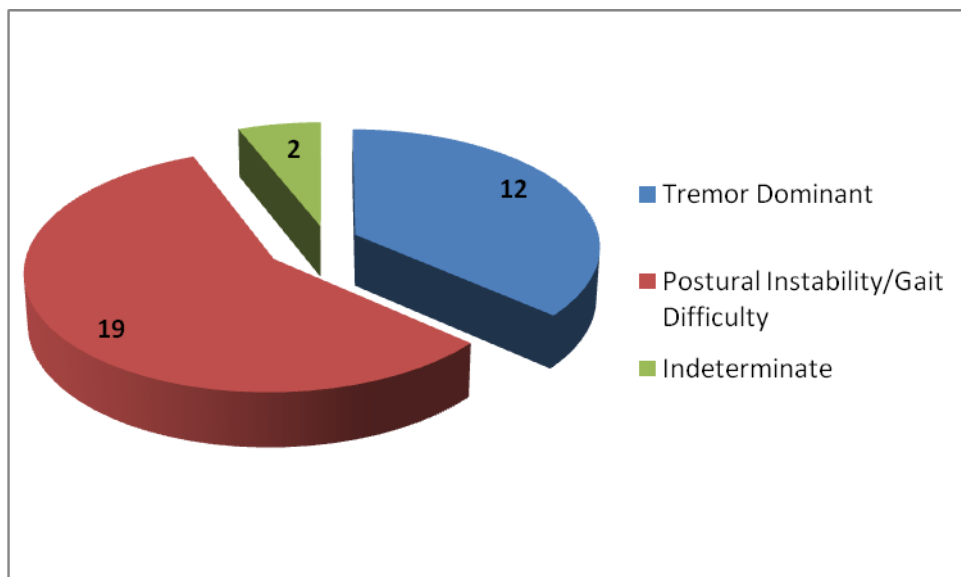
4.II   Patients declining study participation

During the study's recruitment phase a number of participants, identified as being eligible for study entry, declined involvement with the study. Basic demographic information was captured for these patients and where offered, their reason for declining entry into the study was also recorded. Collection of demographic data for non-participants was limited by the constraints of patient confidentiality. The available information is displayed in Table 9 below.

Table 9: Basic demographics of patients declining study participation

Patient ID	Age (years)	Disease Duration (years)	Sex	Dykinesia ?	LED (mg/24hr)	Reason for Declining
1	80	12	F	Yes	390	Not willing to travel to CARU
2	70	35	F	Yes	620	No reason given
3	75	7	M	No	640	No reason given
4	83	1	M	No	200	Recent operation - still recovering
5	61	6	F	Yes	450	No reason given
6	67	16	M	Yes	1316.6	Not willing to travel to CARU
7	77	6	M	Yes	703.3	No reason given
8	79	17	M	No	633.3	No reason given
9	74	8	M	No	500	No reason given
10	74	10	F	Yes	1133.3	Concern regarding duration of time at CARU

The majority of patients who declined study participation did not offer a reason for doing so. Two patients were unwilling to travel to the research facility in Newcastle upon Tyne from their homes in the catchment area of Northumbria. One patient expressed unwillingness to remain at CARU for the duration required due to concern regarding continence problems. Lastly, one patient had recently undergone a major

operation and was still recuperating from this. No patient cited an unwillingness to wear the sensors as reason for their non-participation.

4.III   Comparison between participants and non-participants

Comparison between participants and non-participants is summarised in Table 10 below. Regarding age, no significant difference was demonstrated between those participating and those not (t=-1.66, p=0.104). There was no significant difference seen between the groups when disease duration was considered (t=-0.83, p = 0.413). Levodopa equivalent dose (LED) was not significantly different between the two groups (t=1.98, p = 0.055). There was no statistically significant difference between the groups in terms of gender or frequency of dyskinesia.

Table 10: Comparison of Basic Demographics between study participants and non-participants

Characteristic	Participants (n=34)	Non-participants (n=10)	p value
Age (years)	68.9 (9.1)	74.0 (6.5)	*0.104 (ns)
Sex (% male)	67.6	60.0	**0.472 (ns)
PD Duration (years)	9.8 (5.6)	11.8 (9.5)	*0.413 (ns)
Levodopa equivalent dose (mg/day)	918.1 (372.7)	658.7 (335.6)	*0.055 (ns)
Experience dyskinesia (% yes)	41.2	60.0	**0.714 (ns)

Values expressed as mean (standard deviation) unless expressed otherwise
Statistical tests: * t-test; **Fisher's exact test

4.IV   Discussion

The distribution of H&Y staging across study participants seen in Figure 32 shows that H&Y stage IV patients were under-represented in comparison to H&Y stage I-III patients. There are a number of possible explanations for this. Firstly, this may represent limitations of the Hoehn and Yahr staging system itself. Sato et al. (Sato et

al., 2006) performed a retrospective study with a large cohort of consecutively recruited PD patients that aimed to examine the disease duration to reach different H&Y stages. This work demonstrated that only 13.3% of patients with PD of long duration (≥16 years) were classified as H&Y stage IV, with the majority being classified between stages I-III. This finding suggests that aiming to recruit equal numbers of patients across H&Y stages I-IV is artificial and not likely to be representative of the typical PD population. Comparison of frequency of H&Y stage between the study cohort (mean disease duration of 9.8 years) and a sub-group from Sato et al.'s work with similar disease duration (between 6-10 years) is displayed in Figure 34. These findings suggest that our study cohort contained a larger proportion of patients with early stage disease (H&Y stage I) compared to an unselected PD population of equivalent disease duration. It is important to highlight that 11.1% of data were missing for this particular sub-group in Sato's work and that this included patients who had died, patients who were perhaps more likely to have had more severe disease.

Figure 34: Comparison of frequency of Hoehn and Yahr stages between study cohort and Sato et al. (2006) cohort
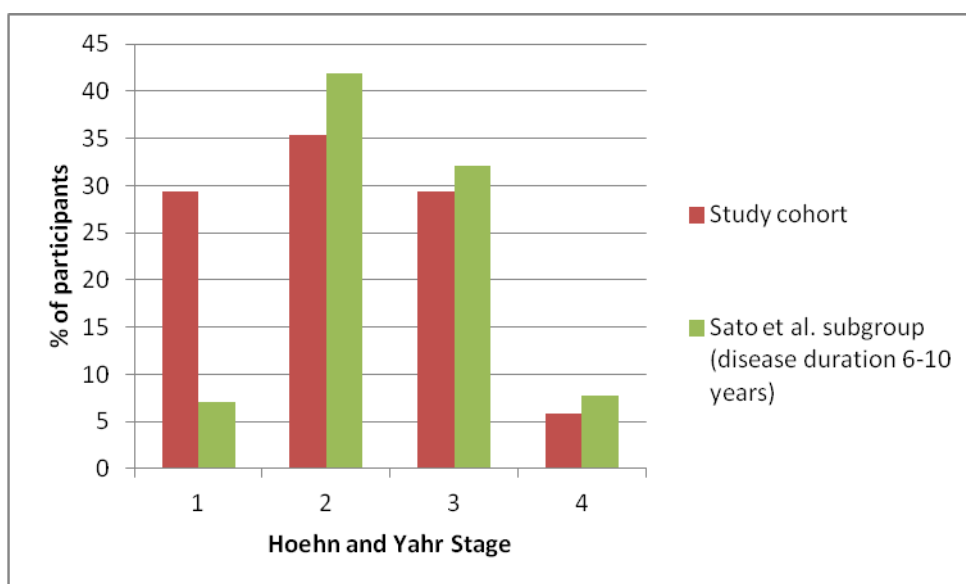
Another possible explanation for the inability to recruit equally across H&Y stages I-IV is that patients classified as H&Y stage IV were perhaps less willing to participate. Figure 34 does suggest that our study cohort contained more patients at an earlier

stage of the disease (H&Y I) and these patients are likely to be younger than those with more advanced disease. Whilst no statistically significant difference was demonstrated between participants and non-participants in terms of age, non-participants were older, with PD of longer duration and were more likely to experience dyskinesia. Whilst not statistically significant, mean LED dose in non-participants was lower than that seen in participants. This observation may be explained by the fact that clinicians, in response to the onset of dyskinesia, may reduce the doses of anti-parkinsonian medications. The majority of patients who were approached agreed to participate, thus the number of patients in the 'non-participants' group was rather small and this may have contributed to the lack of a statistically significant difference between the groups. It is well documented that older people are often underrepresented in clinical trials, with recruitment difficulties often cited as a reason (Buckwalter, 2009).

Chapter 5.   Patient-completed symptom diaries

5.1   Results

All 34 participants completed diaries during phase II of the study with 32 participants doing so for the entire seven day period. Two participants did not complete the seven day period of home monitoring; one withdrew after four days (ATYY: discomfort wearing sensor) and the other after five days (JNQW: unwell). During data transcription from patient completed symptom diary to database, a decision was made for each of the 168 potential hourly entries (seven 24 hour days) as to whether the entry made by the participant was classifiable or not. A response was defined as being unclassifiable when either no entry was made in the diary, or when two or more disease states were selected for a given time period. For each participant the total amount of time classified for each disease state was expressed as a percentage of the seven day monitoring period. For those participants not completing the seven days, the amount of time classified for each disease state was expressed as a percentage of the total time they recorded. The percentage of time for which responses were unclassifiable is also included in this analysis. For each participant an overview of the proportion of time classified in each disease state, including unclassifiable responses, is displayed in Figure 35 below.

The mean percentage of classifiable responses across the study cohort was 82.5% (SD=26.2), suggesting marked variability amongst classification rates within the study cohort. Further analysis was therefore undertaken to attempt to differentiate between diaries with high and low classification rates. A participant was defined as being a 'highly compliant' diarist if ≥90% of their diary responses were classifiable. Participants in whom <90% of diary responses were classifiable were defined as 'poorly compliant' diarists. Applying this cut-off resulted in two sub-groups: 'highly compliant' diarists (n=22) and 'poorly compliant' diarists (n=12). The demographics of these two groups were compared and it was evident that low compliance was related to age, with the

mean age of 'poorly compliant' diarists being 73.0 years (SD=7.5) versus 66.6 years (SD=9.3) in 'highly compliant' diarists [Mann Whitney U test (Z = -2.077, p = 0.038)].

Further analysis was undertaken where the cut-off between 'highly compliant' and 'poorly compliant' diarists was adjusted to 70%. 'Highly compliant' diarists were redefined as those participants whose diaries contained ≥70% of responses that were classifiable. This cut-off produced a group of 29 'highly compliant' diarists and 5 'poorly compliant' diarists. Despite there being only 5 subjects in the 'poorly compliant' diarists group ('CJVB', 'FPYC', 'JNQW', 'YQSE' and 'XKVO') there was a statistically significant difference between the two groups in terms of MMSE score. 'Poorly compliant' diarists had lower MMSE scores compared to 'highly compliant' diarists, 27.2 (SD=1.3) versus 28.8 (SD=1.4) [Mann Whitney U test (Z = -2.275, p =0.023)]. However, when the MoCA was considered, no significant difference was seen between the groups (Z = -1.474, p=0.141). With respect to these two sub-groups, no significant difference was demonstrated between them in terms of age (Z = -1.852, p=0.063). Neither cognitive assessment tool showed any significant difference when the 90% cut-off was applied.

Figure 35: Percentage of time spent in each disease state (from patient recorded diaries) during the home monitoring period
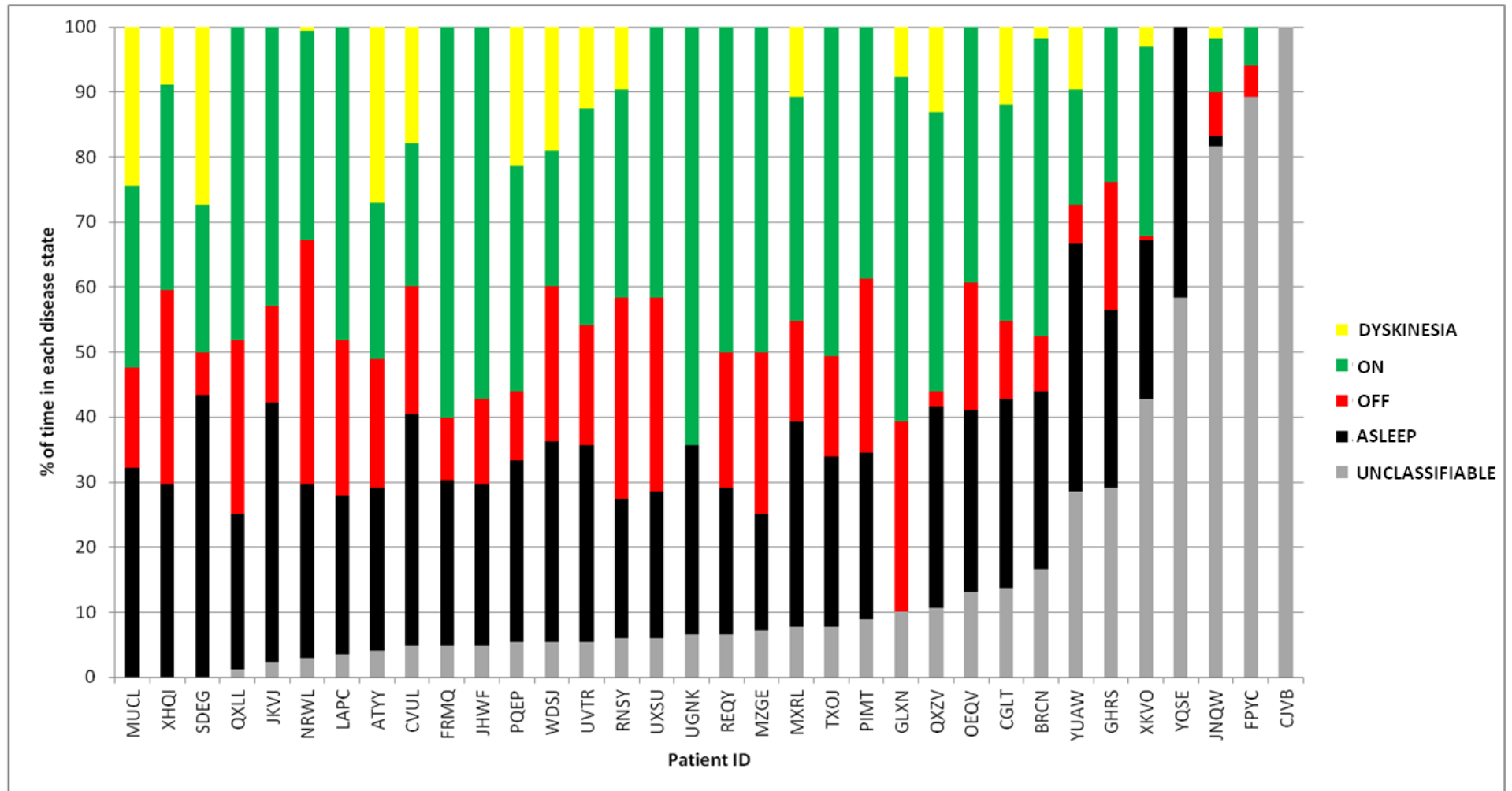
The percentage of classifiable responses was also examined across the entire study cohort on a day by day basis, examining specifically for 'diary fatigue'. 'Diary fatigue' describes a decline in the quality of diary completion over the course of a prolonged diary keeping period and is well recognised within research that requires participants to maintain diaries over multiple days (Tincello et al., 2007). There was no evidence of any increase in the proportion of unclassifiable responses as the week-long monitoring period ensued; in fact a general trend towards a greater proportion of classifiable responses was noted as the week-long monitoring period progressed.

All study participants received a standardised explanation of the tasks required of them during phase II of the study. This orientation took place during phase I of the study and included an introduction to the diaries that were to be used, explanation of the disease state descriptors and what to do in the event of forgetting to complete the diary. Originally it had been planned that participants would leave CARU with sensors and diaries and would commence home monitoring the following day. However this was only achieved in 15 of the 34 participants (44.2%). There were a number of reasons for these delays. Firstly, many participants requested that their monitoring phase be arranged for a different time, simply because the seven days following their attendance at CARU were not convenient for them. Secondly, there were on occasions, insufficient numbers of sensors available to give to the participants due to sensors already being assigned to participants undertaking home monitoring and the need to retain sensors for ongoing phase I visits. Overall, the mean time period between phases I and II of the study was 20.7 days (SD=28.4). 25 participants (73.5%) proceeded to phase II of the study within a month of phase I. No significant correlation was seen between the duration of delay between phases and the percentage of classifiable responses in the home monitoring diaries (Pearson's, r=0.064, p=0.72).

Thus far diary entries have been considered only in terms of whether they were classifiable or not. Further analysis was undertaken to examine the validity of the patient completed symptom diaries i.e. are participants correctly identifying the disease state for a given period? During phase I of the study all participants underwent a full assessment with the MDS-UPDRS. Section IV (motor complications) of the MDS-

UPDRS involves a structured, clinician-led interview during which patients are asked to consider a typical day in the last week. Patients were first asked to estimate the number of hours for which they slept during a typical day in the last week. Patients are then asked to estimate the number of hours, for a typical day in the last week, during which they were in the 'off' state and the number of hours during which they experienced dyskinesia. The number of hours spent in the on state can then be calculated. This data enabled a typical day for each patient to be reconstructed and the proportion of each day spent in each disease state was calculated as a percentage of the 24 hour period. MDS-UPDRS derived estimations of disease state distribution were then compared to diary derived measures. For each patient, all diary data were pooled across the entire seven day period and the proportions of time spent in each state were expressed as a percentage. Blanks in this case were ignored and thus all analyses were based on valid entries scaled to 100% of the time period for which valid entries were made. Pearson's correlation coefficient between data derived from the MDS-UPDRS and from the home diaries are displayed below (Table 11).

Table 11: Correlation between MDS-UPDRS estimations of the proportions of a typical day spent in each disease state with proportions of day spent in each disease state derived from diary entries

Disease State	All Diaries (n=34)		Diaries with ≥90% classifiable responses (n=22)	
	Correlation	p - value	Correlation	p - value
Asleep	0.220	0.211	0.184	0.412
Off	0.233	0.184	0.521	0.013
On	0.304	0.081	0.389	0.074
Dyskinesia	0.386	0.024	0.576	0.005

Firstly, when all diaries completed during phase II were examined (n=34), a moderate but significant correlation was seen between the proportion of time spent dyskinetic as measured by the MDS-UPDRS and by the diaries (r=0.386; p=0.024). No meaningful correlation was seen when any of the other disease states were considered. When 'poorly compliant' diarists were excluded (based on ≥90% cut-off) and the analysis was repeated, a strong and statistically significant correlation was observed between MDS-

119

UPDRS estimates and diary recorded time for both 'off' and 'dyskinesia' states (r=0.521; p=0.013 and r=0.576; p=0.521 respectively). No meaningful correlation was seen when the states 'asleep' and 'on' were considered.

5.II   Discussion

In general the diaries were well completed with high levels of classifiable responses produced by the majority of participants. There were however participants in whom diaries were completed with a large proportion of unclassifiable entries, including one participant who provided no entries at all (CJVB). Dichotomising the study population into 'highly compliant' and 'poorly compliant' diarists revealed that participants who were older and comparatively more cognitively impaired, were less likely to complete diaries to an acceptable standard. As a consequence of the progressive nature of PD, older patients may have had the disease for a longer duration and may experience more severe disease. Patients with advanced, severe disease are likely to experience more frequent fluctuations between motor states and this may render selection of only one disease state more difficult. Furthermore, severe disease may also make the act of maintaining and completing a symptom diary more challenging. However, the significant difference demonstrated in age between 'highly compliant' and 'poorly compliant' diarists cannot be attributed to this, since neither disease duration nor disease severity (as measured using the MDS-UPDRS score or the Hoehn and Yahr staging) showed any significant difference between the two groups, when both 70% and 90% cut-offs were considered. Older patients are also more likely to have other co-morbidities that may impair their ability to maintain a symptom diary. No significant difference in age was evident when 'highly compliant' and 'poorly compliant' diarists were dichotomised by the 70% entry cut-off. This failure to achieve a statistically significant result is most likely explained by the small numbers of patients within the 'poorly compliant' diarists group (n=5).

When the 70% cut-off was applied, the MMSE scores of 'poorly compliant' diarists were found to be significantly lower than those of 'highly compliant' diarists. Similar differences in cognition between 'highly compliant' and 'poorly compliant' diarists

were not replicated when the MoCA was employed. The MMSE is a very commonly used cognitive assessment tool that predominantly tests memory and language; a narrower range of cognitive domains than that which the MoCA addresses (Gill et al., 2008). The process of completing a diary requires the diarist to recall the need to do so at regular time intervals and also to recall the nature of their recent symptoms. Patients with worse memory and recall abilities may find such a task more challenging and this may in part explain this finding.

It is well recognised that patients can find disease state recognition challenging (Goetz et al., 1997). Goetz et al. demonstrated that only 12 out of 32 PD patients achieved >80% agreement between their disease ratings and clinicians' disease ratings during a four hour diary keeping exercise. The standardised training delivered during phase I of the study was designed to improve accuracy of disease state identification by participants during the home monitoring phase. However, the delays between study phases were longer than desired, with approximately a quarter of the study cohort waiting longer than a month to begin the home monitoring period. The standard deviation of the delay between phases was high (28.4 days); this is likely to have been skewed by two particular study participants ('MUCL', 'RNSY'). In these participants the delay between phases exceeded 100 days, in both cases due to intercurrent illness and prolonged hospital admission. Excluding these two participants sees the mean delay between phases I and II drop from 20.7 to 15.2 days. For some of the participants in whom the delay exceeded one month, reorientation to the diaries and sensors was undertaken by a member of the research team prior to commencement of phase two. Goetz et al. (Goetz et al., 2008b) demonstrated that the efficacy of their diary training programme began to decline when reassessed one month after it was delivered. Our findings suggest that there was no significant correlation between the delay between study phases and the proportion of classifiable responses contained within diaries.

Analysis of patient completed symptom diaries thus far has looked simply at whether entries were classifiable or not; that is the presence or absence of a diary entry for a given time slot, and whether or not the diarist had selected one or more than one disease state. It is important to acknowledge the limitations of these criteria. The mere

presence of a diary entry does not necessarily inform us what disease state the patient was truly experiencing for that given time period. Firstly there exists the possibility that the diarist had incorrectly identified their disease state for that given period. This could have been for a number of reasons; perhaps failing to appreciate the presence of dyskinesia or perhaps because of confusion regarding the nomenclature of the disease states. Secondly, patients may retrospectively complete their diary for prolonged periods due to a failure to complete contemporaneously. This may represent honest forgetfulness or deliberate fabrication. Regardless of the cause of delayed diary entry, entries may well be inaccurate due to the difficulties associated with recall.

Correlation between the amounts of time spent in each disease state, as measured by the MDS-UPDRS and the diaries, was examined. Strong, statistically significant correlation was only evident for 'off' and 'dyskinesia' disease states. This finding can be considered weak evidence of criterion validity for these disease states, i.e. the MDS-UPDRS and the diary are measuring similar concepts. The results presented in Table 11 suggest that disease state proportions derived from the MDS-UPDRS should be interpreted with caution. Patient completed symptom diaries, despite their flaws, are considered the current gold standard for prolonged monitoring of symptoms in a home setting. The lack of widespread, strong, significant correlation between the gold standard and the MDS-UPDRS derived estimations of disease state suggest that this method may produce misleading results. No other direct method of validating patient disease state identification was available in this work and the inability to draw meaningful conclusions about the true validity of diary entries is therefore a limitation. The validity of disease state recognition will subsequently be indirectly evaluated (Chapter 7) when the model (constructed using patient reported disease state) is applied to phase I data and its output compared to clinician rating of disease state.

Other studies have compared ratings between patient and clinician during a short period of observation in a clinical environment (Goetz et al., 1997). Such methodology could have been replicated during phase I of the study, as clinician rating of disease state was undertaken at the point of each motor assessment. There are a number of limitations to this approach however. A clinical setting is likely to be relatively free of

distractions and as such participants are likely to be more focussed on the task of rating their disease state compared to when they are at home and have activities of daily living to perform. This also raises the question of whether periods of monitoring in a clinical environment are truly reflective of usual behaviour for a participant and whether conclusions derived from lab-based observation can be applied to home-based periods. Lastly, the methodology employed by Goetz et al. only required disease state rating for a period of four hours. A patient may not exhibit any fluctuation between disease states in such a short time period, making the process of maintaining a diary much simpler. Furthermore, such a short time period is unlikely to result in any 'diary fatigue'. In summary, short dual-rating periods and comparison between patient and clinician may only serve to provide false reassurance about diary validity. Whilst our findings revealed no evidence of diary fatigue in terms of classification of diary entries, the inability to know whether diary entries were truly valid precluded analysis of whether diary fatigue resulted in a decline in the proportion of valid diary entries during the monitoring period. An alternative approach to this issue would be to perform intermittent video-recording of patients during the home monitoring periods. Giuffrida et al. (Giuffrida et al., 2009) provided participants with a computer that recorded them performing clinical assessments during the home monitoring period. Such technology enables clinicians, either in real-time or retrospectively, to review and categorise participants' disease state. This methodology also carries some potential limitations. Firstly, participants may have reservations about being video-recorded in their own home; such a system may be a potential barrier to both ethical approval and to study recruitment. It is also well recognised that patients, when stressed, may exhibit more marked parkinsonian symptoms (Marsden and Owen, 1967, Durif et al., 1999) or even voluntarily suppress symptoms. The act of being observed can influence a person's behaviour, the so-called "Hawthorne effect" (Sedgwick, 2012), resulting in periods of data collection that may not be truly representative of their usual activity.

It is also important to consider whether the diary-keeping habits of the study cohort are representative of an unselected PD population. An exclusion criteria for this study was MMSE score of <24; thus patients with more severe cognitive impairment were

not included in this work. We found that 'poorly compliant' diarists were more likely to have a lower MMSE compared to those deemed to be 'highly compliant' diarists. Classification and validity are likely to decline further as MMSE scores decrease. As such, it is questionable whether patient completed symptom diaries are a viable method of prolonged symptom monitoring for cognitively impaired PD patients. Furthermore, by virtue of the fact they had voluntarily agreed to participate, many of our study cohort appeared highly motivated to complete the diaries. Although patients who complete symptom diaries accurately may reap potential benefits from treatment modification, it is possible that in usual practice patients may be less motivated to complete symptom diaries and as such their quality may decline.

5.II.i   CONCLUSIONS

- The majority of participants completed diaries with a high proportion of classifiable responses.
- The proportion of classifiable responses declined in older, more cognitively impaired patients.
- Delay between phases does not appear to have compromised rates of classifiable diary entries.
- Limited correlation was seen between MDS-UPDRS and diary derived measures of disease state proportions, calling into question the validity of this MDS-UPDRS item.
- Thus far, limited conclusions can be made with regards the validity of disease state recognition by patients during diary keeping periods.

Chapter 6.   Validation of the sensors against patient-completed symptom diaries

6.I   Results

34 participants underwent the home monitoring period. Accelerometer data for four participants was lost due to file corruption (CJVB, YUAW, CGLT, FRMQ). No further analysis was possible for these participants. Participants in whom the sensors functioned but whose diaries contained less than 50% of entries that were classifiable, were also excluded from subsequent analysis, since their diaries were deemed to be too unreliable. There were three such participants (FPYC, JNQW, YQSE) and their respective diaries contained 10.7%, 18.3% and 41.6% of classifiable responses. After these exclusions, 27 participants remained and it is from data derived from this cohort that the artificial neural network (ANN) was constructed. Two participants did fail to complete the entire seven day monitoring period; one withdrew after four days (ATYY: discomfort wearing sensor) and the other after five days (JNQW: unwell). Data from JNQW was excluded (due to a low rate of classifiable diary responses); data from ATYY was still analysed.

An ANN was developed as described in the methodology. Initial training of the ANN was undertaken using all data. Subsequent testing of the ANN was performed using 'leave one out' methods. For each participant the ANN was applied to the sensor data derived during their home monitoring period. Employing the second analytical approach (depicted in Figure 30 and described in 3.V.iii), the ANN produced, for each hour long period, an estimation of the participant's disease state for that time period. The correlation between ANN output and diary entries was examined on a patient-by-patient basis and the results for each participant are displayed in Figure 36 below. To calculate this correlation, diary entries for the patient's entire home monitoring period were converted to numerical values: 1= asleep; 2 = off; 3 = on; 4 = dyskinesia. For each data point the ANN output was calculated via the second analysis approach i.e. the ANN prediction of the disease state for that particular hour. Correlation between diary

values and ANN output values was calculated using Spearman's. The mean correlation value seen was 0.71 (SD = 0.08); these correlation values will be referred to as 'ANN performance' henceforth.

Further analysis was undertaken to examine why higher levels of ANN performance were seen in some participants. A variety of clinical and demographic variables were considered and their relationship to ANN performance examined (Table 12). There was no significant difference in ANN performance between 'old' (≥70 years) and 'young' (<70 years) participants (p=0.471). Dichotomising the group by sex also revealed no significant difference in ANN performance between males and females (p=0.877). There was no significant difference in ANN performance between those with PD of 'long' duration (≥10 years) and those with PD of 'short' duration (<10 years) (p=0.882). Dichotomising the group into those with and those without dyskinesia did not reveal any significant difference in ANN performance between groups (p=0.882).

Figure 36: Correlation between Sensor and Diary Outputs by Patient ID
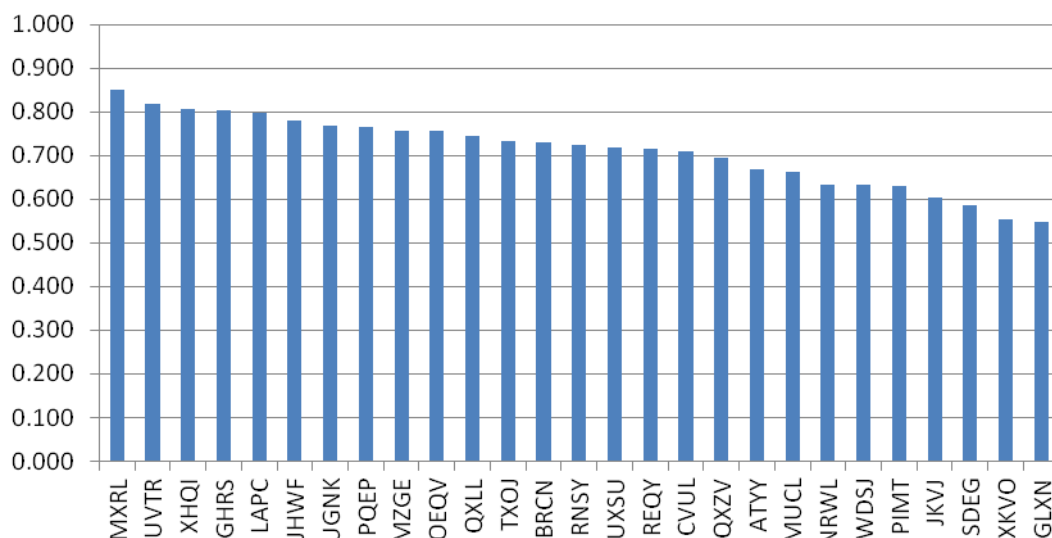
The group was also dichotomised by clinical phenotype (TD and PIGD); no significant difference in ANN performance was observed between the groups (p=0.598). No significant difference in ANN performance was seen between those with 'short' delays between phases I and II (≤30 days) and those with 'long' delays (>30 days) (p=0.162). Consideration was also given to diary classification; 'highly compliant' diarists (≥90% of

responses classifiable) and 'poorly compliant' diarists (<90% of responses classifiable) showed no significant difference in terms of ANN performance (p=0.414). Clinical rating scales and scoring systems were also analysed to examine whether there was any association between these variables and ANN performance. Hoehn and Yahr (p=0.578), MMSE (p=0.346), MoCA (p=0.386), PDSS-2 (p=0.602), AIMS (p=0.539), MDS-UPDRS sections I (p=0.146), II (p=0.976), III (p=0.577), IV (p=0.853) and MDS-UPDRS total (p=0.922) all showed no evidence of a statistically significant relationship with ANN performance.

Table 12: Demographic and clinical variables and their relationship to ANN performance

Variable (dichotomy where relevant)	p-value
Age (≥70 and <70 years)	†0.471 (ns)
Sex	†0.877 (ns)
Disease Duration (≥10 and <10 years)	†0.882 (ns)
Dyskinesia (presence/absence)	†0.882 (ns)
Phenotype (PIGD/TD)	†0.598 (ns)
Delay between phases (≤30 and >30 days)	†0.162 (ns)
Diary Acceptability (≥90 and <90 %)	†0.414 (ns)
Levodopa equivalent dose	*0.122 (ns)
Hoehn and Yahr	‡0.578 (ns)
MMSE	*0.346 (ns)
MoCA	*0.386 (ns)
PDSS2	*0.602 (ns)
MDS-UPDRS 1	*0.146 (ns)
MDS-UPDRS 2	*0.976 (ns)
MDS-UPDRS 3	*0.577 (ns)
MDS-UPDRS 4	*0.853 (ns)
MDS-UPDRS Sum	*0.922 (ns)
AIMS	*0.539 (ns)
Statistical tests: *Pearson; † Mann Whitney U; ‡ ANOVA	
(ns) = non-significant	

Visual representations of the home monitoring period were produced for all participants. These incorporated the patient-recorded symptom diary entries and the

ANN outputs derived from both analytical approaches. For these visual representations (Figures 37-44) the horizontal axis represents the number of five minute epochs during the seven day monitoring period. On the vertical axis numbers represent disease status; 0, asleep; 1, off; 2, on; 3, on with troublesome dyskinesia. The red trace on the plot provides a graphical representation of patient completed symptom diary entries recorded during the seven day monitoring period; un-classifiable entries were left blank. The green and blue traces on the lowermost panel represent the ANN output, in terms of disease status, for the two different analytical approaches employed. Green represents the ANN output when the 'expected value' approach is employed (depicted in Figure 29) and blue represents the ANN output when an additional linear regression model was employed to provide an ANN prediction for hour-long periods (depicted in Figure 30). Visual representations for all participants were inspected and it became evident that a number of recurrent themes existed. In the results below a series of exemplar visual representations are presented for each of the recurrent themes identified. Figure 37 and Figure 38 display examples of home monitoring periods where ANN output and patient reported disease state appear closely aligned.

Figure 37: Visual representation of XHQI's home monitoring period
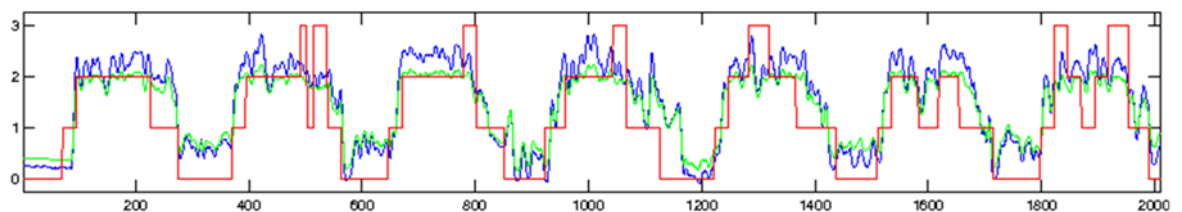
Figure 38: Visual representation of LAPC's home monitoring period

128
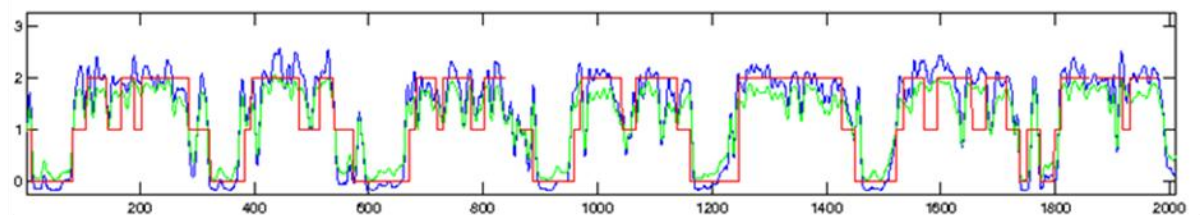
Figure 39 and Figure 40 display examples of home monitoring periods were participants reported dyskinesia but the ANN did not recognise these periods as time with dyskinesia

Figure 39: Visual representation of MUCL's home monitoring period
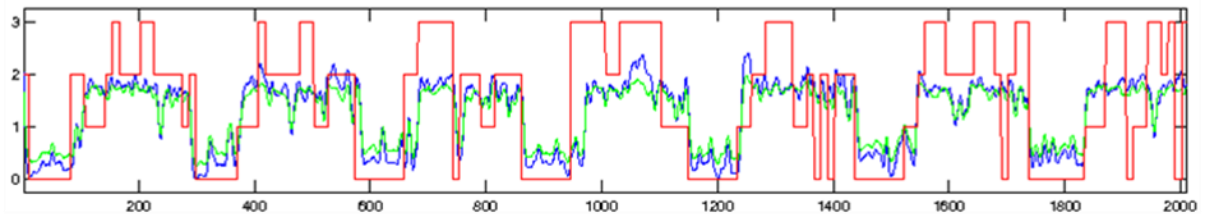
Figure 40: Visual representation of CVUL's home monitoring period
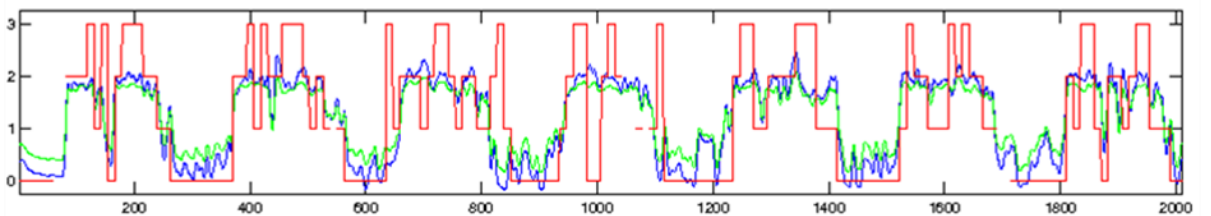
Figure 41 and Figure 42 present examples of home monitoring periods where the ANN failed to discern 'off' and 'on' periods (represented by dashed lines).

Figure 41: Visual representation of JHWF's home monitoring period
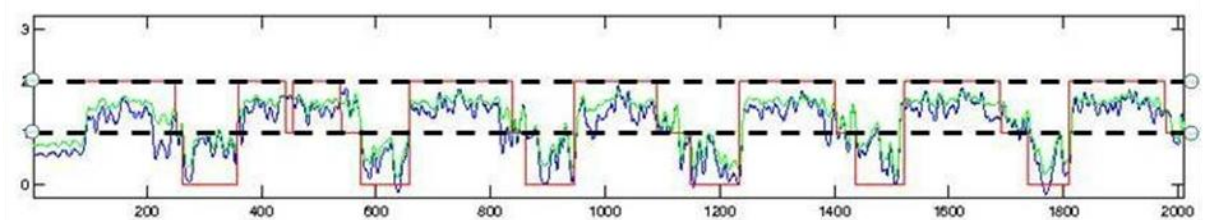
Figure 42: Visual representation of TXOJ's home monitoring period
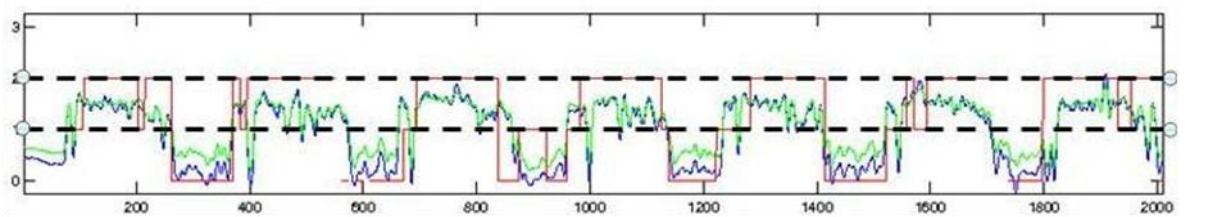
Lastly, Figure 43 and Figure 44 provide examples of home monitoring periods for participants who exhibit numerous motor fluctuations between 'on' and 'off' states during the day. In these examples the ANN appears unable to detect the changes in motor state.

Figure 43: Visual representation of PIMT's home monitoring period
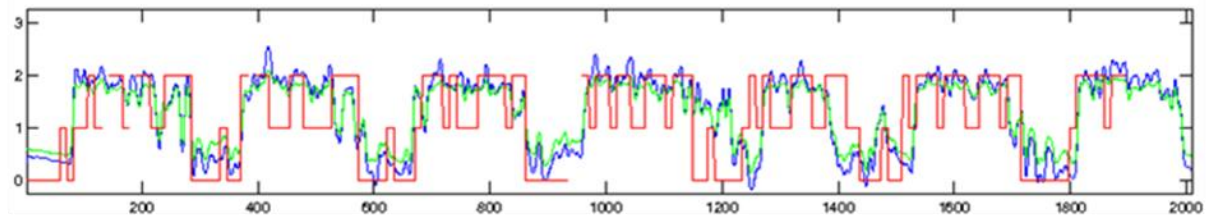
Figure 44: Visual representation of UXSU's home monitoring period
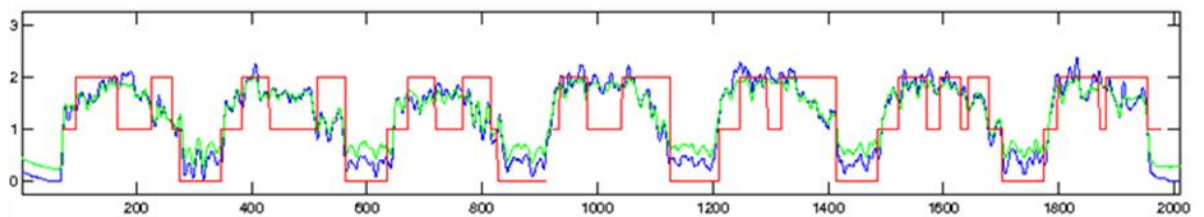
6.II   Discussion

DYSKINESIA

Visual representations of the home monitoring period were developed since it was felt that these were more clinically meaningful and enabled rapid appreciation of the temporal dynamics of a participant's symptoms. Figure 37 provides an example of where the ANN is capable of correctly identifying periods of time as dyskinesia however, as shown in Figure 39 and Figure 40, for many participants this was not the case. Analysis was therefore undertaken to establish why the ANN struggled to detect dyskinesia. In the entire study there were 18 people with dyskinesia. 14 of these participants can be described as having 'validated' dyskinesia, since they were observed exhibiting dyskinesia by a clinician. The dyskinesia reported by the remaining four participants was not witnessed by a clinician and shall therefore be referred to as

'non-validated' dyskinesia. The four participants with non-validated dyskinesia (UXSU, YQSE, RNSY, QXZV) reported, during completion of section IV of the MDS-UPDRS, that they had experienced dyskinesia during a typical day in the last week. When the home diaries of these four participants were examined it was evident that two participants (UXSU, YQSE) in fact recorded zero hours of dyskinesia during the entire home monitoring period. RNSY reported a total of 16 hours of dyskinesia. QXZV reported 22 hours of dyskinesia, although there is a possibility of error in this diary since the number of hours of dyskinesia reported for days one to six ranged from 0-2, yet for day seven a total of 15 hours were reported. Such a large increase in the amount of dyskinesia seen would be unlikely, particularly given that no medication changes were made during the home monitoring period.

Of the 14 participants with 'validated' dyskinesia, data loss for participants CGLT and YUAW meant only 12 remained available for analysis. Consideration was also given to where in the body the dyskinesia was present. Dyskinesia affecting the head, trunk or leg, but not the upper limb, may not be detected by a wrist-worn sensor. The clinical assessment of dyskinesia performed during phase I using the AIMS was reviewed. AIMS assessments include a clinician's evaluation of the severity of dyskinesia in a variety of body parts. Of the 12 remaining 'validated' dyskinesia participants, three were noted to have only "minimal" (AIMS upper limb score = 2) (CVUL) or "no" (AIMS upper limb score = 1) (ATYY, SDEG) dyskinesia in the upper limbs. Of the remaining nine participants with 'validated' upper limb dyskinesia there were concerns regarding the validity of two of the participants' home diaries. JNQW, who withdrew from the study after five days, did complete five diary days but only 18% of the responses provided were classifiable. Data from JNQW was therefore excluded from analysis. 89.9% of the responses provided by GLXN were classifiable but it was noted that this participant reported no sleep at all during the seven day monitoring period, again raising concerns regarding the validity of this diary. A further observation was that, for some patients, the amount of dyskinesia reported in the home diary was much less than expected based on the estimation derived from the MDS-UPDRS. BRCN, for example, estimated during assessment with the MDS-UPDRS section IV, that 29.2% of

their typical day was spent experiencing dyskinesia. Subsequent analysis of BRCN's home diary however, revealed only three hours of dyskinesia in the entire seven day period.

Once these findings have been considered it leaves only eight participants in whom dyskinesia was witnessed by a clinician ('validated'), was prominent in the upper limb (AIMS upper limb score >2) and was reported in diaries with >50% of responses being deemed classifiable (Table 13). In total there were 179 hours of dyskinesia reported during the home monitoring period that can be considered 'good quality' for the reasons outlined above.

Table 13: The eight participants providing 'good quality' dyskinesia to the data set

Patient ID	AIMS Upper Limb Score	Proportion of diary responses classifiable (%)	Hours of reported dyskinesia
MUCL	3	100	41
PQEP	4	95	36
WDSJ	3	95	32
UVTR	4	95	21
MXRL	4	92	18
XHQI	3	100	15
GLXN	3	89	13
BRCN	3	83	3
			179

Originally it was projected that the study would generate 5,712 hours of home monitoring data (34 participants, seven 24 hour days). As described above, the data from four participants was lost and two participants did not complete the entire seven day monitoring period. Consequently the total duration of home monitoring data captured was 4,920 hours. The distribution of this time between the differing disease states reported by participants is displayed in Table 14 below and includes periods during which unclassifiable diary entries were provided. Whilst 327 hours were identified by participants as dyskinesia, the above analysis suggests that only 179 hours (54.7%) can be considered 'good quality'.

Table 14: Home monitoring data by disease state

Diary entry	Number of hours	% of entire data
Asleep	1,281	26.0
Off	849	17.3
On	1,752	35.6
On with troublesome dyskinesia	327	6.6
Unclassifiable	711	14.5

A major limitation of this work relates to the disease state nomenclature that was selected for use; specifically the labelling of the dyskinesia category which introduces a series of problems. Firstly, the presence of the word "troublesome" adds an additional layer of subjectivity. A more active participant may find only very mild dyskinesia 'troubling' whereas a more severely disabled, less active participant may deem equivalent dyskinesia to be 'non-troublesome'. Employing this disease state description raises the possibility that participants who identify that they are experiencing dyskinesia, but deem it to be 'non-troublesome', will record that given time period as 'on' time. The net result of such an occurrence would be decreasing amounts of time labelled as dyskinesia and 'pollution' of true 'on' time with 'non-troublesome' dyskinesia.

The decision to employ the disease categories used in this work was clinically motivated, primarily by the fact that patients do not always interpret time with dyskinesia as being 'bad' (Hauser et al., 2000). Hauser et al. demonstrated that dichotomising the dyskinesia category into troublesome and non-troublesome provided functional separation; 'troublesome dyskinesia' was predominantly considered by patients to be 'bad' time and 'non-troublesome dyskinesia' considered to be 'good' time. This finding, along with that of Reimer et al. (Reimer et al., 2004), who suggested that diaries with three disease state categories were preferable to four due to patient concordance, was the motivation for the category selection. It is arguable that including solely 'on' and 'on with troublesome dyskinesia' categories is logical, since dyskinesia deemed to be 'non-troubling' by a patient is not clinically relevant. A clinician would be unlikely to make treatment changes if a patient reported

experiencing only non-troublesome dyskinesia. On reflection however, a five category diary including the categories 'on with no dyskinesia', 'on with non-troublesome dyskinesia' and 'on with troublesome dyskinesia' (see Figure 4) would have been preferable, assuming that the additional categories did not negatively impact on patient concordance with the diaries.

It is important to highlight that for two of the participants who provided 'good quality' dyskinesia, free-text feedback (provided in evaluation of sensor acceptability: see Chapter 9) contained quotes that call into question the quality of accelerometer data collected due to poorly fitting sensors. Firstly:

"Because I have small wrists the sensors were swinging around and it was difficult to keep them in the upright position. After a couple of hours I used some surgical tape to stick it down where the strap fastens underneath - they are in the same position after 1 week including daily showers" [PQEP]

Whilst the participant PQEP appears to have rectified this problem using surgical tape to secure the sensor, another participant, who also experienced this problem, appears to have been troubled repeatedly by this problem:

"The left, blue sensor did not always stay securely in position and so needed occasional readjustment" [WDSJ]

It is critically important that wearable accelerometers fit to the body securely. A loosely fitting sensor may result in relative motion between the sensor and the body and as a consequence sensor displacement, extraneous signal artefact and declining signal accuracy (Yang and Hsu, 2010).

In summary, there was insufficient accelerometer data labelled as dyskinesia in our data set. An ANN is trained by presenting it with a compendium of labelled data; this enables algorithms to be generated and subsequently new, previously unseen data to be categorised using the same labelling system. In our case, only 6.6% of the data compendium represented time with dyskinesia, with only 54.7% of this felt to represent 'good quality' data. Exposure of the ANN to only small amounts of a given disease state during training is liable to result in the ANN having less 'confidence' in

identifying this particular disease state and thus poor performance. Similar poor performance was seen in work by Tsipouras and colleagues (Tsipouras et al., 2011) where identification of dyskinesia was suboptimal, perhaps in part to the small amount of time with dyskinesia within the data compendium used for ANN training. The reasons for inadequate amounts of dyskinesia data in our work include sensor failure and inadequate diary completion by some participants. Failure of participants to recognise time with dyskinesia is also a possibility and has been described in previous work (Vitale et al., 2001). Absence of dyskinesia specifically in the upper limbs and problems relating to disease state nomenclature may have resulted in a decline in the quality of the data labelled as dyskinesia. The ANN's understanding of what dyskinesia 'is', is likely to have been blurred by the variable quality of data labelled as dyskinesia during training of the ANN, ultimately resulting in compromised ANN performance.

Reliable detection of dyskinesia has also proved challenging for other research groups, with differentiation between mild dyskinesia and 'normal' movements challenging. The lone feature analysis methods employed by Manson et al. had low specificity for mild dyskinesia and a tendency for the feature variable to overlap into the 'normal' range was noted (Manson et al., 2000). Application of an ANN to dyskinesia evaluation resulted in frequent misclassification of mild dyskinesia, particularly when volitional movement was present (Keijsers et al., 2003b). The method by which the data in our work was collected is superior to those employed by these other authors, in terms of the sensor itself and the data collection environment. Manson and colleagues used a bulky restrictive sensor with monitoring performed in a highly controlled environment; Keijsers captured data in a clinical environment and required a series of six body worn sensors.

ON AND OFF

Figure 38 provides an example of a visual representation of a home monitoring period where motor fluctuations were detected by the ANN. Unfortunately this was not the case for many other participants with one of two patterns typically seen: Figure 41 and Figure 42 provide examples of a failure to discern between 'on' and 'off' disease states; Figure 43 and Figure 44 provide examples of a failure to appreciate motor fluctuations.

Three possible explanations exist for why the ANN's performance at discerning on and off disease states was sub-optimal. Firstly, it is evident from Table 14 that 'off' is under-represented in the data set; only 849 hours (17.3%) of home monitoring data was labelled by participants as 'off'. It can be seen in Figure 43 and Figure 44 that the ANN output has a tendency towards the on state in participants exhibiting motor fluctuations. This may be a consequence of there being twice as much 'on' time data in the training compendium presented to the ANN, with 'on' accounting for 35.6% of the data set. Secondly, the problems relating to disease state nomenclature may also have had an adverse effect on ANN performance. As described above there is a possibility that time labelled as 'on' may have been 'polluted' with time deemed to be 'non-troublesome' dyskinesia. The presence of involuntary dyskinetic movements within time labelled as 'on' results in a more nebulous definition of the 'on' state and is likely to result in an ANN less capable of reliably detecting such periods. Thirdly, and potentially most crucially, the inherent subjectivity of disease state rating is likely to have had a large influence on the ANN's ability to differentiate between 'off' and 'on' states. Participants were provided with specific definitions of what constituted each disease state: 'on' was defined as "symptoms of slowness and tremor controlled"; 'off' was defined as "problems with stiffness, slowness and tremor". The difficulty here relates to degree of tolerance of symptoms at the individual level. A person with a low functional level may not deem pronounced motor symptoms to be problematic and thus may describe themselves as being 'on'. Conversely, a person with a high level of function may deem a minimal degree of motor impairment to be unacceptable and therefore report being 'off'. The constellation of symptoms experienced by a patient during periods of 'off' can include motor symptoms, non-motor symptoms (NMS) or both (Stacy et al., 2008). NMS can include low mood and anxiety, and the presence of these may influence a person's perception of the severity of motor symptoms at a given time. A similar concept, whilst not directly transferable to PD, is highlighted by research examining pain diaries which demonstrated that more anxious patients are more likely to report pain (Finan et al., 2008). There is therefore a possibility that diary reports of 'off' included some periods where patients were indeed off, but were predominantly experiencing NMS as opposed to motoric 'off'.

The net result of the issues outlined above is that the ANN's understanding of what 'on' and 'off' constitutes becomes more nebulous. Training an ANN requires serial presentation of labelled examples and ultimately the performance of the ANN will be impaired if examples of a given disease state are not homogenous. In summary, it is perhaps unrealistic to expect high levels of agreement between ANN outputs and patient diaries given their inherent subjectivity.

It is important to acknowledge that the ANN itself has a number of limitations, particularly relating to the process by which a decision on disease status is reached from the data presented. Two different approaches to this problem were employed in this work. The first approach, using the expected value equation, was the more simplistic approach of the two and was reliant on a couple of assumptions. The first assumption was that the difference between 'adjacent' disease categories is the same, thus allowing the disease categories to be translated to consecutive integers. This is likely to be a gross oversimplification; the relationship between disease states is unlikely to be linear and may even vary on an inter-individual level. The second assumption required for this approach is that the model is equally sensitive to each of the different disease states. This is also unlikely to be true given the disproportionate representation of 'on' and 'sleep' within the data compendium used to train the ANN. A further problem with the initial approach taken was that this method treated one minute epochs as distinct entities. This approach means no consideration is given to the time preceding a particular epoch, despite the fact that the disease state immediately before the period of interest is likely to have some influence on the disease state seen during the period of interest. To address this, a second approach was taken that employed a linear regression model to analyse data captured over a more prolonged period. A period of one hour was selected since this was the period of time employed within the patient completed symptom diaries. Whilst this approach is more logical and avoids looking at short periods in isolation, a major limitation remains. The problem relates to how best to move from having probabilistic outputs for each disease state for every minute, to an overall ANN output; a single data point that contains the ANN's prediction of disease state. The method employed here relied

on 20 different features that described the hour long period and used a linear regression model to predict the disease states based on these values.

In this thesis I have highlighted the limitations of simplistic approaches such as linear regression and argued that more complex methodologies such as ANNs may provide superior performance when considering human movement. It is perhaps counter-intuitive that in order to process the huge amounts of data produced by the ANN into a clinically meaningful and interpretable output we have returned to more simplistic approaches. A more complex analytical approach to this problem may produce superior performance. If the same argument for ANNs is applied it follows that construction of an additional ANN may be beneficial. The additional ANN could be presented with the original ANN's probabilistic values for each disease state for a large number of epochs within the hour. To allow training to be undertaken these features would be accompanied by the disease state label assigned by the patient for the relevant period. Ultimately the trained ANN may be able to provide more accurate reflections of disease state based on probabilities produced by the original ANN.

LIMITATIONS

The data provided by seven of the 34 participants (20.6%) was not available for analysis due to inadequate diary completion and sensor failure. Exclusion of the three participants whose diaries contained very low proportions of classifiable entries was implemented because such participants were felt to be less likely to provide valid responses. Exclusion of these participants' data was undertaken in an attempt to minimise the adverse effect this might have ANN performance. An inherent limitation of this work relates to the difference between acceptability and validity of diary entries; the assumption that the presence of classifiable entries and validity are directly associated cannot be proven. Our definition of 'highly compliant' and 'poorly compliant' diarists may not be appropriate when validity of responses is considered; the worst diarists in terms of classification may in fact be providing small number of valid responses. Loss of participants' data due to technical problems is regrettable. In these four cases the accelerometer did appear to have functioned normally during the home monitoring period but the files containing this data were corrupted and thus

rendered inaccessible. Despite analysis of a wide number of demographic and clinical features it was not possible to identify traits that resulted in better sensor performance for some participants compared to others.

6.II.i   CONCLUSIONS

- A strong correlation was seen between ANN output and diary recorded disease status

- Variable ANN performance was seen amongst the study cohort but no factors (demographic, clinical or methodological) were identified to discern between those in whom it worked better or worse.

- Visual representations of home monitoring periods provide a useful way of displaying diary and ANN data, but for many participants disparity was evident between the two.

- It is likely that the majority of the correlation seen between the two measures reflects correct identification of periods of sleep and of on time, both of which dominated the compendium of data upon which the ANN was trained.

- Accelerometers and the ANN employed in this study lack the ability to provide 'real-time' evaluation of PD motor symptoms. Failure to do so likely relates to methodological flaws that restricted both the volume and quality of training data, and to the inherent subjectivity of disease state recognition.

Chapter 7. Validation of the sensors against clinician assessment of patient disease status

7.I   Results

156 abbreviated MDS-UPDRS assessments took place during phase I of the study and as part of this assessment, the rating clinician (JF) made a judgment regarding the disease state of the participant at that time. These clinical assessments were video-recorded but due to technical problems with the digital recording system, recordings were not available for 15 of the 156 assessments. In total, 141 of the 156 assessments (90.4%) had video-recordings available for review. The 15 assessments that were not unavailable included footage from 10 different participants. Despite the loss of video-recordings, every participant had at least one videoed assessment available for review. Video recordings were also reviewed and rated in terms of patient disease status by a second, blinded clinician (RW). Agreement between the two clinicians was evident for all but four of the videos; in these cases the videos were then re-examined in the presence of both clinicians and consensus agreement was reached. On completion of this process the distribution of disease states identified by clinician assessment was as follows: asleep (0, 0%); off (28, 17.9%); on (75, 48.1%); dyskinesia (15, 24.4%); video loss (15, 9.6%). The second, blinded clinician also repeated clinical assessments with the abbreviated MDS-UPDRS and AIMS (upper limb component only) based on the video footage. This was undertaken for a random sample of 50% of the videos and the results compared to those of the first rater. Inter-rater agreement statistics for the clinical assessments are displayed in Table 15 below. Interpretation of kappa values is summarised in Table 16 (Viera and Garrett, 2005).

Table 15: Inter-rater agreement for clinical assessments performed in CARU during phase I of study

Motor Feature	Test	Scale	Sub-section	Kappa Value	P-value
Tremor	Rest Tremor	MDS-UPDRS	3.17	0.86	p<0.001
	Rest Tremor (with distraction)	MDS-UPDRS	(3.17)	0.79	p<0.001
	Postural Tremor	MDS-UPDRS	3.15	0.77	p<0.001
	Kinetic Tremor	MDS-UPDRS	3.16	0.61	p<0.001
Bradykinesia	Finger Taps	MDS-UPDRS	3.4	0.62	p<0.001
	Pronation-Supination	MDS-UPDRS	3.6	0.52	p<0.001
Dyskinesia	Upper Limbs	AIMS	5	0.64	p<0.001

Table 16: Interpretation of the Kappa Statistic (Viera and Garrett, 2005)

Kappa Value	Degree of Agreement
< 0	Less than chance
0.01 - 0.20	Slight
0.21 - 0.40	Fair
0.41 - 0.60	Moderate
0.61 - 0.80	Substantial
0.81 - 0.99	Almost perfect

Inter-rater agreement for the assessment of rest tremor was almost perfect (k=0.86; p<0.001) and remained substantial when rest tremor was assessed whilst subjects were subject to cognitive distraction (k=0.79; p<0.001). Assessment of both postural and kinetic tremors showed substantial inter-rater agreement (k=0.77; p<0.001 and k=0.61; p<0.001 respectively). Assessment of bradykinesia with the finger tapping test showed substantial agreement between raters (k=0.62; p<0.001). For assessment of bradykinesia with the pronation-supination test only moderate agreement was seen (k=0.52; p<0.001). There was substantial inter-rater agreement for assessment of upper limb dyskinesia using the AIMS (k=0.64; p<0.001)

All of the above clinical assessments were time labelled, meaning that the accelerometer data recorded at the time of the assessment could be identified. During analysis, the accelerometer data obtained during the clinical assessments themselves was excluded. This step was taken because the activities performed as part of an abbreviated MDS-UPRDS are highly scripted and do not reflect activities that would realistically be performed by a person during everyday life. Instead, accelerometer

data from the ten minute period before and after the assessment were taken and divided into one minute epochs. It was felt to be unlikely that disease status would change in this period of ±10 minutes. For some participants the initial assessment occurred before they had worn the sensors for ten minutes. Similarly, some participants did not wear the sensors for a full ten minutes after completion of the final assessment because the sensors had been removed and the participant had left the department as planned. As a consequence it was not always possible to derive ten minutes of data collection from before and after every assessment. In these cases the data were still used, regardless of the duration of data collection.

The ANN was presented with data derived from each one minute epoch individually. For each one minute epoch the ANN provided a prediction of the likely disease state for that time. The output of the ANN was probabilistic, in the form of a number (0-1) for each disease state, with the sum of these adding to one. For each one minute period the disease state that was assigned the highest probability value was deemed to be the ANN's overall output for that given time period. In total 2,338 one minute epochs were derived from these assessment periods. For all 2,338 epochs the ANN output was compared to the clinicians' evaluation. A confusion matrix (or 'contingency table') is presented in Table 17 below, and summarises ANN output and clinician rating, with the shaded cells containing the desired outcomes (i.e. agreement between ANN and clinician).

Table 17: Confusion matrix displaying clinical rating of disease status against artificial neural network (ANN) output

		ANN output			
		ASLEEP	OFF	ON	DYSKINESIA
	ASLEEP	0	0	0	0
Clinician Rating	OFF	99	179	194	21
	ON	208	293	702	51
	DYSKINESIA	45	166	239	141

Table 18 displays sensitivity and specificity values for the identification of each disease state by the ANN. To allow these values to be calculated, ratings were adjusted to binary classifications i.e. outcomes were dichotomised, for example 'off' and 'not-off'.

Table 18: Sensitivity and specificity for artificial neural network (ANN) output of disease state

ANN output	Sensitivity	Specificity
ASLEEP	0.0	84.9
OFF	36.3	75.1
ON	56.0	60.1
DYSKINESIA	23.9	95.9

A sensitivity of 0 is quoted for sleep since there were no periods of time in CARU labelled as sleep for the ANN to identify. Specificity of sleep detection was 84.9%. Sensitivity and specificity for other disease states were as follows: off (36.3%, 75.1%), on (56.0%, 60.1%) and dyskinesia (23.9%, 95.9%).

7.II   Discussion

At least moderate inter-rater agreement was evident for all clinical assessments. This finding provides validation of the ability of the CARU-based rater (JF) to perform a motor examination and MDS-UPDRS assessment correctly. Almost perfect levels of agreement were seen for rest tremor assessment. Previous work involving video-recording of motor examination and subsequent evaluation by a clinician, had shown that detection of subtle rest tremor was difficult (Salarian et al., 2007, Van Someren et al., 1998). Consequently efforts were made in this work to ensure that prolonged video-recording was undertaken for rest tremor assessment, with close zooming used to assist visualisation of subtle tremor; this may in part explain the high agreement seen. It was evident that inter-rater agreement for tremor assessment was lowest for kinetic tremor. Frequently it appeared that tremor rated as 1 by the CARU-based clinician, was assigned a score of 0 by the video reviewer. This finding likely reflects the technical difficulty of producing steady images whilst zoomed in on hands that were moving, meaning that subtle tremor was missed on video review. Whilst video-

143

recording does allow a given assessment to be reviewed repeatedly, it is important to note that they provide only a two-dimensional representation of three-dimensional movements and as such inter-rater agreement is unlikely to be perfect. Of all the clinical assessments performed, pronation-supination movements showed the worst inter-rater agreement (k=0.52; "moderate"). This finding likely reflects the degree of subjectivity within this item's descriptors; examination of why agreement was less marked revealed that the raters appeared to differ in their view of what constituted "slight", "mild" and "moderate" slowing of movements.

High levels of agreement were seen between clinicians for disease state assessment, with disagreement seen for only 4/141 videos. The decision was made prior to dual-rating of these videos that the second rater would be given the video clips for each participant in chronological order. Whilst for many video-clips the participant's disease state was obvious (e.g. profound on or off, dyskinesia), it was acknowledged that on occasions it would be extremely difficult to make a judgement on disease status based on an isolated video-clip. As such the blinded clinician was able to review all videos recorded for a participant during their time in CARU and was aware of the order in which they were taken. Whilst this step does partially remove blinding it did enable the second rater to gain an appreciation of the evolution of participants' symptoms over a period of time, thus meaning that their evaluation of disease status was likely to be more accurate.

Based on the confusion matrix presented in Table 17 it can be concluded that the ANN performed sub-optimally when applied to CARU data. The ANN identified numerous periods of time as sleep despite the fact that there were no such periods in the CARU data set. Sensitivity of detection of all disease states was sub-optimal, with the best result being seen for 'on' (56.0%). High specificity values were seen for both dyskinesia (95.9%) and sleep (84.9%). Specificity for the detection of 'off' and 'on' was 75.1% and 60.1% respectively.

Comparison to results achieved by other research groups is difficult since differing methods have been applied. Keijsers et al. (Keijsers et al., 2006) did apply an ANN to this problem but only reported the ANN's training results, quoting sensitivity and

specificity values of 1.00 and 0.98 respectively. Our quoted results are those derived from testing of the ANN i.e. application of the trained ANN to new, previously unseen data (aiming to establish whether the ANN is capable of generalising to a new patient's data). Since Keijsers and colleagues do not present equivalent data, head-to-head comparison is not possible. Keijsers et al. did also perform analysis using single variables and reported sensitivity and specificity values of 0.71-0.74 and 0.78 respectively. In terms of sensitivity and specificity, this analysis approach outperforms ours. It must however be noted that Keijsers work was based on only three hours of recording in a controlled laboratory environment, where participants performed a defined set of activities. Furthermore, data collected during walking was excluded and no attempts were made to discern between 'on' periods and periods with dyskinesia; the two categories were simply amalgamated into 'on'. In contrast, our data was obtained in a truly uncontrolled manner; patients wore the sensors at home for seven days and were completely unsupervised; we had no information regarding the activities that they were performing.

Hoff et al., also using single variable analysis, reported sensitivity and specificity values of 0.60-0.71 and 0.66-0.76 respectively; results superior to ours (Hoff et al., 2004). The analysis methods employed by Hoff were relatively simplistic in comparison to methods employed in this thesis. The question is therefore, why was such disappointing ANN performance seen when CARU data was assessed? A number of potential reasons were identified. The ANN employed in this work was trained using a compendium of data obtained from phase two of the study, the home monitoring phase. During this period participants were asked to go about their usual activities as normal such that the movement data captured was reflective of 'real-life'. Conversely, phase one of the study required participants to attend a research facility (CARU). Participants' time at CARU was, compared to time at home, relatively scripted, with a series of regular assessments to complete. Whilst participants did have periods of free time that were typically spent eating, drinking or completing questionnaires, it is unlikely that data collected during phase one is entirely representative of 'real-life'. Attempts were made to control for this, through exclusion of data obtained during the

abbreviated MDS-UPDRS assessments themselves, since these assessments were considered to be far removed from activities performed in everyday life. Despite this, there remains the possibility that the ANN failed to correctly categorise participants' disease states during CARU, because the CARU data it was presented with 'looked' so different to the home-derived data upon which it was trained.

Examination of the proportions of each disease state recorded in each phase provides further evidence for the argument that the two data sets differ (Figure 45). For all participants, the proportion of time spent in each disease state during the home period was determined from patient-completed symptom diaries. The proportion of time spent in each disease state during the CARU phase was derived from clinician assessments of disease status. "Invalid" data represents diary entries deemed to be unclassifiable and CARU assessments where no video-recording was available.

Figure 45: Proportions of time spent in each disease state, for all participants, during home and CARU phases
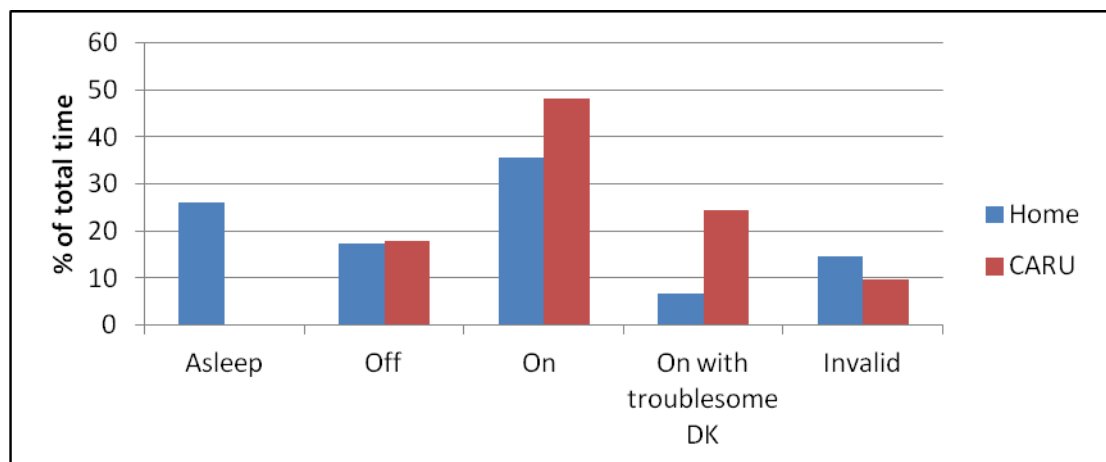
It is evident from Figure 45 that sleep is absent from the CARU data set but forms a significant proportion of the home data set (26.0%). The other striking difference is the disparity between the amount of dyskinesia seen during home and CARU phases (6.7% and 24.4% respectively). There are a number of potential reasons for this finding that will now be discussed.

The first two reasons relate to the possible introduction of systematic bias during the CARU phase. Firstly all participants attending CARU were asked to withhold their first levodopa dose of the day when they were attending; this medication was then taken once the initial clinical assessment had been performed. During the home monitoring phases participants took their medications in their usual manner with no adjustment made to their treatment regimen. It is possible that alteration of the timings of participants' medication may have resulted in different clinical manifestations of their disease during their time in CARU compared to the equivalent time at home. It is likely however that delaying levodopa administration would result in more 'off' time as opposed to increases in time spent with dyskinesia.

Another possible source of systematic bias relates to the time of day that CARU visits were performed. All CARU visits were undertaken in the morning, typically between 0900 and 1300 hours. There is a possibility that perhaps dyskinesia was more prominent amongst participants in the morning, hence the overestimation. To evaluate the validity of this argument the diaries of all participants were re-examined. For each hour long period the total number of hours of dyskinesia reported by the study cohort during the week long monitoring period was calculated (Figure 46).

Figure 46: Total number of hours of dyskinesia reported by the study cohort during the home monitoring period for each hour-long period
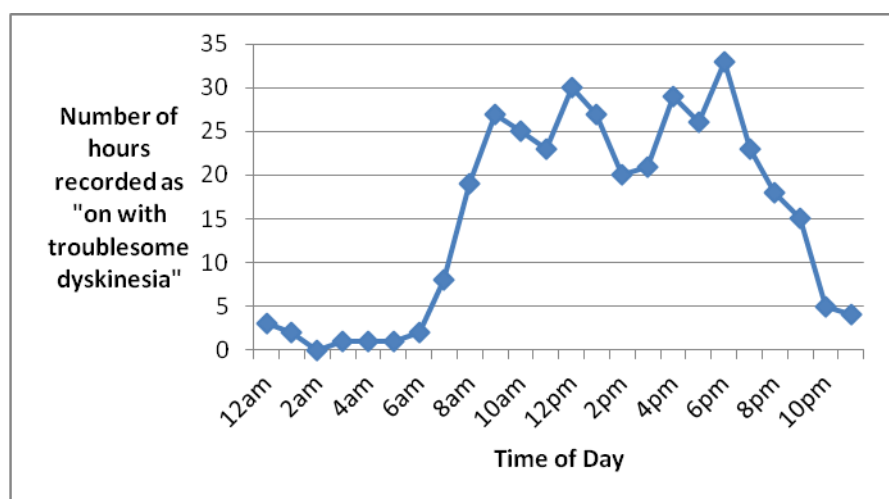
It is evident from Figure 46 that similar amounts of dyskinesia were reported during both morning and evening periods. Consequently, the timing of CARU visits is unlikely to explain the greater amount of reported dyskinesia in the CARU data set compared to the home monitoring data set.

Two other alternative explanations exist for the large disparity between the proportions of dyskinesia seen in the CARU and home datasets. Firstly, as previously described, it is known that patients may be unaware of dyskinesia (Vitale et al., 2001). It may simply be that in some cases patients were oblivious to dyskinesia and therefore did not record it within their diaries. Alternatively it may be that patients did recognise dyskinesia but differentiated between troublesome and non-troublesome dyskinesia. Consequently only dyskinesia considered to be troublesome would have been recorded as dyskinesia in their diaries. Conversely, assessment of dyskinesia during the CARU phase made no evaluation of whether or not the dyskinesia was troublesome or not. This is the most likely reason for the greater proportion of dyskinesia recorded during CARU compared to home.

A major potential contributory factor for the low sensitivity values seen for on and off states, relates to how these states were defined. Disease status was defined by participants during the home phase and by clinicians during the CARU phase. The ANN was therefore trained on the basis of participants' judgement of disease status, yet tested against clinicians' judgement of disease status. Discordance between patients' and clinicians' judgements of disease status has been documented previously (Goetz et al., 1997) and is perhaps attributable to the inherent subjectivity of what constitutes on and off states, and the potential influence of patient psychology on their ratings. As previously stated, Hoff et al. (Hoff et al., 2004), despite employing relatively simplistic analysis methods, demonstrated superior performance in terms of sensitivity to our work. This may reflect the fact that Hoff relied upon participants' judgement of disease status alone. Participants kept a disease state diary for the 24-hour home monitoring period and recorded their disease state (on, off, dyskinesia) on a half hourly basis. There was no method by which the validity of participants' diary entries could be evaluated. This raises an interesting discussion point regarding who is best placed to

pass judgement on disease state. Exactly what constitutes on and off disease states is very much subjective; there is no truly universal definition of what each entails. Whilst explicit guidance was provided to participants, an individual's interpretation of their symptoms is very much dependant on their own, personal frame of reference. The frame of reference through which a clinician makes a judgement on disease status is likely to be heavily influenced by past experiences; they are likely to have seen many PD patients, with differing symptoms. In retrospect, rating of disease state during the CARU phase by both patient and clinician would have been preferable. This addition would have enabled concordance between clinician's and participants' ratings to be examined. Furthermore, comparison of the ANN to both patient and clinician defined disease state could also have been undertaken.

It is important to highlight other possible reasons why Hoff et al.'s simplistic analysis approach produced superior results to our own. Hoff employed a series of accelerometers mounted on the sternum, wrist and thigh. Each accelerometer was wired to a portable, battery-powered activity monitor that was worn on a belt around the patient's waist. In this thesis we relied on two wrist-worn sensors for data capture. The additional sensors employed by Hoff et al. provided them with data from body segments that we were unable to consider. It is possible that data derived from the trunk or thigh sensors proved more useful than wrist-derived data alone, when attempting to discern between differing disease states. Further evidence to support this suggestion comes from the work by Keijsers et al. (Keijsers et al., 2006). Their simplistic ANN relied upon only four input features, the most important of which (in terms of differentiating between on and off states) was derived from a sensor mounted on the trunk (%PF_4: percentage of peak frequencies above 4Hz, where peak frequency is the frequency that occurs with the largest power on Fourier analysis). The feature %PF_4 (trunk) accounted for 96% of network performance with the three other features derived from sensors worn elsewhere on the body providing only minimal influence on performance. The authors postulate that increased prominence of tremor during off periods may explain why this feature is so useful, but fail to explain why the same feature from wrist-worn sensors does not provide comparable results. Similarly,

there is no convincing explanation why this feature should be so effective in patients *without* tremor, since they are unlikely to exhibit movements at greater than 4Hz.

It is important to acknowledge that whilst additional sensors yield more data, there are drawbacks to using more sensors. Hoff et al. described the system they employed as "appropriate for ambulatory monitoring" due to its "small size and weight" (Hoff et al., 2004). The dimensions of the belt-worn activity monitor were 90mm x 150mm x 45mm and it weighed 750g, which, by current day standards, seems somewhat bulky and heavy. Clearly technology has advanced since Hoff's work in 2004, but for comparison, the sensors used in this thesis measured 39mm x 36mm x 12.5mm and weighed only 35g. Hoff et al. made the assumption that their system was "appropriate" for ambulatory monitoring without formally evaluating the acceptability of the system to the wearer. Clearly whether or not the wearer deems the system to be appropriate is critical to its implementation as a home monitoring device. The acceptability of the system we employed will therefore be considered in depth in this thesis (see Chapter 9).

Another possible contributory factor to the difficulty in identifying on and off periods correctly, relates to the fact that clinical features manifested by a patient during off periods are arguably different depending on their disease phenotype. A patient with tremor dominant PD is likely to experience pronounced tremor during off periods whereas a patient with the PIGD phenotype in the off state will exhibit less tremor, but more pronounced bradykinesia. This is exemplified by data presented in Figure 47, which displays accelerometer-derived power spectra for tremor dominant and non-tremor dominant patients (Keijsers et al., 2006).

Figure 47: Accelerometer derived power spectra for PD patients in clinician defined 'off' state (upper panel: tremulous patient; lower panel: non-tremulous patient) [adapted from Keijsers et al.]
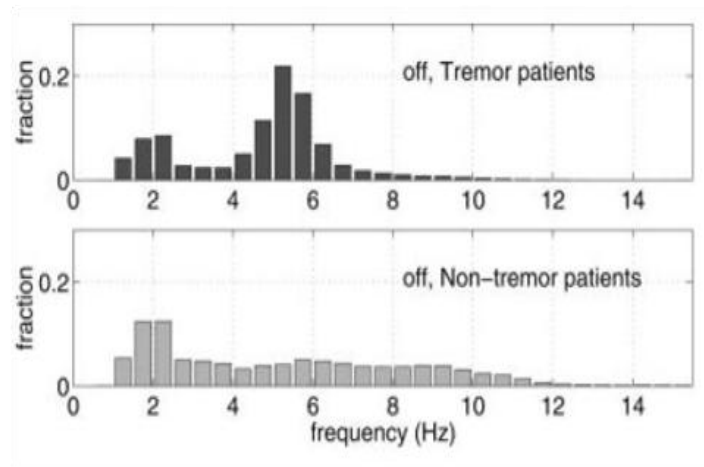
Clear differences in signal composition are evident between the different patient groups. Tremor dominant patients in the 'off' state are seen to exhibit a prominent peak in signal frequency between 4-7Hz'; this is likely to represent prominent Parkinsonian resting tremor. Non-tremor dominant patients in the 'off' state show no such peak in the frequency spectrum. Such a finding further highlights the difficulties associated with developing an ANN to identify motor symptoms in PD. The subjectivity of disease status has already been discussed. The variability of the motor symptoms expressed by patients of differing clinical phenotypes further blurs the ANN's construct of what constitutes 'on' and 'off', and is likely to have further contributed to the suboptimal performance of the ANN.

Application of the ANN to CARU data demonstrated low sensitivity (23.9%) but very high specificity (95.9%) for the detection of dyskinesia. In short, the ANN is poor at identifying those with dyskinesia but excellent at identifying those without dyskinesia. As described earlier, reliable detection of dyskinesia has also proved challenging for other research groups, with particular difficulty seen in attempts to differentiate between mild dyskinesia and 'normal' movements (Keijsers et al., 2003b, Manson et al., 2000). The low sensitivity and high specificity demonstrated in our work may reflect similar difficulties in discerning mild dyskinesia from periods of comparatively 'normal' movements (i.e. on). This finding may be explained by the training process

151

undertaken during development of the ANN. As previously discussed home diary data formed the compendium of data used to train the ANN. The disease state category referring to dyskinesia was termed "on with troublesome dyskinesia". It is therefore likely that participants with dyskinesia, that deemed their dyskinesia to be non-troublesome, indicated such periods as 'on' time. Consequently, the ANN's concept of what dyskinesia 'is', is likely to be skewed towards more severe dyskinesia. The ANN was tested against clinician defined disease state labels that were assigned during the CARU phase. There was no differentiation between troublesome and non-troublesome dyskinesia by the rating clinician; all dyskinesia, regardless of severity, was rated as dyskinesia. It is therefore likely that periods of less severe dyskinesia during the CARU phase, were not recognised as being dyskinesia by the ANN, due to the disparity between the data it was presented with and concept of dyskinesia that it had 'learnt'. Failure of the ANN to recognise periods defined as dyskinesia is the likely explanation for the low sensitivity observed. It is probable that only more severe dyskinesia was detected by the ANN and as such, low numbers of false positives were produced and hence the high specificity value observed.

7.II.i   CONCLUSIONS

- Performance of the ANN, when applied to CARU-derived data, was sub-optimal.
- Differences between CARU and home data sets, both in terms of the proportions of time spent in each disease state and in terms of the activities performed by patients, may have contributed to poor ANN performance.

Chapter 8. Comparing assessment of disease status produced by the MDS-UPDRS, patient-completed symptom diaries and ANN

8.I   Results

Earlier in this work the proportions of time spent in each disease state as measured by the MDS-UPDRS and home diaries were analysed, and correlation between the two measures was assessed. Using the ANN's evaluation of disease state for each hour of home monitoring data, it is now possible, for each participant, to calculate the proportions of time spent in each disease state based on the ANN outputs. For clarity, visual representations of this data, alongside that derived from the MDS-UPDRS and diary, are now presented for two of the study participants. The first (Figure 48) is for MXRL, the participant in whom the highest sensor performance was seen (where performance is defined as the degree of correlation between sensor output and diary entries; in this case 0.851). The second (Figure 49) is for GLXN, the participant in whom the lowest sensor performance was seen (0.547).

Figure 48: Comparison of the percentage of time spent in each disease state as measured by the MDS-UPDRS, patient completed symptom diary and ANN, for participant MXRL (highest sensor performance in cohort)
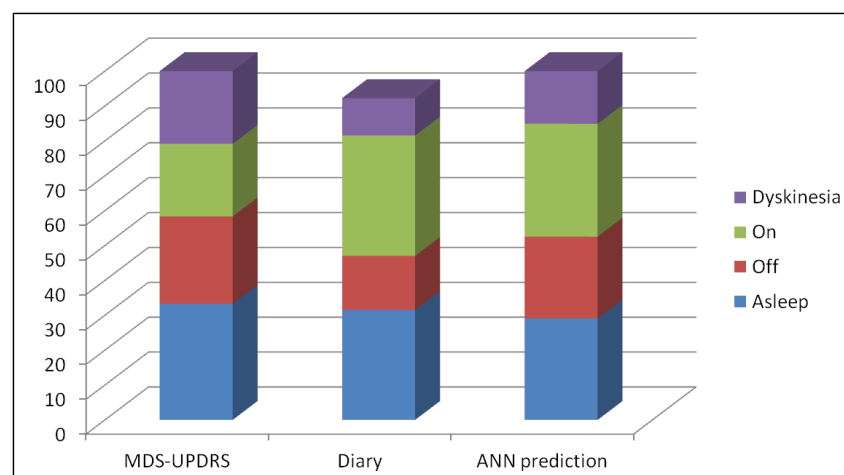
Based on visual inspection of Figure 48 it is evident that, for participant MXRL, the three different assessment methods produced similar values for the proportions of time spent in each disease state. Diary data does not sum to 100% due to the presence of missing (unclassifiable) diary data.

Figure 49: Comparison of the percentage of time spent in each disease state as measured by the MDS-UPDRS, patient completed symptom diaries and ANN for participant GLXN (lowest sensor performance in cohort)
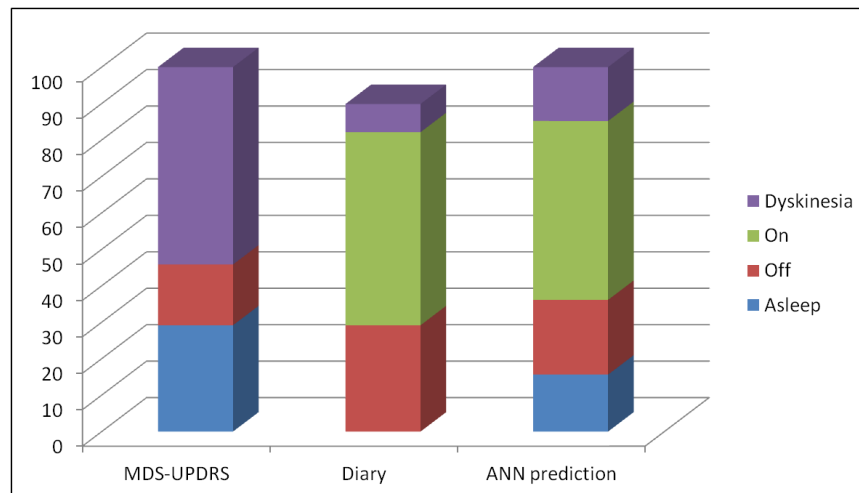
Visual inspection of Figure 49 (participant GLXN) shows disparity between the proportions of time in each disease state as measured by the different assessment methods. It is also evident that, according to the symptom diary, no sleep was recorded during this participant's home monitoring period.

Pearson's coefficient was used to examine correlation between the amounts of time spent in each disease state as measured by two different measurement methods. Comparison between each pair of assessment methods was undertaken for data derived from all eligible participants and the results are presented in Table 19.

Table 19: Correlation between the amounts of time spent in each disease state as measured by the MDS-UPDRS, patient completed symptom diaries and ANN ('Sensor')

Disease state	MDS-UPDRS vs Diary		Diary vs Sensor		MDS-UPDRS vs Sensor	
Consideration of all participants (n=27)						
	Correlation	p-value	Correlation	p-value	Correlation	p-value
Asleep	0.110	0.585	0.424	0.028	0.192	0.336
Off	0.208	0.299	0.603	0.001	0.207	0.299
On	0.372	0.056	0.540	0.004	0.361	0.064
Dyskinesia	0.472	0.013	0.704	<0.001	0.474	0.012

Comparison of diary and ANN-derived data revealed strong correlations between the proportion of time spent 'on' (r=0.540; p=0.004), 'off' (r=0.603; p=0.001) and asleep (r=0.424; p=0.028). A very strong correlation was seen for dyskinesia (r=0.704; p=<0.001). Correlation between the amount of time spent in each disease state as measured by the MDS-UPDRS and the ANN was less marked. A strong, significant correlation was seen for dyskinesia (r=0.474; p=0.012), but examination of all other disease states showed no evidence of a meaningful relationship (asleep: r=0.192; p=0.336; off: r=0.207; p=0.299; on: r=0.361; p=0.064). This analysis was repeated following exclusion of data derived from participants deemed to be 'poorly compliant' diarists (those in whom <90% of diary entries were classifiable); the results are displayed in Table 20.

Table 20: Correlation between the amounts of time spent in each disease state as measured by the MDS-UPDRS, patient completed symptom diaries and ANN (with exclusion of 'poorly compliant' diarists from analysis)

Disease state	MDS-UPDRS vs Diary		Diary vs Sensor		MDS-UPDRS vs Sensor	
Consideration of participants whose diaries contained ≥90% of responses that were classifiable (n=21)						
	Correlation	p-value	Correlation	p-value	Correlation	p-value
Asleep	0.184	0.412	0.161	0.485	0.208	0.336
Off	0.521	0.013	0.727	<0.001	0.280	0.218
On	0.389	0.074	0.704	<0.001	0.336	0.136
Dyskinesia	0.576	0.005	0.718	<0.001	0.442	0.045

Again, correlations between MDS-UPDRS and diary measures are those previously presented (see Table 11) and are reproduced to allow comparison. Comparison of diary and ANN-derived data revealed very strong correlations between the proportion of time spent 'off' (r=0.727; p=<0.001), 'on' (r=0.704; p=<0.001) and dyskinetic (r=0.718; p=<0.001). The correlation for sleep did not achieve statistical significance (r=0.161; p=0.485). Correlation between the amount of time spent in each disease state as measured by the MDS-UPDRS and the ANN was again less marked. A strong, significant correlation was seen for dyskinesia (r=0.442; p=0.045), but examination of all other disease states showed no evidence of a meaningful relationship (asleep: r=0.208; p=0.336; off: r=0.280; p=0.218; on: r=0.336; p=0.136).

8.II   Discussion

It is evident from Figure 48 that the three methods employed to measure the proportions of time spent in each disease state, produced concordant results for this particular participant (MXRL). MXRL was the participant in whom the highest correlation was seen between ANN output and diary data (Figure 36). Similar data is presented in Figure 49 for the participant in whom the lowest correlation between ANN output and diary data was seen (GLXN). Visual inspection of this data shows a degree of concordance between diary and ANN derived results. It also highlights some limitations of the existing assessment methods. Firstly, it is evident that the MDS-UPDRS, for this participant, produced a gross over-estimation of the amount of time with dyskinesia. Secondly, it can be seen that for diary-derived data, this participant recorded no periods of sleep during the entire seven day monitoring period, raising concerns regarding the validity of this diary.

Table 19 and Table 20 present correlations between the amounts of time spent in each disease state as measured by the three assessment methods employed. It was evident that strong, significant correlations existed between diary and ANN-derived data for all disease states. This correlation became more marked for all disease states (except sleep) when 'poor' diarists were excluded from analysis. Limited correlation was seen

when the amounts of time spent in each disease state, as measured by the ANN or diary, were compared to data derived from the MDS-UPDRS.

Diaries are considered to be the current gold standard for home monitoring of motor symptoms in PD. These findings suggest that the ANN developed and employed in this work produced comparable results to those generated by the home diaries; evidence for the criterion-related validity of our assessment method. The proportion of time spent in each disease state is not a meaningless measurement and has real clinical relevance. The amounts of on and off time are frequently employed as outcome measures in PD trials to examine whether any motor benefit is derived from anti-parkinsonian medications. In such trials, the amount of time with dyskinesia is also often used as an outcome measure in order to establish whether any motor benefit comes at the cost of worsening dyskinesia. Our findings also highlight the disparity between MDS-UPDRS and diary-derived estimations of disease status, and should be a note of caution to clinicians relying on the MDS-UPDRS to estimate proportions of time spent in each disease state. As demonstrated previously, cognitive dysfunction can impair a person's ability to complete a symptom diary. Such patients are often those with more advanced PD in whom motor fluctuations and dyskinesia may be more prominent, rendering management more challenging. The lack of an available method to monitor the symptoms of such a patient group is problematic and can be considered an example of the inverse care law (Tudor Hart, 1971). Evaluation of motor symptoms using wrist-worn accelerometry and ANN-based analysis has the potential to provide a viable method of home-monitoring for this patient group. It is well recognised that impaired cognition can be a barrier to involvement with clinical research (Mody et al., 2008); such a method may enable greater participation of cognitively impaired PD patients in medication trials. Whilst the sensor was widely accepted by participants in this study, it cannot be assumed that similar findings will be replicated in a more cognitively impaired cohort. Evaluation of the acceptability of the sensor in such a cohort would form a critical part of any future research project and qualitative approaches, as opposed to questionnaires, would be more likely to generate meaningful data.

8.II.i   CONCLUSIONS

- Estimations of the time spent in each disease state derived from the MDS-UPDRS should be interpreted with caution

- A strong, significant correlation was evident for time spent in each disease state as measured by home diaries and by the ANN.

- ANN-derived results were comparable to those generated by the home diaries; evidence for the criterion-related validity of our assessment method.

- Body-worn sensor systems may allow home based monitoring of PD motor symptoms in patients who are unable, or unwilling, to engage with diary keeping exercises.

Chapter 9.   Acceptability of the Sensors to Participants

9.I   Background

9.I.i   Introduction

The demographic of the United Kingdom's population is changing; people aged 65 and over accounted for 15% of the population in 1985 and this had risen to 17% by 2010 (2012e). It is projected that 23% of the population will be aged 65 and over by 2035. As the UK's population ages, the numbers of patients with chronic conditions such as Parkinson's disease (PD) will increase. Increasing numbers of such patients will inevitably place great strain on the limited resources available to the UK's National Health Service (NHS) due to their complex care needs. Projections suggest that the number of patients with PD will rise from 126,893 in 2009, to around 162,000 by 2020 (2009b). Incorporating the use of new technology into the provision of medical care may be able to improve efficiency of health services and help the NHS to cope with this problem. The NHS of the future is expected to provide increasing proportions of its services in the community as opposed to within costly clinical environments. The use of technology to monitor a patient's condition in their own home may be an intervention that can assist this transition. A recent UK government mandate (2012c) highlighted the need to provide support to patients with chronic disease, by enabling the monitoring and management of their condition in their own homes. This report targeted three million people by 2017, being able to benefit from "telehealth and telecare" in their own homes. The United States of America's Institute of Medicine published a report in 2009 (2009a) that identified remote patient monitoring of chronic conditions as one of their recommendations for areas of high priority research.

Home monitoring of medical conditions is, for some diseases, already well established. The most well known and widely used example is the Holter electrocardiographic monitor. This device has been in use for over 50 years (Mar, 2005) and enables continuous recording of the wearer's electrocardiogram for prolonged periods in a home environment. The prolonged, continuous nature of such a recording enables

159

clinicians to gain an appreciation of heart rate and rhythm during normal daily activities (Binkley, 2003), but also increases the likelihood of capturing rare, short-lived rhythm events that may not be witnessed when a patient attends a clinical setting. It is important to consider the setting in which clinical observations are made. The presence of a patient in a clinical setting may introduce bias to the data collected; a good example being the diagnosis of hypertension. Traditionally, diagnosing hypertension relied upon clinic-based blood pressure measurement. The so-called "white-coat" effect (Verdecchia et al., 2002); a transient blood pressure rise associated with the presence of a doctor, may result in an incorrect diagnosis of hypertension. Recent evidence suggests that prolonged, ambulatory blood pressure monitoring outside of the clinical environment can better inform decisions regarding treatment for hypertension (Hodgkinson et al., 2011). Consequently the use of home blood pressure monitoring is now recommended in the NICE hypertension guideline. In PD it is well recognised that the presence of an observing clinician can alter the clinical signs a patient exhibits. Stress, associated with clinic attendance or social embarrassment, can result in a worsening of a patient's tremor (Marsden and Owen, 1967). Dyskinesia can become more prominent when patients are engaged in mental distraction such as conversing with a clinician (Durif et al., 1999). Conversely patients are sometimes able to suppress motor symptoms such as tremor or dyskinesia. A method of evaluating motor symptoms without the need for clinical observation may remove this observer bias.

Adoption of an effective home monitoring system for a particular condition requires a number of factors to be fulfilled. Firstly, an accurate and reliable device must be developed that is capable of capturing appropriate and relevant data. Secondly, methods must be developed to enable interpretation of the data collected and to allow clinically meaningful conclusions to be drawn. Lastly, and critically, patients must be concordant with the wearing of any such device and remain so for the duration of the required monitoring period. When developing a home monitoring system, failure to adequately address all three of the factors identified may limit the clinical usefulness of the system. In this chapter consideration will be given to the patient

experience of wearing medical technology; factors that may influence patient concordance with sensor wearing will be reviewed and an evaluation of the acceptability of the Axivity-AX3 sensor will be made.

9.I.ii   User-friendliness and wearability

An object may be described as being user-friendly if it is "easy to use, understand or operate" (2013c). The concept of user-friendliness is central to patient concordance with a wearable, medical device and may be influenced by factors relating to both the device and the wearer. The physical properties of a monitoring system have a huge bearing on patient acceptance due to their impact on its wearability, defined as the interaction between the human body and the wearable object (Gemperle et al., 1998). This definition can be further developed to include 'dynamic wearability', which considers the same interaction but when the body is in motion. Gemperle et al. also highlighted the importance of unobtrusive placement of a wearable device; Figure 50 shows areas of the body where wearable objects are considered to be most unobtrusive.

Figure 50: Unobtrusive areas for wearable objects: (a) collar area, (b) rear of upper arm, (c) forearm, (d) rear, side and front of ribcage, (e) waist and hips, (f) thigh, (g) shin, (h) top of foot [adapted from Gemperle et al. 1998]
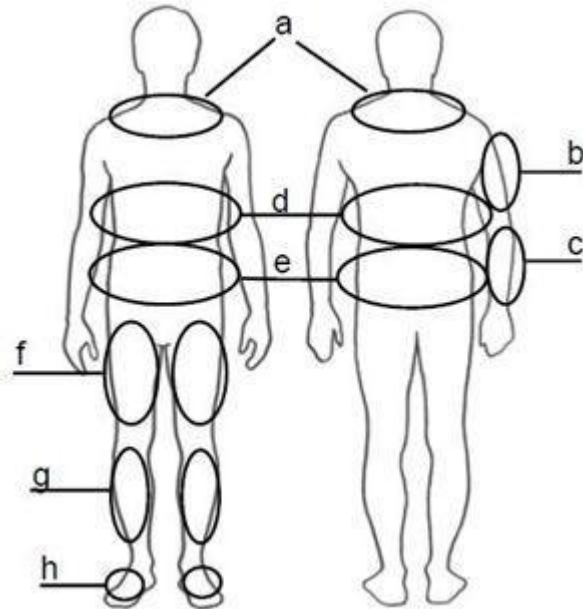
It is argued that the longer a device is worn for, the greater the need for it to be as un-obtrusive as possible if patient concordance is to be maintained (Bonato, 2010). The presence of tremor in patients with PD presents further design challenges since the use of computer peripherals, buttons or switches may prove challenging (Cunningham et al., 2009). It is also recognised that the 'life' of a monitoring system; the period of time before data must be downloaded from the device, or the device itself must be recharged, also impacts on how obtrusive a system is perceived to be (Bonato, 2009).

Human factors also play a significant part in whether such a device is considered to be user-friendly. The need to learn how to use a piece of technology may have an adverse effect on patient willingness to engage with the device (Cunningham et al., 2009). The importance of body-image on concordance should not be underestimated; the wearing of a medical device can have associated negative connotations (Peeters, 2000). Peeters described the stigma associated with the wearing of a neck-worn automatic safety-alarm pager and how elderly patients may reject products they perceive to be "marking them as old". It is important to acknowledge that the converse may be true;

some patients may consider wearable medical devices to be a form of "status symbol", suggestive of a willingness to live a healthy lifestyle (Korhonen et al., 2003). Ultimately, regardless of how well a wearable device may be designed, the wearer is always likely to experience a degree of inconvenience due to obtrusiveness of the device. Whilst designers must strive to minimise this, it is also important to acknowledge that patient motivation plays a significant role in whether the inconvenience is tolerated or not. Providing the wearer with contemporaneous, comprehensible feedback from the monitoring device, can boost motivation to continue wearing it (Korhonen et al., 2003). It is recognised that motivation to wear a medical device is closely linked to the perceived, immediate importance to the wearer's life (Peeters, 2000). Powerful motivating factors to wear such a device include whether the product improved a person's ability to remain living in their own home, living their 'normal' life and to remain integrated in society.

Lehoux (Lehoux, 2004) examined how, from a patient perspective, user-friendliness of medical technology influenced its integration into both the private and social lives of patients. The distinction between private and social spheres highlights how user-friendliness of a device may be location specific, with wearers becoming more self-conscious once outside their "private sphere". This work proposes that the concept of user-friendliness encompasses both user acceptance; the extent to which the user is favourable to using the technology, and user competence; the abilities required to use the technology effectively. Technical and human factors also influence user friendliness, with technical factors predominantly influencing user-competence and human factors predominantly influencing user acceptance. This model is displayed in Figure 51 below.
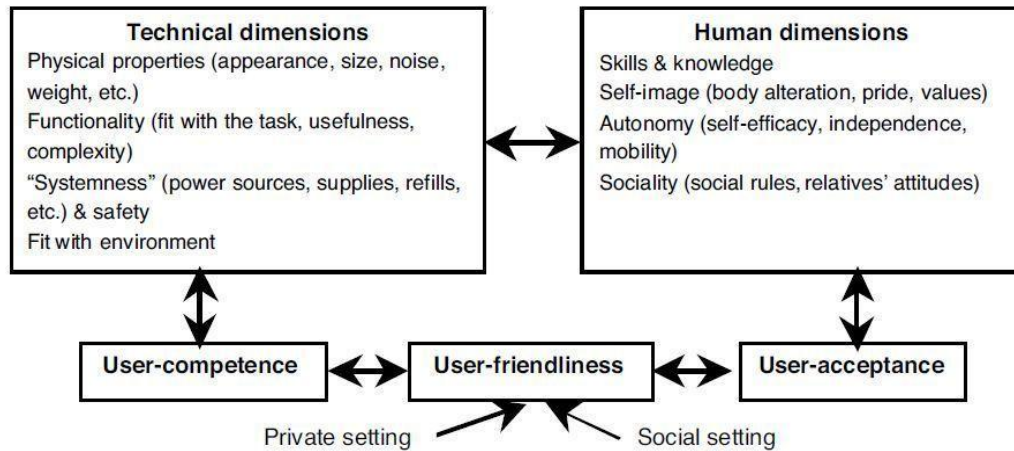
9.I.iii   Consideration of acceptability in a PD population

With regard to the use of wrist-worn sensors in the measurement of upper limb motor symptoms in PD, very little consideration has previously been given to the patient experience of wearing such devices. The overwhelming majority of research in this area makes no reference to any evaluation of patient acceptability. It is telling that a recent review article that examined the use of body-worn sensors in PD (Maetzler et al., 2013), made no reference to any evaluation of the acceptability of the sensors to the wearer. Tsipouras et al. (Tsipouras et al., 2012) did report using a questionnaire to assess the wearability of a series of sensors used to assess dyskinesia in PD patients. This work concluded that the sensors were "considered wearable by all" and that they "did not interfere with normal posture". No results were presented to support this conclusion however, and no details regarding the nature of the questionnaire were given. There is evidence that some authors lack an appreciation of the factors influencing user-friendliness of such a device. Van Someren et al. (van Someren et al., 1993a) advocated home recording of tremor using their wrist-worn sensor, stating that wearing wrist-worn sensors for several weeks would be "no more uncomfortable than wearing a wrist watch". This assumption was proven to be inaccurate within the same piece of research; 2 of the 8 participants in the study "objected" to 24 hour monitoring with the sensor system and completed only 8 hours of data collection. Work by

Giuffrida et al. (Giuffrida et al., 2009) provides the most thorough evaluation of the acceptability of a wrist-worn sensor in a PD cohort. This work used an 8 item questionnaire to capture participants' views on comfort and ease of use of the sensor system, the results of which are summarised in Figure 52.

Figure 52: Results of the questionnaire employed by Giuffrida et al. (Giuffrida et al., 2009) in the evaluation of patient acceptability of their sensor system

Question	Yes
Was the device comfortable to wear?	100%
Did the device feel heavy?	0%
Did the device restrict your arm movement?	0%
Did the device restrict your hand movement?	0%
Would you wear the device at home?	94%
Would you perform arm movements at home similar to those during an office exam?	88%
Would you have trouble putting the device on at home?	3%
Would you wear the device in public?	55%

Responses from 40/60 patients (67%)

The authors concluded that the results indicated "clinical patient acceptance", despite the fact that 45% of participants indicated they would not be willing to wear the device in public. There are several limitations to the questionnaire applied by Giuffrida and colleagues. Firstly the response rate is comparatively low, with a third of participants failing to complete the questionnaire. Secondly, respondents to each question could only select from either 'yes' or 'no', which represents a crude method of capturing data. Furthermore, views were polled after only short-lived recordings that were undertaken in a clinical setting. These results cannot reliably be extrapolated to make judgements on the acceptability of more prolonged, home use of the system since the duration and location of monitoring can influence patient acceptability (Bonato, 2010, Lehoux, 2004). The results presented by Giuffrida and colleagues highlight a disparity between participant willingness to wear the device at home and in public. The ideal home monitoring system would allow unobtrusive evaluation of the wearer during their normal daily living. If participants are less willing to wear the device in public than at home, then there is the possibility that the pattern of activity exhibited by the

individual during the monitoring period may differ to their normal, thus biasing the results. Lastly, Mera et al. published work entitled "Feasibility of home-based automated Parkinson's disease motor assessment" (Mera et al., 2012b). Despite the fact that patient concordance is a critical component of the feasibility of a home monitoring system this work gave no consideration to patient experience, only examining whether or not participants performed the required assessments correctly. This research paper highlights the fact that work in this field has primarily focussed on the agenda of the researchers and that the wearer's agenda has been inadequately addressed. The need for greater consideration of patient feedback in the development of health-care is well recognised (Wensing and Elwyn, 2003).

This chapter presents an in-depth analysis of acceptability of the Activity AX3 body-worn sensor to a population of patients with PD.

9.II   Specific Methods

9.II.i   The Questionnaire

To capture data regarding participants' experiences of wearing the sensor, all participants completed a questionnaire on completion of both phase one and phase two of the study. The questionnaire provided during phase one (CARU) was the final assessment completed by participants, thus ensuring that all participants had worn the sensors for at least four hours before completing the questionnaire. The questionnaire was completed by participants with no input from the research team. The second questionnaire was completed by participants at home on completion of phase two, again with no input from the research team. This questionnaire was identical to that which was provided during phase one and was printed on the final page of participants' home monitoring booklets. During the standardised briefing for the home monitoring phase it was stressed to participants that completion of this questionnaire should be the final task they perform, once the home monitoring period was complete. This request was made in an attempt to ensure that all participants completing this questionnaire did so after having worn the sensors for a week long period. If participants decided that they wished to discontinue home monitoring before the end

of the seven day period, it was requested that they completed the questionnaire at this time. Data generated from such questionnaires was still analysed, regardless of the duration for which participants wore the sensors, as participants withdrawing from the study early may provide invaluable insight into adverse experiences of wearing the sensors. Questionnaires completed at home were returned to the research team in a pre-paid envelope.

The questionnaire included nine statements (items) with a five category Likert-type rating scale. The items developed for this questionnaire are displayed in Table 21 below. Likert scales are frequently used to measure attitudes and provide a range of responses to a given statement, with five categories of response typically being employed (Jamieson, 2004). The five available categories provided in this questionnaire were: strongly agree, agree, neither agree nor disagree, disagree and strongly disagree.

Table 21: Questionnaire Items

Item Number	Statement
1	The sensor looks like it is well made
2	The sensor is comfortable to wear
3	The sensor feels heavy on my arm
4	Performing the assessments was made more difficult by wearing the sensor
5	I would be happy to wear the sensor around the house
6	I would rather keep a regular diary of my symptoms for a week than wear the sensor for a week
7	If the sensor was incorporated into a working wrist watch I would be more likely to wear it
8	The sensor is easy to take on and off
9	I would be happy to wear the sensor in public

The nine items developed for use in the questionnaire were selected to enable evaluation of both the technical and human dimensions recognised as influencing the user-friendliness of a medical device (Figure 51). The Likert items selected for inclusion were based on topics derived from previous research in this area (Giuffrida et al., 2009) and on areas considered to be relevant that had not been addressed in previous

work. Items 1 and 3 were designed to capture participants' opinions regarding the physical properties of the sensor. Items 2, 4 and 8 addressed the functionality of the sensor. The inclusion of items 5 and 9 enabled evaluation of whether the setting (i.e. private or social) influenced acceptability of the sensor. Item 7 was designed to address the physical properties and functionality of the sensor, as well as participants' self-image. Lastly, item 6 provided a direct comparison between the current gold standard method of home monitoring of motor symptoms in PD (patient completed diaries) and the home monitoring system employed.

The questionnaire also included an empty text-box in which participants were invited to "Please write any comments (you have) about the sensor (below)". This opportunity was included because the content of the nine items, whilst derived from research evidence (itself derived from previous patient experience), represents the agenda of the research team. Providing an opportunity for participants to give free-text feedback gives participants the chance to voice their agenda. This may provide the research team reassurance that all relevant issues have been covered or may produce new, unexpected themes not captured by the closed questions. Such data may be used to corroborate Likert-type answers and may also provide illustrative quotes (O'Cathain and Thomas, 2004).

The questionnaire was reviewed by a patient with PD who had expressed an interest in participating in the study. This patient pre-read the questionnaire and provided positive feedback in terms of its readability and its clarity; no changes were made to the questionnaire in light of this. During the development of the questionnaire quantitative evaluation of its readability was undertaken using the Flesch Reading Ease (FRE) formula (Kincaid et al., 1975). This formula generates a numerical value that describes the readability of a document; higher values indicate material that is easier to read, lower values suggest passages that are more challenging to read. Questionnaire items were found to have an FRE value of 78.8, suggestive of 'fairly easy' reading difficulty. Scores greater than 70 suggest text that would be easily understood by 13 year old school children; thus the readability of the questionnaire was deemed to be adequate.

9.II.ii   Assessment of participants' concordance with the sensors

Crude assessment of participants' concordance with the sensor was made by examining the number of days participants reported to have worn the sensor for. The amount of time that the sensors were not being worn during the home monitoring period was also evaluated. Accelerometer data obtained during the home monitoring period was analysed for the patients whose data was used to develop the ANN. Accelerometer data was examined for minute-long periods where no orientation change of the sensor was seen. If ten or more such minutes occurred consecutively, then this period was classified as time where the sensor was not being worn. To avoid inadvertent classification of sleep as periods where sensors were not worn, analysis of accelerometer data was restricted to waking hours. All participants were asked to wear the sensors continuously for the seven day period but were made aware that they were free to remove the sensors at any point. Participants were asked to contact the research team should they wish to discontinue the home monitoring period early. Participants were asked to record timings of any prolonged periods where they had removed the sensors. Despite the sensors being waterproof, participants were invited to remove the sensors during washing/bathing if they preferred to do so. It was estimated that if participants were to remove the sensors during washing for 40 minutes each day, then a total of 280 minutes of "non-wear" time could reasonably be expected over the course of the home monitoring period (4.8% of "waking hours").

The correlation between the proportion of time when the sensor was not worn and the sensor performance for each participant was examined using Pearson's correlation coefficient. As per previous, sensor performance was defined as the degree of correlation between sensor output and diary entries.

9.II.iii   Analysis methods

IBM SPSS statistics software was used to collate responses to Likert-type questions and to produce descriptive statistics. Likert response categories were treated as ordinal

data, since intervals between categories cannot be assumed to be of equal magnitude. The Wilcoxon rank-sum test was employed to examine for significant differences between participants' phase one and two responses.

Analysis of open questions on questionnaires can be challenging due to a lack of clarity regarding the nature of the data collected, being neither strictly qualitative nor quantitative data (O'Cathain and Thomas, 2004). It is suggested that good practice should be to consider formal analysis of such data only if it offers insights or issues not available from data produced by the closed questions. This proved to be the case in this research and thus the analysis method subsequently employed was content analysis. All free-text comments generated by the questionnaire were transcribed verbatim and read by a researcher (JF). Thereafter, a coding framework was developed to describe the thematic content of the comments. Comments were categorised by theme, sentiment (i.e. positive or negative) and by study phase (CARU or home). A second researcher (KGl) who had no prior involvement in this project, but had extensive experience in qualitative research, also performed this process. This researcher was provided with the transcript of participants' comments but was blinded to the coding framework produced. Subsequently the researchers met to compare analyses and consensus opinion was reached on the most appropriate content analysis themes to describe the data captured.

9.III Results

9.III.i Assessment of participants' concordance with the sensors

All 34 participants wore the sensors for the duration of phase one. 32 of the 34 participants reported wearing the sensors for the complete seven day home monitoring period. Two participants did not complete the entire home monitoring period; one withdrew after four days (ATYY: discomfort wearing sensor) and the other after five days (JNQW: unwell). Based on participants' reports alone, the sensors were worn for 233 of a potential 238 days (97.9%).

Accelerometer data obtained during the home monitoring period was analysed for the 27 patients whose data was used to develop the ANN; excluded participants were those in whom sensor failure or extremely low diary completion rate occurred. To avoid inadvertent classification of sleep as periods where sensors were not worn, analysis of accelerometer data was restricted to "waking hours". Variable engagement with patient-completed sleep diaries precluded the use of this data to define "waking hours" for each participant on each day; consequently "waking hours" were defined as 0800 – 2200 (5,880 minutes). The mean duration of time during home monitoring "waking hours" where the sensors were not worn was 228.2 minutes (SD=385.3). The duration of home monitoring "waking hours" where the sensors were not worn is presented for each participant in Table 22 below along with the time expressed as a percentage. The large standard deviation value is in part explained by the results seen for participant ATYY, which represent a clear outlier. ATYY discontinued home monitoring after 4 days citing sensor discomfort; calculation of the percentage of waking hours during which the sensor was not worn was therefore corrected for this participant (i.e. total possible waking hour wear time was 3,360 minutes as opposed to 5,880). When ATYY was excluded, the mean duration of time during home monitoring "waking hours" where the sensors were not worn was 159.7 minutes (SD=150.9).

Table 22: "Non-wear" time during home monitoring period for each study participant

Patient ID	Total duration sensor not worn (minutes)	% of monitoring period "waking hours" during which sensor not worn
GLXN	0	0.00
QXZV	10	0.17
RNSY	39	0.66
XHQI	40	0.68
JHWF	42	0.71
REQY	49	0.83
MUCL	52	0.88
PQEP	61	1.04
MZGE	70	1.19
GHRS	79	1.34
WDSJ	84	1.43
UVTR	84	1.43
LAPC	111	1.89
MXRL	123	2.09
JKVJ	138	2.35
SDEG	157	2.67
NRWL	173	2.94
XKVO	199	3.38
PIMT	202	3.44
UXSU	209	3.55
CVUL	210	3.57
TXOJ	224	3.81
QXLL	243	4.13
UGNK	450	7.65
BRCN	507	8.62
OEQV	596	10.14
ATYY*	2008	59.70
Mean	228.15	4.83
St. Dev.	385.26	11.25
Mean**	159.69	2.72
St. Dev**	150.90	2.57

*denominator corrected for early withdrawal from home monitoring period

**ATYY excluded

The correlation between the proportion of time when the sensor was not worn and the sensor performance for each participant was examined using Pearson's correlation coefficient. No significant correlation between the proportion of time when the sensor was not worn and the sensor performance was evident (r= -0.083; p=0.680). A scatter

plot of this data is presented in Figure 53 below. Note the dashed line which indicates the anticipated percentage of 'non-wear time based on the previously stated estimate of 40 minutes per day.

Figure 53: Sensor performance and percentage of non-wear time
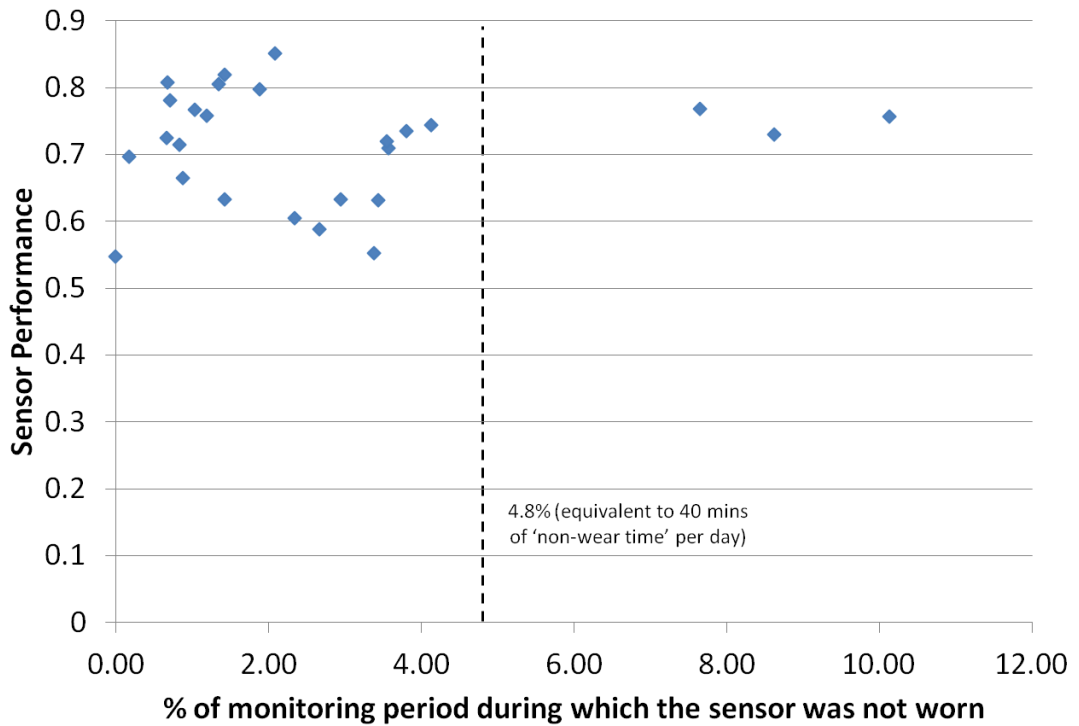
9.III.ii Likert-type questions

All 34 participants completed the questionnaire after both study phases. As described above two participants did not complete phase two but both completed the phase two questionnaire at the point at which they discontinued wearing the sensors. Of a possible 612 responses to the Likert-type questions from both phases, 608 (99.3%) were deemed valid (i.e. participant had selected one category). Of the four invalid responses, three were left blank and one included the selection of multiple categories. The frequency of responses to each question after both the CARU and home monitoring phases of the study are displayed in Table 23 below.

Table 23: Frequency of responses to patient questionnaire after CARU and Home phases

Question		Frequency of response				
		Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
1	CARU	13	21	0	0	0
	HOME	4	28	1	0	1
2	CARU	12	21	1	0	0
	HOME	3	25	3	3	0
3	CARU	0	0	3	19	12
	HOME	0	1	4	15	14
4	CARU	0	0	0	17	16
	HOME*					
5	CARU	15	18	1	0	0
	HOME	9	23	1	1	0
6	CARU	0	0	5	18	11
	HOME	0	1	6	18	9
7	CARU	4	16	5	8	1
	HOME	5	13	7	6	2
8	CARU	8	19	5	1	0
	HOME	5	24	2	2	0
9	CARU	10	23	1	0	0
	HOME	7	22	3	2	0

*question excluded as not of relevance to home phase of study

Analysing questionnaire responses after the home monitoring period revealed that 32/34 participants (94.1%) agreed that the sensor looked like it was well made [item 1] and 28/34 (82.4%) agreed that the sensor was comfortable [item 2]. After a week of wearing the sensors at home, 29/33 respondents (85.3%) agreed that the sensors were easy to take on and off (one participant left this question blank) [item 8]. 1/34 participants (2.9%) reported that the sensor felt heavy on their arm [item 3]. 1/34 (2.9%) participants stated that they would have preferred to have kept a regular diary of their symptoms for a week, as opposed to wearing the sensors [item 6]; this was the participant who withdrew after 4 days due to sensor discomfort. 32/34 (94.1%) agreed that they would be happy to wear the sensors at home [item 5] and 29/34 (85.3%) agreed that they would be happy to wear the sensors in public [item 9]. When asked whether they would be more likely to wear the sensors if it were incorporated into a working wrist-watch [item 7], 18/34 (52.9%) agreed, 8/34 (23.5%) disagreed with 7/34

(20.6%) indicating a neutral response. Statistical analysis was performed to examine for significant differences between participants' CARU and home responses, and is presented in Table 24 below. Item 4 ("Performing the assessments was made more difficult by wearing the sensor) was not analysed for the home phase of the study since there were no clinical assessments for the sensor to potentially interfere with.

Table 24: Examining for differences in questionnaire responses between phases by questionnaire items

Item	Z value	p value
1	-2.970	0.003
2	-3.231	0.001
3	<0.001	>0.999
4		
5	-1.999	0.046
6	-1.387	0.166
7	<0.001	>0.999
8	-0.254	0.799
9	-1.889	0.059

The responses to the following items were found to display a statistically significant difference between study phases: item 1: "The sensor looks like it is well made" (Z=-2.970, p=0.003); item 2: "The sensor is comfortable to wear" (Z=-3.231, p=0.001); item 5: "I would be happy to wear the sensor around the house" (Z=-1.999, p=0.046). The responses to all other items showed no statistically significant difference in their responses between phases. For those items where a statistically significant difference was demonstrated between CARU and home phase responses, further analysis was performed looking at the magnitude of category change for each participant. This data is presented in Table 25 below.

It is evident from Table 25 that the majority of participants responding to items 1, 2 and 5 did not change their response between study phases (24/34 (70.6%), 19/34 (55.9%) and 24/34 (70.6%) respectively). Of those participants that did change their response, the majority did so by only one category and typically this was towards a more negative response. Extreme changes in response (≥2 categories) were rare and

were only expressed by 2/34 (5.9%) of participants to Item 1, 3/34 (8.8%) to Item 2 and 2/34 (5.9%) to Item 5.

Table 25: Examining frequency of category change for items with a statistically significant change in responses between study phases

Change in response		Frequency (%)		
		Item 1	Item 2	Item 5
More positive ->	+4	0 (0)	0 (0)	0 (0)
	+3	0 (0)	0 (0)	0 (0)
	+2	0 (0)	0 (0)	0 (0)
	+1	0 (0)	1 (2.9)	2 (5.9)
No change	0	24 (70.6)	19 (55.9)	24 (70.6)
<- More negative	-1	8 (23.5)	11 (32.4)	6 (17.6)
	-2	1 (2.9)	2 (5.9)	2 (5.9)
	-3	1 (2.9)	1 (2.9)	0 (0)
	-4	0 (0)	0 (0)	0 (0)

For these three items, further analysis was undertaken following contraction of the 5-point Likert scale into a 3-point scale: strongly agree and agree were combined to 'agreement'; strongly disagree and disagree to 'disagreement'; neither agree nor disagree remained unchanged. Analysis using the 3-point scale found no statistically significant change in participants' responses between study phases for items 1 and 5 (p=0.180 and 0.414 respectively). A statistically significant decrease (towards less agreement) in the responses to item 2 (the sensor is comfortable to wear) was evident (p=0.023)

9.III.iii Free-text Feedback

13 participants (38.2%) provided free-text feedback in the questionnaire administered immediately after the CARU phase of the study. 18 participants (52.9%) provided free-text feedback in the questionnaire completed after the home monitoring period. In total, 25 different participants (73.5% of study cohort) provided free-text feedback on at least one occasion during the study.

Content analysis was independently undertaken by two different researchers who subsequently met to discuss their allocation of themes (JF and KGl). Similar themes, and similar allocation of comments to these themes, were evident. Discussion between the two researchers led to a consensus decision on the themes emerging from the free-text data, and the allocation of participants' comments to each. The over-arching themes identified and agreed upon were: 'Appearance', 'Comfort' and 'Useability' (Table 26, Table 27 and Table 28 respectively). 'Appearance' was sub-divided into 'Physical properties' and 'Wearing in public'. 'Comfort' and 'Useability' were sub-divided according to whether their sentiment was deemed to be 'Positive' or 'Negative'. Participants' comments for which no specific theme was identifiable were categorised under 'General comments' (Table 29).

Table 26: Content Analysis – Appearance

APPEARANCE			
Patient ID	Phase	Comment	
GHRS	Home	"Would prefer it to be a little smaller and with watch face as keep thinking it was a watch I was wearing.".	Physical properties
UVTR	Home	"The only problem was that I kept looking to find the time!"	
MZGE	Home	"Wore it for a week, did not cover it up"	Wearing in public
GHRS	Home	"I would not like to wear in warm summer months as more noticeable to people and questions".	
GHRS	CARU	"Happy to wear (in public) but would not like members of public questioning what it is for as illness is private".	

Table 27: Content Analysis - Comfort

		COMFORT	
Patient ID	**Phase**	**Comment**	
CGLT	CARU	"It feels no different to wearing a watch"	
QXLL	CARU	"Feels comfortable, don't mind having it on"	
MXRL	Home	"Found the sensor quite comfortable to wear"	+ve
QXZV	Home	"Very comfortable to wear – just like wearing a watch!"	
MZGE	Home	"No problem. Forgot it was there"	
QXLL	Home	"No problems with sensor, almost forgot it was on."	
UGNK	CARU	"Strap would be more comfortable if leather".	
LAPC	Home	"Velcro slightly uncomfortable".	
FRMQ	Home	"The sensor is slightly scratchy especially when wearing a watch as well"	-ve
ATYY	Home	"It is always on your skin, also, when it gets wet it is very uncomfortable to wear generally and I don't like it very much"	
GHRS	Home	"Comfortable to wear, however, after a week of constant wear feeling a little irritating."	
JKVJ	Home	"If strap were made more comfortable would make wearing very easy".	

178

Table 28: Content Analysis - Useability

Patient ID	Phase	Comment	
		USEABILITY	
MXRL	Home	"I had expected it to interfere with my everyday life but that did not happen"	
WDSJ	CARU	"Someone will do [put on/off] it for me"	
MUCL	CARU	"The sensor was very easy to have on".	+ve
MUCL	Home	"The sensor I found easy to wear".	
ATYY	Home	"The sensor is easy to take on and off"	
PQEP	Home	"Because I have small wrists the sensors were swinging around and it was difficult to keep them in the upright position. After a couple of hours I used some surgical tape to stick it down where the strap fastens underneath – they are in the same position after 1 week including daily showers."	
GHRS	Home	"Found it restricts you wearing tight sleeves on clothes".	
MUCL	Home	"I was a little nervous having a shower".	
WDSJ	Home	"The left, blue sensor did not always stay securely in position and so needed occasional readjustment".	-ve
FRMQ	Home	"I removed them whilst having a bath/shower because they became soggy"	
FRMQ	Home	"For someone with a tremor they are a little awkward"	
BRCN	Home	"Maybe stronger pins in the sensor would help, one came out"	
NRWL	Home	"Sensor a little awkward to fasten strap... (illegible) when feeling off"	

Table 29: Content Analysis - General Comments

Patient ID	Phase	Comment
		GENERAL COMMENTS
YQSE	CARU	"Technology is a wonderful experience".
UVTR	CARU	"B***** clever!"
XKVO	CARU	"The sensor is no problem to me at all".
GLXN	CARU	"I think the working wrist watch is a good idea".
REQY	CARU	"I am happy to wear the sensors to help the study for Parkinson's research".
UXSU	CARU	"I hope the sensor works well in giving essential information in the research of Parkinson's disease".
BRCN	CARU	[Regarding question 7] "no watch".
OEQV	Home	"Sorry the charts [diary] may be a little incomplete but I was really poorly".
JNQW	Home	"The questionnaire [diary] not completed – was ill with sickness and diarrhoea".
CJVB	Home	"I'm looking forward to any feedback"

179

9.IV Discussion

Participant reported concordance with the Axivity AX3 sensor, during both phases of the study, was high. Of the two participants who failed to complete the week-long period of home-monitoring only one withdrew because they were no longer willing to wear the sensor; the other withdrew due to an intercurrent illness.

Formal evaluation of 'wear-time' provides evidence that the majority of participants were highly concordant with sensor wearing. The mean amount of time per day where the sensors were not worn, excluding ATYY, was approximately 22 minutes. This period is less than half of the one hour epoch used for analysis and is therefore unlikely to have influenced results. Only four of the participants exceeded the estimated amount of 'non-wear time' (40 minutes per day; 4.8% of "waking hours"): UGNK, BRCN, OEQV and ATYY. ATYY withdrew early citing sensor discomfort and hence the poor concordance. Free-text feedback provided insight into why some participants were less concordant with sensor wearing. BRCN cited problems with the pins securing the sensor to the strap which may have contributed to an increase in 'non-wear time':

"Maybe stronger pins in the sensor would help, one came out straight away" [BRCN: Home. Useability; negative]

Feedback from OEQV suggested that this participant had been unwell during the home monitoring period but that diary completion had in fact been more problematic than wearing the sensors:

"Sensors were no problem. Sorry the charts may be a little incomplete but I was really poorly." [OEQV: Home. General comments]

Another participant, JNQW, became unwell with diarrhoea and vomiting during the home monitoring period and subsequently withdrew. Concordance with diary completion for this participant was extremely low, with only 18.3% of diary responses being classifiable. Accelerometer data for this participant was examined using the wear-time algorithm. This revealed that JNQW had not worn the sensor for 1583 minutes during the home monitoring period. This equates to 37.7% of the five day monitoring period completed by the participant prior to withdrawing. Whilst concordance with both diaries and sensors by this participant were poor, it is evident

that in terms of time, sensors were found to have superior concordance than diaries. We had anticipated that the sensor would perform less well in participants who were less concordant with wearing of the sensor; this was not the case however. No relationship between sensor performance and duration of 'non-wear' time was seen. This finding suggests that the additional data provided by patients more concordant with monitoring, conferred no benefit in terms of sensor performance on an individual level. The optimal duration of monitoring is currently unknown but in light of this finding, future work involving the sensors may employ shorter periods of home monitoring.

An extremely high response rate to the questionnaire was seen; almost all Likert items were responded to (99.3%) and almost three quarters of participants (73.5%) provided free-text feedback in either questionnaire. An important finding was that after phase one of the study, no respondent to the questionnaire agreed with the statement that "performing the assessments was made more difficult by wearing the sensor" (item 4). This is a reassuring result as any suggestion that the wearing of the sensor interfered with activities performed during phase one would raise doubts over the validity of the data captured. Questionnaire items 1 and 3 were designed to evaluate participants' views regarding some of the physical aspects of the sensor. The overwhelming majority of participants, after both phases of the study, agreed that the sensors "looked like they were well made". 1 participant, after the home phase, strongly disagreed with this item but did not provide further free-text feedback. Analysis of the responses to this item revealed a statistically significant change (towards less agreement) between phases (p=0.003). Table 25 displays the magnitude and frequency of change for this item, demonstrating that in fact the majority of participants showed no change in their responses. The statistically significant changes seen can be in part attributed to a limitation of the Wilcoxon rank-sum test, which excludes from analysis unchanged responses, examining only those that exhibited change. No significant difference was evident for this item when further analysis was undertaken following contraction of the 5-point Likert scale into a 3-point scale.

The majority of participants disagreed with item 3 ("the sensor feels heavy on my arm") after both phases, with no statistically significant difference in responses between phases. Across both phases only one participant reported that the sensor felt heavy on their arm. One participant did provide free-text feedback regarding the dimensions of the sensor:

"Would prefer it to be a little smaller..." [GHRS: Home. Appearance; physical properties]

Another participant commented that wearing of the sensor interfered with items of clothing:

"Found it restricts you wearing tight sleeves on clothes" [GHRS: Home. Useability; negative]

This quote suggests that the profile of the device was too large to allow tightly fitting sleeves to pass over the sensors and to be fastened distal to the sensor. It is evident, in quotes from the same participant presented later, that difficulty in covering the sensor underneath clothing has potential implications in terms of participant body image.

In terms of the functionality of the sensor, participants were asked to complete Likert items pertaining to the ease of putting on and removing the sensors (item 8), and the comfort of wearing them (item 2). The responses to Likert-style questions suggest that, for the majority, putting the sensors on and taking them off was not a problem. Some participants reported having assistance from family or carers. Some participants reported that variation in their motor status affected their ability to do this:

"For someone with a tremor they are a little awkward" [FRMQ: Home. Useability; negative]

"Sensor a little awkward to fasten strap... (illegible) when feeling off" [NRWL: Home. Useability; negative]

After phase one of the study all but one of the participants agreed that the sensor was comfortable to wear. The majority (28/34) still agreed with this statement after phase two of the study, but there was a statistically significant change in responses (towards less agreement) between study phases. Examining the descriptive statistics demonstrates that many participants did not change their views and of those that did, the majority did so by only one category (typically from 'strongly agree' to 'agree'). A small, but significant shift of opinion towards less agreement was maintained when

182

analysis was performed following contraction of the 5-point Likert scale to a 3-point version. This shift in responses towards a more negative sentiment is perhaps to be expected given that participants had worn the sensor for a seven day period. A degree of obtrusiveness is inevitable with even the most well designed sensor and this may be exacerbated by more prolonged periods of wearing (Bonato, 2010). Several participants provided free-text comments supporting the notion that the sensor was comfortable with some describing how they simply forgot they were wearing it:

"No problem. Forgot it was there" [MZGE: Home. Comfort; positive]

Whilst the Likert data provides quantitative evidence to support the acceptability of the sensor in terms of its comfort, further examination of the free-text feedback suggests that some participants did experience problems in this area. In particular, the strap securing the sensor to the body was a source of problems for a number of participants. The strap was made from a synthetic nylon material (see Figure 23) with a Velcro fastener that allowed the strap to be sized to the personal preference of the wearer. Several participants found the nylon material uncomfortable, with one describing it as:

"...scratchy" [FRMQ: Home. Comfort, positive]

Some participants indicated that the comfort of the sensors declined further when they became wet:

"It is always on your skin, also, when it gets wet it is very uncomfortable to wear generally and I don't like it very much" [ATYY: Home. Comfort, negative]

One participant suggested that a leather strap would be much more comfortable. Another recurring theme from the free-text comments related to the sizing of the strap. The sensor could be attached to wrists of varying sizes using an adjustable Velcro fastener. No participants in the study had wrist circumferences that exceeded the maximum dimensions of the strap. Several participants did however report that they were unable to shorten the strap sufficiently to ensure that the sensor was tightly secured:

"Because I have small wrists the sensors were swinging around and it was difficult to keep them in the upright position" [PQEP: Home. Useability; negative]

A loosely fitting sensor may result in relative motion between the sensor and the body and consequently sensor displacement, extraneous signal artefact and declining signal accuracy (Yang and Hsu, 2010). One person reported problems with the pin securing the strap to the sensor (see Figure 25) becoming loose, which may also have adversely affected the quality of signal captured. These findings suggest that the dynamic wearability of the sensor may, for some participants, have been compromised due to problems relating to sizing.

A critical component of this section of the study is establishing whether or not participants would be willing to wear the sensor both at home and in public. Concordance data appears to support willingness to wear the sensor, but does not enable differentiation between settings. Analysis of Likert data suggests that the vast majority of participants were willing to wear the sensor both at home and in public, and widespread agreement with these items was maintained even after a week-long period of wearing the sensors. There was, however, a statistically significant difference in responses to item 5 (wearing the sensor at home), with a change towards less agreement observed after the home monitoring period. Once more however, examination of the descriptive statistics showed that the majority of responses (70%) did not change, and of those that did, the majority did so by only one category (typically from 'strongly agree' to 'agree'). No significant difference was evident for this item when further analysis was undertaken following contraction of the 5-point Likert scale into a 3-point scale. These findings are in contrast with those of Giuffrida et al., where disparity between participants' willingness to use the sensor at home and in public was demonstrated (Giuffrida et al., 2009). Examining free-text feedback provides greater detail into participants' experiences of wearing the sensor, and the extent to which the sensor was acceptable or obtrusive. One participant described how they:

"(I) had expected it to interfere with my everyday life, but that did not happen" [MXRL: Home. Useability; positive]

Another participant's feedback does however suggest that self-consciousness about wearing the sensor in public may have been a factor:

"Happy to wear (in public) but would not like members of public questioning what it is for as illness is private" [GHRS: CARU. Appearance; wearing in public]

Another participant's comments suggested a contrasting experience, perhaps reflecting this person considering wearable medical technology to be a form of "status symbol" (Korhonen et al., 2003):

"Wore it for a week, did not cover it up" [UVTR: Home. Appearance; wearing in public]

One theme that arose from the free-text feedback that had not been considered by the research team beforehand was the effect of season on wearing of the sensors:

"I would not like to wear in summer months as more noticeable to people and questions" [GHRS: Home. Appearance; wearing in public]

Participants were also asked whether keeping a symptom diary for a week would be preferable to a week of wearing the sensor (item 6). This questionnaire item enabled a direct comparison between the current gold standard home monitoring method (diaries) and our system, with participants experiencing both methods and therefore being well placed to comment. Prior to the home monitoring phase, 29/34 participants stated a preference for wearing the sensor for a week, with none doing so for the diaries. There was no significant difference in responses between phase one and two questionnaires, suggesting that participants' views did not change in light of their experiences wearing the sensor. After phase two of the study, one participant did state a preference for the diaries over the sensors; this was the participant who withdrew early from phase two [ATYY]. Comments from two other participants, presented below, illustrate how concordance with paper diaries can be influenced by a person's health and well-being during the monitoring period:

"The questionnaire [diary] not completed – was ill with sickness and diarrhoea" [JNQW: Home. General comments]

"Sorry the charts [diary] may be a little incomplete but I was really poorly" [OEQV: Home. General comments]

These findings provide strong evidence that participants preferred wrist-worn sensors over patient diaries and continued to do so despite a week long period of wearing them.

Questionnaire item 7 asked participants whether the incorporation of the sensor into a working wrist-watch would influence the likelihood of them wearing it. Interestingly a recurring theme in the free-text feedback was participants reporting that they often mistook the sensor for a watch, for example:

"(Would prefer it) with watch face as keep [sic] thinking it was a watch I was wearing" [GHRS: Home. Appearance; physical properties].

Examining the Likert data derived from item 7 shows that whilst many agreed with this statement, the agreement was far from universal. Examination of responses in terms of gender difference yielded little further information: 12/23 (52.2%) of males and 6/11 (54.5%) of females were found to have agreed with this statement; 6/23 (26.1%) of males and 2/11 (18.2%) of females disagreed with this statement. Whilst not evident in the free-text comments, anecdotal reports from some participants to the research team suggest that many of the participants disagreeing with this item did so simply because they did not normally wear a watch. In future there is the possibility that sensors such as these could be incorporated into other everyday objects, perhaps more acceptable for that particular user; jewellery, for example, might prove more acceptable to female wearers. Asada and colleagues (Asada et al., 2003) have previously reported incorporating a wearable photo-plethysmographic sensor into a ring, allowing continuous monitoring of oxygen saturations. Enabling patients to have a choice when selecting the appearance of a medical device has been shown to increase the appeal of the device, even if the choice is simply what colour the device is (Pippin and Fernie, 1997). The straps used in this research were either red or blue, to indicate which arm they were to be worn on. Using more subtle colours (black, grey, skin-tones) or simply offering participants a selection of colours from which to choose, may have further improved concordance and acceptability.

Based on the evidence presented above one can conclude that the sensor used in this research has been widely accepted by participants. Consideration of both the technical

and human dimensions involved with its wearing (Figure 51) reveals further possible explanations as to why the sensor was so widely accepted. The sensor employed in this research was worn in what can be considered an unobtrusive place (Gemperle et al., 1998). The sensor required little in the way of patient training; it was simply worn and recorded data. There was no need for participants to charge the sensor, interact with any buttons or to download any data. Participants were briefed as to the correct orientation of the device when worn, but required no teaching in how to operate the device due to its automated nature; minimising what participants need to learn about the device can help to improve concordance (Cunningham et al., 2009). As previously described, participant motivation plays a large role in the acceptance of the inevitable degree of obtrusiveness associated with the wearing of a medical device. Previous research in PD populations (Goetz et al., 2003a, Valadas et al., 2011) has identified that altruism is a common motivating factor underpinning participants' involvement in research. Several participants in this research were seemingly motivated to wear the sensors by altruistic tendencies:

"I am happy to wear the sensors to help the study for Parkinson's research" [REQY: CARU. General comments]

"I hope the sensor works well in giving essential information in the research of Parkinson's disease" [UXSU: CARU. General comments]

Participants were asked to complete the phase two questionnaire upon completion of the home monitoring period; thus ensuring seven days of wearing before their views were assessed. There is however no way to confirm that participants completed the questionnaire at this point; some may have done so before 7 days, perhaps influencing their attitudes and opinions regarding the sensor. When the study questionnaire was developed, the Likert items selected for inclusion were based on topics derived from previous research in this area (Giuffrida et al., 2009) and from discussions within the research team, which identified areas considered to be relevant that had not been addressed in previous work. It is however recognised that patients and potential research subjects can be an excellent source of items, and that they are frequently overlooked when such tools are being developed (Streiner and Norman, 2008).

Incorporating input from people with PD may have enabled the questionnaire to have been developed in a more rigorous and systematic manner. Discussion with PD patients, via focus groups or key informant interviews, would have elicited their viewpoints regarding the sensor and may have better informed the items selected for inclusion. Another potential improvement would be to have incorporated 'expansion' open questions following each Likert item on the questionnaire, essentially asking the respondent to elaborate on the response given to the closed question (O'Cathain and Thomas, 2004). For example, across both study phases only one person strongly disagreed with the statement that "the device looked well made", but this person did not provide free-text feedback explaining their reasoning. The availability of an expansion question in this situation may have provided further information, thus enabling greater understanding of the reasons underpinning the participant's response.

9.IV.i CONCLUSIONS

- After seven days of continuous wearing, the majority of the study cohort considered the sensor to be acceptable.

- Problems with comfort and dynamic wearability were identified, particularly in relation to the strap.

- Study participants indicated a preference for wearing the sensor instead of completing symptom diaries.

- Findings suggest that participants were happy to wear the sensor in both private and public spheres.

Chapter 10. Future Studies and Final Conclusions

10.I   Future Work

10.I.i   Further clinimetric evaluation of the assessment method

Scientific evaluation of an outcome measurement instrument requires the assessment of three different properties; validity, reliability and responsiveness (Hobart et al., 1996). In this work we have presented evidence for the criterion-related validity of wrist-worn accelerometry and analysis with ANNs, since significant correlation with the current gold standard, home diaries, was demonstrated. To confirm that our system enables clinically meaningful measurement of upper limb motor symptoms in PD, its reliability and responsiveness would need to be evaluated in future work. Test-retest reliability could be established by performing home monitoring with a cohort of patients on two different occasions and then examining the correlation between the data produced. Responsiveness; the ability to detect clinically significant changes, is a critical property to evaluate if the system is to be considered a useful and scientifically sound method of symptom evaluation. There are a number of possible clinical scenarios in which this property could be evaluated. Firstly, newly diagnosed patients could be assessed before and after instigation of treatment, although patients in the early stages of PD may exhibit only mild motor symptoms which may render detection more challenging. Furthermore, consideration must be given to the ethics of delaying initiation of medication for the purposes of research. Alternatively, patients could be assessed before and after adjustment of their medication regimen, be it an increase in the dose or frequency of their anti-parkinsonian medication or the addition of agents such as amantadine to reduce dyskinesia severity.

10.I.ii   Examination of only waking hours

One possible future approach to data analysis that might improve performance of the ANN would be to entirely exclude sleep from the data set. The ANN in this work was trained on continuous data captured over a seven day period. Whilst this did provide

huge amounts of data it can be argued that sleep contributed little to ANN performance. Limited correlation was seen between the amount of time spent sleeping as measured by diaries and by the ANN. Our data suggests that periods of sleep in PD patients are unlikely to simply represent periods with little or no movement. Analysis of patient completed sleep diaries showed a mean number of awakenings per night of 1.3 (SD=0.9) per participant. Responses to the PDSS-2 question "Did you get up at night to pass urine?" showed that 28/34 participants (82.4%) answered 'very often' (6-7 days a week). Furthermore, 12/34 participants (35.3%) reported problems with restless arms/legs disrupting their sleep on at least 2-3 days per week. 12/34 participants (35.3%) reported problems with disturbed sleep due to an urge to move their legs or arms on at least 2-3 days per week. In addition, REM sleep behaviour disorder is well recognised as a feature of PD and can produce violent, involuntary movements of the arms and legs during sleep (Chaudhuri et al., 2005).

The net result is that inclusion of sleep may in fact adversely affect ANN performance. Periods labelled as sleep by participants may in fact include spells of voluntary and involuntary movements that may ultimately render differentiation between sleep and the other disease states more difficult. The proportions of time spent in each disease state during both CARU and home phases, with only waking hours considered, is presented in Figure 54. Whilst the disparity between amounts of dyskinesia still exists, it is evident that in terms of proportions of each disease state within the data compendia, the two sets of data are more similar than when sleep is included (Figure 45). In future studies participants could be instructed to put the sensor on when they wake in the morning and to remove the sensor before going to bed, thus capturing only waking hour data.

Figure 54: Proportions of time spent in each disease state, for all participants, during home phase and CARU phase, when only waking hours are considered
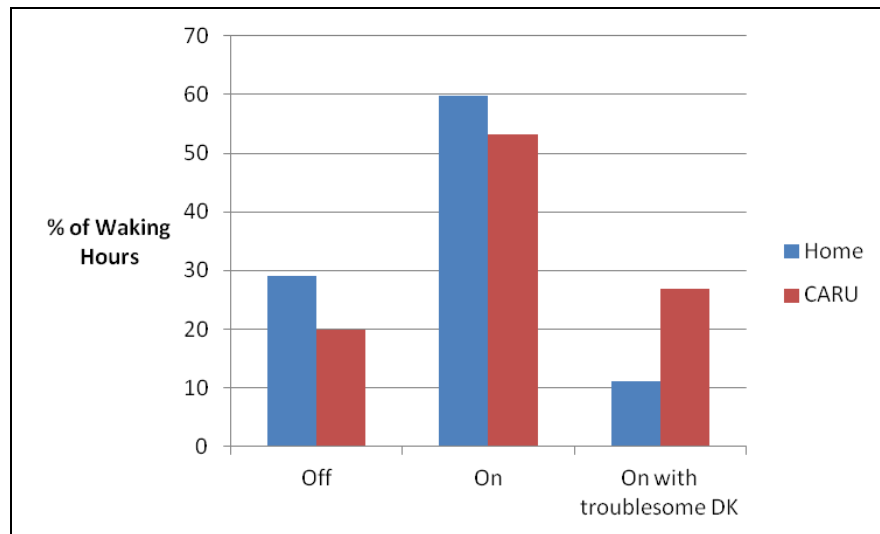
10.I.iii Further development of the ANN

Work is ongoing with the computer science team to improve the performance of the ANN. Many of the changes to the structure of the ANN that are being implemented are complex, computer science-based interventions that are beyond the remit of this clinical thesis. From a clinical viewpoint, one potential change in the analytical approach that may yield improved performance relates to disease phenotype. It is well recognised that PD is not a homogenous disorder. Identification of the differing phenotypes: TD and PIGD, is routine in clinical practice. Furthermore, dividing a PD cohort into TD and PIGD sub-groups has been used in a variety of clinical, aetiological and genetic studies (Stebbins et al., 2013). In the development of the ANN used in this thesis, one of the features employed related specifically to disease phenotype. This feature simply captured a binary decision between TD and PIGD phenotype (i.e. 0 or 1 in mathematical terms). It is evident from the genesis of the ANN that the addition of this feature improved the network's performance because if it had not, it would have been dispensed with during the training phase. It is thought that research into the differing phenotypes of PD may shed new light on the biological and pathological processes involved, and may ultimately lead to tailored treatment strategies (van Rooden et al., 2011). There is therefore an argument to suggest that a separate ANN

191

could be constructed for each disease phenotype. As previously discussed, the disparity in symptoms between a TD and a PIGD patient in the off state translates into vastly different accelerometer data (see Figure 47). The difference between the two sub-groups is further emphasised by the fact that of the 12 TD participants in the study, none displayed dyskinesia. Construction of an ANN for each phenotype, each drawing only on data derived from participants with that particular phenotype, may ultimately produce improved ANN performance.

10.I.iv The Future

The application of artificial neural networks (ANNs) to medical decision making and patient management is rapidly expanding (Dayhoff and DeLeo, 2001). Figure 55 displays the results of a Medline literature search for the phrase "Artificial neural networks". It can be seen that the number of such publications has risen steadily since their first appearance in 1990.

Figure 55: Number of Medline Citations containing "Artificial Neural Networks" (1990-2012)
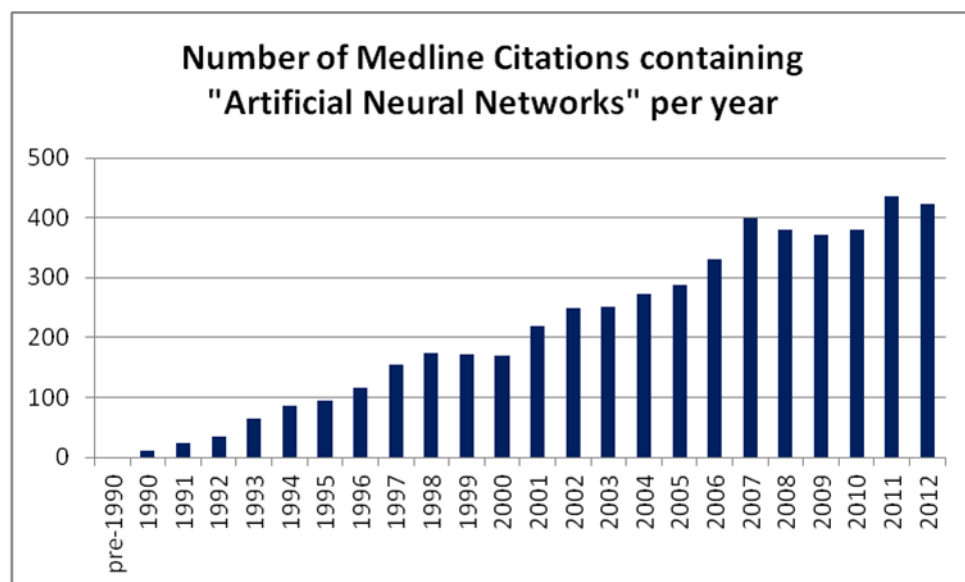
The sensors employed in this work can be termed "data logging" devices; they simply store data that can later be downloaded (Ni Scanaill et al., 2006). Such devices mean that feedback cannot be given to the wearer during the monitoring period and from a

clinical viewpoint it means that an intervention cannot be made until the process of downloading and analysing the data has occurred. The future is likely to see the use of high-speed communication networks to enable rapid transfer of data from the device to the clinical team during the monitoring period, so called "data forwarding" devices (Ni Scanaill et al., 2006). This technology is currently available, but to incorporate it into the sensor employed in this work would place greater demands on both battery life and the device's size, both of which may have a detrimental effect on patient concordance.

Computer technology is forecast to continue to develop rapidly; Moore's law, which states that computer processing power doubles every two years, is expected to hold true until 2050 (Schaller, 1997). As computer technology advances it may become possible for worn sensors to function as "data processing" sensors (Ni Scanaill et al., 2006). Such technology would enable automated, real-time evaluation of the data captured and thus allow rapid feedback to the wearer. It may be possible to give real-time advice to patients on their medication regimens in response to such data; personalised, tailored medication regimens may improve the efficacy of anti-parkinsonian treatments (Binkley, 2003). It has been hypothesised that such technology could be integrated into medication delivery systems such as those used by patients currently taking apomorphine or duodopa (Rodríguez-Molinero et al., 2009). Administration rates of such medications could be automatically adjusted in response to the wearer's disease state as detected by the sensor system.

10.II Final Conclusions

This thesis has presented a system by which upper limb motor symptoms in PD can be evaluated. The system employed two wrist-worn accelerometers that were worn by participants for a seven day period at home. Accelerometer data was analysed using advanced computational algorithms known as artificial neural networks (ANNs). ANNs were developed from a data compendium that was derived from the home monitoring period and were designed to provide outputs in terms of the wearer's disease state. Training of the ANN was undertaken using patient-defined disease states; these were

determined from the patient-completed symptom diaries. The performance of the ANN was evaluated against both patient and clinician defined disease states.

Patient completed-symptom diaries were, on the whole, completed with a high proportion of classifiable responses. It was shown that the proportion of classifiable responses declined in older, more cognitively impaired participants. Delay in initiation of the home monitoring period occurred for some participants, raising concerns that recall of the diary-orientation session undertaken at CARU would be impaired. There was however, no evidence that delay between study phases translated into diaries with fewer classifiable entries. Limited conclusions can be drawn on the validity of diary responses since no external validation of responses was available.

Across the study cohort a strong correlation was seen between ANN output and diary-recorded disease status. Performance of the ANN was however variable. Visual representations of home monitoring periods revealed, for many participants, disparity between ANN output and diary reported disease state. No factors (demographic, clinical or methodological) were identified that allowed differentiation between those in whom the ANN performed better or worse. In summary, accelerometers and the ANN employed in this study lacked the ability to provide 'real-time' evaluation of PD motor symptoms. Failure to do so likely relates to methodological flaws that restricted both the volume and quality of training data, and to the inherent subjectivity of disease state recognition.

Attempts were made to validate the ANN against clinician rated disease status that took place during the CARU phase; performance of the ANN was however sub-optimal. Disparity between the CARU and home data sets, both in terms of the proportions of time spent in each disease state and in terms of the activities performed by participants, may have contributed to poor ANN performance.

Comparison of the proportion of time spent in each disease state, as measured by the ANN, diaries and the MDS-UPDRS, were then compared. Limited correlation was seen between MDS-UPDRS and diary derived measures of disease state proportions, calling into question the validity of this MDS-UPDRS item. A strong, significant correlation was

seen between the measures produced by the home diaries and the ANN. This finding provides evidence for the criterion-related validity of the system employed in this work. It also suggests that accelerometers and ANNs can provide a measure of the proportions of time spent in each disease state that is comparable to that reported by patients via the current gold standard, patient-completed symptom diaries. Further evaluation of this assessment method is required.

To the best of the author's knowledge this is the most in-depth evaluation of the acceptability of body-worn sensors in PD patients. This is an area that has been neglected in the literature previously, with little or no regard given to the patient experience of wearing such a device. The acceptability of body-worn sensors is critical to their applicability as a tool for home monitoring. This thesis presents evidence supporting the acceptability of our system amongst a PD population. Furthermore it demonstrates that reassessment of participants' views after the week long monitoring period did not result in a shift towards negative views on the sensors following prolonged wearing. The majority of participants reported a willingness to wear the sensor in both private and public spheres. Problems were identified in relation to the strap, resulting in an adverse effect on comfort and dynamic wearability for some participants. These design issues would need to be addressed in future iterations of the sensor to improve acceptability further. Of note, study participants expressed a preference for wearing the sensor as opposed to completing symptom diaries.

The MDS-UPDRS is the current gold standard PD assessment tool. Despite having undergone thorough clinimetric evaluation the MDS-UPDRS retains an element of subjectivity at the level of the individual passing judgment on a given item. The current gold standard for home monitoring of PD symptoms, the symptom diary, remains shrouded in subjectivity. There is therefore a great need for an objective assessment tool in PD. The ubiquity, simplicity and low cost of accelerometers make them an attractive potential option. The ability of accelerometers to yield huge quantities of data makes them an ideal source of data for ANNs, which require large amounts of data to be provided for their development and training. Algorithms such as those employed by ANNs are already in use in many walks of life (Steiner, 2012) and their

use in healthcare is projected to expand rapidly in the next decade. The heterogeneity of PD and the degree of symptom variation between different disease phenotypes does present challenges to the development of such systems. There is a growing argument for dichotomising PD cohorts by disease phenotype during research work.

The ageing population and the increasing burden of chronic diseases such as PD, present healthcare services with huge challenges. Appetite exists amongst Western governments to increase provision of care in the community and remote monitoring of patient symptoms using body-worn sensors may help to achieve this. Demonstration of the validity of such an assessment tool is central to its integration into clinical practice. However, a clinician-centric approach that fails to consider the acceptability of the system to the wearer must be avoided. Evaluation of the patient experience with such technology must not be neglected.

Chapter 11. References

1997. *European Parkinson's Disease Association Charter* [Online]. Available: http://www.epda.eu.com/ru/archive/past-projects-completed/epda-charter/ [Accessed 20th April 2013.

2004. Levodopa and the Progression of Parkinson's Disease. *New England Journal of Medicine,* 351, 2498-2508.

2006. Parkinson's Disease: Diagnosis and management in primary and secondary care [CG35]. *In:* (ed.) *National Institute for Health and Clinical Excellence.* London: National Institute for Health and Clinical Excellence

2009a. *Initial National Priorities for Comparative Effectiveness Research*, The National Academies Press.

2009b. Parkinson's prevalence in the United Kingdom. London: Parkinson's UK.

2010. Diagnosis and pharmacological management of Parkinson's disease: A national clinical guideline [113]. *Scottish Intercollegiate Guidelines Network.* Edinburgh: Scottish Intercollegiate Guidelines Network.

2011. Consultant physicians working with patients: The duties, responsibilities and practice of physicians in medicine. *In:* PHYSICIANS, R. C. O. (ed.) 5th ed. London: RCP.

2012a. *Ageing in the UK Datasets, 2011 Pensions Act update - January 2012* [Online]. Office for National Statistics. Available: http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-248389 [Accessed 2nd July 2012 2012].

2012b. *Engineering: the nature of problems* [Online]. Available: http://labspace.open.ac.uk/mod/resource/view.php?id=420014 [Accessed 4th September 2012].

2012c. The Mandate: A mandate from the Government to the NHS Commissioning Board: April 2013 to March 2015. *In:* HEALTH, D. O. (ed.). London: Williams Lea.

2012d. *Parkinson's prevalence in the United Kingdom* [Online]. Parkinson's UK. Available: http://www.parkinsons.org.uk/pdf/ParkinsonsprevalenceUK.pdf [Accessed 2nd July 2012 2012].

2012e. Population Ageing in the United Kingdom, its Constituent Countries and the European Union. *In:* STATISTICS, O. F. N. (ed.).

2013a. *Axivity AX3 3 Axis Logging Accelerometer: Data Sheet* [Online]. Available: http://axivity.com/v2/products/AX3/AX3-1.7-Datasheet.pdf [Accessed 21st December 2013.

2013b. *Axivity AX3 3 Axis Logging Accelerometer: User Guide* [Online]. Available: http://axivity.com/v2/products/AX3/AX3-User-Guide.pdf [Accessed 19th February 2013.

2013c. *Longman English Dictionary* [Online]. Available: http://www.ldoceonline.com/dictionary/user-friendly [Accessed 16th August 2013.

ABBOTT, R. D., PETROVITCH, H., WHITE, L. R., MASAKI, K. H., TANNER, C. M., CURD, J. D., GRANDINETTI, A., BLANCHETTE, P. L., POPPER, J. W. & ROSS, G. W. 2001. Frequency of bowel movements and the future risk of Parkinson's disease. *Neurology,* 57, 456-462.

AGOSTINO, R., BERARDELLI, A., ACCORNERO, N. & MANFREDI, M. 1992. Sequential arm movements in patients with Parkinson's disease, Huntington's disease and dystonia. *Brain,* 115, 1481-1495.

AGOSTINO, R., CURRA, A., GIOVANNELLI, M., MODUGNO, N., MANFREDI, M. & BERARDELLI, A. 2003. Impairment of Individual Finger Movements in Parkinson's Disease. *Movement Disorders,* 18, 560-565.

ALLEN, D. P., PLAYFER, J. R., ALY, N. M., DUFFEY, P., HEALD, A., SMITH, S. L. & HALLIDAY, D. M. 2007. On the Use of Low-Cost Computer Peripherals for the Assessment of Motor Dysfunction in Parkinson's Disease - Quatification of Bradykinesia Using Target Tracking Tasks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* 15, 286-294.

ALY, N. M., PLAYFER, J. R., SMITH, S. L. & HALLIDAY, D. M. 2007. A novel computer-based technique for the assessment of tremor in Parkinson's disease. *Age & Ageing,* 36, 395-399.

ANDERSON, R. C., FAGAN, M. J. & SEBASTIAN, J. 2001. Teaching students the art and science of physical diagnosis. *The American Journal of Medicine,* 110, 419-423.

ANTONINI, A., ISAIAS, I. U., CANESI, M., ZIBETTI, M., MANCINI, F., MANFREDI, L., DAL FANTE, M., LOPIANO, L. & PEZZOLI, G. 2007. Duodenal levodopa infusion for advanced Parkinson's disease: 12-month treatment outcome. *Movement Disorders,* 22, 1145-1149.

ASADA, H. H., SHALTIS, P., REISNER, A., SOKWOO, R. & HUTCHINSON, R. C. 2003. Mobile monitoring with wearable photoplethysmographic biosensors. *Engineering in Medicine and Biology Magazine, IEEE,* 22, 28-40.

ASCHERIO, A., ZHANG, S. M., HERNÁN, M. A., KAWACHI, I., COLDITZ, G. A., SPEIZER, F. E. & WILLETT, W. C. 2001. Prospective study of caffeine consumption and risk of Parkinson's disease in men and women. *Annals of Neurology,* 50, 56-63.

BACHER, M., SCHOLZ, E. & DIENER, H. C. 1989. 24 Hour continuous quantification based on EMG recording. *Electroencephalography and Clinical Neurophysiology,* 72, 176-183.

BAIN, P. G., FINDLEY, L. J., ATCHISON, P., BEHARI, M., VIDAILHET, M., GRESTY, M., ROTHWELL, J. C., THOMPSON, P. D. & MARSDEN, C. D. 1993. Assessing Tremor Severity. *Journal of Neurology, Neurosurgery and Psychiatry,* 56, 868-873.

BAJAJ, N. P. S. 2010. SWEDD for the General Neurologist. *Advances in Clinical Neuroscience and Rehabilitation,* 10, 30-31.

BAJAJ, N. P. S., KNÖBEL, M., GONTU, V. & BAIN, P. G. 2011. Can spiral analysis predict the FP-CIT SPECT scan result in tremulous patients? *Movement Disorders,* 26, 699-704.

BAJAJ, N. P. S., WANG, L., GONTU, V., GROSSET, D. G. & BAIN, P. G. 2012. Accuracy of subjective and objective handwriting assessment for differentiating Parkinson's disease from tremulous subjects without evidence of dopaminergic deficits (SWEDDs): an FP-CIT-validated study. *Journal of Neurology,* 259, 2335-40.

BENAMER, H. T. S., PATTERSON, J., GROSSET, D. G., BOOIJ, J., DE BRUIN, K., VAN ROYEN, E., SPEELMAN, J. D., HORSTINK, M. H. I. M., SIPS, H. J. W. A., DIERCKX, R. A., VERSIJPT, J., DECOO, D., VAN DER LINDEN, C., HADLEY, D. M., DODER, M., LEES, A. J., COSTA, D. C., GACINOVIC, S., OERTEL, W. H., POGARELL, O., HOEFFKEN, H., JOSEPH, K., TATSCH, K., SCHWARZ, J. & RIES, V. 2000. Accurate differentiation of parkinsonism and essential tremor using visual assessment of [123I]-FP-CIT SPECT imaging: The [123I]-FP-CIT study group. *Movement Disorders,* 15, 503-510.

BENNETT, D. A., SHANNON, K. M., BECKETT, L. A., GOETZ, C. G. & WILSON, R. S. 1997. Metric properties of nurses' ratings of parkinsonian signs with a modified Unified Parkinson's Disease Rating Scale. *Neurology,* 49, 1580-1587.

BEUTER, A., BARBO, E., RIGAL, R. & BLANCHET, P. J. 2005. Characterization of subclinical tremor in Parkinson's disease. *Movement Disorders,* 20, 945-950.

BEUTER, A., DE GEOFFROY , A. & CORDO, P. 1994. The measurement of tremor using simple laser systems. *Journal of Neuroscience Methods,* 53, 47-54.

BIGLAN, K. M., VOSS, T. S., DEUEL, L. M., MILLER, D., EASON, S., FAGNANO, M., GEORGE, B. P., APPLER, A., POLANOWICZ, J., VITI, L., SMITH, S., JOSEPH, A. & DORSEY, E. R. 2009. Telemedicine for the care of nursing home residents with Parkinson's disease. *Movement Disorders,* 24, 1073-1076.

BINDER, S., DEUSCHL, G. & VOLKMANN, J. 2009. *Effect of cabergoline on Parkinsonian tremor assessed by long-term actigraphy*, European Neurology. 61 (3) (pp 149-153), 2009. Date of Publication: February 2009.

BINKLEY, P. F. 2003. Predicting the potential of wearable technology. *Engineering in Medicine and Biology Magazine, IEEE,* 22, 23-27.

BOLGER, N., DAVIS, A. & RAFAELI, E. 2003. Diary Methods: Capturing Life as it is Lived. *Annual Review of Psychology,* 54, 579-616.

BONATO, P. Clinical Applications of Wearable Technology. 31st Annual International Conference of the IEEE IBMS, 2009 Minneapolis, Minnesota, USA. 6580-3.

BONATO, P. Advances in wearable technology and its medical applications. Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE, Aug. 31 2010-Sept. 4 2010 2010. 2021-2024.

BOOIJ, J., TISSINGH, G., BOER, G. J., SPEELMAN, J. D., STOOF, J. C., JANSSEN, A. G., WOLTERS, E. C. & VAN ROYEN, E. A. 1997. [123I]FP-CIT SPECT shows a pronounced decline of striatal dopamine transporter labelling in early and advanced Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry,* 62, 133-140.

BOOSE, A., SPIEKER, S., JENTGENS, C., KLOCKGETHER, T., SCHOLZ, E. & DICHGANS, J. 1994. Assessing tremor severity with long-term tremor recordings. *Journal of Neurology, Neurosurgery & Psychiatry,* 57, 397.

BRAAK, H., TREDICI, K. D., RÜB, U., DE VOS, R. A. I., JANSEN STEUR, E. N. H. & BRAAK, E. 2003. Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiology of Aging,* 24, 197-211.

BRADBURN, N., RIPS, L. & SHEVELL, S. 1987. Answering autobiographical questions: the impact of memory and inference on surveys. *Science,* 236, 157-161.

BREWER, B. R., PRADHAN, S., CARVELL, G. & DELITTO, A. 2009. Application of Modified Regression Techniques to a Quantitative Assessment for the Motor Signs of Parkinson's Disease. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on,* 17, 568-575.

BRONTE-STEWART, H. M., DING, L., ALEXANDER, C., ZHOU, Y. & MOORE, G. P. 2000. Quantitative Digitography (QDG): A Sensitive Measure of Digital Motor Control in Idiopathic Parkinson's Disease. *Movement Disorders,* 15, 36-47.

BRUNDIN, P., LI, J.-Y., HOLTON, J. L., LINDVALL, O. & REVESZ, T. 2008. Research in motion: the enigma of Parkinson's disease pathology spread. *Nat Rev Neurosci,* 9, 741-745.

BUCKWALTER, K. C. 2009. Recruitment of older adults: an ongoing challenge. *Research in gerontological nursing* 2, 265-6.

BURKHARD, P. R., SHALE, H., LANGSTON, J. W. & TETRUD, J. W. 1999. Quantification of Dyskinesia in Parkinson's Disease: Validation of a Novel Instrumental Method. *Movement Disorders,* 14, 754-763.

BURN, D. J., ROWAN, E. N., MINETT, T., SANDERS, J., MYINT, P., RICHARDSON, J., THOMAS, A., NEWBY, J., REID, J., O'BRIEN, J. T. & MCKEITH, I. G. 2003. Extrapyramidal features in Parkinson's disease with and

without dementia and dementia with Lewy bodies: A cross-sectional comparative study. *Movement Disorders,* 18, 884-889.

BURTON, C., WELLER, D. & SHARPE, M. 2007. Are electronic diaries useful for symptoms research? A systematic review. *Journal of Psychosomatic Research,* 62, 553-561.

BUSSE, M. E., PEARSON, O. R., VAN DEURSEN, R. & WILES, C. M. 2004. Quantified measurement of activity provides insight into motor function and recovery in neurological disease. *Journal of Neurology, Neurosurgery & Psychiatry,* 75, 884-888.

CALIGIURI, M. P. & PETERSON, S. 1993. A quantitative study of levodopa-induced dyskinesia in Parkinson's disease. *Journal of Neural Transmission - Parkinson's Disease and Dementia Section,* 6, 89-98.

CALIGIURI, M. P. & TRIPP, R. M. 2004. *A portable hand-held device for quantifying and standardizing tremor assessment*, Journal of Medical Engineering and Technology. 28 (6) (pp 254-262), 2004. Date of Publication: Nov 2004.

CEREDA, E., PEZZOLI, G. & BARICHELLA, M. 2010. Role of an electronic armband in motor function monitoring in patients with Parkinson's disease. *Nutrition,* 26, 240-242.

CHARCOT, J. M. & SIGERSON, G. 1879. On paralysis agitans. *Lectures on the diseases of the nervous system, delivered at La Salpêtrière, 2nd ed.* Philadelphia, PA, US: Henry C. Lea.

CHASTIN, S. F. M., BAKER, K., JONES, D., BURN, D. J., GRANAT, M. H. & ROCHESTER, L. 2010. The Pattern of Habitual Sedentary Behaviour Is Different in Advanced Parkinson's Disease. *Movement Disorders,* 25, 2114-2120.

CHAUDHURI, K. R., PRIETO-JURCYNSKA, C., NAIDU, Y., MITRA, T., FRADES-PAYO, B., TLUK, S., RUESSMANN, A., ODIN, P., MACPHEE, G., STOCCHI, F., ONDO, W., SETHI, K., SCHAPIRA, A. H. V., CASTRILLO, J. C. M. & MARTINEZ-MARTIN, P. 2010. The nondeclaration of nonmotor symptoms of Parkinson's disease to health care professionals: An international study using the nonmotor symptoms questionnaire. *Movement Disorders,* 25, 704-709.

CHAUDHURI, K. R., YATES, L. & MARTINEZ-MARTIN, P. 2005. The non-motor symptom complex of Parkinson's disease: A comprehensive assessment is essential. *Current Neurology and Neuroscience Reports,* 5, 275-283.

CHUA, R., CRAIG, J., WOOTTON, R. & PATTERSON, V. 2001. Randomised controlled trial of telemedicine for new neurological outpatient referrals. *Journal of Neurology, Neurosurgery & Psychiatry,* 71, 63-66.

CLARKE, C. E., PATEL, S., IVES, N., RICK, C., JENKINSON, C., WHEATLEY, K., WILLIAMS, A. & GRAY, R. 2012. 1.092 A LARGE RANDOMISED TRIAL ASSESSING QUALITY OF LIFE IN PATIENTS WITH LATER PD: RESULTS FROM PD MED LATER. *Parkinsonism & Related Disorders,* 18, Supplement 2, S33.

COLOSIMO, C., MARTINEZ-MARTIN, P., FABBRINI, G., HAUSER, R. A., MERELLO, M., MIYASAKI, J., POEWE, W., SAMPAIO, C., RASCOL, O., STEBBINS, G. T., SCHRAG, A. & GOETZ, C. G. 2010. Task Force Report on Scales to Assess Dyskinesia in Parkinson's Disease: Critique and Recommendations. *Movement Disorders,* 25, 1131-1142.

COOPER, S. E., MCINTYRE, C. C., FERNANDEZ, H. H. & VITEK, J. L. 2013. ASsociation of deep brain stimulation washout effects with parkinson disease duration. *JAMA Neurology,* 70, 95-99.

COSTA, J., GONZALEZ, H. A., VALLDEORIOLA, F., GAIG, C., TOLOSA, E. & VALLS-SOLE, J. 2010. Nonlinear dynamic analysis of oscillatory repetitive movements in Parkinson's disease and essential tremor. *Movement Disorders,* 25, 2577-86.

COWELL, T. K., OWEN, D. A. L. & MARSDEN, C. D. 1965. Objective Measurement of Parkinsonian Tremor. *The Lancet,* 2, 1278-1279.

CRISS, K. & MCNAMES, J. Video Assessment of Finger Tapping for Parkinson's Disease and Other Movement Disorders. 33rd Annual International Conference of the IEEE EMBS, 2011 Boston, Massachusetts. USA., 7123-7126.

CROSBY, R. D., KOLOTKIN, R. L. & WILLIAMS, G. R. 2003. Defining clinically meaningful change in health-related quality of life. *Journal of Clinical Epidemiology,* 56, 395-407.

CUBO, E., GABRIEL-GALÁN, J. M., MARTÍNEZ, J. S., ALCUBILLA, C. R., YANG, C., ARCONADA, O. F. & PÉREZ, N. M. 2012. Comparison of office-based versus home Web-based clinical assessments for Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society,* 27, 308-311.

CUNNINGHAM, L., MASON, S., NUGENT, C., MOORE, G., FINLAY, D. & CRAIG, D. 2011. Home-Based Monitoring and Assessment of Parkinson's Disease. *IEEE Transactions on Information Technology in Biomedicine,* 15, 47-53.

CUNNINGHAM, L. M., NUGENT, C. D., FINLAY, D. D., MOORE, G. & CRAIG, D. 2009. A review of assistive technologies for people with Parkinson's disease. *Technology and Health Care,* 17, 269-279.

DALRYMPLE-ALFORD, J. C., MACASKILL, M. R., NAKAS, C. T., LIVINGSTON, L., GRAHAM, C., CRUCIAN, G. P., MELZER, T. R., KIRWAN, J., KEENAN, R., WELLS, S., PORTER, R. J., WATTS, R. & ANDERSON, T. J. 2010. The MoCA: well-suited screen for cognitive impairment in Parkinson disease. *Neurology,* 75, 1717-1725.

DANEAULT, J.-F., CARIGNAN, B., CODÈRE, C. É., SADIKOT, A. F. & DUVAL, C. 2013. Using a smart phone as a standalone platform for detection and monitoring of pathological tremors. *Frontiers in Human Neuroscience,* 6.

DAYHOFF, J. E. & DELEO, J. M. 2001. Artifical Neural Networks. *Cancer,* 91, 1615-1635.

DICKSON, D. W., BRAAK, H., DUDA, J. E., DUYCKAERTS, C., GASSER, T., HALLIDAY, G. M., HARDY, J., LEVERENZ, J. B., DEL TREDICI, K., WSZOLEK, Z. K. & LITVAN, I. 2009. Neuropathological assessment of Parkinson's disease: refining the diagnostic criteria. *The Lancet Neurology,* 8, 1150-1157.

DODEL, R. C., BERGER, K. & OERTEL, W. H. 2001. Health-Related Quality of Life and Healthcare Utilisation in Patients with Parkinson's Disease. *Pharmaco-economics,* 19, 1013-1038.

DORSEY, E., VENKATARAMAN, V., GRANA, M. J. & ET AL. 2013. RAndomized controlled clinical trial of "virtual house calls" for parkinson disease. *JAMA Neurology,* 70, 565-570.

DORSEY, E. R., DEUEL, L. M., VOSS, T. S., FINNIGAN, K., GEORGE, B. P., EASON, S., MILLER, D., REMINICK, J. I., APPLER, A., POLANOWICZ, J., VITI, L., SMITH, S., JOSEPH, A. & BIGLAN, K. M. 2010. Increasing access to specialty care: A pilot, randomized controlled trial of telemedicine for Parkinson's disease. *Movement Disorders,* 25, 1652-1659.

DOTCHIN, C., JUSABANI, A., GRAY, W. K. & WALKER, R. 2012. Projected numbers of people with movement disorders in the years 2030 and 2050: Implications for sub-Saharan Africa, using essential tremor and Parkinson's disease in Tanzania as an example. *Movement Disorders,* 27, 1204-1205.

DOTCHIN, C., MSUYA, O., KISSIMA, J., MASSAWE, J., MHINA, A., MOSHY, A., ARIS, E., JUSABANI, A., WHITING, D., MASUKI, G. & WALKER, R. 2008. The prevalence of Parkinson's disease in rural Tanzania. *Movement Disorders,* 23, 1567-1672.

DUNCAN, G. W., YARNALL, A. J., MARRINAN, S. & BURN, D. J. 2013. New horizons in the pathogenesis, assessment and management of movement disorders. *Age and Ageing,* 42, 2-10.

DUNNEWOLD, R. J. W., HOFF, J. I., VAN PELT, H. C. J., FREDRIKZE, P. Q., WAGEMANS, E. A. H. & VAN HILTEN, B. J. J. 1998. Ambulatory Quantitative Assessment of Body Position, Bradykinesia, and Hypokinesia in Parkinson's Disease. *Journal of Clinical Neurophysiology Intraoperative Evoked Potentials,* 15, 235-242.

DUNNEWOLD, R. J. W., JACOBI, C. E. & VAN HILTEN, J. J. 1997. Quantitative assessment of bradykinesia in patients with parkinson's disease. *Journal of Neuroscience Methods,* 74, 107-112.

DURIF, F., VIDAILHET, M., DEBILLY, B. & AGID, Y. 1999. Worsening of Levodopa-Induced Dyskinesias by Motor and Mental Tasks. *Movement Disorders,* 14, 242-245.

EICHHORN, T. E., GASSER, T., MAI, N., MARQUARDT, C., ARNOLD, G., SCHWARZ, J. & OERTEL, W. H. 1996. Computational Analysis of Open Loop Handwriting Movements in Parkinson's Disease: A Rapid Method to Detect Dopaminergic Effects. *Movement Disorders,* 11, 289-297.

ELBLE, R., SINHA, R. & HIGGINS, C. 1990. Quantification of tremor with a digitizing tablet. *Journal of Neuroscience Methods,* 32, 193-198.

ESPAY, A. J., GIUFFRIDA, J. P., CHEN, R., PAYNE, M., MAZZELLA, F., DUNN, E., VAUGHAN, J. E., DUKER, A. P., SAHAY, A., KIM, S. J., REVILLA, F. J. & HELDMAN, D. A. 2011. Differential response of speed, amplitude, and rhythm to dopaminergic medications in Parkinson's disease. *Movement Disorders,* 26, 2504-2508.

EVANS, J. R. & BARKER, R. A. 2011. Defining meaningful outcome measures in trials of disease-modifying theapies in Parkinson's disease. *Expert Opinion on Pharmacotherapy,* 12, 1249-1258.

FAHN, S., ELTON, R. L. & MEMBERS, U. P. 1987. *Recent Developments in Parkinson's Disease,* Florham Park, NJ, Macmillan Healthcare Information.

FARKAS, Z., CSILLIK, A., SZIRMAI, I. & KAMONDI, A. 2006. Asymmetry of tremor intensity and frequency in Parkinson's disease and essential tremor. *Parkinsonism & Related Disorders,* 12, 49-55.

FAVA, G. A., TOMBA, E. & SONINO, N. 2012. Clinimetrics: the science of clinical measurements. *International Journal of Clinical Practice,* 66, 11-15.

FINAN, P., ZAUTRA, A. & TENNEN, H. 2008. Daily diaries reveal influence of pessimism and anxiety on pain prediction patterns. *Psychology & Health,* 23, 551-568.

FOERSTER, F. & SMEJA, M. 1999. Joint amplitude and frequency analysis of tremor activity. *Electromyography & Clinical Neurophysiology,* 39, 11-9.

FOERSTER, F., THIELGEN, T., FUCHS, G., HORNIG, A. & FAHRENBERG, J. 2002. *Tremor assessment of patients with parkinson's disease: 24-h-monitoring by calibrated accelerometry. [German]*, Neurologie und Rehabilitation. 8 (3) (pp 117-121), 2002. Date of Publication: 2002.

FOLSTEIN, M. F. & FOLSTEIN, S. E. 1975. "Mini-Mental State". A Practical Method for Grading the Cognitive State of Patients for the Clinician. *Journal of Psychiatric Research,* 12, 189-198.

FOLTYNIE, T., BRAYNE, C. & BARKER, R. A. 2002. The heterogeniety of idiopathic Parkinson's disease. *Journal of Neurology,* 249, 138-145.

GALNA, B., LORD, S., DAUD, D., ARCHIBALD, N., BURN, D. & ROCHESTER, L. 2012. Visual sampling during walking in people with Parkinson's disease and the influence of environment and dual-task. *Brain Research,* 1473, 35-43.

GARCIA RUIZ, P. J. & BERNADOS, V. S. 2008. Evaluation of ActiTrac (ambulatory activity monitor) in Parkinson's Disease. *Journal of the Neurological Sciences,* 270, 67-69.

GASSER, T., HARDY, J. & MIZUNO, Y. 2011. Milestones in PD genetics. *Movement Disorders,* 26, 1042-1048.

GEMPERLE, F., KASABACH, C., STIVORIC, J., BAUER, M. & MARTIN, R. Design for wearability. Wearable Computers, 1998. Digest of Papers. Second International Symposium on, 19-20 Oct. 1998 1998. 116-122.

GHIKA, J., WIEGNER, A. W., FANG, J. J., DAVIES, L., YOUNG, R. R. & GROWDON, J. H. 1993. Portable system for quantifying motor abnormalities in Parkinson's disease. *Biomedical Engineering, IEEE Transactions on,* 40, 276-283.

GIBB, W. R. & LEES, A. J. 1988. The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry,* 51, 745-752.

GILL, D. J., FRESHMAN, A., BLENDER, J. A. & RAVINA, B. 2008. The montreal cognitive assessment as a screening tool for cognitive impairment in Parkinson's disease. *Movement Disorders,* 23, 1043-1046.

GIOVANNONI, G., VAN SCHALKWYK, J., FRITZ, V. U. & LEES, A. J. 1999. Bradykinesia akinesia inco-ordination test (BRAIN TEST): an objective computerised assessment of upper limb motor function. *Journal of Neurology, Neurosurgery & Psychiatry,* 67, 624-629.

GIUFFRIDA, J. P., RILEY, D. E., MADDUX, B. N. & HELDMAN, D. A. 2009. Clinically deployable Kinesia™ technology for automated tremor assessment. *Movement Disorders,* 24, 723-730.

GOETZ, C. G., FAHN, S., MARTINEZ-MARTIN, P., POEWE, W., SAMPAIO, C., STEBBINS, G. T., STERN, M. B., TILLEY, B. C., DODEL, R., DUBOIS, B., HOLLOWAY, R., JANKOVIC, J., KULISEVSKY, J., LANG, A. E., LEES, A., LEURGANS, S., LEWITT, P. A., NYENHUIS, D., OLANOW, C. W., RASCOL, O., SCHRAG, A., TERESI, J. A., VAN HILTEN, J. J. & LAPELLE, N. 2007. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Process, format, and clinimetric testing plan. *Movement Disorders,* 22, 41-47.

GOETZ, C. G., JANKO, K., BLASUCCI, L. & JAGLIN, J. A. 2003a. Impact of placebo assignment in clinical trials of Parkinson's disease. *Movement Disorders,* 18, 1146-1149.

GOETZ, C. G., LEURGANS, S., HINSON, V. K., BLASUCCI, L. M., ZIMMERMAN, J., FAN, W., NGUYEN, T. & HSU, A. 2008a. Evaluating Parkinson's disease patients at home: Utility of self-videotaping for objective motor, dyskinesia, and ON–OFF assessments. *Movement Disorders,* 23, 1479-1482.

GOETZ, C. G., LEURHANS, S., HINSON, V. K., BLASUCCI, L. M., ZIMMERMAN, J., FAN, W., NGUYEN, T. & HSU, A. 2008b. Evaluating Parkinson's Disease Patients at Home: Utility of Self-Videotaping for Objective Motor, Dyskinesia and ON-OFF Assessments. *Movement Disorders,* 23, 1479-1482.

GOETZ, C. G., LEWITT, P. A. & WEIDENMAN, M. 2003b. Standardized training tools for the UPDRS activities of daily living scale: Newly available teaching program. *Movement Disorders,* 18, 1455-1458.

GOETZ, C. G., NUTT, J. G. & STEBBINS, G. T. 2008c. The Unified Dyskinesia Rating Scale: Presentation and clinimetric profile. *Movement Disorders,* 23, 2398-2403.

GOETZ, C. G. & STEBBINS, G. T. 2004. Assuring interrater reliability for the UPDRS motor section: Utility of the UPDRS teaching tape. *Movement Disorders,* 19, 1453-1456.

GOETZ, C. G., STEBBINS, G. T., BLASUCCI, L. M. & GROBMAN, M. S. 1997. Efficacy of a patient-training videotape on motor fluctuations for on-off diaries in parkinson's disease. *Movement Disorders,* 12, 1039-1041.

GOETZ, C. G., STEBBINS, G. T., CHMURA, T. A., FAHN, S., KLAWANS, H. L. & MARSDEN, C. D. 1995. Teaching tape for the motor section of the unified Parkinson's disease rating scale. *Movement Disorders,* 10, 263-266.

GOETZ, C. G., STEBBINS, G. T., SHALE, H. M., LANG, A. E., CHERNIK, D. A., CHMURA, T. A., AHLSKOG, J. E. & DORFLINGER, E. E. 1994. Utility of an objective dyskinesia rating scale for Parkinson's disease: Inter- and intrarater reliability assessment. *Movement Disorders,* 9, 390-394.

GOETZ, C. G., STEBBINS, G. T., WOLFF, D., DELEEUW, W., BRONTE-STEWART, H., ELBLE, R., HALLETT, M., NUTT, J., RAMIG, L., SANGER, T., WU, A. D., KRAUS, P. H., BLASUCCI, L. M., SHAMIM, E. A., SETHI, K. D., SPIELMAN, J., KUBOTA, K., GROVE, A. S., DISHMAN, E. & TAYLOR, C. B. 2009. Testing objective measures of motor impairment in early Parkinson's disease: Feasibility study of an at-home testing device. *Movement Disorders,* 24, 551-556.

GOETZ, C. G., TILLEY, B. C., SHAFTMAN, S. R., STEBBINS, G. T., FAHN, S., MARTINEZ-MARTIN, P., POEWE, W., SAMPAIO, C., STERN, M. B., DODEL, R., DUBOIS, B., HOLLOWAY, R., JANKOVIC, J., KULISEVSKY, J., LANG, A. E., LEES, A., LEURGANS, S., LEWITT, P. A., NYENHUIS, D., OLANOW, C. W., RASCOL, O., SCHRAG, A., TERESI, J. A., VAN HILTEN, J. J. & LAPELLE, N. 2008d. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Movement Disorders,* 23, 2129-2170.

GORDON, A. M. & REILMANN, R. 1999. Getting a grasp on research: does treatment taint testing of parkinsonian patients? *Brain,* 122, 1597-1598.

GRAY, R., PATEL, S., IVES, N., RICK, C., JENKINSON, C., WHEATLEY, K., WILLIAMS, A. & CLARKE, C. E. 2012. 1.091 A LARGE RANDOMISED TRIAL ASSESSING QUALITY OF LIFE IN PATIENTS WITH EARLY PD: RESULTS FROM PD MED EARLY. *Parkinsonism & Related Disorders,* 18, Supplement 2, S32.

GRIFFITHS, R. I., KOTSCHET, K., ARFON, S., XU, Z. M., JOHNSON, W., DRAGO, J., EVANS, A., KEMPSTER, P., RAGHAV, S. & HORNE, M. K. 2012. Automated Assessment of Bradykinesia and Dyskinesia in Parkinson's Disease. *Journal of Parkinson's Disease,* 2, 47-55.

GRIMALDI, G. & MANTO, M. 2010. Neurological Tremor: Sensors, Signal Processing and Emerging Applications. *Sensors,* 10, 1399-1422.

GROSSET, D. G., GROSSET, K. A., OKUN, M. S. & FERNANDEZ, H. H. 2009. *Parkinson's Disease: Clinician's Desk Reference,* London, Manson Publishing.

GROSSET, D. G. & SCHAPIRA, A. H. 2008. Timing the initiation of treatment in Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry,* 79, 615.

GUY, W. 1976. Abnormal Involuntary Movement Scale. ECDEU assessment manual for psychopharmacology. Washington DC: US Government Printing Office. .

GUYATT, G., WALTER, S. & NORMAN, G. 1987. Measuring change over time: Assessing the usefulness of evaluative instruments. *Journal of Chronic Diseases,* 40, 171-178.

HALLIDAY, G., HELY, M., REID, W. & MORRIS, J. 2008. The progression of pathology in longitudinally followed patients with Parkinson's disease. *Acta Neuropathologica,* 115, 409-415.

HALLIDAY, G., LEES, A. & STERN, M. 2011. Milestones in Parkinson's disease—Clinical and pathologic features. *Movement Disorders,* 26, 1015-1021.

HAUSER, R. A., AUINGER, P. & ON BEHALF OF THE PARKINSON STUDY, G. 2011a. Determination of minimal clinically important change in early and advanced Parkinson's disease. *Movement Disorders,* 26, 813-818.

HAUSER, R. A., CANTILLON, M., POURCHER, E., MICHELI, F., MOK, V., ONOFRJ, M., HUYCK, S. & WOLSKI, K. 2011b. Preladenant in patients with Parkinson's disease and motor fluctuations: a phase 2, double-blind, randomised trial. *The Lancet Neurology,* 10, 221-229.

HAUSER, R. A., DECKERS, F. & LEHERT, P. 2004. Parkinson's Disease Home Diary: Further Validation and Implications for Clinical Trials. *Movement Disorders,* 19, 1409-1413.

HAUSER, R. A., FRIEDLANDER, J., ZESIEWICZ, T. A., ADLER, C. H., SEEBERGER, L. C., O'BRIEN, C. F., MOLHO, E. S. & FACTOR, S. A. 2000. A Home Diary to Assess Functional Status in Patients with Parkinson's Disease with Motor Fluctuations and Dyskinesia. *Clinical Neuropharmacology,* 23, 75-81.

HAUSER, R. A., ZESIEWICZ, T. A., FACTOR, S. A., GUTTMAN, M. & WEINER, W. J. 1997. Clinical trials of add-on medications in Parkinson's disease: Efficacy versus usefulness. *Parkinsonism & Related Disorders,* 3, 1-6.

HERNÁN, M. A., TAKKOUCHE, B., CAAMAÑO-ISORNA, F. & GESTAL-OTERO, J. J. 2002. A meta-analysis of coffee drinking, cigarette smoking, and the risk of Parkinson's disease. *Annals of Neurology,* 52, 276-284.

HINDLE, J. V., HINDLE, C. M. & HOBSON, P. 2007. Co-morbidity and the frequency of general practitioner consultations in Parkinson's disease in the United Kingdom. *Movement Disorders,* 22, 1054-1055.

HOBART, J. C., LAMPING, D. L. & THOMPSON, A. J. 1996. Evaluating neurological outcome measures: the bare essentials. *Journal of Neurology, Neurosurgery & Psychiatry,* 60, 127-130.

HODGKINSON, J., MANT, J., MARTIN, U., GUO, B., HOBBS, F. D. R., DEEKS, J. J., HENEGHAN, C., ROBERTS, N. & MCMANUS, R. J. 2011. Relative effectiveness of clinic and home blood pressure monitoring

compared with ambulatory blood pressure monitoring in diagnosis of hypertension: systematic review. *BMJ,* 342.

HOEHN, M. M. & YAHR, M. D. 1967. Parkinsonism: onset, progression and mortality. *Neurology,* 17, 427-442.

HOFF, J. I., VAN DEN PLAS, A. A., WAGEMANS, E. A. & VAN HILTEN, J. J. 2001a. Accelerometric assessment of levodopa-induced dyskinesias in Parkinson's disease. *Movement Disorders,* 16, 58-61.

HOFF, J. I., VAN DER MEER, V. & VAN HILTEN, J. J. 2004. *Accuracy of Objective Ambulatory Accelerometry in Detecting Motor Complications in Patients With Parkinson Disease. [References]*, Clinical Neuropharmacology. Vol.27(2), Mar-Apr 2004, pp. 53-57.

HOFF, J. I., VAN HILTEN, B. J. & ROOS, R. A. 1999. A review of the assessment of dyskinesias. *Movement Disorders,* 14, 737-43.

HOFF, J. I., WAGEMANS, E. A. & VAN HILTEN, B. J. 2001b. Ambulatory objective assessment of tremor in Parkinson's disease. *Clinical Neuropharmacology,* 24, 280-3.

HOFFMANN, T., RUSSELL, T., THOMPSON, L., VINCENT, A. & NELSON, M. 2008. Using the Internet to assess activities of daily living and hand function in people with Parkinson's disease. *NeuroRehabilitation,* 23, 253-261.

HUBBLE, J. P., PAHWA, R., MICHALEK, D. K., THOMAS, C. & KOLLER, W. C. 1993. Interactive video conferencing: A means of providing interim care to parkinson's disease patients. *Movement Disorders,* 8, 380-382.

HUGHES, A. J., DANIEL, S. E., BEN-SHLOMO, Y. & LEES, A. J. 2002. The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. *Brain,* 125, 861-870.

HUGHES, A. J., DANIEL, S. E., KILFORD, L. & LEES, A. J. 1992. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *Journal of Neurology, Neurosurgery & Psychiatry,* 55, 181-184.

HUNG, S. W., ADELI, G. M., ARENOVICH, T., FOX, S. H. & LANG, A. E. 2010. Patient perception of dyskinesia in Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*.

INGVARSSON, P. E., JOHNELS, B., STEBBINS, G. T. & OLSSON, T. 1999. Objective assessment in Parkinson's disease: optoelectronic movement and force analysis in clinical routine and research. *Advances in Neurology,* 80, 447-458.

JAMIESON, S. 2004. Likert scales: how to (ab)use them. *Medical Education,* 38, 1217-1218.

JANKOVIC, J. 2005. Motor fluctuations and dyskinesias in Parkinson's disease: Clinical manifestations. *Movement Disorders,* 20, S11-S16.

JANKOVIC, J. & FROST, J. D. 1981. Quantitative assessment of parkinsonian and essential tremor: Clinical application of traxial accelerometry. *Neurology,* 31, 1235-1240.

JANKOVIC, J., MCDERMOTT, M., CARTER, J., GAUTHIER, S., GOETZ, C., GOLBE, L., HUBER, S., KOLLER, W., OLANOW, C., SHOULSON, I., STERN, M., TANNER, C., WEINER, W. & GROUP, P. S. 1990. Variable expression of Parkinson's disease: A base-line analysis of the DAT ATOP cohort. *Neurology,* 40, 1529.

JEDYNAK, C. P., BONNET, A. M. & AGID, Y. 1991. Tremor and idiopathic dystonia. *Movement Disorders,* 6, 230-236.

JENNER, P., MORI, A., HAUSER, R., MORELLI, M., FREDHOLM, B. B. & CHEN, J. F. 2009. Adenosine, adenosine A_{2A} antagonists and Parkinson's disease. *Parkinsonism & Related Disorders,* 15, 406-413.

JOHNELS, B., INGVARSSON, P. E., THORSELIUS, M., VALLS, M. & STEG, G. 1989. Disability profiles and objective quantitative assessment in Parkinson's disease. *Acta Neurologica Scandinavica,* 79, 227-238.

JOUNDI, R. A., BRITTAIN, J.-S., JENKINSON, N., GREEN, A. L. & AZIZ, T. 2011. Rapid tremor frequency assessment with the iPhone accelerometer. *Parkinsonism & Related Disorders,* 17, 288-90.

KALE, R. & MENKEN, M. 2004. Who should look after people with Parkinson's disease? *BMJ,* 328, 62-63.

KANDORI, A., YOKOE, M., SAKODA, S., ABE, K., MIYASHITA, T., OE, H., NARITOMI, H., OGATA, K. & TSUKADA, K. 2004. Quantitative magnetic detection of finger movements in patients with Parkinson's disease. *Neuroscience Research,* 49, 253-260.

KATAYAMA, S. 2001. Actigraphic Analysis of Diurnal Motor Fluctuations during Dopamine Agonist Therapy. *European Journal of Neurology,* 46, 11-17.

KATZENSCHLAGER, R., SCHRAG, A., EVANS, A., MANSON, A., CARROLL, C. B., OTTAVIANI, D., LEES, A. J. & HOBART, J. 2007. Quantifying the impact of dyskinesias in PD. THe PDYS-26: A patient-based outcome measure. *Neurology,* 69, 555-563.

KEIJSERS, N. L. W., HORSTINK, M. W. & GIELEN, C. C. 2003a. Online monitoring of dyskinesia in patients with Parkinson's disease. *IEEE Engineering in Medicine and Biology Magazine,* 22, 96-103.

KEIJSERS, N. L. W., HORSTINK, M. W. I. M. & GIELEN, S. C. A. M. 2003b. Automatic assessment of levodopa-induced dyskinesias in daily life by neural networks. *Movement Disorders,* 18, 70-80.

KEIJSERS, N. L. W., HORSTINK, M. W. I. M. & GIELEN, S. C. A. M. 2006. *Ambulatory motor assessment in Parkinson's disease*, Movement Disorders. 21 (1) (pp 34-44), 2006. Date of Publication: Jan 2006.

KEIJSERS, N. L. W., HORSTINK, M. W. I. M., VAN HILTEN, J. J., HOFF, J. I. & GIELEN, C. C. A. M. 2000. *Detection and assessment of the severity of levodopa-induced dyskinesia in patients with Parkinson's disease by neural networks*, Movement Disorders. 15 (6) (pp 1104-1111), 2000. Date of Publication: 2000.

KEMPSTER, P., HURWITZ, B. & LEES, A. J. 2007. A new look at James Parkinson's *Essay on the Shaking Palsy*. *Neurology,* 69, 482-485.

KHLEBTOVSKY, A., RIGBI, A., MELAMED, E., ZIV, I., STEINER, I., GAD, A. & DJALDETTI, R. 2012. Patient and caregiver perceptions of the social impact of advanced Parkinson's disease and dyskinesias. *Journal of Neural Transmission,* 119, 1367-1371.

KHOO, T. K., YARNALL, A. J., DUNCAN, G. W., COLEMAN, S., O'BRIEN, J. T., BROOKS, D. J., BARKER, R. A. & BURN, D. J. 2013. The specturm of nonmotor symptoms in early Parkinsons disease. *Neurology,* 80, 276-281.

KINCAID, P., FISHBURNE, R., ROGERS, R. & CHISSOM, B. 1975. Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. National Technical Information Service, Springfield, Virginia 22151

KORHONEN, I., PARKKA, J. & VAN GILS, M. 2003. Health monitoring in the home of the future. *Engineering in Medicine and Biology Magazine, IEEE,* 22, 66-73.

KRAUS, P. H. & HOFFMANN, A. 2010. Spiralometry: Computerized assessment of tremor amplitude on the basis of spiral drawing. *Movement Disorders,* 25, 2164-2170.

KUPRYJANOW, A., KUNKA, B. & KOSTEK, B. UPDRS Tests for Diagnosis of Parkinson's Disease Employing Virtual-Touchpad. Database and Expert Systems Applications (DEXA), 2010 Workshop on, Aug. 30 2010-Sept. 3 2010 2010. 132-136.

LAI, B. C. L., MARION, S. A., TESCHKE, K. & TSUI, J. K. C. 2002. Occupational and environmental risk factors for Parkinson's disease. *Parkinsonism & Related Disorders,* 8, 297-309.

LAKIE, M. & MUTCH, W. J. 1989. Finger tremor in Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry,* 52, 392-394.

LANG, A. E., MELAMED, E., POEWE, W. & RASCOL, O. 2013. Trial designs used to study neuroprotective therapy in Parkinson's disease. *Movement Disorders,* 28, 86-95.

LAURETANI, F., MAGGIO, M., SILVESTRINI, C., NARDELLI, A., SACCAVINI, M. & CEDA, G. P. 2012. Parkinson's disease (PD) in the elderly: An example of geriatric syndrome (GS)? *Archives of Gerontology and Geriatrics,* 54, 242-246.

LEE, C. S., SCHULZER, M., MAK, E., HAMMERSTEAD, J. P., CALNE, S. & CALNE, D. B. 1995. Patterns of asymmetry do not change over the course of idiopathic parkinsonisn: implications for pathogenesis. *Neurology,* 45, 435-439.

LEHOUX, P. 2004. Patients' perspectives on high-tech home care: a qualitative inquiry into the user-friendliness of four technologies. *BMC health services research,* 4, 28.

LEMOYNE, R., MASTROIANNI, T., COZZA, M., COROIAN, C. & GRUNDFEST, W. 2010. Implementation of an iPhone for characterizing Parkinson's disease tremor through a wireless accelerometer application. *Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine & Biology Society,* 2010, 4954-8.

LEWIS, S. J. G., FOLTYNIE, T., BLACKWELL, A. D., ROBBINS, T. W., OWEN, A. M. & BARKER, R. A. 2005. Heterogeneity of Parkinson's disease in the early clinical stages using a data driven approach. *Journal of Neurology, Neurosurgery & Psychiatry,* 76, 343-348.

LITVAN, I., AGID, Y., CALNE, D., CAMPBELL, G., DUBOIS, B., DUVOISIN, R. C., GOETZ, C. G., GOLBE, L., GRAFMAN, J., GROWDON, J., HALLETT, M., JANKOVIC, J., QUINN, N., TOLOSA, E. & ZEE, D. S. 1996. Clinical research criteria for the diagnosis of progressive supranulear palsy (Steele-Richardson-Olszewski syndrome): Report of the NINDS-SPSP International Workshop. *Neurology,* 47, 1-9.

LIU, X., CARROLL, C., WANG, S.-Y., ZAJICEK, J. & BAIN, P. 2005. Quantifying drug-induced dyskinesias in the arms using digitised spiral-drawing tasks. *Journal of Neuroscience Methods,* 144, 47-52.

LIU, X., OSTERBAUER, R., AZIZ, T., MIALL, R. & STEIN, J. 2001. Increased response to visual feedback of drug-induced dyskinetic movements in advanced Parkinson's disease. *Neuroscience Letters,* 304, 25-28.

LIU, X., S.A., T., AZIZ, T. Z., MIALL, R. C. & STEIN, J. F. 1999. Effects of visual feedback on manual tracking and action tremor in Parkinson's disease. *Experimental Brain Research,* 129, 477-481.

LLORET, S. P., ROSSI, M., CARDINALI, D. P. & MERELLO, M. 2010. Actigraphic Evaluation of Motor Fluctuations in Patients with Parkinson's Disease. *International Journal of Neuroscience,* 120, 137-143.

LOUIS, E. D., LYNCH, T., MARDER, K. & FAHN, S. 1996. Reliability of patient completion of the historical section of the unified Parkinson's disease rating scale. *Movement Disorders,* 11, 185-192.

LYONS, K. E. & PAHWA, R. 2007. Electronic motor function diary for patients with Parkinson's disease: a feasibilty study. *Parkinsonism & Related Disorders,* 13, 304-307.

MACHOWSKA-MAJCHRZAL, A., PIERZCHATA, K. & PIETRASZEK, S. 2007. Analysis of selected parameters of tremor recorded by a biaxial accelerometer in patients with parkinsonian tremor, essential tremor and cerebellar tremor. *Neurologia i Neurochirurgia Polska,* 41, 241-250.

MACMAHON, D. G. 1999. Parkinson's disease nurse specialists: an important role in disease management. *Neurology,* 52, S21-5.

MAETZLER, W., DOMINGOS, J., SRULIJES, K., FERREIRA, J. J. & BLOEM, B. R. 2013. Quantitative wearable sensors for objective assessment of Parkinson's disease. *Movement Disorders,* 28, 1628-1637.

MANGIONE, S. & PEITZMAN, S. 1996. Physical diagnosis in the 1990s. *Journal of General Internal Medicine,* 11, 490-493.

MANSON, A. J., BROWN, P., O'SULLIVAN, J. D., ASSELMAN, P., BUCKWELL, D. & LEES, A. J. 2000. An ambulatory dyskinesia monitor. *Journal of Neurology, Neurosurgery & Psychiatry,* 68, 196-201.

MAR, B. D. 2005. The History of Clinical Holter Monitoring. *Annals of Noninvasive Electrocardiology,* 10, 226-230.

MARCONI, R., LEFEBVRE-CAPARROS, D., BONNET, A., VIDAILHET, M., DUBOIS, B. & AGID, Y. 1994. Levodopa-Induced Dyskinesias in Parkinson's Disease Phenomenology and Pathophysiology. *Movement Disorders,* 9, 2-12.

MARSDEN, C. D. & OWEN, D. A. L. 1967. Mechanisms underlying emotional variation in parkinsonian tremor. *Neurology,* 17, 711-715.

MARSDEN, C. D. & PARKES, J. D. 1976. "ON-OFF" EFFECTS IN PATIENTS WITH PARKINSON'S DISEASE ON CHRONIC LEVODOPA THERAPY. *The Lancet,* 307, 292-296.

MARSDEN, C. D. & SCHACHTER, M. 1981. Assessment of extrapyramidal disorders. *British Journal of Clinical Pharmacology,* 11, 129-151.

MARTINEZ-MARTIN, P. & FORJAZ, M. J. 2006. Metric attributes of the unified Parkinson's disease rating scale 3.0 battery: Part I, feasibility, scaling assumptions, reliability, and precision. *Movement Disorders,* 21, 1182-1188.

MARTINEZ-MARTIN, P., RODRIGUEZ-BLAZQUEZ, C., KURTIS, M. M., CHAUDHURI, K. R. & ON BEHALF OF THE, N. V. G. 2011. The impact of non-motor symptoms on health-related quality of life of patients with Parkinson's disease. *Movement Disorders,* 26, 399-406.

MARTINEZ-MARTIN, P., SCHAPIRA, A. H. V., STOCCHI, F., SETHI, K., ODIN, P., MACPHEE, G., BROWN, R. G., NAIDU, Y., CLAYTON, L., ABE, K., TSUBOI, Y., MACMAHON, D., BARONE, P., RABEY, M., BONUCCELLI, U., FORBES, A., BREEN, K., TLUK, S., OLANOW, C. W., THOMAS, S., RYE, D., HAND, A., WILLIAMS, A. J., ONDO, W. & CHAUDHURI, K. R. 2007. Prevalence of nonmotor symptoms in Parkinson's disease in an international setting; Study using nonmotor symptoms questionnaire in 545 patients. *Movement Disorders,* 22, 1623-1629.

MCHORNEY, C. A. & TARLOV, A. R. 1995. Individual-Patient Monitoring in Clinical Practice: Are Available Health Status Surveys Adequate? *Quality of Life Research,* 4, 293-307.

MDS 2003. The Unified Parkinson's Disease Rating Scale (UPDRS): Status and recommendations. *Movement Disorders,* 18, 738-750.

MERA, T. O., BURACK, M. A. & GIUFFRIDA, J. P. Quantitative assessment of levodopa-induced dyskinesia using automated motion sensing technology. Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE, Aug. 28 2012-Sept. 1 2012 2012a. 154-157.

MERA, T. O., BURACK, M. A. & GIUFFRIDA, J. P. 2013. Objective Motion Sensor Assessment Highly Correlated with Scores of Global Levodopa-Induced Dyskinesia in Parkinson's Disease. *Journal of Parkinson's Disease,* 3, 399-407.

MERA, T. O., HELDMAN, D. A., ESPAY, A. J., PAYNE, M. & GIUFFRIDA, J. P. 2012b. Feasibility of home-based automated Parkinson's disease motor assessment. *Journal of Neuroscience Methods,* 203, 152-156.

MITCHELL, S. L., HARPER, D. W., LAU, A. & BHALLA, R. 2000. Patterns of outcome measurement in Parkinson's disease clinical trials. *Neuroepidemiology,* 19, 100-8.

MODY, L., MILLER, D. K., MCGLOIN, J. M., FREEMAN, M., MARCANTONIO, E. R., MAGAZINER, J. & STUDENSKI, S. 2008. Recruitment and Retention of Older Adults in Aging Research. *Journal of the American Geriatrics Society,* 56, 2340-2348.

MORRIS, H. R. 2005. Genetics of Parkinson's disease. *Annals of Medicine,* 37, 86-96.

MÜLLER, T. & BENZ, S. 2002. Quantification of the dopaminergic response in Parkinson's disease. *Parkinsonism & Related Disorders,* 8, 181-186.

MULLER, T., SCHAFER, S., KUHN, W. & PRZUNTEK, H. 2000. Correlation between Tapping and Inserting of Pegs in Parkinson's Disease. *The Canadian Journal of Neurological Sciences,* 27, 311-315.

MUTHURAMAN, M., HOSSEN, A., HEUTE, U., DEUSCHL, G. & RAETHJEN, J. 2011. A new diagnostic test to distinguish tremulous Parkinson's disease from advanced essential tremor. *Movement Disorders,* 26, 1548-1552.

NAYAK, S., NAYAK, V. C. & MENEZES, R. 2012. Telemedicine: The Future of Healthcare Delivery. *International Journal of Computer Science and Management Research,* 1, 4-11.

NAZEM, S., SIDEROWF, A. D., DUDA, J. E., TEN HAVE, T., COLCHER, A., HORN, S. S., MOBERG, P. J., WILKINSON, J. R., HURTIG, H. I., STERN, M. B. & WEINTRAUB, D. 2009. Montreal Cognitive Assessment Performance in Patients with Parkinson's Disease with "Normal" Global Cognition According to Mini-Mental State Examination Score. *Journal of the American Geriatrics Society,* 57, 304-308.

NI SCANAILL, C., CAREW, S., BARRALON, P., NOURY, N., LYONS, D. & LYONS, G. M. 2006. A Review of Approaches to Mobility Telemonitoring of the Elderly in Their Living Environment. *Annals of Biomedical Engineering,* 34, 547-563.

NORMAN, K. E., EDWARDS, R. & BEUTER, A. 1999. The measurement of tremor using a velocity transducer: comparison to simultaneous recordings using transducers of displacement, acceleration and muscle activity. *Journal of Neuroscience Methods,* 92, 41-54.

NUSSBAUM, R. L. & ELLIS, C. E. 2003. Alzheimer's Disease and Parkinson's Disease. *The New England Journal of Medicine,* 348, 1356-64.

NYHOLM, D., KOWALSKI, J. & AQUILONIUS, S.-M. 2004. Wireless Real-Time Electronic Data Capture for Self-Assessment of Motor Function and Quality of Life in Parkinson's Disease. *Movement Disorders,* 19, 446-451.

O'CATHAIN, A. & THOMAS, K. 2004. "Any other comments?" Open questions on questionnaires – a bane or a bonus to research? *BMC Medical Research Methodology,* 4, 1-7.

O'SUILLEABHAIN, P. E. & DEWEY JR, R. B. 2001. Validation for Tremor Quantification of an Electromagnetic Tracking Device. *Movement Disorders,* 16, 265-271.

O'SULLIVAN, S. S., WILLIAMS, D. R., GALLAGHER, D. A., MASSEY, L. A., SILVEIRA-MORIYAMA, L. & LEES, A. J. 2008. Nonmotor symptoms as presenting complaints in Parkinson's disease: A clinicopathological study. *Movement Disorders,* 23, 101-106.

OBESO, J. A., RODRIGUEZ-OROZ, M. C., RODRIGUEZ, M., DELONG, M. R. & OLANOW, C. W. 2000. Pathophysiology of levodopa-induced dyskinesias in Parkinson's disease: problems with the current model. *Annals of Neurology,* 47, S22-32.

OKUNO, R., YOKOE, M., AKAZAWA, K., ABE, K. & SAKODA, S. Finger Taps Movement Acceleration Measurement System for Quantitative Diagnosis of Parkinson's disease. Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE, Aug. 30 2006-Sept. 3 2006 2006. 6623-6626.

OLANOW, C. W., WUNDERLE, K. B. & KIEBURTZ, K. 2011. Milestones in movement disorders clinical trials: Advances and landmark studies. *Movement Disorders,* 26, 1003-1014.

OLIVER, C. M., HUNTER, S. A., IKEDA, T. & GALLETLY, D. C. 2013. Junior doctor skill in the art of physical examination: a retrospective study of the medical admission note over four decades. *BMJ Open,* 3.

OWEN, D. A. L. & MARSDEN, C. D. 1965. Effect of Adrenergic β-blockade on Parkinsonian tremor. *The Lancet,* 286, 1259-1262.

PAL, P. K., LEE, C. S., SAMII, A., SCHULZER, M., STOESSL, A. J., MAK, E. K., WUDEL, J., DOBKO, T. & TSUI, J. K. C. 2001. Alternating two finger tapping with contralateral activation is an objective measure of clinical severity in Parkinson's disease and correlates with PET[18F]-DOPA Ki. *Parkinsonism & Related Disorders,* 7, 305-309.

PAPAPETROPOULOS, S. M. D. P., JAGID, J. R. M., SENGUN, C. M., SINGER, C. M. & GALLO, B. V. M. 2008. Objective monitoring of tremor and bradykinesia during DBS surgery for Parkinson disease. *Neurology,* 70, 1244-1249.

PARKINSON, J. 2002. An essay on the shaking palsy. *The Journal of Neuropsychiatry and Clinical Neurosciences* 14, 223-236.

PATEL, S., LORINCZ, K., HUGHES, R., HUGGINS, N., GROWDON, J., STANDAERT, D., AKAY, M., DY, J., WELSH, M. & BONATO, P. 2009. Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors. *IEEE Transactions on Information Technology in Biomedicine,* 13, 864-73.

PÉCHEVIS, M., CLARKE, C. E., VIEREGGE, P., KHOSHNOOD, B., DESCHASEAUX-VOINET, C., BERDEAUX, G., ZIEGLER, M. & ON BEHALF OF THE TRIAL STUDY, G. 2005. Effects of dyskinesias in Parkinson's disease on quality of life and health-related costs: a prospective European study. *European Journal of Neurology,* 12, 956-963.

PEETERS, P. H. F. 2000. Design criteria for an automatic safety-alarm system for elderly. *Technology and Health Care,* 8, 81-91.

PETERSON, M. C., HOLBROOK, J. H., HALES, D., SMITH, N. L. & STAKER, L. V. 1992. Contributions of the History, Physical Examination, and Laboratory Investigation in Making Medical Diagnoses. *Western Medical Journal,* 156, 163-165.

PETO, V., JENKINSON, C., FITZPATRICK, R. & GREENHALL, R. 1995. The development and validation of a short measure of functioning and well being for individuals with Parkinson's disease. *Quality of Life Research,* 4, 241-248.

PIMLOTT, R. M., GIBSON, J. M., BROWN, I. T. & KENNARD, C. 1983. A new ambulatory monitoring system for tremor. *Electroencephalography and Clinical Neurophysiology,* 56, 694-695.

PIPPIN, K. & FERNIE, G. R. 1997. Designing devices that are acceptable to the frail elderly: a new understanding based upon how older people perceive a walker. *Technology and Disability,* 7, 93-102.

PORTER, B., MACFARLANE, R., UNWIN, N. & WALKER, R. 2006. The prevalence of Parkinson's disease in an area of North Tyneside in the North-East of England. *Neuroepidemiology,* 26, 156-161.

POST, B., MERKUS, M. P., DE BIE, R. M. A., DE HAAN, R. J. & SPEELMAN, J. D. 2005. Unified Parkinson's Disease Rating Scale Motor Examination: Are Ratings of Nurses, Residents in Neurology , and Movement Disorders Specialists Interchangeable? *Movement Disorders,* 20, 1577-1584.

POSTUMA, R. B., AARSLAND, D., BARONE, P., BURN, D. J., HAWKES, C. H., OERTEL, W. & ZIEMSSEN, T. 2012. Identifying prodromal Parkinson's disease: Pre-Motor disorders in Parkinson's disease. *Movement Disorders,* 27, 617-626.

POSTUMA, R. B., GAGNON, J. F., VENDETTE, M., FANTINI, M. L., MASSICOTTE-MARQUEZ, J. & MONTPLAISIR, J. 2009. Quantifying the risk of neurodegenerative disease in idiopathic REM sleep behaviour disorder. *Neurology,* 72, 1296-1300.

POTVIN, A. R., TOURTELLOTTE, W. W., SNYDER, D. N., HENDERSON, W. G. & ALBERS, J. W. 1975. Validity of Quantitative Tests Measuring Tremor. *American Journal of Physical Medicine,* 54, 243-252.

POWELL, C. 2008. Frailty and Parkinson's disease: Theories and clinical implications. *Parkinsonism & Related Disorders,* 14, 271-272.

PULLMAN, S. 1998. Spiral Analysis: A new technique for measuring tremor with a ditizing tablet. *Movement Disorders,* 13, 85-89.

QUINN, N. 1989. Mulitple System Atrophy - the nature of the beast. *Journal of Neurology, Neurosurgery & Psychiatry,* 52, 78-89.

RAJARAMAN, V., JACK, D., ADAMOVICH, S. V., HENING, W., SAGE, J. & POIZNER, H. 2000. A novel quantitative method for 3D measurement of Parkinsonian tremor. *Clinical Neurophysiology,* 111, 338-343.

RASCOL, O. 2006. Defining a minimal clinically relevant difference for the unified Parkinson's rating scale: An important but still unmet need. *Movement Disorders,* 21, 1059-1061.

RASCOL, O., BROOKS, D. J., KORCZYN, A. D., DE DEYN, P. P., CLARKE, C. E. & LANG, A. E. 2000. A Five-Year Study of the Incidence of Dyskinesia in Patients with Early Parkinson's Disease Who Were Treated with Ropinirole or Levodopa. *New England Journal of Medicine,* 342, 1484-1491.

RASCOL, O., LOZANO, A., STERN, M. & POEWE, W. 2011. Milestones in Parkinson's disease therapeutics. *Movement Disorders,* 26, 1072-1082.

REIMER, J., GRABOWSKI, M., LINDVALL, O. & HAGELL, P. 2004. Use and interpretation of on/off diaries in Parkinson's disease. *J Neurol Neurosurg Psychiatry,* 75, 396-400.

RICHARDS, M., MARDER, K., COTE, L. & MAYEUX, R. 1994. Reliability of symptom onset assessment in Parkinson's disease. *Movement Disorders,* 9, 340-342.

RIGAS, G., TZALLAS, A. T., TSALIKAKIS, D. G., KONITSIOTIS, S. & FOTIADIS, D. I. 2009. Real-time quantification of resting tremor in the Parkinson's disease. *Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine & Biology Society,* 2009, 1306-9.

RIGAS, G., TZALLAS, A. T., TSIPOURAS, M. G., BOUGIA, P., TRIPOLITI, E. E., BAGA, D., FOTIADIS, D. I., TSOULI, S. G. & KONITSIOTIS, S. 2012. Assessment of Tremor Activity in the Parkinson’s Disease Using a Set of Wearable Sensors. *Information Technology in Biomedicine, IEEE Transactions on,* 16, 478-487.

RIVIERE, C., REICH, S. & THAKOR, N. 1997. Adaptive Fourier modeling for quantification of tremor. *Journal of Neuroscience Methods,* 74, 77-87.

RODRÍGUEZ-MOLINERO, A., PÉREZ-MARTÍNEZ, D. A., CATALÀ, A., CABESTANY, J. & YUSTE, A. 2009. Treatment of Parkinson's disease could be regulated by movement sensors: subcutaneous infusion of varying apomorphine doses according to the intensity of motor activity. *Medical hypotheses,* 72, 430-433.

ROINE, R., OHINMAA, A. & HAILEY, D. 2001. Assessing telemedicine: a systematic review of the literature. *Canadian Medical Association Journal,* 165, 765-771.

ROSS, G. W., ABBOTT, R. D., PETROVITCH, H., TANNER, C. M., DAVIS, D. G., NELSON, J., MARKESBERY, W. R., HARDMAN, J., MASAKI, K., LAUNER, L. & WHITE, L. R. 2006. Association of olfactory dysfunction with incidental Lewy bodies. *Movement Disorders,* 21, 2062-2067.

ROSS, G. W., PETROVITCH, H., ABBOTT, R. D., TANNER, C. M., POPPER, J., MASAKI, K., LAUNER, L. & WHITE, L. R. 2008. Association of olfactory dysfunction with risk for future Parkinson's disease. *Annals of Neurology,* 63, 167-173.

ROY, S. H., COLE, B. T., GILMORE, L. D., DE LUCA, C. J., THOMAS, C. A., SAINT-HILAIRE, M. M. & NAWAB, S. H. 2013. High-resolution tracking of motor disorders in Parkinson's disease during unconstrained activity. *Movement Disorders*, n/a-n/a.

SAINANI, K. 2010. The Importance of Accounting for Correlated Observations. *Physical Medicine and Rehabilitation,* 2, 858-861.

SAITO, N., YAMAMOTO, T., SUGIURA, Y., SHIMIZU, S. & SHIMIZU, M. 2004. Lifecorder: A New Device for the Long-term Monitoring of Motor Activities for Parkinson's Disease. *Internal Medicine,* 43, 685-692.

SALARIAN, A., RUSSMANN, H., WIDER, C., BURKHARD, P. R., VINGERHOETS, F. J. G. & AMINIAN, K. 2007. Quantification of Tremor and Bradykinesia in Parkinson's Disease Using a Novel Ambulatory Monitoring System. *Biomedical Engineering, IEEE Transactions on,* 54, 313-322.

SATO, K., HATANO, T., YAMASHIRO, K., KAGOHASHI, M., NISHIOKA, K., IZAWA, N., MOCHIZUKI, H., HATTORI, N., MORI, H. & MIZUNO, Y. 2006. Prognosis of Parkinson's disease: Time to stage III, IV, V, and to motor fluctuations. *Movement Disorders,* 21, 1384-1395.

SAUERMANN, S., STANDHARDT, H., GERSCHLAGER, W., LANMULLER, H. & ALESCH, F. 2005. Kinematic evaluation in Parkinson's disease using a hand-held position transducer and computerized signal analysis. *Acta Neurochirurgica,* 147, 939-45; discussion 945.

SAUNDERS-PULLMAN, R., DERBY, C., STOCCHI, F., FLOYD, A., BRESSMAN, S., LIPTON, R., DELIGTISCH, A., SEVERT, L., YU, Q., KURTIS, M. & PULLMAN, S. 2008. Validity of spiral analysis in early Parkinson's disease. *Movement Disorders,* 23, 531-537.

SAWYER, J. & BENNETT, A. 2006. Comparing the Level of Dexterity offered by Latex and Nitrile SafeSkin Gloves. *Annals of Occupational Hygiene,* 50, 289-296.

SCHALLER, R. R. 1997. Moore's law: past, present and future. *Spectrum, IEEE,* 34, 52-59.

SCHAPIRA, A. H. & JENNER, P. 2011. Etiology and pathogenesis of Parkinson's disease. *Movement Disorders,* 26, 1049-1055.

SCHLESINGER, I., BENYAKOV, O., ERIKH, I., SURAIYA, S. & SCHILLER, Y. 2009. Parkinson's disease tremor is diminished with relaxation guided imagery. *Movement Disorders,* 24, 2059-62.

SCHOLZ, E., BACHER, M., DIENER, H. C. & DICHGANS, J. 1988. Twenty-four-hour tremor recordings in the evaluation of the treatment of Parkinson's disease. *Journal of Neurology,* 235, 475-84.

SCHRAG, A., BEN-SHLOMO, Y. & QUINN, N. 2002. How valid is the clinical diagnosis of Parkinson's disease in the community? *Journal of Neurology, Neurosurgery & Psychiatry,* 73, 529-534.

SCHRAG, A., MÜNCHAU, A., BHATIA, K. P., QUINN, N. P. & MARSDEN, C. D. 2000. Essential tremor: an overdiagnosed condition? *Journal of Neurology,* 247, 955-959.

SCHRAG, A., SAMPAIO, C., COUNSELL, N. & POEWE, W. 2006. Minimal clinically important change on the unified Parkinson's disease rating scale. *Movement Disorders,* 21, 1200-1207.

SEDGWICK, P. 2012. The Hawthorne effect. *BMJ,* 344.

SHANNON, K. M., KESHAVARZIAN, A., DODIYA, H. B., JAKATE, S. & KORDOWER, J. H. 2012. Is alpha-synuclein in the colon a biomarker for premotor Parkinson's Disease? Evidence from 3 cases. *Movement Disorders,* 27, 716-719.

SHIMA, K., TSUJI, T., KAN, E., KANDORI, A., YOKOE, M. & SAKODA, S. Measurement and Evaluation of Finger Tapping Movements Using Magnetic Sensors. 30th Annual International IEEE EMBS Conference, 2008 Vancouver, Britisih Columbia, Canada. 5628-5631.

SIEBNER, H. R., CEBALLOS-BAUMANN, A., STANDHARDT, H., AUER, C., CONRAD, B. & ALESCH, F. 1999. Changes in Handwriting Resulting from Bilateral High-Frequency Stimulation of the Subthalamic Nucleus in Parkinson's Disease. *Movement Disorders,* 14, 964-971.

SKIDMORE, F. M., MACKMAN, C. A., PAV, B., SHULMAN, L. M., GARVAN, C., MACKO, R. F. & HEILMAN, K. M. 2008. Daily ambulatory activity levels in idiopathic Parkinson disease. *Journal of Rehabilitation Research & Development,* 45, 1343-1348.

SMEJA, M., FOERSTER, F., FUCHS, G., EMMANS, D., HORNIG, A. & FAHRENBERG, J. 1999. 24-h Assessment of Tremor Activity and Posture in Parkinson's Disease by Multi-Channel Accelerometry. *Journal of Psychophysiology,* 13, 245-256.

SPIEKER, S., BOOSE, A., BREIT, S. & DICHGANS, J. 1998. Long-Term Measurement of Tremor. *Movement Disorders,* 13, 81-84.

SPIEKER, S., JENTGENS, C., BOOSE, A. & DICHGANS, J. 1995. Reliability, specificity and sensitivity of long-term tremor recordings. *Electroencephalography and Clinical Neurophysiology,* 97, 326-331.

SPIEKER, S., STROLE, V., SAILER, A., BOOSE, A. & DICHGANS, J. 1997. Validity of Long-Term Electromyography in the Quantification of Tremor. *Movement Disorders,* 12, 985-991.

STACY, M. A., MURPHY, J. M., GREELEY, D. R., STEWART, R. M., MURCK, H. & MENG, X. 2008. The sensitivity and specificity of the 9-item Wearing-off Questionnaire. *Parkinsonism & Related Disorders,* 14, 205-212.

STAM, J. & CREVEL, H. 1990. Reliability of the clinical and electromyographic examination of tendon reflexes. *Journal of Neurology,* 237, 427-431.

STANLEY, K., HAGENAH, J., BRÜGGEMANN, N., REETZ, K., SEVERT, L., KLEIN, C., YU, Q., DERBY, C., PULLMAN, S. & SAUNDERS-PULLMAN, R. 2010. Digitized spiral analysis is a promising early motor marker for Parkinson Disease. *Parkinsonism & Related Disorders,* 16, 233-234.

STEBBINS, G. T., GOETZ, C. G., BURN, D. J., JANKOVIC, J., KHOO, T. K. & TILLEY, B. C. 2013. How to identify tremor dominant and postural instability/gait difficulty groups with the movement disorder society unified Parkinson's disease rating scale: Comparison with the unified Parkinson's disease rating scale. *Movement Disorders,* 28, 668-670.

STEG, G., INGVARSSON, P. E., JOHNELS, B., VALLS, M. & THORSELIUS, M. 1989. Objective measurement of motor disability in Parkinson's disease. *Acta Neurologica Scandinavica,* 126, 67-75.

STEINER, C. 2012. *Automate This: How Algorithms Came to Rule Our World,* New York, USA, Penguin.

STOESSL, A. J. 2010. Scans without evidence of dopamine deficiency: The triumph of careful clinical assessment. *Movement Disorders,* 25, 529-530.

STOESSL, A. J., BROOKS, D. J. & EIDELBERG, D. 2011. Milestones in neuroimaging. *Movement Disorders,* 26, 868-978.

STONE, A. A., SHIFFMAN, S., SCHWARTZ, J. E., BRODERICK, J. E. & HUFFORD, M. R. 2002. Patient non-compliance with paper diaries. *British Medical Journal,* 324, 1193-1194.

STONE, A. A., SHIFFMAN, S., SCHWARTZ, J. E., BRODERICK, J. E. & HUFFORD, M. R. 2003. Patient compliance with paper and electronic diaries. *Controlled Clinical Trials,* 24, 182-199.

STREINER, D. L. & NORMAN, G. R. 2008. *Health Measurement Scales: A Practical Guide to their Developement and Use*, Oxford University Press.

TAVARES, A. L. T., JEFFERIS, G. S. X. E., KOOP, M., HILL, B. C., HASTIE, T., HEIT, G. & BRONTE-STEWART, H. 2005. Quantitative Measurements of Alternating Finger Tapping in Parkinson's Disease Correlate With UPDRS Motor Disability and Reveal the Improvement in Fine Motor Control From Medication and Deep Brain Stimulation. *Movement Disorders,* 20, 1286-1298.

TETRUD, J. W. & LANGSTON, J. W. 1989. MPTP-induced parkinsonism as a model for Parkinson's disease. *Acta Neurologica Scandinavica,* 80, 35-40.

THIELGEN, T., FOERSTER, F., FUCHS, G., HORNIG, A. & FAHRENBERG, J. 2004. Tremor in Parkinson's disease: 24-hr monitoring with calibrated accelerometry. *Electromyography & Clinical Neurophysiology,* 44, 137-46.

TINCELLO, D. G., WILLIAMS, K. S., JOSHI, M., ASSASSA, R. P. & ABRAMS, K. R. 2007. Urinary Diaries: A comparison of data collected for three days versus seven days. *Obstetrics & Gynaecology,* 109, 277-280.

TODES, C. J. & LEES, A. J. 1985. The pre-morbid personality of patients with Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry,* 48, 97-100.

TOMLINSON, C. L., STOWE, R., PATEL, S., RICK, C., GRAY, R. & CLARKE, C. E. 2010. Systematic review of levodopa dose equivalency reporting in Parkinson's disease. *Movement Disorders,* 25, 2649-2653.

TRENKWALDER, C., KOHNEN, R., HÖGL, B., METTA, V., SIXEL-DÖRING, F., FRAUSCHER, B., HÜLSMANN, J., MARTINEZ-MARTIN, P. & CHAUDHURI, K. R. 2011. Parkinson's disease sleep scale—validation of the revised version PDSS-2. *Movement Disorders,* 26, 644-652.

TSIPOURAS, M. G., TZALLAS, A. T., FOTIADIS, D. I. & KONITSIOTIS, S. On automated assessment of Levodopa-induced dyskinesia in Parkinson's disease. Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE, Aug. 30 2011-Sept. 3 2011 2011. 2679-2682.

TSIPOURAS, M. G., TZALLAS, A. T., RIGAS, G., TSOULI, S., FOTIADIS, D. I. & KONITSIOTIS, S. 2012. An automated methodology for levodopa-induced dyskinesia: Assessment based on gyroscope and accelerometer signals. *Artificial Intelligence in Medicine,* 55, 127-135.

TUCHA, O., MECKLINGER, L., THOME, J., REITER, A., ALDERS, G. L., SARTOR, H., NAUMANN, M. & LANGE, K. W. 2006. Kinematic analysis of dopaminergic effects on skilled handwriting movements in Parkinson's disease. *Journal of Neural Transmission,* 113, 609-623.

TUDOR HART, J. 1971. The Inverse Care Law. *The Lancet,* 297, 405-412.

VALADAS, A., COELHO, M., MESTRE, T., GUEDES, L. C., FINISTERRA, M., NORONHA, A., ROSA, M. M., SAMPAIO, C. & FERREIRA, J. J. 2011. What motivates Parkinson's disease patients to enter clinical trials? *Parkinsonism & Related Disorders,* 17, 667-671.

VAN HILTEN, J. J., BRAAT, E. A. M., VAN DER VELDE, E. A., MIDDELKOOP, H. A. M., VAN DIJK, J. G., LIGTHART, G. J. & ROOS, R. A. C. 1995. Hypokinesia in Parkinson's disease: Influence of age, disease severity, and disease duration. *Movement Disorders,* 10, 424-432.

VAN HILTEN, J. J., HOFF, J. I., MIDDELKOOP, H. A. M. & ROOS, R. A. C. 1994. The clinimetrics of hypokinesia in Parkinson's disease: subjective versus objective assesment. *Journal of Neural Transmission,* 8, 117-121.

VAN HILTEN, J. J., HOOGLAND, G., VAN DER VELDE, E. A., MIDDLEKOOP, H. A. M., KERKHOF, G. A. & ROOS, R. A. 1993a. Diurnal effects of motor activity and fatigue in Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry,* 56, 874-877.

VAN HILTEN, J. J., HOOGLAND, G., VAN DER VELDE, E. A., VAN DIJK, J. G., KERKHOF, G. A. & ROOS, R. A. C. 1993b. Quantitative Assessment of Parkinsonian Patients by Continuous Wrist Activity Monitoring. *Clinical Neuropharmacology,* 16, 36-45.

VAN HILTEN, J. J., KABEL, J. F., MIDDLEKOOP, H. A. M., KRAMER, C. G. S., KERKHOF, G. A. & ROOS, R. A. C. 1993c. Assessment of response fluctuations in Parkinson's disease by ambulatory wrist activity monitoring. *Acta Neurologica Scandinavica,* 87, 171-177.

VAN HILTEN, J. J., MIDDELKOOP, H. A., KERKHOF, G. A. & ROOS, R. A. 1991. A new approach in the assessment of motor activity in Parkinson's disease. *Journal of neurology, neurosurgery, and psychiatry,* 54, 976-979.

VAN HILTEN, J. J., MIDDLEKOOP, H. A. M., KUIPER, S. I. R., KRAMER, C. G. S. & ROOS, R. A. C. 1993d. Where to record motor activity: an evaluation of commonly used sites of placement for activity monitors. *Electroencephalography and Clinical Neurophysiology,* 89, 359-362.

VAN HILTEN, J. J., VAN EERD, A. A., WAGEMANS, E. A. H., MIDDELKOOP, H. A. M. & ROOS, R. A. C. 1998. Bradykinesia and hypokinesia in Parkinson's disease: what's in a name? *Journal of Neural Transmission,* 105, 229-237.

VAN ROODEN, S. M., COLAS, F., MARTÍNEZ-MARTÍN, P., VISSER, M., VERBAAN, D., MARINUS, J., CHAUDHURI, R. K., KOK, J. N. & VAN HILTEN, J. J. 2011. Clinical subtypes of Parkinson's disease. *Movement Disorders,* 26, 51-58.

VAN SOMEREN, E. J. 1997. Actigraphic monitoring of movement and rest-activity rhythms in aging, Alzheimer's disease, and Parkinson's disease. *IEEE Transactions on Rehabilitation Engineering,* 5, 394-8.

VAN SOMEREN, E. J., VAN GOOL, W. A., VONK, B. F., MIRMIRAN, M., SPEELMAN, J. D., BOSCH, D. A. & SWAAB, D. F. 1993a. Ambulatory monitoring of tremor and other movements before and after thalamotomy: a new quantitative technique. *Journal of the neurological sciences,* 117, 16-23.

VAN SOMEREN, E. J. W., LAZERON, R. H. C., VONK, B. F. M., MIRMIRAN, M. & SWAAB, D. F. 1996. Gravitational artefact in frequency spectra of movement acceleration: implications for actigraphy in young and elderly subjects. *Journal of Neuroscience Methods,* 65, 55-62.

VAN SOMEREN, E. J. W., PTICEK, M. D., SPEELMAN, J. D., SCHUURMAN, P. R., ESSELINK, R. E. & SWAAB, D. F. 2006. New Actigraph for Long-Term Tremor Recording. *Movement Disorders,* 21, 1136-1143.

VAN SOMEREN, E. J. W., VAN GOOL, W. A., VONK, B. F. M., MIRMIRAN, M., SPEELMAN, J. D., BOSCH, D. A. & SWAAB, D. F. 1993b. Ambulatory monitoring of tremor and other movements before and after thalamotomy: A new quantitative technique. *Journal of the Neurological Sciences,* 117, 16-23.

VAN SOMEREN, E. J. W., VONK, B. F. M., THIJSSEN, W. A., SPEELMAN, J. D., SCHUURMAN, P. R., MIRMIRAN, M. & SWAAB, D. F. 1998. A new actigraph for long-term registration of the duration and intensity of tremor and movement. *Biomedical Engineering, IEEE Transactions on,* 45, 386-395.

VERDECCHIA, P., STAESSEN, J. A., WHITE, W. B., IMAI, Y. & O'BRIEN, E. T. 2002. Properly defining white coat hypertension. *European Heart Journal,* 23, 106-109.

VIERA, A. J. & GARRETT, J. M. 2005. Understanding Interobserver Agreement: The Kappa Statistic. *Family Medicine,* 37, 360-363.

VITALE, C., PELLECCHIA, M. T., GROSSI, D., FRAGASSI, N., CUOMO, T., DI MIAO, L. & BARONE, P. 2001. Unawareness of dyskinesias in Parkinson's and Huntington's diseases. *Neurological Sciences,* 22, 105-106.

WAGHORN, A. & MCKEE, M. 1999. Surgical outpatient clinics: are we allowing enough time? *International journal for quality in health care : journal of the International Society for Quality in Health Care / ISQua,* 11, 215-219.

WAGLE SHUKLA, A., OUNPRASEUTH, S., OKUN, M. S., GRAY, V., SCHWANKHAUS, J. & METZER, W. S. 2012. Micrographia and related deficits in Parkinson's disease: a cross-sectional study. *BMJ Open,* 2.

WALKER, R. W., HAND, A., JONES, C., WOOD, B. H. & GRAY, W. K. 2010. The prevalence of Parkinson's disease in a rural area of North-East England. *Parkinsonism & Related Disorders,* 16, 572-575.

WALTON, J. 1997. Handwriting changes due to aging and Parkinson's syndrome. *Forensic science international,* 88, 197-214.

WEINTRAUB, D. & BURN, D. J. 2011. Parkinson's disease: The quintessential neuropsychiatric disorder. *Movement Disorders,* 26, 1022-1031.

WENNING, G. & GESER, F. 2004. Diagnosis and Treatment of Multiple System Atrophy: an Update. *Advances in Clinical Neuroscience and Rehabilitation,* 3, 5-9.

WENSING, M. & ELWYN, G. 2003. Methods for incorporating patients' views in health care. *BMJ,* 326, 877-879.

WHITE, D. K., WAGENAAR, R. C., DEL OLMO, M. E. & ELLIS, T. 2007. Test-Retest Reliability of 24 Hours of Activity Monitoring in Individuals With Parkinson's Disease in Home and Community. *Neurorehabilitation and Neural Repair,* 21, 327-340.

WHITE, D. K., WAGENAAR, R. C. & ELLIS, T. 2006. Monitoring Activity in Individuals with Parkinson Disease: A Validity Study. *Journal of Neurological Physical Therapy,* 30, 12-20.

WONG, W. Y., WONG, M. S. & LO, K. H. 2007. Clinical applications of sensors for human posture and movement analysis: A review. *Prosthetics and Orthotics International,* 31, 62-75.

YANG, C.-C. & HSU, Y.-L. 2010. A Review of Accelerometry-Based Wearable Motion Detectors for Physical Activity Monitoring. *Sensors,* 10, 7772-7788.

ZANNI, G. R. 2007. Patient Diaries: Charting the Course. *The Consultant Pharmacist,* 22, 472-482.

ZEUNER, K. E. M., SHOGE, R. O., GOLDSTEIN, S. R. M., DAMBROSIA, J. M. P. & HALLETT, M. M. 2003. Accelerometry to distinguish psychogenic from essential or parkinsonian tremor. *Neurology,* 61, 548-550.

ZWARTJES, D. G. M., HEIDA, T., VAN VUGT, J. P. P., GEELEN, J. A. G. & VELTINK, P. H. 2010. Ambulatory Monitoring of Activities and Motor Symptoms in Parkinson's Disease. *Biomedical Engineering, IEEE Transactions on,* 57, 2778-2786.

Chapter 12. Appendices

12.I   Study Documentation

12.I.i   Participant Information Sheet

Participant Information Sheet

Objective Assessment of Upper Limb Function in Parkinson's Disease

Protocol Reference Number: 11/NE/0264

We would like to invite you to take part in our research study. Before you decide, we would like you to understand why this research is being done and what it would involve for you.

Please take time to read this sheet carefully and do not hesitate to discuss with us if you would like some more information or if there is anything you do not understand.

- Part 1 tells you the purpose of this study and what will happen to you if you decide to take part.

- Part 2 gives you more detailed information about the conduct of the study.

PART ONE

**What is the purpose of the study?**

Parkinson's Disease (PD) is the second most common neurodegenerative condition in the UK and the main symptoms experienced by sufferers are resting tremor (shaking), rigidity (stiffness) and bradykinesia (slowness of movement).

Currently, assessment of treatment response for patients is based on scales such as the Unified Parkinson's Disease Rating Scale (UPDRS), and their change over time. Patient diaries are also used to help assess the effect medication has on symptoms. These methods of measuring symptoms are subjective, and there is great need for a more objective (accurate) assessment method of arm symptoms in PD

Our area of interest involves the use of accelerometry (motion detecting) sensors (figure 1) that are worn on the wrist (figure 2) much like a wrist watch. We believe that our sensor will be able to measure patients' symptoms more accurately. We hope its smaller size, when compared to previous such sensors, will mean that patients will find it less of a nuisance to wear. Our goal is to confirm that the sensor can be used to accurately and reliably monitor patients' symptoms while in their own home. We hope this will give us more information on how PD affects patients'

ability to perform everyday activities and also how well their medications are controlling their symptoms.
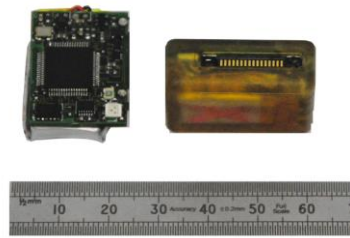
Figure 1

Figure 2

Why have I been invited?

You have been invited as you are a patient currently under the care of the Northumbria Parkinson's Disease service, the team organising patient recruitment into this study. We feel you would be an ideal candidate to potentially trial this sensor, as your condition means you experience the symptoms we are looking to examine using the sensor.

Do I have to take part?

No. It is up to you to decide whether or not to take part. If you do decide to take part we will give you a copy of this information sheet to keep. We will then ask you to sign a consent form. You are free to withdraw at any time and would not have to give a reason. A decision to withdraw at any time, or a decision to not take part, will not affect the standard of the care you receive.

What will happen to me if I take part?

If you are interested in taking part, one of the research team (Dr James Fisher) will be available to answer any further questions you may have about the project when you attend clinic. You will receive a copy of the information sheet to keep. We will then ask you to sign a consent form and give you a copy to keep. On this consent form we will, with your permission, record your contact telephone number. This is to allow us to get in touch with you to arrange your subsequent visit and to contact you during the home monitoring stage of the study to make sure you are not having any problems.

We plan to assess up to 40 people with PD for a period of 7 days while wearing the sensor at home. Firstly participants will attend a morning session (around 5 hours) where the sensor will be explained to them (and their carer if required) and they will receive training in its use.

This will take place in the Human Movement Laboratory at the Clinical Ageing Research Unit (CARU), located at the site of the former Newcastle General Hospital. We will provide you with free taxis to and from the unit, as well as lunch.

Participants will be asked to attend at 9am, having not taken their morning immediate-release levodopa PD medication (this will be explained in more detail beforehand). In CARU we will apply a sensor on each wrist. After taking initial readings we will then ask you to take your

medications (this will allow us to record changes in your symptoms before and after your tablets). Over the course of the morning you will be asked to perform a series of simple actions while wearing the sensor, for example walking or writing. While this is happening we will also, with your permission, take a video recording of the actions performed. These recordings will be examined by an expert doctor in movement problems at a later time, to provide an independent assessment of the severity of your symptoms. During the morning we will also ask you a series of questions, perform a brief examination of you (similar to that which takes place on a clinic visit) and ask for some feedback about the sensor itself.

After this initial visit we will then arrange a convenient time for you to wear the sensor while at home for a period of 7 days. During this period you would be asked to go about your normal daily activities and also to record a diary of symptoms and medications timings. During this period we will ask you to replicate some of the activities that were performed in CARU. At the end of the 7 day period the sensor can be posted back to us in the envelope provided, or collected by the researcher, and the data collected will be analysed.

We plan to carry out the initial assessments between October and December 2011. The results of the study will be available by late 2012. At this time all patients involved in the study will be offered a summary of our findings.

What are the possible risks/disadvantages of taking part?

The sensor itself is worn on the wrist, very much like a wrist watch, and is sterilised between patients; therefore use of the sensor itself carries no risks. The actions we will be asking you to perform at CARU will be low intensity and will be supervised at all times.

What are the possible benefits of taking part?

We would adjust your medications if your patient diary (completed during the home monitoring stage) suggested that your current medication could be adjusted to control your symptoms better. We cannot promise that this study will directly help you, but we hope that the information we gain from this research will help us to develop a better method of assessing the symptoms of PD patients like yourself.

Contact details:

Dr James Fisher. North Tyneside General Hospital. Rake Lane. North Shields. NE29 8NH
Telephone: 0191 293 2709

This completes Part 1 of the Information Sheet.

If the information in Part 1 has interested you, please continue to read the additional information in Part 2 before making any decision.

PART TWO

What will happen if I do not want to carry on with the study?

You can let us know at any time if you do not wish to be a part of the study. Your data will be entirely removed from our records if you request us to do so.

What if there is a problem?

With this type of study, problems are rare, as we will not be experimenting with pills or procedures. It is therefore normal practice in this type of study not to arrange any special arrangements for compensation.

If you have any cause for concern regarding your participation in the study please do not hesitate to contact the researcher (contact details listed in part one). If you still have concerns after this you may wish to complain formally. The NHS operated Patient Advice and Liaison Service (PALS) can provide guidance and assistance with any complaints or concerns: Freephone 0800 0320202. Email: northoftynepals@nhct.nhs.uk. Text/SMS: 01670511098

What happens if I were to lose the sensor?

If you take part in this study you would not be held liable and would not be asked to pay anything should the sensor be lost, stolen or accidentally damaged whilst in your possession.

Will my taking part in this study be confidential?

All information collected about you in this study will be kept confidential. It is our duty in law to protect your personal information and all of our procedures for handling, processing, storing and destroying data are compliant with the Data Protection Act (1998).

Our interview with you will be recorded onto a digital video-recorder. The recording will then be uploaded onto a computer by the researcher. It is important to stress that all of the electronic data will be stored on an encrypted university or NHS computer kept in a locked office.

We will give each recording a code to make it as anonymous as possible. We are unable to completely anonymise the recordings as we need identifiable information so we can match it up to data obtained from the wrist-worn sensor. We will keep copies of the recordings for 2 years after the last person has been assessed (until December 2014). After this they will be destroyed.

Will my General Practitioner (GP) be involved?

We will inform your GP of your participation in our study. If you would prefer us not to do so then your wishes will be respected.

What will happen to the results of the research study?

We will offer all the participants in the study a summary of our findings. The results of the study will be written up as a part of a higher degree at Newcastle University and in addition to this we hope to publish our results in scientific journals and to give oral presentations on our work at conferences in the UK and abroad.

Who is organising and funding the research?

The study is being undertaken as a part of Dr James Fisher's MD degree at the University of Newcastle upon Tyne and also for Mr Nils Hammerla's PhD. The funding is via the Northumberland, Tyne and Wear Comprehensive Local Research Network, Northumbria Healthcare NHS Foundation Trust and Newcastle University's Computing Science Department

Who has reviewed the study?

All research in the NHS is looked at by a Research Ethics Committee (an independent group of people). They are there to ensure that your safety, rights, wellbeing and dignity are protected. This study has been given a favourable opinion by the County Durham and Tees Valley Research and Ethics Committee.

Further information and contact details

You can get more information on this study by contacting Dr James Fisher: North Tyneside General Hospital, Rake Lane. North Shields. NE29 8NH

Telephone: 0191 293 2709

More information regarding research can be found online at the National Research Ethics Service website: http://www.nres.npsa.nhs.uk/patients-and-the-public/

For further independent information about being involved in a research study please contact the Patient Advice and Liaison Service (PALS): Freephone: 0800 0320202; Email: northoftynepals@nhct.nhs.uk ; Text/SMS: 01670511098

12.I.ii   Consent Form

Consent Form

Project Title: **Objective Assessment of Upper Limb Function in Parkinson's Disease**

Name of Researcher: Professor Richard Walker, North Tyneside General Hospital.

Contact number: 0191 293 2709

Please initial boxes

1)   I confirm that I have read and understand the information sheet dated 22nd November 2011, Version 1.5 for the above study. I have had the opportunity to consider the information, ask questions and have had any questions answered satisfactorily.

2)   I understand that my participation in the study is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected.

3)   I understand that in this study I will be recorded onto a digital video recorder.

4)   I agree to withhold the morning dose of my immediate-release levodopa medication on the morning of my attendance at CARU.

5)   I agree to being contacted on the following telephone number.

     ....

6)   I understand that data collected during the study may be looked at by individuals from regulatory authorities or from the NHS Trust where it is relevant to my taking part in this research. I give permission for these individuals to have access to my data

7)   I agree to my GP being informed of my participation in the study

8)   I agree to take part in the above study.

.

....
Name of patient   Date   Signature

....
Name of person taking consent   Date   Signature
   (if different from researcher)

....
Researcher   Date   Signature

12.II   CARU Phase Documentation

12.II.i   MDS-UPDRS

The Movement Disorder Society prohibit reproduction of the MDS-UPDRS.

The MDS-UPDRS can be freely accessed on the internet at the following URL:

http://www.movementdisorders.org/UserFiles/New%20UPDRS%207%203%208%20final.pdf

12.II.ii Modified AIMS

Rate highest severity observed; code:

1   None

2   Minimal, may be extreme normal

3   Mild

4   Moderate

5   Severe

Extremity Movements:

- Upper (arms, wrists, hands, fingers). Include choreic movements (i.e., rapid, objectively purposeless, irregular, spontaneous), athetoid movements (i.e., slow, irregular, complex, serpentine). Do NOT include tremor (i.e., repetitive, regular, rhythmic).
- Lower (legs, knees, ankles, toes). (E.g., lateral knee movement, foot taping, heel-dropping, foot squirming, inversion and eversion of foot.)

Trunk Movements:

- Neck, shoulders, hips (e.g., rocking, twisting, squirming, pelvic gyrations)

Global judgments:

- Severity of abnormal movements:

12.II.iii PDSS-2

Please rate the severity of the following based on your experiences during the past week (7 days). Please make a cross in the answer box.

Choices include: Very often (6 to 7 days a week); Often (4 to 5 days a week); Sometimes (2 to 3 days a week); Occasionally (1 day a week); Never.

	Very often	Often	Sometimes	Occasionally	Never
1. Overall, did you sleep well during the last week?	0	1	2	3	4
2. Did you have difficulty falling asleep each night?	4	3	2	1	0
3. Did you have difficulty staying asleep?	4	3	2	1	0
4. Did you have restlessness of legs or arms at nights causing disruption of sleep?	4	3	2	1	0
5. Was your sleep disturbed due to an urge to move your legs or arms?	4	3	2	1	0
6. Did you suffer from distressing dreams at night?	4	3	2	1	0
7. Did you suffer from distressing hallucinations at night (seeing or hearing things that you are told do not	4	3	2	1	0
8. Did you get up at night to pass urine?	4	3	2	1	0
9. Did you feel uncomfortable at night because you were unable to turn around in bed or move due to	4	3	2	1	0
10. Did you feel pain in your arms or legs which woke you up whilst sleeping at night?	4	3	2	1	0
11. Did you have muscle cramps in your arms or legs which woke you up whilst sleeping at night?	4	3	2	1	0
12. Did you wake early in the morning with painful posturing of arms and legs?	4	3	2	1	0
13. On waking, did you experience tremor?	4	3	2	1	0
14. Did you feel tired and sleepy after waking in the morning?	4	3	2	1	0
15. Did you wake up at night due to snoring or difficulties with breathing?	4	3	2	1	0

12.II.iv MMSE

Reproduction prohibited. Please refer to: Folstein, M.F. and S.E. Folstein, *"Mini-Mental State". A Practical Method for Grading the Cognitive State of Patients for the Clinician.* Journal of Psychiatric Research, 1975. **12**: p. 189-198.

12.II.v Hoehn and Yahr Staging

Stage 0 - no signs of disease

Stage 1 - symptoms on one side only (unilateral)

Stage 2 – symptoms on both sides (bilateral) but no impairment of balance

Stage 3 - balance impairment. Mild to moderate disease. Physically independent

Stage 4 - severe disability, but still able to walk or stand unassisted

Stage 5 - needing a wheelchair or bedridden unless assisted.

12.II.vi MoCA

Reproduction prohibited. MoCA accessible via: mocatest.org

12.III   Home Phase Documentation

12.III.i Home Diary Data Collection Document (abridged: editing removed)

Patient ID:_____

This booklet is your guide to help you during the 7 day period of the study where you will be at home and wearing the movement sensors.

This booklet contains:

- Contact details of the researcher (in case you have any questions or problems)
- Instructions for completing your symptoms, medications, sleep and activity diaries
- A questionnaire to be completed at the end of the week

If you have any problems or questions during the study the researcher, James Fisher, is contactable via 0191 293 2709

Symptom Diary

An important part of this study is understanding what symptoms you have over the course of the 7 day period. Your symptoms diary will help record this. The symptom diary is divided into 1 hour periods. Each day's diary starts at midnight and ends at midnight the following evening. We would like to know how much time you spend in the different Parkinson's states. We would also like you to record when you were asleep.

What are the options to choose from?

- ON: Symptoms of slowness and tremor controlled
- ON with Troubling Dyskinesia: Problems with involuntary twisting, turning movements. (These movements are different from the rhythmic shaking in Parkinson's disease)
- OFF: Problems with stiffness, slowness and tremor
- ASLEEP: Time spent asleep

On the next page is an example of how this can be recorded: [Not reproduced (see Figure 20)] Mark each box with an "X". Remember that you should only mark one answer for each hour period

Medication Diary

Another important part of this study is seeing how your symptoms react to your medications. To get this information we will need to know when you are taking your medications. We would like you to record at what time you take this medication on each day using this diary. We would like you to put the time you

took the medication in the box labelled 'time taken'. Below is an example of how this information should be recorded [Not reproduced (see Figure 21)]

Sleep Diary

We know that sleep disturbances are very common in patients who have Parkinson's disease. We hope to gain information on sleep during this project using the sensors you will be wearing. To do this accurately we will need to know what time you turned the lights off to go to sleep and at what time you got out of bed in the morning. We would also like to know if you had any sleep disturbances during the night, for example getting up to go to the toilet. Below is an example of how this information should be recorded: [Not reproduced (see Figure 22)]

Blank diaries for each day of home monitoring follow thereafter

[not reproduced]

12.III.ii Supplementary Pictures Provided During Home Monitoring Phase
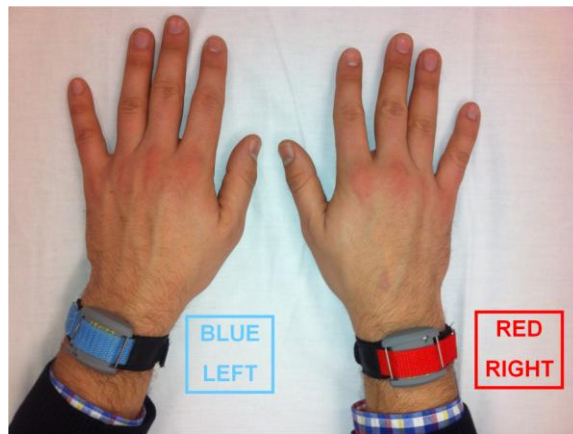
12.III.iii Sensor Questionnaire (used after both study phases)

Sensor Questionnaire

Please complete at the end of the 7 day period

1) **The sensor looks like it is well made**

Strongly Agree   Agree   Disagree   Strongly Disagree

2) **The sensor is comfortable to wear**

Strongly Agree   Agree   Disagree   Strongly Disagree

3) **The sensor feels heavy on my arm**

Strongly Agree   Agree   Disagree   Strongly Disagree

4) **Performing the assessments was made more difficult by wearing the sensor**

Strongly Agree   Agree   Disagree   Strongly Disagree

5) **I would be happy to wear the sensor around the house**

Strongly Agree   Agree   Disagree   Strongly Disagree

6) **I would rather keep a regular diary of my symptoms for a week than wear the sensor for a week**

Strongly Agree   Agree   Disagree   Strongly Disagree

7) **If the sensor was incorporated into a working wrist watch I would be more likely to wear it**

Strongly Agree   Agree   Disagree   Strongly Disagree

8) **The sensor is easy to take on and off**

Strongly Agree   Agree   Disagree   Strongly Disagree

9) **I would be happy to wear the sensor in public**

Strongly Agree   Agree   Disagree   Strongly Disagree

10) **Please write below any comments you have about the sensor:**