Objective evaluation of Parkinson's disease bradykinesia

Dr Jane Elizabeth Alty
MA (Cantab) MB BChir MRCP (UK)

Submitted in accordance with the requirements for the degree of
Doctor of Medicine (MD)

University of Hull and University of York
Hull York Medical School

April 2014

Abstract

Bradykinesia is the fundamental motor feature of Parkinson's disease - obligatory for diagnosis and central to monitoring. It is a complex clinical sign that describes movements with slow speed, small amplitude, irregular rhythm, brief pauses and progressive decrements. Clinical ascertainment of the presence and severity of bradykinesia relies on subjective interpretation of these components, with considerable variability amongst clinicians, and this may contribute to diagnostic error and inaccurate monitoring in Parkinson's disease. The primary aim of this thesis was to assess whether a novel non-invasive device could objectively measure bradykinesia and predict diagnostic classification of movement data from Parkinson's disease patients and healthy controls. The second aim was to evaluate how objective measures of bradykinesia correlate with clinical measures of bradykinesia severity. The third aim was to investigate the characteristic kinematic features of bradykinesia. Forty-nine patients with Parkinson's disease and 41 healthy controls were recruited in Leeds. They performed a repetitive finger-tapping task for 30 seconds whilst wearing small electromagnetic tracking sensors on their finger and thumb. Movement data was analysed using two different methods - statistical measures of the separable components of bradykinesia and a computer science technique called evolutionary algorithms. Validation data collected independently from 13 patients and nine healthy controls in San Francisco was used to assess whether the results generalised. The evolutionary algorithm technique was slightly superior at classifying the movement data into the correct diagnostic groups, especially for the mildest clinical grades of bradykinesia, and they generalised to the independent group data. The objective measures of finger tapping correlated well with clinical grades of bradykinesia severity. Detailed analysis of the data suggests that a defining feature of Parkinson's disease bradykinesia called the sequence effect may be a physiological rather than a pathological phenomenon. The results inform the development of a device that may support clinical diagnosis and monitoring of Parkinson's disease and also be used to investigate bradykinesia.

List of Contents

List of Tables

List of Figures

Dedication

To Carsten

Without your enduring love and patience I could not have done it.

Acknowledgements

A project of this size requires a team effort and I am immensely grateful to everyone who has supported me.

First and foremost thank you to my husband Carsten who has lovingly supported me throughout the process and endured my frequent disappearances to the university library. Thank you to my children Elizabeth and Thomas who have provided lots of laughter, affection, and perspective. Thank you to Mum and Dad for always encouraging me in my aspirations.

I am grateful to the research participants and their families who have given up their time to take part in the study. Without their altruism this thesis would not have been possible.

I am grateful to my supervisors, Dr Stephen Smith and Dr Stuart Jamieson who have supported and guided me through the journey of the MD – from the first meeting when we discussed the potential application of evolutionary algorithms to movement disorders through to the completion of the thesis they have always been encouraging and approachable. They have opened my eyes to the potential benefits of objectively measuring clinical signs and made me question the reliability of diagnostic tests and clinical rating scales; no doubt these skills will remain with me throughout my clinical practice and shape my research career. In particular I have very much enjoyed my brain storming meetings with Steve over some good cups of Costa coffee in York and have appreciated Stuart's careful reading of my chapters and tactful suggestions for improving my grammar.

Thank you to Prof Charles Lacy who has chaired my Thesis Advisory Panel meetings and supported the timely progression and overall direction of the work. His insightful questions and candour have made this a clearer thesis.

Thank you to Dr Mic Lones who has educated me on evolutionary algorithms and classifiers and had immense patience in answering my repetitive requests for clarification on the terminology and interpretation. He has demystified some of the language of computer science for me and always responded very quickly and thoroughly to my queries.

I am grateful to Stuart Lacy whose sharp intellect, attention to detail and perceptive questions about bradykinesia have led to the sequence effect analysis. With our mixed backgrounds in electronics and medicine, we have had some fascinating email conversations and used inter-disciplinary approaches to investigate movement disorders.

I am indebted to Prof Kate Possin at University California San Francisco for being my unofficial third supervisor. She has read through all the abstracts and journal paper submissions and her challenging questions and objective view of the research has improved my understanding of it no end. With regular emails and phone calls she has been central to the research and her background expertise in psychology has opened up new directions that the research may progress in the future. She has also introduced her colleague Prof Norbert Schuff to the research team who has helped with some of the statistical analyses and his plans for correlating the bradykinesia measurements with brain imaging is another exciting future direction.

Thank you to Dr Jeremy Cosgrove for progressing the research further by pursuing some of the studies outlined in Chapter 7. He has also patiently read through various chapters of my thesis.

I am grateful to Kathy Oldfield and Jose Bendezares on the neurology day case unit at Leeds General Infirmary who supported the patients attending the six-hour assessment periods for the dyskinesia study. Thank you to Philippa Duggan-Carter and Paddy Harris who have helped with scoring the patient videos and the community assessments for the dyskinesia study. Thank you too to Christina Harrison and Michelle Charnock who have provided administrative support.

I acknowledge the support of the National Institute for Health Research, through the Dementias and Neurodegenerative Diseases Research Network. Permissions were gratefully received to use the MDS-UPDRS, UDysRS, PDYS-26 and MoCA for the research outlined in this thesis.

Collaborations and roles in research

I acknowledge the specific contributions made my colleagues within the forthcoming chapters of the thesis. In addition I outline their contributions to the research here:

Dr Stephen Smith and Dr Stuart Jamieson
Devised the finger tapping study protocol and gained ethical approval
Read through all chapters of the thesis and suggested edits to the text

Dr Mic Lones
Production of the evolutionary algorithm classifier scores
Production of Figures 23, 41, 45, 46 and 47

Stuart Lacy
Production of the separable component measures
Production of Figures 21, 22, 24, 25, 51 and 52
Execution of the movement component statistical analyses in Chapter 6

Professor Kate Possin and Professor Norbert Schuff
Coordination of the finger tapping study in UCSF
Read through Chapter 6 and suggested edits to the text

Professor Tony Bland
Reviewed the statistical methods used in Chapter 6
Initial statistical analysis for FT study and also LID study

Philippa Duggan –Carter
Allocated clinical grades to the dyskinesia study videos

Cheryl McGurk
Recruitment and assessment of the first five Leeds participants

Author's declaration

I confirm that this work is original and that if any passages or diagrams have been copied from academic papers, books, the internet or any other sources these are clearly identified by the use of quotation marks and the references fully cited. I certify that other than where indicated, this is my own work and does not breach the regulations of HYMS, the University of Hull or the University of York regarding plagiarism or academic conduct in in examinations. I have read the HYMS Code of Practice on Academic Misconduct, and state that this piece of work is my own and does not contain any unacknowledged work from any other sources. I confirm that any patient information obtained to produce this piece of work has been appropriately anonymised.

Chapter 1 Introduction

Overview of Parkinson's disease and aims and objectives of thesis

1.1 What is Parkinson's disease?

Parkinson's disease (PD) is a progressive and incurable neurodegenerative condition that manifests as a clinical syndrome characterised by disordered movement. The predominant motor features are slowness, stiffness, shaking and impaired balance and these are encompassed by the medical terms bradykinesia, rigidity, tremor and postural instability respectively. Distinct pathological changes occur in the brain marked by degeneration of dopaminergic neurons and Lewy body (LB) deposition. The cause is unknown hence PD is interchangeably used with the term 'idiopathic' PD (IPD).

1.1.1 Historical perspective

The oldest surviving references to a syndrome consistent with PD are probably in Ancient Indian Sanskrit texts dating from around 3000 BC.[1] There are also anecdotal reports in ancient Chinese, Greek and Biblical texts[2] but the first detailed clinical descriptions were made in 1817 by an English surgeon called James Parkinson. In An Essay of the Shaking Palsy he described a series of six people, three of whom he had examined and three whom he had simply observed in the streets of London, emphasising that excessive "*shaking*" movements (i.e. tremor) occurred in combination with "*palsy*" or reduced movement (i.e. bradykinesia) (Parkinson, 1817). The early neurologists Charcot, Trousseau, Gowers and Erb subsequently added to and refined the clinical phenotype but James Parkinson's pioneering contribution was recognised and the eminent French neurologist, Jean-Martin Charcot re-named Shaking Palsy to 'Parkinson's disease' in 1877.

[1] In Ayurveda, the Indian system of medicine, there are descriptions of a condition called
[2] Ecclesiastes 12:3 "In the day when the keepers of the house shall tremble…"

Over the next century, the pathological basis for PD was gradually uncovered. In 1912 a German neurologist called Friedrich Lewy described inclusion bodies as the pathological hallmark of PD and in 1919 a Russian neuropathologist, Konstantin Tretiakoff described degeneration of pigmented substantia nigra cells in the brainstem of encephalitic PD cases. In the 1950s Carlsson made the link between loss of dopamine (DA) in the basal ganglia (BG) and PD (Carlsson, 1959).

1.1.2 Clinical phenotype

PD classically presents in people over the age of 60 with the symptoms and signs associated with bradykinesia and at least one of the following: tremor, rigidity and postural instability[3]. Various clinical sub-types of PD have been described based on the age of onset or predominant signs (outlined in Chapter 2) but a general overview of the clinical syndrome is presented in this section.

Bradykinesia is the cardinal motor feature of PD and the only clinical sign that is obligatory for diagnosis. 'Brady - kinesia' translates literally to 'slow - movement' but the gold standard definition of bradykinesia is considerably more complex and also describes movements that are small, dysrhythmic and progressively get slower and smaller with repetition (Gibb and Lees, 1988). Depending on the body part involved, bradykinesia may result in slowed gait with small shuffling steps, poor dexterity, lack of gesticulation, reduced blink frequency and facial expression, a quiet soft voice and difficulty swallowing. As bradykinesia is a multi-faceted clinical sign it may be difficult for clinicians to ascertain whether it is definitely present, especially in the early stages of PD and the reliance on subjective interpretation of this key motor feature has potentially serious ramifications on diagnostic accuracy of PD (Bajaj, Gontu et al., 2010); these issues are discussed further in Chapter 2.

[3] postural instability describes balance problems particularly with the automatic righting reflexes that keep humans standing steady.

Tremor describes a regular rhythmic oscillating movement and affects approximately 70% of people with PD (PwPD). It usually presents in the hands but may also afflict the arms, legs, jaw and tongue. Typically the tremor has a frequency of 4-6 Hz, is most prominent at rest, reduces with action and re-emerges after a brief latent period when a posture is held (Hoehn and Yahr, 1967).

Rigidity affects 89-99% of PwPD and is probably the major contributor to the characteristic flexed posture of PD (Hoehn and Yahr, 1967; Hughes et al. 1992). Rigidity is felt as a steady resistance when the examiner passively moves a patient's limb or other body part and is likened to the sensation of bending a lead pipe. Interestingly the resistance increases further when the patient moves a contralateral limb. A combination of tremor and rigidity in the same body part is described as 'cogwheel rigidity' as there is a jerky rhythmic resistance to passive movement.

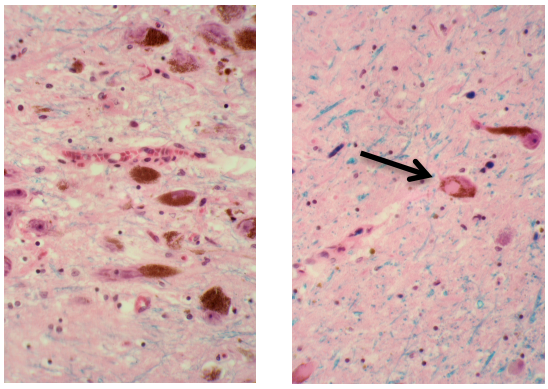
The onset of PD is so insidious that the individual may attribute the early signs to normal ageing. Motor features typically present solely, or predominantly, in one limb and this asymmetry persists throughout the disease course with the side initially affected remaining more symptomatic. The ipsilateral limb becomes symptomatic within a year and then the contralateral limbs approximately two or three years later (Hoehn and Yahr, 1967). For more than a hundred years PD was classified purely as a movement disorder but over the last two decades the non-motor features have been increasingly recognised too. These include hyposmia[4], depression, anxiety, hallucinations, sleep disturbance, autonomic dysfunction, and cognitive impairment. Consequently some clinicians now consider PD to be a neuro-psychiatric disorder.

[4] hyposmia is a reduced sense of smell and may occur many years before the onset of motor signs

1.1.3 Pathology

The two pathological hallmarks of PD are loss of pigmented dopaminergic neurons in the brainstem and Lewy Body (LB) deposition (Figure 1). The pigmented nuclei include the substantia nigra (SN), locus ceruleus and dorsal nucleus of the vagus but SN degeneration is thought to be the predominant cause of PD motor features. LBs are inclusions of alpha-synuclein (α-SN), ubiquitin and neuro-filament proteins contained within the cytoplasm of neurons. The distribution of LBs in PD is specific with only the autonomic nervous system and medium to large monoaminergic and cholinergic neurons affected. Braak et al. have delineated six stages of evolving LB deposition and associated cell death; the first two describe pathology in the dorsal motor nucleus of the vagus nerve and anterior olfactory structures, stages 3 and 4 encompass spread of pathology to the midbrain and BG, and stages 5 and 6 include cortical spread (Braak, Del Tredici et al., 2003).

Figure 1 Parkinson's disease pathology



(a) Normal (b) Parkinson's disease

Legend: Histology of substantia nigra (SN) using haematoxylin and eosin stain and x200 magnification. 1(a) shows normal SN pathology and 1(b) demonstrates neuronal loss and Lewy body (arrow) deposition in a case of PD. *Images provided by Dr Ismail, Consultant Neuro-pathologist at Leeds Teaching Hospitals NHS Trust.*

The trigger for this degenerative process remains unknown but is probably an interaction of environmental and genetic factors causing α-SN to

transform into a toxic protein that aggregates inside neurons. It is hypothesised that the cell responds by forming LBs around α-SN but sometimes this defensive process fails and the neuron dies.

Debate remains as to whether PD is an accelerated form of physiological ageing as LBs have been reported in healthy elderly people with no clinical evidence of PD. Furthermore the number of SN cells diminishes with advancing age in all subjects but in PD the numbers fall faster to approximately 30% of age-matched controls (Fearnley and Lees, 1991). However it is noteworthy that the pattern of cell loss is different in PwPD to healthy controls with the lateral ventral SN more affected in patients and the medial ventral and dorsal aspects more affected in non-PD controls (Fearnley and Lees, 1991).

More recently it has also been recognised that the clinical manifestations of PD are not solely due to a dopaminergic deficit and other neurotransmitters such as acetylcholine and serotonin are also important.

Most cases of PD are sporadic, known as IPD, but first degree relatives have a two-fold increased relative risk for developing PD (Gasser, 1998). Few single-gene mutations that have been identified in familial parkinsonism. It is thought that genetic and environmental factors contribute to final common mechanisms in PD pathogenesis, namely failure of mitochondria and oxidative stress within neurons leading to cell death. It is unclear whether this is due to neurotoxins (i.e. environmental) or impaired anti-oxidant stress systems (i.e. genetic) such as a malfunctioning ubiquitin-proteasome system that normally clears excess or mis-folded proteins, or mitochondrial defects.

1.1.4 Epidemiology

The prevalence of PD in industrialised countries is estimated to be approximately 1-2 per 1000 of the entire population and 1-2 % of the population aged over 60 years (Nussbaum and Ellis, 2003). Men are more frequently affected than women (de Lau and Breteler, 2006). The incidence

increases with advancing age and the mean age of onset is 60 years old but 5% of cases present before the age of 40. Worldwide all ethnic groups and countries are affected although PD seems more common in Caucasians than Asians and Africans (de Lau and Breteler, 2006).

A study using NHS General Practice Research Database records[5] calculated the 2009 UK PD prevalence rate to be 27.4 cases per 10,000 people (30.9 and 24.1 for men and women respectively). This equates to 126,893 PD cases in the UK[6] and is equivalent to approximately 1 in 500 people or 1 in 100 of those aged over 65 as having PD (Parkinson's UK, 2011). With an ageing population the UK prevalence is predicted to increase by almost 30% by 2020 to 161,165 cases. These figures are based on existing medical records and include only those who have sought medical attention. This means that the true prevalence could be considerably higher as door-to-door community studies have shown that up to 25% of PD cases remain undiagnosed (deRijk, Tzourio et al., 1997).

1.1.5 Economics

PD results in economic costs for the patient, their carers, the NHS and social services. Findley, et al. estimated that in the UK just under £600 million was spent each year on the direct costs of 100,000 people with PD which equated to £5993 per patient (Findley, Aujla et al., 2003). As the study was undertaken more than a decade ago and was based on a smaller population of patients than current estimates it is likely that present day costs are considerably higher than this.

Certain sub-groups of PD patients have significantly higher costs: motor fluctuations and dyskinesia double the direct costs (Dodel, Berger et al., 2001) and co-existent dementia triples them (Vossius, Larsen et al., 2011) compared to PD patients without such complications. Advancing age and

[5] This is the world's largest database of anonymised longitudinal medical records, comprising approximately 3.4 million people's records

[6] Approximate number of cases per country: 108,000 in England, 10,000 in Scotland, 5,900 in Wales and 3,000 in Northern Ireland.

increasing severity of disease are also associated with rising costs (Dodel, Singer et al., 1997, LePen, Wait et al., 1999, Whetten-Goldstein, Sloan et al., 1997) in part due to the fact that elderly and more disabled PwPD have increased risk of falls, dementia, hallucinations and residential home admissions (Kempster, O'Sullivan et al., 2010).

There are also indirect costs of PD on the patient, their family and wider society consequent to loss of economic productivity through early retirement of both the patient and their carer. It is crucial that management of PD focuses on reducing functional progression of the disease in order to minimise the economic impact on the individual and the national medical and social care budgets.

1.2 Management of PD

1.2.1 Diagnosis

National UK guidelines recommend that people with symptoms suggestive of PD should be referred untreated to a specialist for diagnosis (NICE, June 2006, SIGN, January 2010). Diagnosis is based on clinical interpretation of the history and examination findings with a focus on eliciting the key motor features, especially bradykinesia, and is described further in Chapter 2.

A number of conditions can mimic PD and it is not always straightforward to make a confident diagnosis, especially in the early stages of the disease. Even amongst movement disorder experts the diagnostic accuracy for differentiating PD from other parkinsonian condition is only about 85% (Bajaj, Gontu et al., 2010, Hughes, Daniel et al., 2002). Functional brain imaging may aid diagnosis in selected cases but the scans are not specific for PD, involve ionising radiation, require an additional visit to another hospital department, and are expensive. Thus there remains a need for a non-invasive diagnostic tool that could be used in clinics to aid diagnosis.

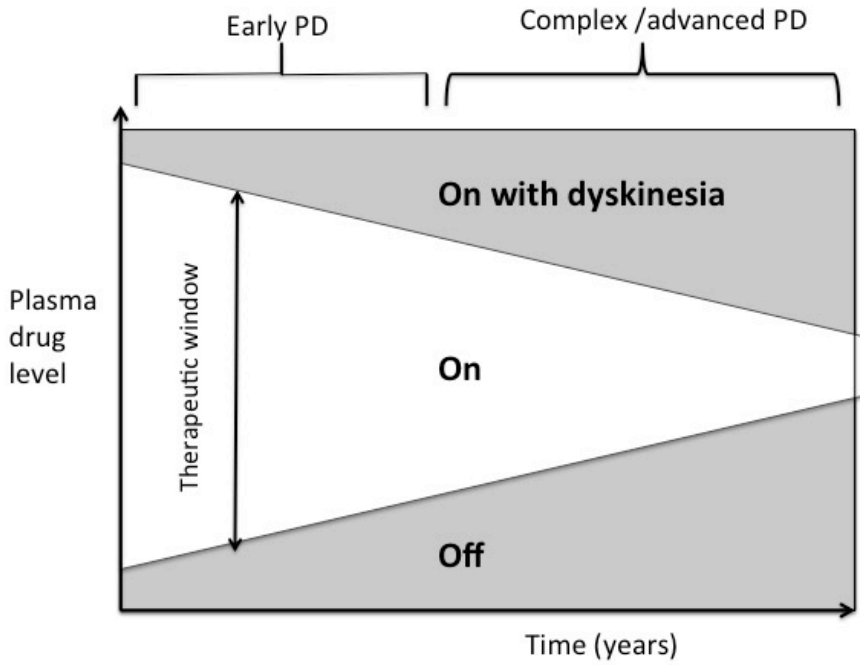
1.2.2 Treatment

A multidisciplinary team approach, with good communication between healthcare professionals and patients, is paramount for effective management of PD (Department of Health, 2005). Usually the neurologist, geriatrician or PD nurse specialist will lead the team and coordinate additional specialist input when necessary. Other professionals involved may include general practitioners (GPs), physiotherapists, occupational therapists, psychiatrists, psychologists, speech and language therapists, dieticians and palliative care physicians.

Currently all PD treatments are symptomatic and none have clinically meaningful neuro-protective effects. The mainstay of PD management is pharmacological with the general aim being to improve symptoms in order to reduce functional disability and maximise quality of life (QOL). This is primarily done through giving drugs to provide maximal time *on* and minimal time *off*. The term *on* describes a clinical state when the PwPD has few or no PD motor symptoms and *off* describes a clinical state when motor features are noticeable.

The 'therapeutic window' refers to the range of drug levels in the blood that provide *on* time for the PwPD (Figure 2). Outside this window the patient will be *off* (sub-therapeutic levels) or *on* with troublesome adverse effects (AEs) (supra-therapeutic levels) and the window becomes progressively narrower as the disease advances. Hence physicians do not attempt to 'normalise' patients by eradicating every symptom as higher doses of medication are associated with an increased risk of short and long-term AEs.

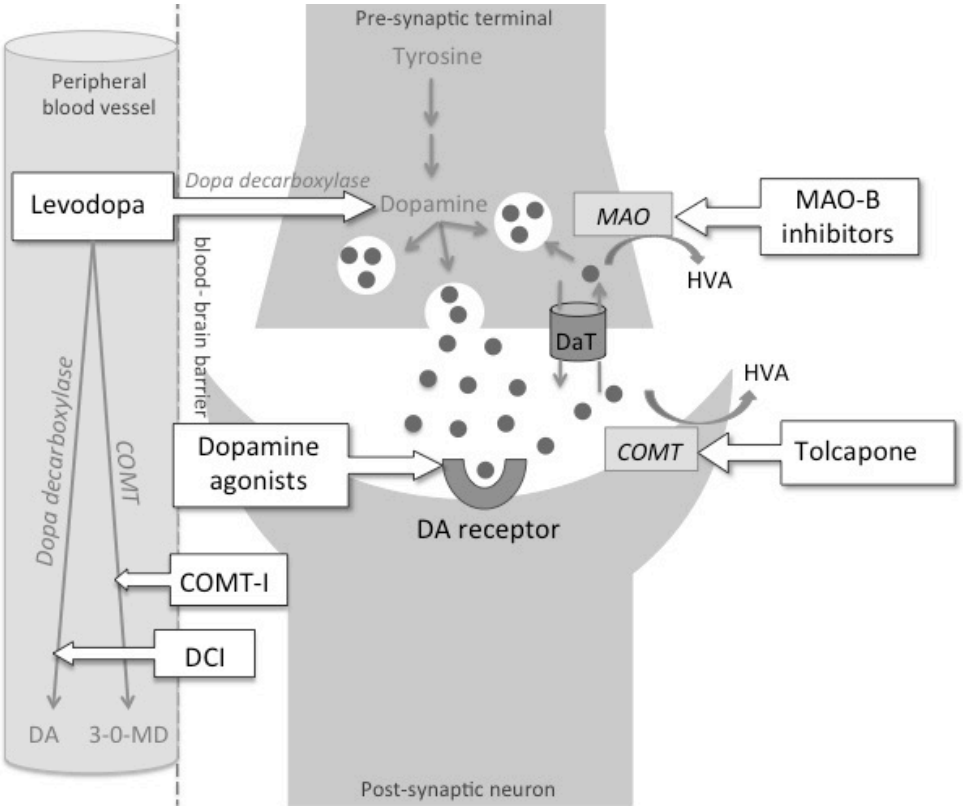
Figure 2 Therapeutic window in Parkinson's disease



Legend: Schematic diagram illustrating the therapeutic window becoming narrower as PD progresses. In advanced stages of the disease the dose of dopaminergic medication may need to be reduced at the cost of losing some motor function because AEs are so troublesome.

The anatomy of the BG are described in Chapter 2 but, in summary, the majority of PD motor impairments are thought to be due to a deficiency of DA in the BG secondary to dopaminergic SN neuronal degeneration. Hence most of the drugs used to treat PD increase stimulation of the post-synaptic dopaminergic receptors although they do this through different mechanisms. These include levodopa, dopamine agonists, monoamine oxidase inhibitors (MAOI) and catechol-O-methytransferase (COMT) inhibitors and their mechanisms of action are outlined in Figure 3.

Figure 3 Schematic diagram of dopaminergic synapse



Legend: DA is synthesised from tyrosine and packaged into synaptic vesicles. When the neuron fires the DA is released into the synaptic space where it activates post-synaptic DA receptors. DA is principally cleared from the synapse by the dopamine active transporter (DaT) and metabolised into homovallinic acid (HVA) by monoamine oxidase-B (MAO-B) and catechol-O-methyltransferase (COMT).

Each type of drug has a different potency and AE profile and the individual's response to a particular drug is not entirely predictable. Consequently there cannot be a 'one size fits all' protocol of PD treatment and the National Institute for Health and Care Excellence (NICE) guidelines remain necessarily ambiguous concluding:

"there is no single drug of choice in the initial pharmacotherapy of early PD" and

"the choice of drug should take into account clinical and lifestyle characteristics of the patients and their preference once they have been informed of the short and long-term benefits and drawback of the drug classes."

(NICE, June 2006)

This emphasises the importance of specialist input throughout the course of the disease in order to monitor progression and tailor effective management to the individual.

The non-motor symptoms of PD such as disorders of mood, sleep, cognition and autonomic function often require separate treatments to the dopaminergic drugs outlined above, and are sometimes exacerbated by the drugs used to treat the motor deficits. In cases of more advanced PD it may be necessary to consider delivering dopaminergic drugs via enteric or subcutaneous routes or referring for deep brain stimulation surgery.

1.2.3 Motor complications

One of the most challenging aspects of managing PD is matching dosing regimens and classes of drug to the individual's symptoms, lifestyle and co-morbidities. This requires a balance between the need to increase doses in order to treat the progressive motor deficits and the need to limit doses in order to reduce the risk of AEs. Levodopa is the most effective drug for reducing motor disability and nearly all patients will eventually require long-term treatment with it. Unfortunately most patients taking this drug develop adverse motor complications and some of these are irreversible.

Motor complications may be sub-divided into motor fluctuations and dyskinesia. The former describes how motor symptoms respond in an unpredictable and undesirable manner to each dose of levodopa so that the period *on* may be delayed, may not occur at all, or may switch suddenly to being *off* again. Dyskinesias, also called 'levodopa induced dyskinesias' (LID), are troublesome involuntary writhing movements that may occur in

any part of the body and be localised or generalized. Approximately 10% of PwPD per year who take levodopa will develop LID (Ahlskog and Muenter, 2001) and these movements are initially dose dependent but with long-term use they occur at progressively lower doses. They lead to considerable disability, reduced QOL and increased costs.

Another important principle of PD management involves the concept of continuous dopaminergic stimulation (CDS). This hypothesis states that late stage motor complications develop due to pulsatile high DA levels stimulating the depleted neurons, leading to erratic firing and involuntary unpredictable movements (Olanow, Obeso et al., 2006). Hence, delivering steady dopaminergic stimulation replicates the normal physiological state more closely and should prevent motor complications. CDS has become an important influence on PD management and is reflected in current practice aiming to prolong the action of dopaminergic stimulation or use frequent but small doses of drugs in an attempt to limit large fluctuations of dopamine in the brain. Certainly, since the recognition that early treatment with high dose levodopa leads to early motor complications, the general shift in the UK has been towards more conservative management of PD.

1.2.4 Monitoring

Regular specialist review is required in order to minimise functional disability through optimising management. The clinician needs to assess, treat and monitor an array of motor and non-motor symptoms and signs, and carefully balance the management of each. Frequently the drugs given for one symptom exacerbate another; for example DA agonists improve bradykinesia but exacerbate hallucinations and levodopa improves tremor but exacerbates dyskinesia, so the clinician must carefully weigh up the need for each medication and its overall effects on the PwPD. Monitoring is usually conducted through six-monthly outpatient clinic appointments and the limitations of this method are discussed in Chapter 2.

1.3 Progression of Parkinson's disease

A PwPD will progress through several clinical phases that may be described as diagnosis, maintenance, complex and palliative or quantified using the Hoehn and Yahr (HY) stages (Table 1 and Figure 4). Although the *order* of progression is predictable the period of time spent in each stage varies depending on PD clinical sub-type and patient characteristics such as age and co-morbidities.

Table 1 Hoehn and Yahr (HY) stages of Parkinson's disease

HY stage	Clinical description
1	Symptoms on one side of body
2	Symptoms on both sides of body
3	Balance impairment and moderate disease
4	Severe disability but able to stand unassisted
5	Wheelchair bound unless assisted

Adapted from Hoehn and Yahr 1967 (Hoehn and Yahr, 1967)

1.3.1 Diagnosis phase

During this phase the patient recognises symptoms, consults the GP and is referred to a specialist. If the clinical signs are subtle there may be a period of watchful waiting, investigations, and diagnostic uncertainty before the diagnosis of PD is made. A period of denial or other psychological adjustment may ensue before the person accepts the diagnosis and decides to commence treatment. Typically the diagnosis period lasts six to 24 months.

1.3.2 Maintenance phase

The commencement of symptomatic therapy usually improves motor symptoms markedly thereby enabling patients to maintain a relatively uncompromised QOL for five to ten years. Patients in this phase have not developed postural instability yet and thus equate to HY stages 1 and 2.

1.3.3 Complex phase

As PD progresses the patient enters a complex phase characterised by a combination of postural instability, dyskinesia, autonomic problems[7], dysphagia[8] and significant psychiatric manifestations. This phase typically lasts 3-5 years during which patients spend increasingly longer periods in the *off* or '*on* with LID' states and hence functional disability progresses (HY stages 3 and 4). Clinicians try to keep the drug effects within the therapeutic window and to increase *on* time (without LID) by splitting drug doses into smaller but more frequent aliquots. Indeed it would not be unusual for patients in this phase to be taking 6-10 doses of three or four different drugs per day. Some patients in this phase may be referred for subcutaneous or intrajejunal infusions of drugs or neurosurgery.

1.3.4 Palliative phase

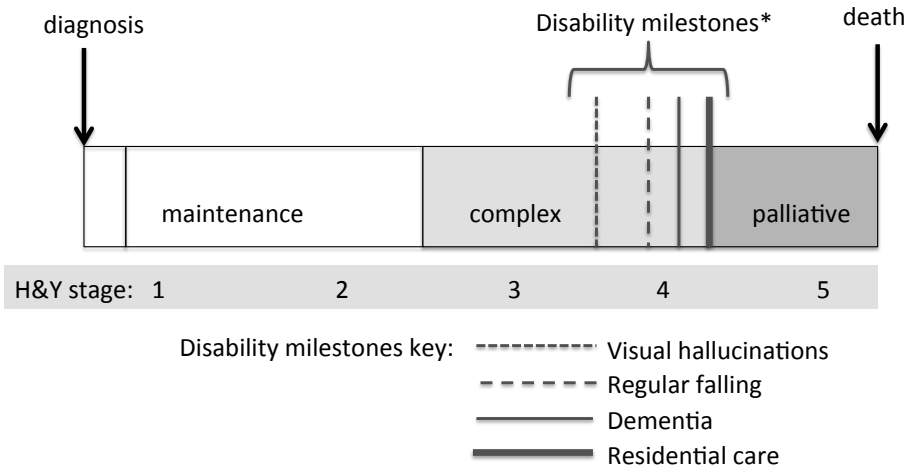
Certain clinical events signal the beginning of the palliative stages of PD: hallucinations, regular falls, dementia and need for residential care and these typically occur within five years of death (Figure 4) (Kempster, O'Sullivan et al., 2010). The palliative phase usually lasts one or two years and is characterised by the patient's inability to tolerate adequate dopaminergic therapy due to a combination of AEs, advanced co-morbidities and dysphagia. Dopaminergic medication may need to be withdrawn due to severe psychiatric side effects at the expense of increased motor disability.

After 15 years of diagnosis 70% of patients will have died and of those that survive 80% have developed frequent falling, 50% dementia, 50% choking, and 40% need residential care (Hely, Morris et al., 2005). The commonest cause of death in PD patients is pneumonia (Diem-Zangerl, Seppi et al., 2009, Hely, Morris et al., 1999, Pennington, Snell et al., 2010) with approximately 45% of PD patients dying from this (Pennington, Snell et al., 2010), presumably related to dysphagia. Cardiovascular disease and malignancy are less common causes of death in PD patient compared to background population rates (Pennington, Snell et al., 2010). Inconsistent

[7] Manifesting as poor control of bladder, bowels and blood pressure
[8] Swallowing problems with consequent increased risk of aspiration pneumonia

documentation of PD on death certificates[9] means that the cause of death in PD may be imprecise though (Pennington, Snell et al., 2010).

Figure 4 Clinical phases of Parkinson's disease



Adapted from Kempster et al. (Kempster, O'Sullivan et al., 2010)

Legend: A schematic diagram illustrating how PD progresses through diagnosis, maintenance, complex and palliative clinical phases and these are approximately aligned with the corresponding HY stage. The time spent in each stage depends on age of onset, PD subtype and other co-morbidities. * There are four key disability milestones that occur within the last five years of life independent of the age of onset.

1.3.5 Rate of progression

Rate of disease progression varies considerably between individuals and is difficult to measure during interval clinical assessments due to the confounding effects of treatment (Maetzler, Liepelt et al., 2009). Examining patients in the placebo arms of drug trials has shown that motor deficits progress faster in untreated than treated cases (Fahn, Shoulson et al., 2004, Siderowf and Parkinson Study, 2004). The rate of progression is probably

[9] Pennington et al found that only 63% of 143 deceased PD patients had PD documented anywhere on their death certificate (20).

not linear though as more rapid progression of symptoms occurs in the earlier stages of disease i.e 5% per year for HY stages 1 to 2.5 and 0.35% per year for HY stages 3-5) (Hely, Morris et al., 2005, Schrag, Dodel et al., 2007) and is consistent with reports of exponential declines in SN counts in PD over time (Fearnley and Lees, 1991). In the Sydney Multicenter Study of PD (Hely, Morris et al., 1999) the time to HY stage 4 was approximately ten years and similar to pre-levodopa cohorts (Marttila and Rinne, 1977) but other post-levodopa era studies suggest much longer latencies, presumably reflecting the heterogeneity of PD clinical subtypes (Table 2).

Prior to the introduction of levodopa the mean survival time of PD was 9.4 years after diagnosis (Hoehn and Yahr, 1967) but in the post-levodopa era the survival of PwPD has been roughly similar to the general population for the first ten years (Hely, Morris et al., 1999). Beyond that period the study results are conflicting but suggest a rise in mortality of between 1.3 (Bennett, Beckett et al., 1996, Diem-Zangerl, Seppi et al., 2009) to 2.5 (Hely, Reid et al., 2008) times the background rate after 20 years.

Table 2 Rate of progression of Parkinson's disease

	HY stage				
Author, year	**1**	**2**	**3**	**4**	**5**
Pre-levodopa era					
Hoehn & Yahr 1967	3	6	7	9	14
Marttila & Rinne, 1977	-	2.9	5.5	7.5	9.7
Post-levodopa era					
Hoehn 1983	-	9	12	12	18
Hely, 1999	-	-	3.5	8	10
Muller, 2000	-	3	5.5	14	15

Adapted from Poewe et al. (Poewe and Mahlknecht, 2009).

Legend: The average number of years to reach each HY stage is shown for several longitudinal studies.

1.4 Aims and objectives of thesis

This chapter has highlighted the need for new tools to aid early diagnosis and accurate monitoring of PD. The incidence of PD is predicted to rise as the population ages and early diagnosis enables early treatment, reduced functional disability and improved QOL. Accurate monitoring of PD facilitates a drug regimen tailored to each individual that maximises therapeutic response without exacerbating side effects.

Bradykinesia is the fundamental movement abnormality in PD - it is the only motor feature obligatory for diagnosis and is central to monitoring. However bradykinesia is a complex clinical sign and ascertainment of its presence and severity relies on subjective interpretation and thus may be imprecise. It remains an incompletely understood clinical feature as there have been few kinematic studies investigating bradykinesia in PD.

The aim of the thesis is:

- to objectively evaluate PD bradykinesia in order to inform the development of a device that could potentially aid clinical diagnosis, monitoring and investigation of PD.

A study was undertaken in Leeds using a commercial movement sensor device to record finger-tapping (FT) movements performed by PD patients and healthy controls (HC). The movement data was analysed using standard statistical measures and purpose written software at the University of York called evolutionary algorithms (EA). The same study was undertaken on a small validation sample of PD and HC individuals in San Francisco and this enabled assessment of whether the Leeds results generalised to an independent group.

The objectives of the thesis are to precisely measure FT movements in PwPD and HC in order to:

- Find out which individual component measures of bradykinesia best discriminate PD patients movements from HC movements.

- Develop composite bradykinesia models using logistic regression to classify movement data into the correct diagnostic group

- Develop classifiers using EA to classify movement data into the correct diagnostic group

- Compare classification accuracy of individual component measures of bradykinesia, composite models of bradykinesia and EA induced classifiers

- Correlate objective measurements of bradykinesia with subjective clinical grades of bradykinesia severity

- Correlate objective measurements of bradykinesia with clinical variables of PD progression

- Investigate the clinically defined characteristic features of PD bradykinesia using objective measures

Chapter 2 Diagnosis and monitoring of Parkinson's disease

2.1 Diagnosis

2.1.1 Introduction

Eight thousand new cases of PD are diagnosed every year in the UK and with an aging population the incidence is predicted to rise (Parkinson's UK, 2011). Accurate diagnosis of PD is the cornerstone of effective management. For the patient it provides information on prognosis, access to appropriate interventions (drugs, physiotherapy) and services (PD clinic, specialist nurse, benefit payments) with subsequent improvement in symptoms and QOL. For health systems it provides efficiency through targeting the correct therapies at appropriate patients and for clinical research it is critical in order to better understand the disease. *Early* diagnosis is preferable but it is during the preliminary stages of the disease that diagnostic accuracy is most uncertain as the characteristic signs, or any atypical signs, tend not to have evolved yet. This chapter outlines how PD is diagnosed, the differential diagnosis, the accuracy of clinical diagnoses, the consequences of mis-diagnosing PD and the diagnostic tools that may aid clinical assessment.

2.1.2 How is Parkinson's disease diagnosed?

2.1.2.1 Overview

Diagnosis of PD is based on clinical interpretation of symptoms and signs elicited through history taking and examination. The focus is on detecting progressive asymmetrical parkinsonism without any atypical features. Parkinsonism is an umbrella term for a clinical syndrome comprising bradykinesia and at least one of the following additional signs: tremor, rigidity and postural instability. Parkinsonism therefore describes the clinical phenotype without specifying the cause and several conditions other than PD may manifest with parkinsonism, or be described as 'parkinsonian' as outlined in section 2.1.5. Diagnosing PD can be very straightforward when at least two cardinal motor signs are present but may be much more difficult in the early stages when the signs are subtle and the various parkinsonian syndromes have more clinical overlap.

2.1.2.2 Diagnostic criteria

The UK Parkinson's Disease Society Brain Bank Clinical Diagnostic Criteria (UKBBDC) were first proposed by Gibb and Lees in 1989 after they analysed the medical notes and pathological findings from 78 subjects considered to have clinical evidence of PD and 191 subjects who had clinical evidence of other parkinsonian syndromes (Gibb and Lees, 1989). In order to meet the diagnosis of "definite clinical PD" the UKBBDC (Table 3) specifies that the subject must have parkinsonism, none of the exclusion criteria and have three or more of the supportive criteria. Clearly some of the supportive criteria cannot be met until several years after initial diagnosis, e.g. 'levodopa response for at least 5 years' and 'clinical course for at least 10 years', perhaps reflecting the fact they arose from clinic-pathological studies and supporting the assertion that diagnostic accuracy improves with the passage of time.

The UKBBDC have become the gold standard for clinical diagnosis of PD and improve clinicopathological correlation when strictly applied (Hughes, Daniel et al., 2001). However clinicopathological (Hughes, Daniel et al., 2002, Hughes, Daniel et al., 2001), longitudinal (Bajaj, Gontu et al., 2010) and imaging (Marshall, Reininger et al., 2009) studies have shown that stringent use of the UKBBDC still does not result in 100% diagnostic accuracy though, and these studies are discussed further in section 2.1.7. It is not known how carefully or frequently the UKBBDC are used in routine clinical practice. Clinical acumen requires interpretation of the individual's history and examination in order to reach *the most likely* diagnosis so pragmatic management may be commenced rather than applying inclusion and exclusion criteria rigidly. It is conceivable that parkinsonian patients could have UKBBDC exclusion criteria but the clinician nevertheless diagnoses PD; for example the patient may have cerebellar signs but the clinician interprets these as related to a history of excessive alcohol, or they may be found to have a cerebral tumour on imaging but this is considered incidental to the clinical presentation.

Table 3 UK Parkinson's Disease Society Brain Bank Clinical Diagnostic Criteria

Step 1: Diagnosis of a parkinsonian syndrome
Bradykinesia **and** ≥ 1 of the following: Muscular rigidity 4-6 Hz rest tremor Postural instability
Step 2: Exclusion criteria for Parkinson's disease
History of strokes with stepwise progression of parkinsonism
History of repeated head injury or definite encephalitis
Oculogyric crises
Neuroleptic treatment at onset of symptoms
More than one affected relative
Sustained remission
Strictly unilateral features after 3 years
Supranuclear gaze palsy or cerebellar signs
Early severe autonomic involvement
Early severe dementia
Babinski sign
Cerebral tumour or communicating hydrocephalus on imaging
Negative response to large doses of levodopa
MPTP exposure
Step 3: Supportive prospective positive criteria for PD (≥ 3 required for diagnosis of definite PD)
Unilateral onset
Rest tremor
Progressive disorder
Persistent asymmetry affecting side of onset most
Excellent response to levodopa
Levodopa induced chorea
Levodopa response for ≥ 5 years
Clinical course of ≥ 10 years

Adapted from Hughes AJ et al. 1992 (Hughes, Daniel et al., 1992)

These sorts of decisions are commonplace in the practice of medicine and the ability to weigh up evidence for and against the likely diagnosis is an essential skill of the accomplished clinician. Nevertheless expert opinions are still subjective and potentially fallible as discussed further in section 2.1.7.

2.1.3 Consequences of misdiagnosing Parkinson's disease

From a patient's perspective a false negative (missed or delayed) diagnosis of PD results in inaccurate prognosis and progressive disability due to a lack of treatment, or inappropriate treatment with potential AEs. This will become even more important if neuroprotective drugs for PD become available. A false positive diagnosis of PD also results in inaccurate prognosis and the risks of anti-parkinsonian drugs being given inappropriately. The consequences of both over- and under-diagnosing PD have a subsequent negative impact on the economics of the health and social systems. In order to minimise misdiagnosis rates, NICE and the Scottish Intercollegiate Guidelines Network (SIGN) recommend that people in the UK with a possible diagnosis of PD are referred untreated to a specialist in movement disorders (NICE, June 2006, SIGN, January 2010).

2.1.4 Clinical subtypes of Parkinson's disease

The clinical manifestations of PD vary considerably and this is acknowledged by the UKBBDC with no requirements to have *all* the features in steps one and three. The diversity of PD phenotypes is such that Marras and Lang have proposed that "*Parkinson's diseases*" may be a more fitting term to encompass the various clinical entities (Marras and Lang, 2008). Several PD subtypes have been described based on motor presentation (tremor dominant and akinetic-rigid), age of onset (young and late onset), cognitive status (dementia, mild cognitive impairment and normal cognition) and rate of progression (rapid and slow).

Two methods have been used to define subtypes – data-driven cluster analysis and empirical classification. The former involves using mathematical techniques to examine the associations between variables in

patient cohorts and then an equation or algorithm developed that groups together clusters of strongly associated variables. The algorithm ideally should then be validated on an independent sample to predict the same clinical subtype. The second method, empirical classification, involves proposing common features of a particular subtype based on clinical observations.

Only two data-driven cluster analysis studies have validated the subtype algorithm on independent samples. In 2002 Gasporili et al. defined rapid and slow progression groups. The latter had an older age of onset, symmetrical parkinsonism and the predominant presenting signs were bradykinesia-rigidity and disturbed gait (Gasparoli, Delibori et al., 2002). In 2006 Schrag et al. defined subtypes based on age of onset with the older onset patient groups having more rapid progression, less dyskinesia and less motor fluctuations (Schrag, Quinn et al., 2006).

The most commonly used subtypes based on empirical classification are 'early versus late onset' using 40 years old as the cut-off, and 'tremor dominant versus non-tremor dominant' with the latter being labelled 'akinetic-rigid' by some authors and 'postural instability gait disorder' (PIGD) by others. Patients may be allocated to motor sub-groups based on clinical judgement or using Unified Parkinson's Disease Rating Scale (UPDRS) motor examination scores. For example Jankovic et al. based the tremor dominant (TD) subtype definition on the UPDRS tremor item score divided by UPDRS 'PIGD' items score (postural instability and gait using examination, and walking, freezing and falls by history) with PIGD defined by a ratio of <=1.0, TD by a ratio of >1.5 and an 'indeterminate' group by a ratio of 1.0 -1.5 (Jankovic, McDermott et al., 1990).

But even this may be an oversimplification. It is intriguing that when Lui et al. examined how data driven subtypes were related to empirically classified motor subtypes that the TD, PIGD, and indeterminate subtypes (as defined by Jankovic) did not classify easily into the four data driven clusters of TD, non-TD, rapid progression and young onset (Liu, Feng et al., 2011). Whilst

the data driven approach may be considered more objective, the traditional motor subtyping derived from empirical classification is most frequently used as they are simpler to apply, rely on commonly used clinical measures and can be easily applied retrospectively to cohorts.

The differential diagnoses of PD differ for the various clinical subtypes as discussed in section 2.1.5. Despite acknowledged clinical heterogeneity of PwPD, most studies assessing PD diagnostic accuracy or response to treatment have not sub-typed the subjects. Herein lies a major drawback of research into aspects of PD diagnosis: the accuracy of clinical skills, UKBBDC or ancillary tests for diagnosing one particular PD subtype may be diluted by the mixture of clinical phenotypes within a cohort.

2.1.5 Differential diagnosis of Parkinson's disease

The conditions most commonly misdiagnosed as PD vary depending on the expertise of the clinician: specialists tend to confuse PD with other degenerative atypical parkinsonian syndromes (Hughes, Daniel et al., 2002) whereas non-specialists tend to mistake PD for Essential Tremor and secondary causes of parkinsonism (Meara, Bhowmick et al., 1999). These findings may reflect referral bias though and are discussed further in Section 2.1.7. The conditions frequently included in the differential diagnosis of PD are described below and the discriminating clinical features highlighted.

2.1.5.1 Degenerative causes of parkinsonism

There are a number of progressive neurodegenerative conditions that present with parkinsonism and may mimic PD. The commonest of these are progressive supranuclear palsy, multiple system atrophy, cortical basal degeneration, dementia with Lewy bodies and Alzheimer's disease. Apart from Alzheimer's disease, they are all rarer than PD but may be clinically indistinct in the early stages of disease.

The diagnosis typically becomes clearer over time as these conditions respond poorly, or only temporarily to dopaminergic drugs, have a more rapid rate of clinical deterioration and additional signs that are atypical of PD emerge. Clinicopathological series suggest that 60% of cases with the final clinical diagnosis of degenerative parkinsonian syndrome had had their diagnosis revised, and 60% of the revised cases were initially diagnosed with PD (Hughes, Daniel et al., 2002). This section is not exhaustive and there are many other degenerative causes of parkinsonism such as Huntington's disease, Wilson's disease, fragile X tremor ataxia syndrome, neuroacanthocytosis, and spinocerebellar ataxias but these are not included as they are much less commonly misdiagnosed as PD, or visa versa.

Progressive supranuclear palsy

Progressive supranuclear palsy (PSP) is characterised by symmetrical parkinsonism, cognitive impairment, vertical supranuclear gaze palsy, early falls, and bulbar dysfunction. Two clinical phenotypes are recognised – typical PSP (Richardson syndrome), and PSP-parkinsonism, PSP-P. The latter closely resembles akinetic-rigid PD especially in the early stages when the sole presentation may be asymmetrical parkinsonism that partly responds to dopaminergic drugs (Williams, de Silva et al., 2005). Also the age of onset is similar to PD with most cases beginning in the seventh decade. Two large pathological series showed that early bradykinesia was reported in 75% (Williams, de Silva et al., 2005) and 88% (Litvan, Agid et al., 1996) of patients with pathologically confirmed PSP, but closer inspection of bradykinesia in PSP has shown that it is subtly different to bradykinesia in PD as there does not seem to be any decrementing amplitude with repeated movements (Ling, Massey et al., 2012). Other differentiating features of PSP-P are marked axial rigidity and a broad-based gait. Nevertheless these differences may be indistinct in the early stages of the disease and clinicopathological and epidemiological studies show that PSP is frequently confused with PD (Daniel, Debruin et al., 1995).

Multiple system atrophy

There are two clinical phenotypes of multiple system atrophy (MSA): MSA-P with early predominant parkinsonism, and MSA-C with early predominant cerebellar signs. Both subtypes have a degree of autonomic failure and over time cerebellar and parkinsonian signs progress. Early MSA-P is clinically similar to PD with asymmetric parkinsonism, tremor, a positive response to levodopa and LID often present and age of onset is typically 55-60 years. However there are several atypical features that evolve over time: pyramidal and cerebellar signs, early falls, stridor, abnormal respiratory pattern, sleep disturbance and non-sustained levodopa response. In contrast to PD it is also unusual for a patient with MSA to develop dementia or hallucinations (Edwards, Quinn et al. 2008)

Corticobasal degeneration

Corticobasal degeneration (CBD) may initially be confused with PD as it typically presents with rigidity and parkinsonism of one limb in people aged over 60 years, sometimes with a jerky tremor. However as CBD progresses atypical features such as early falls, fixed dystonia, myoclonus, the alien-limb phenomenon and asymmetrical cortical syndromes develop such as primary progressive aphasia, apraxia and cortical sensory deficits. There is a rapid progression of symptoms and poor response to levodopa (Edwards, Quinn et al. 2008).

Dementia with Lewy bodies

Dementia with Lewy bodies (DLB) is a progressive dementia syndrome with prominent attentional and visuospatial deficits, marked cognitive fluctuation, visual hallucinations and parkinsonism. PD with dementia (PDD) and DLB share clinical and pathological overlap and debate remains about whether they are different clinical manifestations on the spectrum of LB disease or two separate diseases (Edwards, Quinn et al. 2008). According to the Movement Disorders Society Consensus, in DLB the dementia must occur before, or within the first year, of the onset of parkinsonism (Emre, Aarsland et al., 2007). Therefore good history taking

about the temporal relationship of cognitive and motor symptoms should discriminate DLB from PDD.

Alzheimer's disease

Alzheimer's disease (AD) is the commonest cause of dementia and typically presents initially with progressive cognitive deficits in learning, recall and language. Parkinsonism is well reported in AD (Hughes, Daniel et al., 1992, Morris, Drazner et al., 1989) and conversely AD pathology is frequently found in PD cases. Indeed some authors argue that AD and PD may be two clinical manifestations on a spectrum of neurodegenerative disorders. In AD the motor signs usually develop after the onset of dementia so thorough history taking should differentiate the condition from PD or PDD. Nevertheless atypical forms of AD or the presence of AD and PD as co-existent conditions could cause diagnostic confusion.

2.1.5.2 Secondary causes of parkinsonism

Drug-induced parkinsonism

Drug induced parkinsonism (DIP) is probably the second most common cause of parkinsonism after PD (Wenning, Kiechl et al., 2005). DA receptor blocking drugs, such as anti-psychotic and anti-emetic drugs are the commonest causes, but anti-epileptic and calcium channel blocking drugs can also cause DIP (Table 4). Usually there is a symmetrical akinetic rigid syndrome, but in 40-50% of cases the presentation is asymmetrical and includes rest tremor (Hardie and Lees, 1988, Hassin-Baer, Sirota et al., 2001). Early orofacial dyskinesia is particularly suggestive of DIP over IPD. The symptoms typically improve within a few weeks of withdrawing the offending drug but it may take up to six months or more for complete resolution.

DIP is frequently misdiagnosed as PD even by specialists (Esper and Factor, 2008). For example, Esper et al. showed that neurologists diagnosed PD in

42% of DIP cases and started dopaminergic drugs in these subjects without withdrawing the offending drug (Esper and Factor, 2008).

Table 4 Drugs that can cause parkinsonism

Antipsychotics	Anti-emetics	Others
Chlorpromazine	Metoclopramide	Tetrabenazine
Haloperidol	Prochlorperazine	Lithium
Flupenthixol		Sodium Valproate
Quetiapine		Diltiazem
Clozapine		Flunarizine
Risperidone		Cinnarizine
Olanzapine		
Amisulpride		

Adapted from Edwards M, Quinn N, Bhatia K. Parkinson's Disease and Other Movement Disorders. Oxford University Press. 2008. p217.

Vascular parkinsonism

Vascular parkinsonism (VP) is probably the third commonest cause of parkinsonism but the diagnosis remains somewhat controversial because BG infarcts are commonly found in elderly people with no evidence of parkinsonism. This means that it is difficult to prove the relationship between vascular damage and parkinsonism. Moreover definite diagnosis requires exclusion of LBs as well as demonstration of BG vascular damage so technically VP is a pathological diagnosis. Nevertheless a syndrome of sudden onset of parkinsonism, sometimes with a step-wise progression, with gait disproportionately impaired compared to upper limbs, giving rise to the name 'lower half parkinsonism', is recognised in people with small vessel disease and deep subcortical infarcts. A wide based gait, frequent freezing, symmetrical parkinsonism and early cognitive impairment may point towards a diagnosis of VP rather than PD (Wenning, Kiechl et al., 2005).

Other secondary causes of parkinsonism

There are other less common causes of parkinsonism such as toxins (e.g. carbon monoxide, manganese, solvents), metabolic disorders (e.g. Wilson's disease, hypoparathyroidism and hypoxia), infections (e.g. influenza, HIV and Epstein Barr virus) and normal pressure hydrocephalus (NPH) that may cause symmetrical parkinsonism. These conditions can usually be discriminated from PD by their accompanying neurological or systemic symptoms and a history of illness or exposure prior to the onset of parkinsonism. Space occupying lesions (SOL) such as tumours and vascular malformations of the BG may result in contralateral parkinsonism, often with additional pyramidal signs.

2.1.5.3 Tremulous movement disorders

The classical 4-6 Hz asymmetrical 'pill-rolling' tremor of PD occurs at rest, re-emerges on posture and disappears on action (Jankovic, Schwartz et al. 1999). It contrasts sharply with the 8-10Hz fine amplitude symmetrical postural tremor of Essential tremor (ET) and the typically jerky large amplitude task-specific tremor of dystonia. Therefore one would expect the discrimination of PD from ET or dystonia to be straightforward and often it is. However the tremors may present with less 'stereotypical' phenomenology and severe tremor may also interfere with interpretation of bradykinesia assessments, making it remarkably difficult to clinically separate the disorders especially in the elderly or when mono-symptomatic (Jankovic, Schwartz et al. 1999, Bajaj Schneider et al. 2010)

Essential tremor

ET is a syndrome of bilateral, largely symmetrical tremor of 8-10Hz that affects the hands and forearms and is predominantly postural and to a lesser extent kinetic. Sometimes in severe cases it is also present at rest. It is three times more common than PD and the average age of onset is in the fifth decade (Bain, Findley et al., 1994, Rajput, Offord et al., 1984). The tremor usually occurs immediately on outstretching the arms whereas in PD there may be latency for tremor occurrence; Jankovic termed this a "*re-emergent tremor"* (Jankovic, Schwartz et al., 1999) and Schwingenschuh found this

had 100% specificity (but only 56% sensitivity) for PD compared to other tremulous movement disorders (Schwingenschuh, Schneider et al., 2010). However clinical discrimination of ET from PD on this basis is contentious with Brooks et al. (Brooks, Playford et al., 1992) concluding that the postural upper limb tremors of ET and PD were "*clinically and electrophysiologically indistinguishable*", and Bajaj et al. not finding any significant difference in latency of postural tremor between PD and ET cases (Bajaj, Gontu et al., 2010). To complicate matters further a resting tremor, asymmetry and mild indeterminate parkinsonian features can occur in ET (Rajput, Rozdilsky et al., 1993, Chaudhuri, Buxton-Thomas et al., 2005) and conversely an immediate 6-8 Hz postural tremor may occur in PD.

Hence ET and TD-PD can be difficult to distinguish and one study reported that 20% of ET patients were misdiagnosed with PD (Louis, Levy et al., 2001). Reciprocally, when patients with a previous diagnosis of ET were assessed by a specialist, 18% had their diagnosis revised to PD; these patients were more likely to have leg tremor or asymmetric arm tremor than those whose diagnosis of ET was confirmed (Jain, Lo et al., 2006). Other clinical clues in subjects with apparent mono-symptomatic tremor that point towards a diagnosis of ET are alcohol sensitivity, an autosomal dominant family history and vocal tremor (Bain, Findley et al., 1994).

Dystonic tremor

Dystonic tremor (DT) is typically an asymmetric jerky postural and kinetic tremor that affects one upper limb or the neck muscles and is associated with dystonia. However DT may mimic PD as the dystonia may be subtle, the tremor may occur at rest and the time to re-emergence on posture is not significantly different in DT and PD patients (Bajaj, Gontu et al., 2010).

It may be more difficult to interpret true bradykinesia in PD from slowness of movement in DT due to the tremor interrupting the normal flow of movement. Furthermore slow and small movements have been reported in DT patients, but in contrast to PwPD there is no decrement in size or speed of movements (Bhatia, Schneider et al., 2010, Schneider, Edwards et al.,

2007). Conversely, dystonia can occur in untreated PD patients, particularly in young onset autosomal recessive genetic parkinsonism (Khan, Graham et al., 2003).

Scans without evidence of dopaminergic deficit (SWEDDs)

Three large drug trials that used functional dopamine imaging (see sections 2.1.8) as markers of PD progression found unexpectedly that 4-15% of recruited patients with early PD had 'scans without evidence of dopaminergic deficit' (SWEDDs) (Fahn, Shoulson et al., 2004, Parkinson Study Group, 2002, Whone, Watts et al., 2003). Considerable debate ensued as to the likely explanation for this – did these subjects have PD but the scans had not been able to detect the nigrostriatal degeneration or did they have PD without nigrostriatal degeneration or had they been erroneously clinically diagnosed with PD and had a different condition altogether? These subjects have been followed up longitudinally but in the vast majority of cases the scans have remained normal.

Careful examination of the SWEDD subjects has since shown that there is considerable clinical overlap with PD. Swingenschuh et al. compared 25 SWEDDs with asymmetric rest tremor to 25 TD PD patients and the most discriminating clinical features were fatiguing of amplitude and velocity during repetitive finger and leg tasks: a combination of both resulted in a sensitivity/specificity of 0.84/0.96 for PD. Interestingly, dystonia and other components of bradykinesia (reduced speed and amplitude) only had moderate diagnostic accuracy (Swingenschuh, Schneider et al., 2010). In summary SWEDDs is an umbrella term that is applied to subjects with tremulous movement disorders who have overlapping features with PD but normal functional dopamine scans and SWEDD cohorts probably comprise mostly DT and ET subjects.

2.1.6 How is diagnostic accuracy measured?

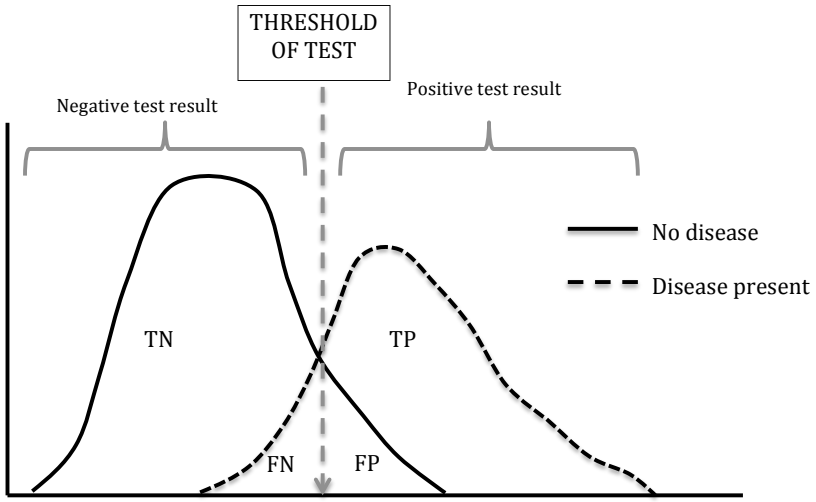
2.1.6.1 True positives and negatives

The diagnostic accuracy of a test is the ability to discriminate diseased cases from normal cases. It is unusual for a clinical test to perfectly separate

subjects with the disease from subjects without the disease as most physiological variables have a range in the diseased group that overlaps with the range in the healthy group. Additionally there may be imperfect discrimination of diseased and healthy cases due to the test having limited ability to detect subtle differences between the groups.

The threshold level is a cut-off point that defines a positive versus negative test result. Ideally it would perfectly distinguish between healthy and diseased groups by allocating a positive test result to all subjects with the disease and a negative result to all subjects without the disease. More often the following scenario occurs though (Figure 5): some of the subjects with the disease will test positive and these are known as true positives (TP), but some subjects *without* the disease also test positive, and these are known as false positives (FP). Similarly there will be some subjects without the disease who have a negative test, known as true negatives (TN) but also some subjects *with* disease who test negative, known as false negatives (FN). Table 5 compares test outcomes with disease states to summarise the main aspects of diagnostic accuracy (Altman and Bland, 1994c).

Figure 5 The trade off between sensitivity and specificity



Legend: The threshold of a test determines when a test result is positive i.e. signifying disease is present and when it is negative signifying disease is not present. Higher thresholds (i.e. moving to the right) results in lower

sensitivity (because some true positives are not 'detected') but higher specificity (because more true negatives are included). The opposite is true with lower threshold test cut off points.

Table 5 Comparison of test outcome with disease state

		Disease	
		Present	Absent
Test	Positive	TP	FP
	Negative	FN	TN

Abbreviations: TP, true positive; FP, false positive; FN, false negative; TN, true negative

2.1.6.2 Sensitivity and specificity

The diagnostic accuracy of a test is frequently expressed by the sensitivity and specificity. Sensitivity is the probability of the test being positive when disease is present and can be calculated as:

$$\text{Sensitivity} = TP/ (TP + FN)$$

A test with perfect sensitivity will detect all diseased subjects and this would be described as having 100% sensitivity, or a sensitivity of 1.0. Alternatively the probability of the test being negative when disease is not present is termed the 'specificity' and can be calculated as follows:

$$\text{Specificity} = TN/ (TN + FP)$$

A test with perfect (100% or 1.0) specificity will detect all healthy subjects, allocating a negative test result to them. There is a trade-off between sensitivity and specificity depending on the threshold level. In Figure 5 it can be seen that a higher threshold (moving the level towards the right)

would result in less TP and FP, and more TN and FN. This would have the following effect:

$$\text{Sensitivity} = \downarrow TP/(\downarrow TP + \uparrow FN) = \text{lower}$$

$$\text{Specificity} = \uparrow TN/(\uparrow TN + \downarrow FP) = \text{higher}$$

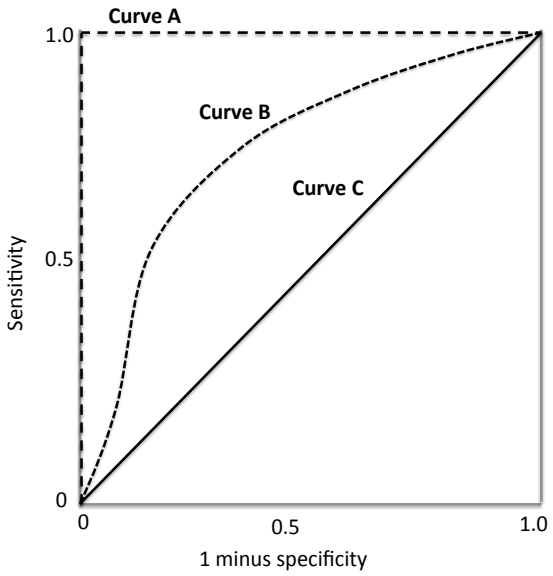
This may be appropriate if it was important to not erroneously diagnose a particular disease in a healthy subject, but the trade off is that some subjects with the disease would be 'missed' and receive a FN result. The reverse pattern would occur if the threshold were moved to the left in Figure 5 i.e. a lower level, resulting in more diseased cases testing positive (i.e. higher sensitivity) but some healthy cases also testing positive (lower specificity) (Altman and Bland, 1994c).

2.1.6.3 Receiver operating characteristic (ROC) curves

Receiver operating characteristic (ROC) curves summarise the sensitivity and specificity trade-offs for the complete range of possible threshold levels and hence can be a succinct method of presenting comprehensive information on diagnostic accuracy.

In a ROC curve the TP rate (sensitivity) is plotted against the FP rate (1 minus specificity) for different cut-off points (Figure 6). Each point on the curve is a sensitivity/specificity pair for a particular cut-off point. This means that the diagnostic accuracies of a test can be evaluated at a range of thresholds. Additionally the area under the ROC curve (AUC) quantifies overall how well the test distinguishes between two groups – it is measured between 0.5 to 1.0 with an AUC of 1.0 implying perfect sensitivity and specificity and an AUC of 0.5 implying no better than chance (50% sensitivity and specificity); this means that direct comparisons may be made between multiple diagnostic tests (Altman and Bland, 1994b).

Figure 6 Receiver Operating Characteristic curves



Adapted from Zou K et al. 2007 (Zou, O'Malley et al., 2007)

Legend: Curve A represents the perfect test with 100% sensitivity and 100% specificity and an AUC of 1.0. Curve B is a typical curve for a diagnostic test showing the trade off between sensitivity and specificity at a range of thresholds, and an AUC of 0.85. Curve C represents a test that has 50% sensitivity and 50% specificity, and an AUC of 0.5.

2.1.6.4 Positive and negative predictive values

The limitation of using measures of sensitivity and specificity is that they do not provide information on how likely a test result gives the correct diagnosis. This is critical information for a diagnostic test to be useful and is given by positive predictive value (PPV) and negative predictive value (NPV). The PPV is the proportion of positive test results that are TP and it may be calculated as:

$$PPV = TP/ (TP + FP)$$

The NPV is the proportion of subjects with a negative test result that are TN and may be calculated as:

$$NPV = TN/ (TN + FP)$$

However PPV and NPV depend on the prevalence of the condition so these straightforward calculations can only be applied when TP and FP are derived from population-based studies. In other types of studies, the prevalence needs to be incorporated into the calculations as follows:

$$PPV = \frac{(sensitivity)(prevalence)}{(sensitivity)(prevalence) + (1\text{-}specificity)(1\text{-}prevalence)}$$

$$NPV = \frac{(specificity)(1\text{-}prevalence)}{(specificity)(1\text{-}prevalence) + (1\text{-}sensitivity)(prevalence)}$$

Thus PPV is directly proportional, and NPV is inversely proportional, to the prevalence of the disease. As PD affects approximately 1% of people aged over 60 years the PPV of a test used for screening in the general population will tend to be low, and the NPV high, even if the sensitivity and specificity are very high. Ideally a test would have high PPV and a high NPV so that one would have confidence that a positive result was indicative of a disease being present and a negative result indicated that the disease was absent but like sensitivity and specificity there tends to be a trade off between PPV and NPV depending on the threshold chosen and the prevalence of the disease in the population being tested (Altman and Bland, 1994a).

2.1.7 Previous studies of clinical diagnostic accuracy in PD

Looking at PD specifically there are two main types of diagnostic error namely over-diagnosis, due to erroneously diagnosing PD in non-PD cases (FP), and under-diagnosis due to *not* diagnosing PD in cases of PD (FN).

The approaches used to study the accuracy of clinical diagnosing PD include assessing clinical diagnosis against pathological findings, expert clinical assessment, and/or imaging and these studies are discussed further in this section.

2.1.7.1 Clinical diagnosis assessed against pathological diagnosis

The gold standard diagnosis of PD is pathological and depends upon finding reduced numbers of pigmented SN neurons and the presence of LBs. Five studies comprising 507 patients have assessed the accuracy of the final clinical diagnosis given by movement disorders specialists before death against the pathological diagnosis (Hughes, Daniel et al., 2002, Hughes, Daniel et al., 1992, Hughes, Daniel et al., 2001, Litvan, MacIntyre et al., 1998, Rajput, Rozdilsky et al., 1991) and two of these, Litvan et al. (Litvan, MacIntyre et al., 1998) and Rajput et al. (Rajput, Rozdilsky et al., 1991) also compared the *initial* clinical diagnosis to the PM findings.

In the 1990s Rajput et al. (Rajput, Rozdilsky et al., 1991) and Hughes et al. (Hughes, Daniel et al., 1992) both published studies that showed clinical diagnosis of PD made by consultant neurologists was confirmed in 76% of cases at autopsy. Rajput et al. (Rajput, Rozdilsky et al., 1991) examined the brains of 65 subjects diagnosed with parkinsonism in life and confirmed the diagnoses pathologically in 31/41 subjects diagnosed with PD. Hughes et al. (Hughes, Daniel et al., 1992) compared the pathological findings of 100 patients diagnosed with PD and confirmed the diagnosis in 76 of these; 24 cases had other pathological diagnoses such as PSP, MSA, AD, and vascular disease. In a later study Hughes et al. retrospectively applied the UKBBDC to the same cases and 11 cases were found not to fulfil the diagnostic criteria for PD and 8 of these had non-PD pathological findings thereby improving the diagnostic accuracy to 82% (Hughes, Daniel et al., 2001). However 16 of the 24 cases with non-PD pathological findings still fulfilled the criteria, highlighting the limitations of diagnosing PD clinically, even with stringent diagnostic criteria. It also raises the question of whether the clinical syndromes of PD may have separable pathologies.

The five pathological studies taken together suggest that clinical diagnosis in the early stages of PD has good sensitivity (at least 0.90) but poor specificity (0.42-0.77) (Hughes, Daniel et al., 2002, Hughes, Daniel et al., 1992, Hughes, Daniel et al., 2001, Litvan, MacIntyre et al., 1998, Rajput, Rozdilsky et al., 1991). Over time, and with progression of the disease the diagnostic accuracy improves to a final diagnosis sensitivity of 0.91-0.94 and specificity of 0.62-0.98. Whilst pathological diagnosis is considered the gold standard reference it is prudent to note that referral bias is a confounding factor in clinic-pathological studies as disproportionately high numbers of atypical parkinsonian cases and akinetic-rigid syndromes are referred.

2.1.7.2 Clinical diagnosis assessed against specialist assessment

In two community studies, movement disorder specialists clinically assessed people previously diagnosed with PD to determine the accuracy of the initial diagnosis (Meara, Bhowmick et al., 1999, Schrag, Ben-Shlomo et al., 2002). In 1999 Meara et al. (Meara, Bhowmick et al., 1999) examined 402 people with presumed PD who were identified through North Wales GP practice electronic records as recipients of anti-parkinsonian medication. Cases of DIP due to neuroleptic drug treatment of mental illness were excluded. GPs had made the diagnosis in 59% of cases and a specialist in the other 41%. The mean PD duration of all cases was 8 years. The subjects were assessed by a specialist in movement disorders and their diagnosis based on history, examination, medical records review and application of UKBBDC.

A diagnosis of definite PD was confirmed in 53% of cases and 'parkinsonism' in 74% of all cases. Of the parkinsonism cases 71% were classified as 'clinically probable PD' (but not strictly fulfilling UKBBDC) 16% as 'other parkinsonian degenerative conditions' such as PSP and MSA, and 13% as DIP. Twenty six per cent of subjects had no evidence of parkinsonism though and the diagnosis was revised to ET, VP and AD in 48%, 36% and 16% of these cases respectively. This study highlighted the

high misdiagnosis rates of PD and parkinsonism in the community despite fairly long disease durations.

In 2002 Schrag et al. (Schrag, Ben-Shlomo et al., 2002) published a similar community study. The electronic records from several London GP practices were used to find all subjects diagnosed with PD or parkinsonism, all people taking anti-parkinsonism medications and all patients with tremor aged over 50 years. Patients with dementia prior to the onset of parkinsonism or those who had received neuroleptic drugs within six months of the onset of parkinsonism were excluded. Two hundred and two patients were identified with 131 of these diagnosed with PD. Seventy eight per cent of subjects diagnosed with PD had previously seen a specialist.

A movement disorders specialist assessed the subjects using history-taking, examination, questionnaires and application of UKBBDC then confirmed the diagnosis of PD in 85% of the 131 patients and rejected it in 15%. The most common conditions that had been diagnosed as PD were VP, PSP, non-PD tremor, and MSA. Seventy eight per cent of patients with non-PD diagnoses had this confirmed by the specialist. This study showed that although the majority of patients had previously seen a specialist, other parkinsonian disorders and tremulous movement disorders were frequently confused with PD.

In both studies there are several limitations. Firstly, the method of ascertainment did not account for undiagnosed subjects (and also untreated subjects in Meara's study) and door-to-door prevalence studies have shown that this group may account for 12-60% of all people with PD (deRijk, Tzourio et al., 1997, Schoenberg, Anderson et al., 1985). Also the exclusion criteria will have resulted in under-representation of AD, DLB and DIP. The reference standard was specialist clinical diagnosis rather than pathological or other objective assessments such as functional DA imaging. This means that the standard against which clinical diagnoses have been assessed is also subjective and imperfect: the clinicopathological studies described in section 2.1.7.1 demonstrated that even movement disorders specialists

strictly applying UKBBDC do not have 100% diagnostic accuracy. Indeed in Bajaj et al. a study (Bajaj, Gontu et al., 2010), which is outlined below, the inter-rater reliability of two movement disorders specialists was remarkably low (kappa coefficient 0.24) even when accounting for the fact that the patients were 'difficult' clinically indeterminate cases and diagnoses were based on videoed examinations.

2.1.7.3 Clinical diagnosis assessed against functional dopamine imaging

Dopamine active transporter (DaT) scans give a measure of the number of presynaptic dopaminergic terminals in the striatum. The uptake of dopamine ligands is reduced in PD and other parkinsonian conditions relative to age matched controls and are discussed further in section 2.1.8.

In 2009 Marshall et al. published the results of a longitudinal multi-centre study that compared clinical diagnosis to DaT scan results. Ninety-nine patients who had clinically indeterminate parkinsonism or tremor were assessed using a videoed UPDRS motor examination and a DaT scan at 0, 18, and 36 months after referral (Marshall, Reininger et al., 2009). The reference diagnosis was a consensus opinion of two movement disorders specialists (blinded to DaT scan results) made at 36 months using the videoed UPDRS examinations and all previous clinical information. The authors compared the initial clinical diagnosis (0 months) made by two raters of the videoed assessment (who were blinded to the DaT scan result) to this reference diagnosis. They showed that 54% of patients diagnosed as non-PD at 3 years were diagnosed with PD initially. The sensitivity/ specificity values for initial clinical diagnosis were 0.93/ 0.46 and for initial DaT scan imaging were 0.78/ 0.97. So compared to imaging the initial clinical assessment had higher sensitivity but lower specificity resulting in PD being over-diagnosed. The sensitivity and specificity of the diagnosis improved at 18 and further at 36 months confirming that clinical diagnosis becomes more accurate over time.

Bajaj et al. examined how accurately two movement disorders specialists distinguished TD PD patients from other tremulous movement disorders

(Bajaj, Gontu et al., 2010). The specialists (blinded to history and investigation results) viewed videos of 38 subjects being examined using the motor section of UPDRS and allocated a diagnosis of PD or non-PD with the latter diagnostic group sub-divided into ET, DT, mono-symptomatic rest tremor and atypical tremor. These diagnoses were compared to the clinical diagnosis made by a third movement disorder expert, Bajaj, who had based it on history, examination, DaT scan results, long term follow up over three years and response to dopaminergic medication.

The sensitivity/ specificity for diagnosing PD was 0.72/ 0.86 and 0.93/ 0.79 for each specialist respectively and they had lower inter-rater agreement. The most common reason for erroneously diagnosing PD, in 8 out of 10 false positive diagnoses, was mis-interpretation of the bradykinesia items of the UPDRS especially in DT cases. It is likely that this study under-estimated the diagnostic accuracy of specialists though as these cases were not representative of routinely seen clinical cases, all being clinically indeterminate initially. Additionally the experts were hindered by the lack of 'hands on' physical examination which meant they could not examine for rigidity or indeed had any details from the history.

2.1.7.4 Diagnostic accuracy assessed with longitudinal follow-up

It is not surprising that the accuracy of clinical diagnoses improves with the passage of time as symptoms and signs increase in severity and frequency. SIGN highlights this point explicitly to doctors working in NHS Scotland:

"Clinicians should be aware of the poor specificity of a clinical diagnosis of Parkinson's disease in the early stages of the disease, and consider this uncertainty when giving information to the patient and planning management."

(SIGN, January 2010)

Therefore another approach to assessing diagnostic accuracy of PD is to compare the initial diagnosis to diagnoses made during longitudinal follow up. Many of the studies described above have incorporated an element of

this comparing and often using the later diagnosis as the gold standard. For example, Rajput et al. (Rajput, Rozdilsky et al., 1991) retrospectively compared the clinical notes to pathologically confirmed cases of PD and found that the clinical diagnosis was correct in 65% of cases when it was made within five years of symptom onset but this improved to 76% of cases 12 years after symptom onset. In the community studies (Meara, Bhowmick et al., 1999, Schrag, Ben-Shlomo et al., 2002) it could also be argued that some of the initial diagnoses were revised not solely because of the expertise or thoroughness of the specialist assessor but rather because they benefitted from evaluating the disease at a more progressed stage.

Clinical trials also facilitate longitudinal clinical information to be extracted for large groups of subjects: for example in the Deprenyl and Tocopherol Antioxidative Therapies for Parkinson Disease (DATATOP) study 800 patients with mild early parkinsonism were diagnosed with PD and enrolled but 8.9% of these were later reported to have a different diagnosis (Jankovic, Rajput et al., 2000). These studies highlight the importance of reviewing the diagnosis periodically especially for atypical signs that may take some time to evolve.

2.1.7.5. Diagnostic accuracy according to expertise of clinician

In Schrag's study 74% of PD diagnoses had been made by a specialist (neurologist or geriatrician) and 26% by a GP and she showed that GPs overall had lower diagnostic accuracy (sensitivity/specificity of 0.74/0.79) than specialists (0.94/0.65) (Schrag, Ben-Shlomo et al., 2002). The GPs diagnoses of PD had a lower PPV than specialists (0.74 vs. 0.89) but similar NPV (0.77 vs. 0.79). However the authors speculate that the discrepancy in diagnostic accuracy between the groups may be even greater than this because the specialists were likely to have been referred disproportionately more of the difficult cases. On the other hand the specialists were likely to have seen the cases later on and hence have the benefit of the signs being more progressed.

The importance of expertise was emphasised in Hughes et al.'s 2002 clinicopathological study (Hughes, Daniel et al., 2002). In this, 143 brains of patients who had been diagnosed by 'super-specialist' neurologists working at a tertiary referral centre were examined and the pathological diagnosis compared to the clinical diagnosis. The diagnostic accuracy was much higher than in previous series with the clinical diagnosis of PD confirmed in 72/73 cases (sensitivity of 0.91) and there were only seven FNs, (specificity of 0.90).

This study is clearly not representative of most hospitals, or even regional neuroscience centres, though as approximately half the cases had non-PD parkinsonian diagnoses. The authors succinctly summarise the benefits of clinical expertise over diagnostic criteria:

"It is interesting that the diagnostic accuracy exceeded that claimed for most clinical diagnostic criteria and suggests that neurologists with experience in movement disorders are better at correctly eliciting and interpreting key clinical features. This study implies that neurologists with particular expertise in the field of movement disorders may be using a method of pattern recognition for diagnosis which goes beyond that inherent in any formal set of diagnostic criteria."

(Hughes, Daniel et al., 2002)

Furthermore the *types* of diagnostic errors made by specialists and GPs have been shown to be different. In the community studies GPs tended to confuse PD with non-parkinsonian conditions, such as ET, whereas specialists tended to confuse PD with other neurodegenerative parkinsonian conditions (Meara, Bhowmick et al., 1999, Schrag, Ben-Shlomo et al., 2002). The former is considered a more serious diagnostic error because the conditions respond to different treatments and hence patients would be inappropriately treated.

2.1.8 Diagnostic tools

The various parkinsonian conditions and tremulous movement disorders have most overlap during the early stages so it may be difficult to make a confident diagnosis of PD based solely on clinical assessment. One approach is 'watchful waiting' for a period of time (often 6 to 12 months) until the symptoms have progressed enough to make the diagnosis with more certainty. In cases where this method is unacceptable to the patient or clinician, or the interval review has not helped, ancillary tests may be employed to supplement clinical assessment and aid diagnosis.

In this section the diagnostic tools are discussed in the context of the studies that were undertaken to assess them. It is important to note that most tools have been tested in subjects with established PD so the diagnostic accuracy in early disease may not have been confirmed. Also most subjects involved in the studies have not had their clinical diagnoses confirmed at autopsy, rather the gold standard diagnosis has been established through longitudinal clinical follow up and/or blinded specialist assessment.

For patients in England and Wales NICE guidelines recommend:

"PD should be diagnosed clinically and based on the UK Parkinson's Disease Society Brain Bank Criteria"

but advises that the following ancillary tests may be considered:

"123 I-FP-CIT SPECT should be considered for people with tremor where essential tremor cannot be clinically differentiated from parkinsonism."
"*Structural MRI may be considered for the differential diagnosis of parkinsonian syndromes."*

All other tests outlined below are not recommended in routine clinical setting and remain research tools.

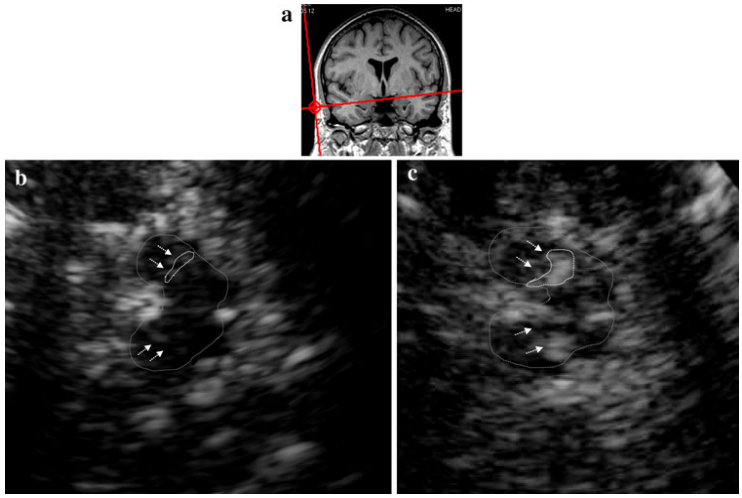
2.1.8.1 Structural imaging

CT brain scans are not routinely recommended as they are usually normal in PD or reveal incidental BG calcification or age-related changes such as generalised atrophy. They may be a helpful investigation if there are other neurological signs suggesting NPH or SOL. Structural MRI brain scans can sometimes help to differentiate PD from other degenerative parkinsonian conditions such as MSA, PSP and CBD. However characteristic MRI changes usually only occur at later stages of the disease, when the clinical diagnosis is already apparent from the atypical physical signs.

Diffusion weighted imaging (DWI) MRI has demonstrated abnormalities in the water content of the putamen in 90% of MSA and PSP cases but is normal in PD (Nicoletti, Lodi et al., 2006). MSA has also been discriminated from PSP on DWI by increased signal in the superior and middle cerebellar peduncles respectively (Paviour, Thornton et al., 2007). These studies suggest that DWI sequences may have potential as a future ancillary diagnostic tool. Other newer MRI techniques have demonstrated abnormalities of the SN of PD patients: 7 Tesla MRI revealed changes in the morphology (Politis, Oertel et al., 2011), and diffusion tensor imaging MRI detected abnormalities of water flow (Vaillancourt, Spraker et al., 2009) but these methods currently remain research tools only.

Transcranial ultrasound (TUS) imaging has revealed increased signal or 'hyperechogenicity' of the SN in 90% of PD patients compared to 10% of HCs (Berg, Siefker et al., 2001); see Figure 7. The test lacks specificity though and other degenerative causes of parkinsonism show high rates of hyperechogenicity too e.g. 40% PSP (Behnke, Berg et al., 2005) and 88% CBD (Walter, Dressler et al., 2004). Also TUS is a highly operator-dependent imaging technique with subjective variation in practice and interpretation and approximately 10% of Caucasians have a temporal bone that does not allow TUS assessment (Berg, 2011).

Figure 7 Transcranial ultrasonography of substantia nigra



Reproduced with permission from Berg D, 2011 (Berg, 2011)

Legend: (b) The midbrain (surrounded by a dotted line) of a HC with normal echogenicity of the SN (encircled ipsilaterally with dotted line and marked with arrows contralaterally); **(c**) In a PwPD the area of hyperechogenicty at the anatomical site of the SN is enlarged (encircled ipsilaterally with dotted line and marked with arrows contralaterally);

2.1.8.2 Functional imaging techniques

Over the last decade, single photon emission computed tomography (SPECT) imaging of the striatal DaT activity has played an important role in supporting, or refuting, a clinical diagnosis of PD. The scan, involves injecting a radionuclide into the patient's vein and then 3-6 hours later performing a SPECT head scan to image the uptake in the nigrostriatal (presynaptic) nerve endings. Two tracers are commonly used – beta-CIT and iodine^{123} labelled ioflupane (FP-CIT).[10] The patient must also be given potassium bromide two hours before and 24 hours after the radionuclide injection to minimise thyroid uptake of radio-iodine and thereby protect against iatrogenic hypothyroidism. The radiation dose per scan is equivalent to one year's background radiation exposure or one CT chest scan. Patients

[10] A DatSCAN is the trade name for SPECT imaging of FP-CIT

are required to stop certain drugs for several days before the scan as they may interfere with FP-CIT uptake and each scan costs approximately £1000 (Bajaj, Hauser et al. 2013, UCL 2013).

The clinical diagnosis of PD is supported if there is reduced specific binding of FP-CIT in the contralateral striatum to the parkinsonian limb. The scans are sensitive to the early stages of PD as there is typically 40 – 50% loss of striatal DaT before the motor symptoms of PD present (Bernheim.H, Birkmaye.W et al., 1973, Kaufman and Madras, 1991), (Benamer, Oertel et al., 2003)). Furthermore abnormal SPECT scans in subjects with olfactory deficit (Ponsen, Stoffers et al., 2004) or rapid eye movement sleep behaviour disorder (RBD)[11] (Postuma, Gagnon et al., 2009, Stiasny-Kolster, Doerr et al., 2005) who later develop PD suggests they are also sensitive in the premotor phase.

However they are not specific for PD and may be positive in other degenerative parkinsonian conditions and also in VP if there is a striatal dopaminergic deficit (Table 6). Some studies have suggested that FP-CIT SPECT scans can differentiate PSP, CBD, MSA and DLB from PD but generally the subjects assessed did not have early stage disease and signs were well developed (Filippi, Manni et al., 2006, Matsui, Udaka et al., 2005) so they are not used clinically for differential diagnosis of degenerative parkinsonian conditions. NICE and SIGN recommend that FP-CIT SPECT imaging is only used to aid diagnosis if there is clinical uncertainty between PD and non-degenerative parkinsonism or non-parkinsonian tremor. Studies report that scans distinguish PD from ET with sensitivity/specificity in the range of 95-97/93-100 (Benamer, Oertel et al. 2003, Marshall, Reininger et al. 2009).

[11] Approximately 50% of people who are diagnosed with REM-sleep behavior disorder in adulthood will develop neurodegenerative parkinsonism with mean latency of 12 years (61)

Table 6 Conditions that have an abnormal FP-CIT SPECT scan

Abnormal scan	Normal scan
PD	ET
PSP	Dystonic tremor
MSA	Drug-induced parkinsonism
CBD	Toxin-induced parkinsonism
DLB	Normal pressure hydrocephalus
VP*	Dopa-responsive dystonia
Huntington's disease	Psychogenic parkinsonism

Table adapted from Brooks DJ. 2012 (Brooks, 2012).

*VP may also have a normal scan depending on the position of the infarcts.

The interpretation of the scan is ultimately subjective as they rely on a radiologist's visual inspection and decision regarding whether there is any asymmetry in the striatal uptake and if the uptake is abnormally low. Although there is not a definite 'cut off' point for abnormality the inter-rater (kappa 0.82-0.92) and intra-rater (0.92-1.00) agreements are good (Benamer, Oertel et al. 2003, Marshall, Reininger et al. 2009) and new quantification computer packages have been developed to improved on this further. The lack of specificity, ionizing radiation dose, and cost warrant caution though when considering who should be imaged. Furthermore, not all clinicians have easy access to scans and certainly on a worldwide basis few countries have any access at all.

SPECT imaging of iodine -123 labelled meta-iodobenzylguanidine (MIBG) binding to cardiac postganglionic sympathetic nerves may be used to differentiate MSA and PSP from PD. Several studies have reported reduced tracer binding in PD but preservation in MSA and PSP and sensitivity/specificity is at least 0.90/0.90 (Yoshita, 1998).

Positron emission tomography (PET) has also been used as a research tool in PD but is not recommend by NICE as a diagnostic tool. The most

commonly used PET isotope is fluorine-18-labelled-dopa (^{18}F fluorodopa) and this has a reduced asymmetric pre-synaptic striatal uptake in PD. PET also has identified pre-clinical nigrostriatal impairment in first degree relatives of PwPD, some of whom developed disease on follow up (Piccini, Burn et al., 1999).

PET imaging using an isotope called 18-flurodeoxyglucose (FDG) shows promise as a diagnostic tool in PD – it is normal in PD but demonstrates striatal hypometabolism in 80% of PSP and MSA cases (Antonini, Kazumata et al., 1998). Another study showed that PET could differentiate PD from MSA, PSP and CBD with sensitivity/specificity of 0.86/0.91, and could differentiate MSA, PSP and CBD from each other with sensitivity/specificity of approximately 0.75/0.92-0.97 (Hellwig, Amtage et al., 2012). However PET scans are expensive, involve high doses of ionizing radiation, and are largely inaccessible to most clinicians.

2.1.8.3 Other tests

Most cases of PD are though to be sporadic due to interactions of environmental stimuli and predisposing genes. However there are an increasing number of genetic mutations recognised that lead to familial parkinsonism. These tests are expensive and generally only used in research except for recessive Parkin gene mutation tests in young onset patients (positive in 5% of those younger that 40 years old) and LRRK2 tests in those with autosomal dominant pedigrees (positive in 5-6% of these cases) (Edwards, Quinn et al. 2008).

Some studies suggest that apomorphine and levodopa challenges may differentiate PD from atypical parkinsonian syndromes (Hughes, Lees et al., 1990, Hughes, Lees et al., 1991). However the diagnostic accuracy of these methods is poor with non-PD conditions such as VP, PSP and MSA sometimes responding well to dopaminergic stimulation. A sustained good response to levodopa is supportive of PD according to UKBBDC but negative results of a drug challenge are clinically not useful for diagnosis

and NICE guidelines advise against using them for purposes of diagnosis. (NICE, June 2006)

Ninety per cent of PwPD have olfactory dysfunction and this is less common in other parkinsonian conditions. A systematic review of two studies that used two objective smell tests, University of Pennsylvania Smell Identification Test (UPSIT) and Sniffin Sticks, concluded that olfactory testing had only moderate sensitivity (0.77) and specificity (0.85) for differentiating PD from other degenerative causes of parkinsonism (McKinnon, Demaerschalk et al., 2007).

Autonomic failure is part of the diagnostic criteria for MSA but as it also occurs frequently in PD, tests of orthostatic hypotension, urodynamics or other autonomic function tests are not discriminatory. Sphincter electromyography (EMG) and heart rate variability may help differentiate PD from MSA but generally neurophysiology tests are not specific for PD and not routinely used to aid diagnosis (Edwards, Quinn et al. 2008).

2.1.9 Summary of diagnosis of Parkinson's disease

There is good evidence that PD is over-diagnosed especially in the early stages of the disease. Community studies suggest that at least 15-26% of people with a diagnosis of PD are FPs and the most common mimics are other degenerative parkinsonian conditions, late onset tremor and VP (Meara, Bhowmick et al., 1999, Schrag, Ben-Shlomo et al., 2002). Pathological studies (Hughes, Daniel et al., 1992, Hughes, Daniel et al., 2001, Rajput, Rozdilsky et al., 1991) broadly support these findings with 10-24% of clinical PD cases not confirmed pathologically and PSP, MSA, VP and AD the most common revised diagnoses. Clinical trials suggest that even when strict clinical diagnostic criteria are fulfilled, 4-15% of patients diagnosed with early PD did not have positive DaT scans. Diagnostic accuracy improves with clinician expertise and the passage of time (Hughes, Daniel et al., 2002, Meara, Bhowmick et al., 1999).

Misdiagnosis may lead to inappropriate management and incorrect prognostication. FP-CIT SPECT scans are helpful for providing evidence of striatal degeneration to support a diagnosis of PD but they are not specific for PD, involve ionising radiation, are expensive and not universally available. Few other diagnostic tests aid early diagnosis of PD and there remains a need for a quick non-invasive objective test that could be administered in a clinical setting to support a clinical working diagnosis of PD.

2.2 Monitoring of Parkinson's disease

2.2.1 Why do Parkinson's disease patients need monitoring?

PD is a heterogeneous condition in terms of clinical manifestations and rate of progression. Also, patients vary in terms of how they respond to the therapeutic and adverse effects of medications, how other comorbidities interact with the disease and its treatment, and how well they can recognise and report the signs and symptoms of PD. In view of such diversity it is impossible to devise one uniform 'optimal PD treatment formula' for all patients – rather the management plan must be tailored to the individual based on their concerns, expectations, functional disability, drug response, disease stage and co-morbidities.

Clinical monitoring is the process that enables tailored management to be achieved - it involves specialist review of the patient with a focus on assessing clinical status, checking response to interventions, vigilance for side effects and surveillance of disease progression and complications. Therefore the first and foremost reason that PwPD are monitored is to optimise their clinical management in order to minimise functional disability.

The second reason why PD patients are monitored is for research. There is no biomarker to determine the severity of PD so monitoring for research purposes typically involves serial detailed clinical measurements of impairment and disability using formal clinical rating scales (see section

2.2.7) in large numbers of PD patients over a period of time. This quantification of clinical status enables aspects of PD such as disease progression and the response to new drugs or surgical procedures to be studied. There are several devices capable of objectively measuring movement characteristics in PD (outlined in section 2.3.7) but these have been used in only a few research studies to aid monitoring.

2.2.2 How are Parkinson's disease patients monitored?

Several methods are used for monitoring PD patients. For clinical purposes the most common method is interval assessment in outpatient clinics. For research studies clinical rating scales are almost exclusively used.

2.2.3 Clinical assessment in outpatient clinics

The vast majority of PD monitoring is conducted through clinical assessments undertaken in regular outpatient clinics. NICE guidelines recommend that PD patients are reviewed by a PD specialist at least every 6-12 months for *"clinical monitoring and medication adjustment."* (NICE, June 2006) The 'specialist' is usually a consultant (or specialist registrar) in neurology or geriatric medicine, or a PD nurse specialist (PDNS). The frequency of review will depend in part on the severity of symptoms and disease stage - when new drugs are commenced or the individual has progressed to complex phase PD more frequent clinic appointments are usually required. The 2011 Parkinson's UK National Audit suggested that most patients meet the NICE recommendations but 12% of neurology patients, and 6% of geriatric patients are seen less frequently than this (Parkinson's UK, 2011).

Typically each patient is allocated a 15 to 20 minute appointment and during this time the specialist will gather information about the severity and impact of PD symptoms, response to drugs and any new relevant co-morbidities. This information is obtained through the clinical skills of history taking and examination. The former describes a process of listening to the patient's description of his symptoms and then asking a series of

questions to obtain clarification and detail. If a family member has accompanied the patient they may provide additional details that the patient may not have noticed or remembered.

The main examination technique used in consultations is probably observation – from the moment the patient walks into the room the specialist closely watches and analyses their movements specifically looking for slowed gait, resting tremor, impassive face, quiet speech or dyskinesia. Therefore most of the examination can usually be done concurrently during history-taking but sometimes supplementary techniques are used to assess certain motor signs; for example the clinician may examine for rigidity by passively moving a limb, or for bradykinesia by asking the patient to repeatedly tap their finger and thumb together as rapidly as possible.

A skilful clinician can thus gather an enormous amount of clinical information during a relatively brief review but this method of monitoring does have some limitations. Physical clinical assessments performed at intervals can only provide a brief ‘snapshot’ of the patient’s status at that particular moment in time and may not accurately reflect their impairment on a day-to-day basis. Increasing the frequency or length of consultations will increase the likelihood of obtaining a more representative view of the patient but still may miss pertinent clinical information. This is because the signs and symptoms of PD fluctuate over hours and days, reflecting characteristics of the disease, response to medications, and other influences such as emotional upset or co-morbidities. This longitudinal view of the patient in their typical environment simply cannot be replicated in a clinic. Making a management decision based on a brief interaction in clinic may have a profound effect on how the patient functions for the next six months or so and yet the treating clinician will not often learn of the response until the next appointment and then only with the same limitations of a brief consultation.

History taking may provide some useful longitudinal information about how PD impairments change through the course of the day or with certain activities. Unfortunately many patients lack insight or objectivity into their clinical status though – for example many are not aware of physical signs such as LID (Vitale, Pellecchia et al., 2001), others may perceive them as disproportionately severe due to low mood, and others may simply not remember details due to the passage of time or cognitive impairment. Therefore even thorough history taking does not guarantee reliable information on which to base important management decisions.

A second drawback is that all assessments within a clinical consultation are subjective; the clinician may be a specialist but nevertheless his interpretation of symptoms and signs is still subjective. This is not to say that subjective assessments are inferior to objective ones, but rather that they are vulnerable to variation between individual clinicians and over time. Even movement disorders specialists struggle to agree on whether movements are bradykinetic or not and the characteristics of tremor (Bajaj, Gontu et al., 2010).

Surveillance of disease progression and response to medications over months and years is thus imprecise with interval subjective assessments. A clinician may document a description of the clinical findings such as ‘mild bradykinesia of right arm’ and it is difficult to know at the next review whether this sign has deteriorated as there is no objective quantification to compare it to. This problem is further compounded if a different clinician reviews the patient at their next consultation.

2.2.4 Patient symptom diaries

Patient-completed symptom diaries are sometimes used in order to gather more clinical information over the course of hours, days or weeks that can supplement the clinic appointment review. They may be useful in a small subset of patients who can diligently document pertinent features at regular intervals. However generally diaries are considered unhelpful as many patients simply find them too onerous to fill out regularly and the data has been shown to correlate poorly with clinician assessments (Golbe and Pae, 1988), perhaps because many PwPD are unaware of their own clinical signs (Amanzio, Monteverdi et al., 2010, Vitale, Pellecchia et al., 2001).

2.2.5 Clinical assessment at home

The 2011 National PD audit suggested that approximately 78% and 91% of patients reviewed in geriatric medicine and neurology clinics respectively have access to a PDNS (Parkinson's UK, 2011). Some PDNSs visit patients at home to clinically assess them and community geriatricians and matrons may provide a similar form of monitoring, especially when patients are in the complex or palliative stages of PD. This provides insight into how individuals function in their own home environments and usually the duration of the monitoring is longer than a clinic appointment. Nevertheless this form of monitoring is still a subjective assessment over a brief period of time.

2.2.6 Inpatient monitoring

Sometimes patients are admitted to the hospital for a period of monitoring over several days if severe parkinsonism or drug side effects cannot be managed through outpatient clinic consultations. Inpatient monitoring enables more detailed and longitudinal assessment of the individual patient. Commonly ‘on-off’ charts are completed hourly so that the temporal relationship of a patient’s clinical state can be correlated with the drug regimen. This enables a more informed decision to be made regarding drug adjustments in order to increase therapeutic effects and reduce AEs.

However this method is very expensive with each day on an NHS ward costing several hundred pounds. Secondly the accuracy of the on-off chart documentation relies on the experience and training of nursing staff to recognise the clinical states accurately and their other roles on a busy ward may preclude hourly documentation. Sometimes semi-quantitative assessments, such as hourly UPDRS motor examinations (see 2.2.7), Timed Get up and Go Tests (TGUGT) or Purdue Pegboard Tests (PPB) are also undertaken but these can be quite burdensome to the patient and time-consuming for staff. Finally the clinical environment may not reflect he patient's usual level of impairment at home.

2.2.7 Clinical rating scales

2.2.7.1 Overview

Since the 1960s a number of clinical rating scales have been devised to quantify the severity of PD. Table 7 summarises the scales that assess motor features. There are also scales that specifically measure the non-motor features of PD such as depression, anxiety, cognition, sleep, apathy and psychosis but these are not discussed further in this thesis.

The Movement Disorders Society Sponsored revision of the UPDRS – the MDS-UPDRS - is considered the gold standard clinical assessment of PD. Apart from HY and previous versions of the UPDRS all of the other clinical rating scales in Table 7 are rarely used nowadays. The rest of section 2.2.7 will focus on evaluating the MDS-UPDRS and UPDRS.

Table 7 Clinical rating scales of Parkinson's disease motor features

Disability	Impairment	Disability & Impairment
Hoehn and Yahr	Webster Scale	UPDRS, MDS-UPDRS
Schwab and England	Columbia University Rating Scale	New York University Scale
	PD Impairment Scale	Short PD Evaluation Scale
		University of California Los Angeles PD Disability Scale

2.2.7.2 Unified Parkinson's Disease Rating Scale (UPDRS)

The UPDRS was published in 1987 as a comprehensive instrument for evaluating impairment and disability in PD patients. This landmark development 'unified' many elements from the array of previously used PD scales in order to allow comparison of study outcomes. It became the most widely used clinical rating scale for PD and a reference for the development of all other measures (Ramaker, Marinus et al., 2002).

In 2008 the UPDRS was revised to the MDS-UPDRS in order to improve clarity of the instructions, reduce ambiguities of grades and incorporate non-motor symptoms (Goetz, Fahn et al., 2007). It is an expansive scale with assessments of disability and impairment in a wide range of motor and non-motor domains and is considered the gold standard clinical assessment of PD. The MDS-UPDRS comprises four parts that grade different aspects of PD:

Part I: non-motor experiences of daily living
Part II: motor experiences of daily living
Part III: motor examination, and
Part IV: motor complications.

Within each part there are a series of items that are graded on a five-point scale of zero to four[12] with higher scores indicating more disability or impairment. Sixty five items are rated with a distribution of 13, 13, 33 and 6 items in parts I to IV respectively, giving a maximum total score of 260. Some items are rated by the patient through self-administered questionnaires or responding to questions administered by the clinician, and other items are clinician-rated based on observation and physical examination. In other words the scale provides information about PD from the patient and clinician perspectives.

The four part MDS-UPDRS scores are the primary outcome measure in most clinical trials of PD therapeutics. Sometimes solely part III, the motor examination, or individual items within Parts III and IV (such as the 'bradykinesia items') are used as the primary outcome measures. The scale is undergoing an extensive translation programme and so far the Spanish, Italian and Estonian versions have been validated and published (Antonini A, Abbruzzese G et al., 2012, Martinez-Martin, Rodriguez-Blazquez et al., 2013).

Overall the scale has been shown to have good clinimetric characteristics. Goetz et al. assessed 877 PD patients with UPDRS and MDS-UPDRS and found the two scales correlated well for total scores (r = 0.96) as well as individual parts (range r = 0.76 (part 1) to 0.96 (part III). There was minimal floor or ceiling effects and factor analysis confirmed that items on the scale clustered in clinically relevant domains (Goetz, Tilley et al., 2008). In 2010 Goetz et al. demonstrated that for part III three movement disorders experts had inter-rater agreement scores (Kendall coefficient of concordance)

[12] 0 = normal, 1 = slight, 2 = mild, 3 = moderate, 4 = severe

ranging from 0.72 for facial expression to 0.99 for gait and postural instability (Goetz, Stebbins et al., 2010). In 2013 Martinez-Martin et al. demonstrated that the MDS-UPDRS subscales correlated well with disease duration and HY disability stage and the motor examination additionally correlated well with age (Martinez-Martin, Rodriguez-Blazquez et al., 2013).

There is no doubt that the UPDRS and MDS-UPDRS have improved the quality of trial outcomes and enabled better comparison of results. However the scales do have several limitations.

2.2.7.3 Limitations

Administration time

The comprehensiveness of the scale necessitates a lengthy administration time of approximately 30 minutes for the whole scale and 15 minutes for the motor examination (Goetz, Fahn et al., 2007). This makes it unwieldy for routine clinical use. Many items will not apply to each individual and the consequent item redundancy results in inefficient use of clinician time and can provoke anxiety in patients who may worry they are eventually going to develop every complication that they have been asked about.

Training

Training is recommended before administering the MDS-UPDRS in order to "*enhance consistency of data acquisition and interpretation"* (Goetz, Stebbins et al., 2010). This is particularly important for research studies when serial patient assessments are typically conducted by different clinicians. The online training course and assessment takes approximately three hours to complete (author's personal experience) and the grades that the trainee allocates to four videoed assessment subjects must be within the 95% confidence intervals of the grades allocated by three movement disorders experts in order to pass and become certified (Goetz, Stebbins et al., 2010).

This formal training still does not assure adequate application of the scale though. Goetz et al. reported that only 55% of 226 raters (comprising 57% professors of neurology, 31% physicians involved with PD and 12% study coordinators) passed the training assessment on the first occasion *after* watching the UPDRS[13] training video (Goetz and Stebbins, 2004).

Also the video assessment does not include any rigidity items as they require physical examination; this means that one cannot ensure uniformity of the rigidity item grades which account for 15% of the total motor examination score. It is also important to note that there was a significant difference between the training video assessment pass rates in North American (62%) and European raters (41%) (Goetz and Stebbins, 2004) and this hints at heterogeneity of scale application in different countries.

Reliability

The reliability of a test is a measure of how similar the results are when the test is repeated either by a different clinician (inter-rater) or by the same clinician on a different occasion (intra-rater). Reliability coefficients express the degree of reliability with perfect agreement designated 1.0 and agreement no better than chance designated as zero.

The bradykinesia-related items of the UPDRS have been shown to have the lowest reliability coefficients of all items with finger tapping less than 0.5 (Camicioli, Grossmann et al., 2001, Henderson, Kennard et al., 1991, Martinezmartin, Gilnagel et al., 1994). When the scores from the bradykinesia items are combined[14] the inter-rater reliability coefficients ranged from 0.0 (Camicioli, Grossmann et al., 2001) to 0.69 (Siderowf, McDermott et al., 2002).

[13] There are no published results for the MDS-UPDRS training video assessment
[14] Extracting motor sub-scores is unvalidated practice but commonly done

There are likely to be several reasons for this variability. Firstly the interpretation of the signs and symptoms are ultimately subjective. Even for the motor examination part - which is often erroneously described as 'objective' - the grades allocated rely on the clinician's subjective interpretation of motor performance. This is expressly acknowledged as a general point in the MDS-UPDRS instructions for this section:

"*The investigator should "rate what you see". Admittedly, concurrent medical problems such as stroke, paralysis, arthritis, contracture, and orthopaedic problems ... may interfere with individual items in the motor examination."* (Goetz, Tilley et al., 2008)

So this means that resolution or progression of these concurrent medical problems between consecutive assessments may result in improving or deteriorating UPDRS scores despite *no* true change in PD signs. Furthermore the definitions of the grades are also open to interpretation. This point is demonstrated by item 3.4 of the motor examination that assesses bradykinesia using a finger tapping (FT) examination administered as follows:

"Instruct the patient to tap the index finger on the thumb 10 times as quickly AND as big as possible. Rate each side [hand] separately, evaluating speed, amplitude, hesitations halts and decrementing amplitude."

(Goetz, Tilley et al., 2008)

The grades are then described as:

0: Normal: *No problems.*

1: Slight: *Any of the following: a) the regular rhythm is broken with one or two interruptions or hesitations of the tapping movement; b) slight slowing; c) the amplitude decrements near the end of the 10 taps.*

2: Mild: *Any of the following: a) 3 to 5 interruptions during tapping; b) mild slowing; c) the amplitude decrements midway in the 10-tap sequence.*

3: Moderate: *Any of the following: a) more than 5 interruptions during tapping or at least one longer arrest (freeze) in ongoing movement; b) moderate slowing; c) the amplitude decrements starting after the first tap.*

4: Severe: *Cannot or can only barely perform the task because of slowing, interruptions or decrements.*

The item acknowledges several components of bradykinesia such as reduced amplitude, reduced speed and dysrhythmia but nevertheless instructs the rater to combine these components into an overall composite score[15]. This means that differential response of the individual components may well be diluted.

Heldman et al. demonstrated that the weight placed on each component when allocating an overall bradykinesia item grade varies considerably between individuals with most clinicians basing their grade on the degree of reduced amplitude, then rhythm and hardly at all on speed, but some consistently based it on rhythm or speed (Heldman, Giuffrida et al., 2011). This suggests that, for example, if an individual's FT amplitude decrements near the end of ten taps ('slight' definition) but has mild slowing ('mild' definition) most clinicians would allocate a 'sight' grade, but some would give a 'mild' grade. In other words different combinations of components could be allocated the same grade, and conversely the same combination of components could be allocated different grades by different clinicians.

Furthermore one person's interpretation of 'mild' (2 points) slowing may be the same as another's for 'slight' (1 point) slowing and this was confirmed

[15] All five of the limb bradykinesia assessments have the same descriptions of grades.

by Goetz et al. who found that the UPDRS inter-rater reliability was lowest for grading patients with parkinsonism at the two ends of the scale's spectrum – i.e. with only slight, or very severe impairment - as many clinicians over-rate the severity of slight signs and under-rate the severity of severe signs (Goetz and Stebbins, 2004). Additionally less experienced assessors tend to allocate higher grades and have greater inter-rater variability than movement disorders specialists (Post, Merkus et al., 2005).

Discontinuous data

The scale results in semi-quantification of PD signs and symptoms rather than a continuous measure of impairment and disability for a number of reasons. Firstly each item sub-score compresses a wide range of performance into a coarse grained scale resulting in just five steps between 'normal' and 'severe'. Secondly we do not know if the intervals between the steps are the same and hence whether the scale is continuous or ordinal i.e. we know that increased grades reflect increased impairment but is the *degree* of impairment the same between 'slight' and 'mild' grades as between 'moderate' and 'severe'? As we do not know how wide the range of impairment is within a particular grade, it is difficult to confidently compare continuous measures (e.g. drug dose, age, disease duration) to the graduated steps of the rating scale.

Interval assessment

Any clinical rating scale only gives a measure of the patient's state at that particular moment in time or 'a snapshot' of how they might be during their daily activities. A motor assessment, however comprehensive cannot give information about motor fluctuations over the course of hours or days or weeks. Repeated interval assessments provide more representative longitudinal information but this may be cumbersome for the patient, take considerable clinician time and still do not provide continuous monitoring over time.

Costs

The MDS-UPDRS scale is free for individual clinical use but beyond this remit there are several costs incurred: a $1000 charge for use in funded research and £20,000 for industry-funded research. The online training programme costs $1,000-1,500 for each clinician who wishes to use it in clinical trials and $250-500 for non-members of the MDS who wish to undertake training for personal use or for non-profit research (Movement Disorders Society, 2013). Additional costs in terms of time and money to employ clinicians to administer the scale need to be considered also.

2.2.8 Summary of Parkinson's disease monitoring

Most PwPD are monitored through infrequent interval clinical assessments undertaken during consultations in outpatient departments. The subjective and imprecise nature of these assessments may lead to poorly-informed management decisions, sub-optimal functional outcome and inefficient use of resources. Clinical rating scales provide detailed semi-quantification of PD clinical status that enable more accurate longitudinal follow up and are the primary outcome measure in most clinical research studies of PD. However they are expensive to use, take considerable time and remain vulnerable to inter- and intra-rater reliability issues. There remains a need for an accurate objective measure of PD motor signs that is quick to administer in order to improve the quality of monitoring for clinical and research purposes.

2.3 Bradykinesia

2.3.1 Definitions

Bradykinesia is the cardinal motor feature of PD. It is obligatory for diagnosis, central to monitoring and frequently used as an outcome measure in research studies. In clinical practice the term 'bradykinesia' is often used interchangeably with 'hypokinesia' and 'akinesia' but the literal meanings are slowness of movement (Greek; brady – slow, kinesis- movement), small

amplitude of movement (Latin; hypo – deficient) and absence of expected movement (Greek; a – without) respectively. The gold standard clinical diagnostic criteria, the UKBBDC, encompass the latter terms within the official definition of bradykinesia as:

"slowness of initiation with progressive reduction in speed and amplitude of repetitive action" (Gibb and Lees, 1988)

Interestingly, the gold standard rating scale for monitoring, the MDS-UPDRS suggests a slightly different definition as for each bradykinesia motor examination item the examiner is instructed to:

"evaluate speed, amplitude, hesitations, halts and decrementing amplitude" (Goetz, Tilley et al., 2008)

Few studies have objectively evaluated the clinically defined components of bradykinesia in PD but those that have tended to define the core components as follows: reduced amplitude, reduced speed, impaired rhythm and the sequence effect (SE) (Abdo et al., 2010, Espay et al., 2011, Ling et al., 2012). The SE describes progressive decrement of the amplitude and/or speed of repetitive voluntary movements and is discussed further in Chapter 6.

Although bradykinesia typically manifests in PwPD as impaired limb movements, causing slowed gait and reduced dexterity for example, it is important to note that it extends to all motor activities including tongue movements (Van Lieshout, Steele *et al*., 2011) and eye saccades (Matsumoto, Terao et al., 2011).

2.3.2 Overview of neural basis of voluntary motor control in humans

The physiology of human motor control is complex but in brief involves descending motor pathways that originate in the brain and activate spinal cord neurons that connect via nerves to muscle. When the muscles contract a movement ensues. Reflex movements are simple stereotyped actions that occur in response to specific stimuli whereas voluntary movements are more complex and usually initiated as a result of internal cognitive processes rather than external stimulation.

Input pathways to the motor cortex

The motor areas of the cortex generate 'motor programs' that direct movements. The primary motor cortex (PMC) is in the pre-central gyrus and several areas of motor association cortex lie just anterior to the PMC including the pre-motor cortex laterally and the supplementary motor areas (SMA) medially. These regions are involved in higher order motor planning and project to the PMC.

These areas receive input from several sources. The association sensory cortex in the parietal lobe transmits sensory information about the position of the body to the SMA and premotor areas so that a motor plan can be developed e.g. which muscles need to contract and with what force. During planning and execution of movements the cerebellum and BG act as motor control systems to further refine the motor program. As the movement occurs there are also feedback loops from the occipital-parietal visual pathways and thalamic ascending somatosensory pathways that monitor the action (Edwards, Quinn et al. 2008).

Output pathways from the motor cortex

The motor plan is implemented through the PMC sending commands to descending pathways in the brainstem and spinal cord. They can be divided into lateral and medial motor systems based on location in the spinal cord. The two lateral motor systems are the corticospinal tract and rubrospinal tract and these control the extremities. The lateral corticospinal tract is essential for rapid fine movements at individual joints or digits, such as FT.

The four medial motor systems are the anterior corticospinal tract, the vestibulospinal tract, reticulospinal tract and tectospinal tract. These control the proximal and axial muscles involved in postural tone, balance, orientation of the head and neck, and automatic gait related movements.

2.3.3 Pathophysiological models of basal ganglia

Structural anatomy of the basal ganglia

The BG are a group of grey matter nuclei located deep within the white matter of the cerebral hemispheres and comprise the following structures:

caudate nucleus, } 'striatum'
putamen,
internal globus pallidus (GPi), } lentiform nucleus
external globus pallidus (GPe),
nucleus accumbens,
ventrolateral nucleus of the thalamus (VLT),
subthalamic nucleus (STN),
substantia nigra pars compacta (SNc),
and substantia nigra pars reticulate (SNr).

The caudate nucleus is a C-shaped nucleus positioned next to the lateral ventricles and the putamen lies inferior and slightly lateral to it. The caudate and putamen are joined together by several cellular bridges that appear as stripes or striations in histological sections so 'striatum' is a term commonly used to describe the caudate and putamen together. The anterior portion of the putamen fuses with the head of the caudate forming the ventral striatum and this is largely made up of the nucleus accumbens. The globus pallidus (GP) lies medial to the putamen and has external and internal sections, GPe and GPi. The lentiform nucleus is a term used to describe the putamen and GP together. The thalamus lies medial to the GP and the small spindle-shaped STN lies inferior to the thalamus.

The anterior limb of the internal capsule passes between the caudate in the lentiform nucleus and the posterior limb of the internal capsule (containing myelinated corticobulbar and corticospinal tracts) runs between the lentiform nucleus and the thalamus. The SN is the largest nucleus in the midbrain and lies dorsal to the cerebral peduncles. It is split into a ventral portion, SNr, which contains cells very similar to the GPi, and a dorsal portion called the SNc containing pigmented dopaminergic neurons.

Functional anatomy of the basal ganglia

The BG nuclei are connected to each other via intrinsic circuits and connected to other brain structures via extrinsic BG loops.

Extrinsic basal ganglia loops

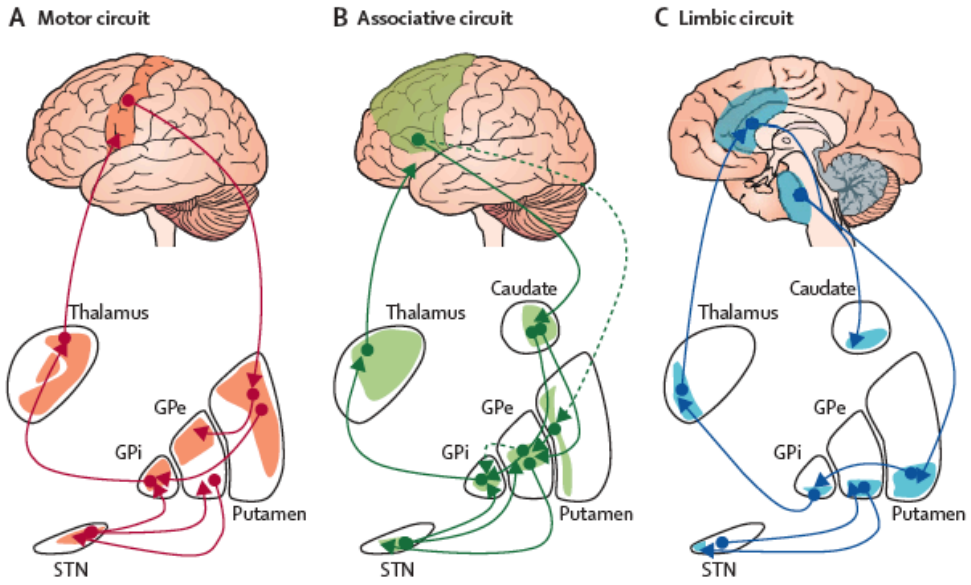
The BG receive a large number of input neurons from the cerebral cortex and have a relatively small number of output neurons that project back to the cortex and brainstem. BG inputs arrive via the striatum and the outputs leave via the GPi and SNr.

A major role of the BG are to mediate movement, processing information from the cortex before it is passed onto the brainstem and back to the cortex in a 'motor loop'. There are a number of other parallel BG loops that are anatomically and physiologically subdivided (Figure 8): the motor, oculomotor, associative and limbic territories take information from different parts of the cortex and are important for the control of movement, eye movements, executive function and emotion, respectively (Obeso, Cruz Rodriguez-Oroz et al., 2008).

Focussing on the motor circuits, the main inputs to the striatum comes from the motor cortex projections that are mostly glutaminergic (i.e. excitatory). There are also dopaminergic (excitatory and inhibitory for different parts of the striatum) projections from the SNc, glutaminergic inputs from intralaminar thalamic nuclei, and serotinergic inputs from the brainstem raphe nuclei.

The main BG outputs are from the SNr for the head and neck motor control, and GPi for the rest of the body motor control and use the inhibitory neurotransmitter, GABA. The outputs project to the VLT and ventral anterior nucleus of the thalamus, and then onto the pre-motor cortex, SMA and PMC. There are also BG outputs to the descending reticulospinal and tectospinal tracts via the brainstem reticular formation and superior colliculus respectively.

Figure 8 Functional organisation of the basal ganglia



Legend: The BG are divided into motor (A), associative (B), and limbic (C) which have extrinsic loops with the motor cortex, prefrontal cortex nd cingulate cortex respectively. *Reproduced from Obeso et al. 2008 with permission (Obeso, Cruz Rodriguez-Oroz et al., 2008).*

Intrinsic basal ganglia circuits

There are intrinsic connections between the various BG structures called 'circuits'. There is still debate about the exact function of the intrinsic circuits, and how dysfunction of them may result in various movement disorders with several models of BG circuits being proposed. Probably the best known model is the classical pathophysiological model of the BG

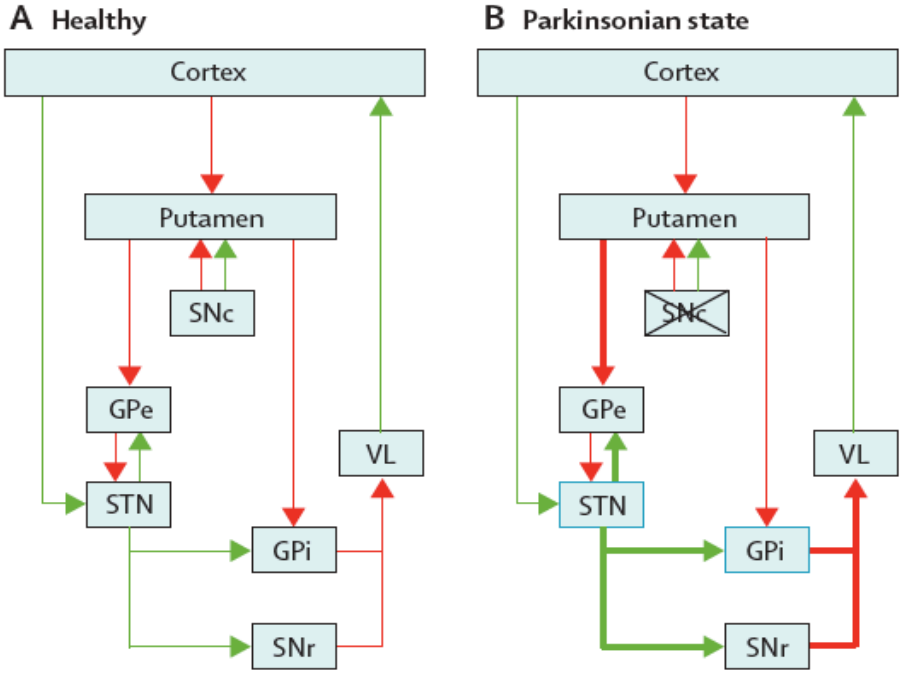
which describes two sets of major circuits called the direct and indirect pathways (Obeso, Cruz Rodriguez-Oroz et al., 2008).

The direct pathway is a direct connection between the striatum to the GPi and SNr via inhibitory neurons. This pathway inhibits GPi output and hence activates the thalamus and cortex (by reducing the inhibitory output from the GPi). There is also a direct excitatory connection between the cortex and the STN called the 'hyperdirect pathway' (Figure 9). The direct and hyperdirect pathways result in more movement as they increase activity in the thalamus and cortex.

The indirect pathway comprises inhibitory neurons projecting from the striatum to the GPe that in turn projects inhibitory neurons to the STN. By inhibiting the usual inhibitory output from the GPe, there is an increase in the STN output and hence more activity of the GPi. As the output of GPi is inhibitory, the net result is a reduction in thalamus and cortex activity and hence less movement. In summary, the direct pathway is the 'go' pathway as the net output is activation of movement and the indirect pathway is the 'stop' pathway as movement is inhibited (Figure 9).

The SNc projects dopaminergic neurons to the striatum and these neurons have opposite effects on the direct and indirect pathways. This is because the direct pathway has D1 dopamine (DA) receptors that are activated by DA but the indirect pathway had D2 receptors that are inhibited by DA. DA activates the direct pathway and inhibits the indirect pathway with both of these responses resulting in more movement.

Figure 9 Classical pathophysiological model of the basal ganglia



Legend: The direct and indirect pathways in healthy and PD states are shown with green arrows indicating excitatory activity and red arrows inhibitory activity. *Adapted from Rodriguez-Oroz et al. (Rodriguez-Oroz, Jahanshahi et al., 2009)*

Limitations of the classical pathophysiological basal ganglia model

The classical pathophysiological model of the BG in PD suggests that bradykinesia is primarily due to DA deficiency in the SNc cells (secondary to degeneration) which leads to increased activation of the indirect pathway, and increased inhibition of VLT nucleus and cortex (Figure 9). Therefore, the theory follows that the greater the deficiency of DA projections from the SNc, the greater the BG inhibitory output and the greater the severity of bradykinesia. This theory is supported by increased resting GPi firing rates in experimental parkinsonian monkey models (Delong, 1990) and also by human microelectrode recordings performed during neurosurgery that have shown a reduced firing rate of GPi (secondary to apomorphine infusions) correlates with less bradykinesia (Merello, Lees et al., 1999).

However lesioning the GPi has minimal effect on bradykinesia and Marsden described this as the "*paradox of surgery*" (Marsden and Obeso, 1994). Also neurophysiology studies of patients during deep brain stimulation (DBS) of the GPi have shown that stimulation of the ventral GPi abolishes the anti-bradykinesia effect of levodopa and stimulation of the dorsal GPi results in reduced bradykinesia but could sometimes induce dyskinesia too (Krack, Pollak et al., 1998). These studies suggest that the effect of reduced GPi output on dyskinesia and bradykinesia can be separated and the classical pathophysiological model of BG has some limitations. Furthermore it is clear from kinematic studies that not all aspects of bradykinesia are responsive to dopaminergic stimulation (Espay, Giuffrida et al., 2011). Taken together these studies suggest that the currents models are likely to be an oversimplification of parkinsonism.

2.3.4 Pathophysiology of bradykinesia

Whilst the pathophysiology of bradykinesia remains incompletely understood, there is good evidence that impaired motor planning leads to an inappropriately small movement being made, followed by a series of extra 'catch-up' movements to compensate. Berardelli succinctly summarises these deficits as:

"an underscaling of internally generated movements."

(Berardelli, Rothwell et al., 2001).

This underscaling is primarily thought to be due to a combination of BG pathology leading to reduced SMA activity and compensatory mechanisms occurring in the lateral pre-motor cortical areas. The next sections outline the evidence for this.

Impaired preparation of movements

A number of studies suggest an impairment of internally generated movement commands in PD. Research using functional Magnetic Resonance Imaging (fMRI) and PET scans has shown that PD patients performing a range of movements using a joystick have reduced activity in

midline structures such as the SMA, thalamus and BG and increased activity in lateral cortical areas such as the premotor cortex (Jahanshahi, Jenkins et al., 1995, Jenkins, Fernandez et al., 1992). These studies are consistent with the theory that bradykinesia is primarily due to problems with planning movement and/or the selecting correct movement as tests on HCs suggest that the SMA and prefrontal cortex activation are involved in planning the next movement. Injection of apomorphine (a DA agonist) increases SMA and prefrontal cortex activity and this correlates to improvements in clinical scores of bradykinesia (Jenkins, Fernandez et al., 1992).

Further evidence of reduced midline, and compensatory increased lateral, cortical activity in bradykinesia comes from neurophysiology studies of PD patients who have undergone pallidotomy or DBS. These have shown that post-operatively there is markedly increased activity in the lateral prefrontal cortex and a slight increase in SMA activity. This is associated with an improvement in the execution of movement rather than improving preparation to move (measured by BSP – see below) and suggests that the surgery has preferentially removed the inhibitory GPi output to the lateral cortical areas (Grafton, Waters et al., 1995).

Underscaling of movement

Internally generated movements may be under-scaled due to a problem with planning the correct movement, or selecting the correct motor programme. Brain activity during movement reflects motor and sensory components so many studies have focused on electroencephalogram (EEG) changes prior to movement so that sensory input will be minimized.

The Bereitschaftspotential (BSP) is a pre-movement EEG change that may be sub-divided into BSP1 that occurs 1-2 seconds before a self-paced voluntary movement, and BSP2 that occurs 650 milliseconds before. Studies have shown that BSP1 is smaller, and BSP2 larger in PD patients than HCs (Dick, Rothwell et al., 1989) and that BSP1 reflects SMA activity and BSP2 reflects premotor cortex activity (Ikeda, Luders et al., 1992). The BSP normalises in PD patients when they perform a task triggered by an

external cue (Jahanshahi, Jenkins et al., 1995). These BSP studies supporting the theory that bradykinesia is due to primary problem of under-scaling movement and there are compensatory mechanisms in lateral structures.

In PD there is an abnormality of another EEG movement-related phenomenon called 'event related de-synchronisation'. Normally the power in the alpha (10Hz) and beta (20Hz) regions of EEG are attenuated approximately one second before a planned movement and stay lower than rest levels throughout the movement. It is thought that neurons in parts of the cortex that are relatively inactive have idling rhythms that tend to synchronise in the 10-20Hz regions and de-synchronisation of these rhythms may be considered a marker of cortical activation. Wang et al. showed that in PD the period of de-synchronisation occurring prior to movement is shorter and the pattern of attenuation abnormal but these changes normalise with dopaminergic therapy and correlate to clinical improvements in bradykinesia (Wang, Lees et al., 1999). This suggests that there may be a failure of the BG to *release* cortical areas from idling rhythms just prior to, and during, a movement occurring, which may contribute to a smaller or delayed movement being made.

Impaired sensorimotor integration

The under-scaling of movements may be partly due to impaired sensorimotor processing. On prehension (reach and grasp) tasks PwPD take longer to develop peak grip force and grip harder than HCs which suggests impaired sensorimotor processing (Fellows, Noth et al., 1998). Experiments examining contingent negative variation (CNV) lend further support to this theory. The CNV is a negative EEG potential recorded over the frontocentral regions that occurs between a warning stimulus (S1) and an imperative stimulus (S2) and seems to reflect planning forthcoming movements and anticipating S2. The amplitude of the CNV is reduced in PD and the degree of impairment is related to the severity of disease (Ikeda, Shibasaki et al., 1997). Usually the CNV is more affected in PD than BSP possibly because S2 may act as an external cue so the CNV may occur with

no input from internally generated movement commands at all if the patients fails to prepare a movement internally between S1 and S2, whereas BSP occurs with self-paced movements and hence will always have some degree of contribution from self generated motor command. In other words CNV may represent purely a measure of response to a sensory input, and if this theory is correct, the CNV abnormalities provide some evidence for deficits in sensorimotor processing in PD (Berardelli, Rothwell et al., 2001).

Related to this is the observation that cueing may improve bradykinesia and motor arrest known as 'freezing'. If internally generated movements depend on the BG influence on SMA and premotor areas, external triggering may become dominant when this influence is reduced: so cues such as stepping over an object, or listening to a metronome, are forms of sensory stimuli that may help PwPD initiate or regulate gait because there is a relative excess influence of external sensory control over movement when the internally driven movement centres are impaired (Berardelli, Rothwell et al., 2001).

Impaired timing mechanisms

Clinically and kinematically it has been shown that bradykinetic movements in PD lack rhythm (Espay, Giuffrida et al., 2011, Taylor Tavares, Jefferis et al., 2005). This may be due to impaired timing mechanisms as rhythm generation and time estimation is abnormal in parkinsonian rodents with increased striatal expression of D2 receptors (Drew, Simpson et al., 2007). Also neuronal recordings in primate and rodent models of PD have shown impaired temporal processing of spatial information (Leblois, Meissner et al., 2006).

Under-recruitment of muscle contractile force

Direct stimulation of the PMC has shown that connections between the motor cortex and motor neurons are normal in PD; hence bradykinesia is not due to a deficit in the final PMC output pathway (Dick, Cowan et al., 1984). However, the electromyogram (EMG) patterns in bradykinetic PwPD are

abnormal with a loss of the normal tri-phasic EMG pattern during ballistic movements: the first agonist muscle burst is smaller than normal with subsequent bursts much larger than expected (Hallett and Khoshbin, 1980) and this effect is improved, but not normalised by levodopa (Baroni, Benvenuti et al., 1984). If the patient anticipates a larger amplitude movement the first EMG burst may normalise which suggests there is no problem recruiting muscles, but rather the muscle power is underscaled for the task (Hallett, Shahani et al. 1975).

Some studies suggest that tremor and rigidity may be minor contributing factors to the pathophysiology of bradykinesia too. For example voluntary movements performed by PwPD may be entrained to the frequency of any action tremor when they attempt to move at frequencies close to that of their tremor, particularly if the tremor amplitude is large (Logigian, Hefter et al., 1991). Rigidity may also be a contributing element as latency stretch reflexes are enhanced in PwPD and reflexes elicited in antagonist muscles are suppressed less than normal with the degree of abnormality related to the amount of clinical bradykinesia (Johnson, Kipnis et al., 1991, Rothwell, Obeso et al., 1983).

2.3.5 Methods used to evaluate bradykinesia

A cheap quick objective measure of bradykinesia that could be used in clinics would be very useful to aid the diagnosis and monitoring of PD. Such a device could also have a role in screening in epidemiological studies and for early treatment if neuro-protective medications become available. This section outlines the current methods used to measure bradykinesia, the types of devices that can quantify movements, and the previous studies that have evaluated bradykinesia through quantifying FT.

Finger tapping test

Bradykinesia has been measured through a variety of methods including quantification of prehension, reaction times, movement times, gait and handwriting but probably the most commonly used assessment is the FT task. FT has been used for almost a century to assess motor ability in an

array of conditions (Hollingworth, 1914) including PD (Espay, Giuffrida et al., 2011, Ling, Massey et al., 2012, Yokoe, Okuno et al., 2009), ataxia (Notermans, Vandijk et al., 1994), stroke recovery (Emara, Moustafa et al., 2010) and pre-manifest HD (Biglan, Ross et al., 2009).

Versions of the FT test include tapping a single finger, or alternate fingers against the table and tapping multiple fingers sequentially against the thumb. The MDS-UPDRS motor examination FT test requires repeated tapping the index finger and thumb together as fast as possible for a sequence of ten taps with each hand separately (Goetz, Fahn et al., 2007).

FT is a sensitive test of bradykinesia with PD patients disproportionately impaired for FT compared to other UPDRS upper limb bradykinesia tests (Agostino, Berardelli et al., 1998, Agostino, Curra et al., 2003). However the inter-rater agreement for grading speed and amplitude during FT is low (Bajaj, Gontu et al., 2010, Espay, Beaton et al., 2009) and FT is one of the least reliable items in the whole UPDRS motor examination (Goetz and Stebbins, 2004, Martinezmartin, Gilnagel et al., 1994, Richards, Marder et al., 1994).

This may be related to the fact that human visual quantification of movement is only approximate and it is difficult to attend to grading speed and amplitude simultaneously. It is likely that the five point UPDRS scale is too coarse to reflect subtle clinical changes too, particularly as the grade allocated is a composite of all the movement components that comprise bradykinesia. So whilst FT is a sensitive clinical test for bradykinesia the clinical rating may lack definition to fully evaluate the various movement components.

Handwriting

Visual assessment of handwriting is frequently used as a simple measure of bradykinesia in clinical practice. Micrographia (small handwriting) is common in PD so may support diagnosis. Serial documentation provides

recordable evidence of movements with reduced amplitude that may aid monitoring of bradykinesia.

This method has the advantage of being quick and requires no equipment other than pen and paper. However micrographia is not specific for PD and also occurs in parietal lobe lesions (Kim, Lee et al., 2005, Scolding and Lees, 1994), Huntington's disease (Iwasaki, Ikeda et al., 1999) and atypical parkinsonism (Ling, Massey et al., 2012). Even looking for 'decrementing micrographia' is not particularly sensitive for PD as studies suggest that only 15% of PD patients exhibit a progressive reduction in size (McLennan, Tyler et al., 1972). Some studies have employed EM digitising tablets to quantify handwriting in PD more precisely. These have shown that there is good test-retest reliability for handwriting tests (Banaskiewicz, Rudzinska et al., 2009), tremor does not interfere with quantification of bradykinesia (Banaskiewicz, Rudzinska et al., 2009) and both PD and DIP exhibit similar slowness but only PD has an abnormal jerky acceleration pattern (Caligiuri, Teulings et al., 2006).

Reaction time and movement time

The long established PPB and the TGUGT are basic measures of movement time (MT) defined as the period between the first movement being made and the target being reached. The PPB involves placing nine pegs in nine holes as fast as possible and the TGUGT requires the subject to stand up, walk 10 metres, turn round, walk back and then sit down again. The MT measure is simply how many seconds these tasks take to complete and it is clear that other non-bradykinesia factors such as musculoskeletal or visual problems may alter the results. Computer generated stimuli and touch pads have been used to quantify more precisely the 'reaction time' (RT), defined as the time lapsed between a 'go' signal and MT in PD. These studies have shown that RT and MT are prolonged in PD but respond differently to interventions: for example changing the target position unexpectedly only reduced the RT whereas levodopa administration only improved the MT. (Dunnewold, Jacobi et al., 1997). Gait analysis laboratories have been used in research studies to quantify separable movement components of

bradykinesia such as amplitude (stride length), speed and rhythm (variation in stride length or speed) (Chee, Murphy et al., 2009, Morris, Iansek et al., 1996) but these are expensive and not routinely available in clinical practice.

Prehension

Skill acquisition is impaired in PD but prehension is a naturally developed movement that does not require skill acquisition and is a useful paradigm for investigating motor control deficits in PD. Measurements of prehension have been made using optokinetic systems (Majsak, Kaminski et al., 2008) (Jackson, Jackson et al., 1995), resistive bend sensors (Schettino, Rajaraman et al., 2004) and pressure sensors (Alberts, Tresilian et al., 1998). These studies have shown that HCs execute the components in parallel i.e. pre-shaping the hand whilst reaching, but in PD there is a tendency to execute the reach and grasp components sequentially. The reaching component improves considerably more than grasping with visual cues (Majsak, Kaminski et al., 1998, Majsak, Kaminski et al., 2008) levodopa (Negrotti, Secchi et al., 2005) and STN DBS (Dafotakis, Fink et al., 2008) and these physiological assessments have revealed important information about sensorimotor integration for motor control. However prehension is not used commonly as a clinical test of bradykinesia perhaps because the deficits are too subtle to detect clinically.

2.3.6 The need for an objective assessment of bradykinesia

Clinical assessment of bradykinesia can be difficult as clinicians need to integrate all the movement components during a dynamic test. This is particularly difficult if other musculoskeletal co-morbidities or pain are present that may reduce the size and speed of movements. Imprecise clinical ascertainment of bradykinesia may lead to inaccurate diagnosis and monitoring of PD. For example in a study looking at how movement disorders specialists discriminated tremulous PD patients from patients with other tremulous conditions, the most common reason for erroneously

diagnosing PD was misinterpretation of bradykinesia and the authors concluded:

"The diagnosis of bradykinesia was particularly challenging".

(Bajaj, Gontu et al., 2010)

Regarding monitoring, the gold standard clinical rating scales may not accurately reflect changes in bradykinesia either. Espay et al. found that clinicians tend to weight their overall UPDRS bradykinesia grade based on the degree of reduced amplitude rather than rhythm and speed, and succinctly summarised the problems of using a single 'bradykinesia' grade:

"By combining multiple movement features into a single score, the UPDRS not only dilutes the power of finding true changes but may result in a differential response becoming unnoticed when evaluating the overall ''bradykinesia'' outcome of clinical trials."

(Espay, Giuffrida et al., 2011)

2.3.7 How can devices quantify bradykinesia in Parkinson's disease?

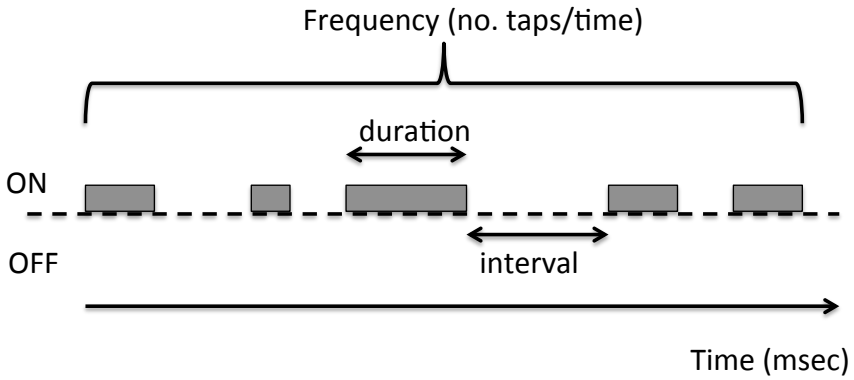
Over the last two decades, devices such as contact sensors, accelerometers, gyroscopes and EM sensors have been used to measure the motor signs of PD. Modern technology has facilitated the manufacture of such devices on a miniature scale, usually out of silicon, and these instruments typically measuring only 1-2 mm are called micro-electromechanical systems (MEMS). Movement sensor MEMS are integrated into many everyday objects such as smart phones, cameras and cars but few have been used in mainstream clinical practice. This section focuses on quantification of FT using movement sensor MEMS but many of the same principles can be applied to other methods of assessing bradykinesia.

Contact sensors and computer keyboards

Computer keyboards, metal plates, touch pads and electrical switches all work on the same principle of simple contact sensors – when the key is

pressed an electrical circuit is connected. The timing of the 'on-off' output allows several FT measures to be calculated such as frequency, duration, interval, rhythm, and accuracy of strike (Figure 10). Computer keyboards have the advantage that they are usually already available in clinics but these methods are limited by their inability to measure amplitude, speed or 3D movements. Also other disorders of upper limb function such as tremor and arthritis may affect the FT results if the target keys are several centimetres apart.

Figure 10 Quantifying finger tapping using contact sensors

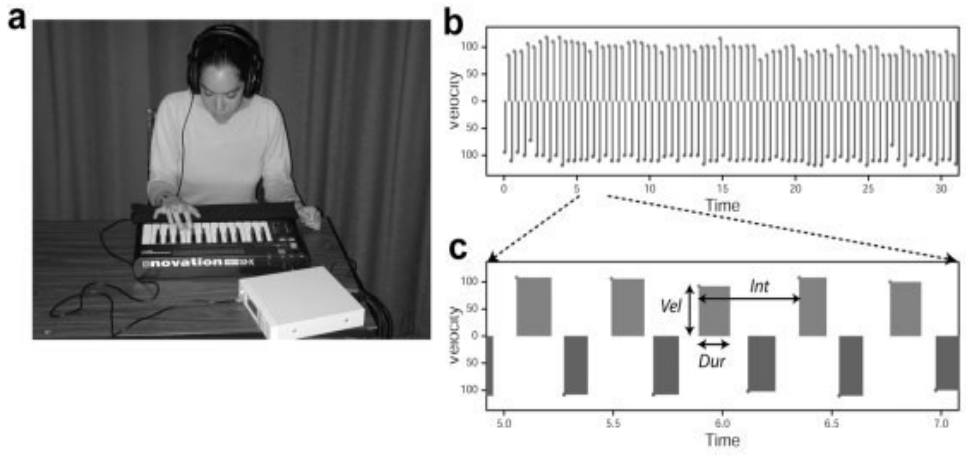


Legend: Schematic illustration of how contact sensors enable quantification of FT movement components such as frequency (number of taps per second), duration (period of time that sensor is on), interval (period of time between taps) and rhythm (variation in duration or interval).

Musical keyboards

Musical instrument digital interface (MIDI) was developed 20 years ago to facilitate combining different instrument sounds together. When MIDI is interfaced with an electric piano keyboard a number of FT parameters can be calculated such as speed and duration of key strike, frequency and rhythm (Figure 11).

Figure 11 Quantifying finger tapping using musical keyboards



Legend: (a) subject performing an alternating (index and middle finger) FT task with auditory and visual feedback, (b) example trace of key strike velocity data, (c) calculation of FT parameters such as velocity (vel) and duration (dur) of keystrikes and the interval (Int) between keystrikes can be calculated from the velocity data. *Figure reproduced from Taylor-Taveres et al. 2005 with permission.*

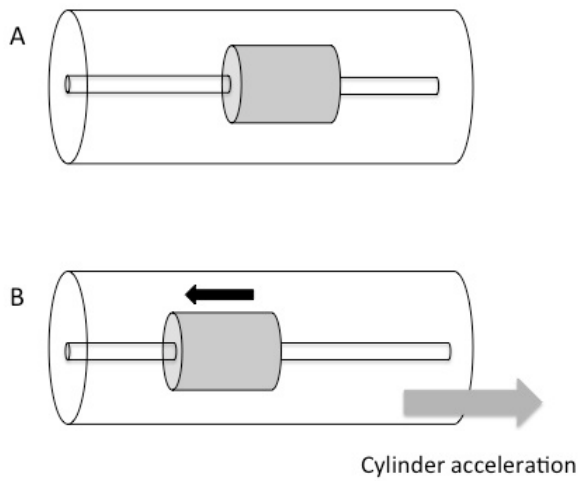
The main advantages of using MIDI to quantify FT include: high reliability, can be used with or without visual and auditory feedback, does not require any wires to be attached directly to the subject, and it has already been 'tried and tested' in the musical industry. The main drawbacks are that it cannot measure FT amplitude or 3D motion, the equipment is rather large and the resolution of key strike speed is limited by the range of integers in the MIDI volume scale with movement data outside this range not accurately quantified (Taylor Tavares, Jefferis et al., 2005).

Accelerometers

Accelerometers are electromechanical devices that measure static acceleration forces, such as the magnitude of gravity on an object thereby providing information about the tilt, and dynamic acceleration forces such as vibrations. They work on the basic principle of a mass suspended by a spring inside a container. When the container is accelerated the mass moves

relative to it and the distance moved is proportional to the output signal (Figure 12).

Figure 12 Basic principle of an accelerometer



Legend: (A) the mass remains still when the cylinder is static or moving with constant velocity, (B) inertia causes the mass to lag behind when the cylinder accelerates. The relative distance moved by the mass, or the force of the supporting springs is proportional to the output signal.

Different methods convert acceleration to electricity including optical or capacitive systems to detect movements of a small mass within a silicon sensor, and piezo-electric and piezo-resistive materials that convert the force applied by the mass to an electrical charge or resistance respectively. Acceleration data may then be integrated to provide velocity and amplitude measurements of movement. Their small size means movements are not dampened and equipment portable but disadvantages include complex processing and calibration and relative expensive.

Electromagnetic tracking sensors

EM sensor systems work on the principle that when an electrical current is passed through a coiled wire a magnetic field is set up around it and the magnitude of the field is inversely proportional to the square of the distance

from the coil. Most systems use a transmitter and a sensor that both contains three coils on orthogonal axes. When an alternating current (AC) passes through the transmitter a pulse of a magnetic field is sequentially emitted from each coil that oscillates at the same frequency as the AC. The magnetic field inducts a voltage in the sensor's coils and the magnitude of this within each orthogonal coil enables the sensor's distance from the transmitter to be accurately measured in 3D space. The derivatives of distance may then be calculated to enable velocity and acceleration to be measured too.

For FT measurements the transmitter can be attached to the finger and the sensor to the thumb or the transmitter can be placed on a table and two sensors attached to the digits. The advantages of EM sensors include the fact they are small, lightweight, cheap, easy to calibrate, and enable 3D movements to be measured. Disadvantages are that the sensors are not wireless, they must remain within a defined range of the transmitter and other nearby electrical equipment may interfere with the data.

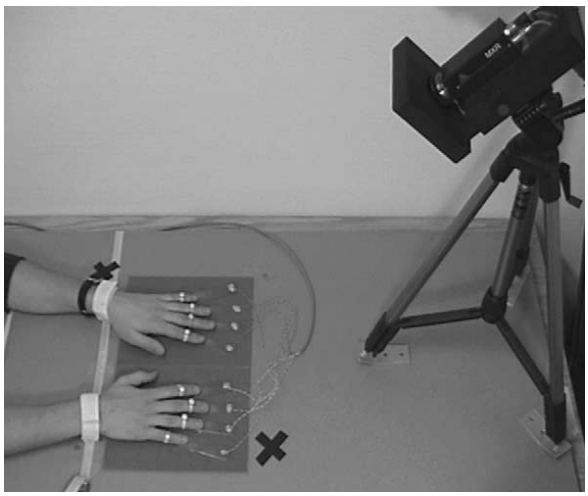
Gyroscopes

Gyroscopes measure orientation and rotation. Mechanical gyroscopes are based on the principle of a rapidly spinning wheel mounted inside two rings that enable the wheel's axle or 'spin axis' to assume any orientation: if the gyroscope is tipped the rings try to re-orientate to keep the spin axis in the same position and the degree of resistance in the spinning wheel will be proportional to the tilt applied. Most modern gyroscopes use a related phenomenon: they comprise a tuning fork configuration of two masses oscillating and moving in opposite directions and when the device is rotated there are opposite forces on each mass and the resultant capacitance change is proportional to the angular velocity applied. Gyroscopes may be combined with tri-axial accelerometers to measure movements with six degrees of freedom (3 orthogonal and 3 rotational) but their measurements can be complicated to understand and they are generally relatively expensive compared to other MEMS movement sensors.

Optokinetic systems

Optokinetic systems (OKS) comprise passive markers such as light emitting diodes (LEDs) attached to body parts and an infra-red camera to detect the positions of the markers (Figure 13). The markers are small and lightweight thereby not impeding movement. Their trajectories can be calculated from position-time functions thus enabling 3D amplitude, velocity, acceleration, rhythm (and derivatives of these) to be measured. The main drawbacks of OKS are their expense and the need for complex data processing to extract 3D spatial information from a 2D image.

Figure 13 Optokinetic system equipment



Legend: OKS equipment with reflective markers attached to the fingers and a table with metal stripes and bands around the wrists assure the reproducible hand-camera arrangement. *Image reproduced from Jobbagy et al. 2005 with permission. (Jobbagy, Harcos et al., 2005)*

2.3.8 Studies that objectively measured finger tapping in PD

The development of a device that can objectively evaluate bradykinesia in PD has been pursued for several decades as motion sensors can objectively measure multiple components of a movement simultaneously whereas clinical scales are vulnerable to inter- and intra-rater variability. This section summarises the medical studies that have measured FT in PD using movement sensors.

Probably the first study to objectively assess FT was Shimoyama et al's in 1990; they used an electrocardiogram (ECG) machine to quantify FT in 111 HCs, 14 PwPD, 17 people with cerebellar disorders and 14 with hemiparesis. Subjects repeatedly pressed a switch on the ECG machine as rapidly as possible for 15 seconds with their index finger. The machine was connected to a computer and the tapping frequency, intervals between FT, means, standard deviation (SD) and coefficient of variation (SD/mean) were calculated. They found that tap frequency was significantly reduced with advancing age, for women, and in the non-dominant hand. Tap frequency distinguished HCs from the patient groups but did not distinguish between any of the pathologies (Shimoyama, Ninchoji et al., 1990).

In 1995 Muir et al. used an electronic touchpad with four metal finger plates to assess seven PwPD *on* and ten HCs. Three different tapping experiments were performed using single and multiple FT protocols with maximum speed tests performed for 10 seconds with and without an 84 gram weight attached to finger. In the dominant hand single- and two-finger alternating FTs were 10% and 48% slower respectively in PD than HC. Like Shimoyama they found that the tap frequency reduced in the non-dominant hand and with age but only in the HC group. Intriguingly the weight resulted in PwPD slowing by 19%, and HC speeding up by 9%, relative to their initial FT speed (Muir, Jones et al., 1995).

Muir et al. also assessed paced FT using 30 visual and auditory pacing stimuli at various frequencies. FT frequency was plotted against pacing frequency and with all frequencies of auditory pacing PD and HC subjects exhibited a 'hastening phenomenon' - they speeded up to reach a fixed FT frequency that was near to maximum tapping frequency, but for visual pacing most subjects slowed down. The pacing frequency at which PwPD lost synchronisation was lower than HCs and this difference was more marked for visual pacing compared to auditory.

In 1997 Dunnewold et al. examined how well accelerometers measure bradykinesia compared to touch pads and clinical scores, and also how tremor affects the bradykinesia measurements (Dunnewold, Jacobi et al., 1997). 33 PwPD and 29 HCs wore three uni-axial accelerometers on their wrists and alternatively pressed two touch pads placed 30 cm apart for 30 seconds using one hand. The PD group had significantly lower tap rate and longer MT than HC and resting tremor did not influence the assessment of bradykinesia. There was good correlation ($r > 0.90$) between the accelerometer scores, FT rate and MT but low correlations between the UPDRS bradykinesia item scores and the tap rate ($r = 0.29$), MT and accelerometer measurements. Dunnewold summarises the problem of comparing an objective measurement to a clinical score and suggests that comparing two objective measures may be a better assessment of validity:

"The UPDRS is a subjective clinical assessment, whereas the other two assessments are objective. Where a 'gold standard' is not available, weaker evidence of the validity of an assessment procedure can be found in its tendency to agree with other tests that are used with the same purpose".

(Dunnewold, Jacobi et al., 1997)

In 1999 Giovannoni et al. described a new Bradykinesia Akinesia Incoordination Test (BRAIN TEST©) that involved subjects alternatively pressing two computer keys 15 cm apart, for 60 seconds 'as fast and as accurately as possible' with their dominant index finger (Giovannoni, van Schalkwyk et al., 1999). The test was repeated using the non-dominant hand. Thirty-five PwPD, 12 patients with cerebellar dysfunction and 27 HCs were assessed and four movement variables calculated: a kinesia score (KS), analogous to tapping frequency, akinesia time (AT) which was the cumulative time that any key was depressed for longer than 17 milliseconds, an incoordination score (IS) which measured the variance in time intervals between keystrokes (i.e. rhythmicity) and a dysmetria score (DS), which was adjusted for speed and reflected the number of incorrectly hit keys.

Like Shimoyama and Muir, the authors found that HCs had significantly lower tapping frequencies (KS) with advancing age but, in contrast, did not find any differences for gender. PwPD had significantly lower tap frequency and higher dysrhythmia than HCs when dominant hand scores were compared (Figure 14). Measures of speed, rhythm and akinesia correlated significantly with the total UPDRS bradykinesia item motor scores but did not predict diagnostic categories as there was such large overlap between the group's ranges.

Figure 14 BRAIN test scores in Parkinson's disease and controls

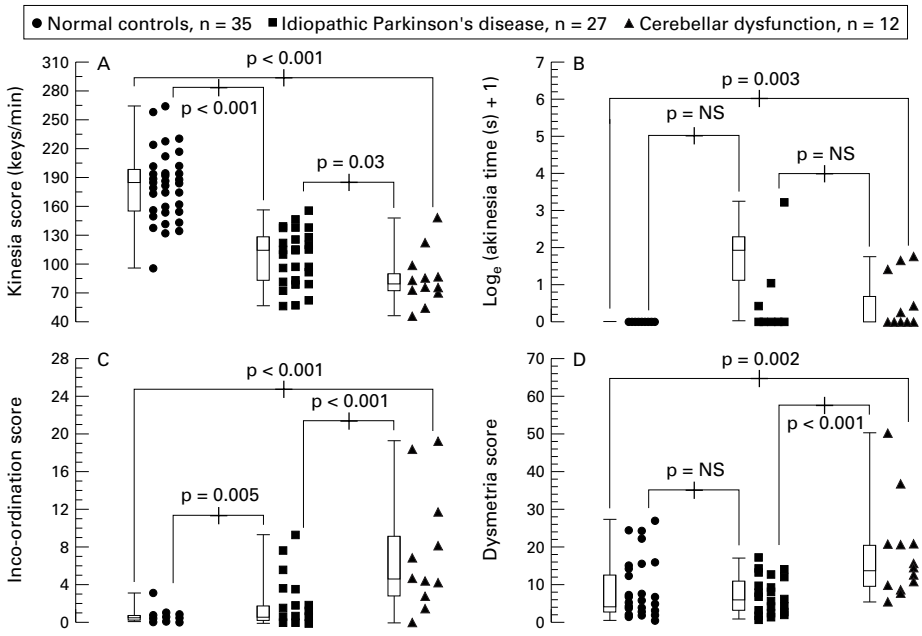


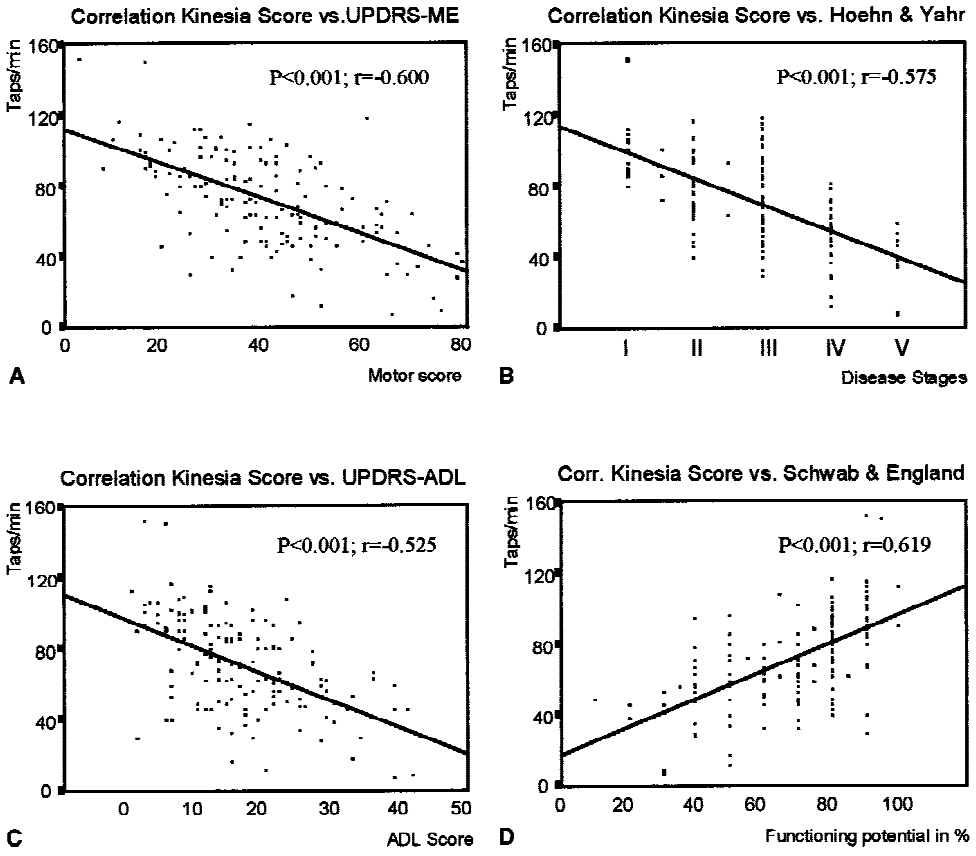
Figure 2 BRAIN TEST results. Scatter and box and whisker plots of the (A) kinesia score, (B) akinesia time, (C) inco-ordination score, and (D) dysmetria score. The p values were derived from a post hoc analysis of an ANOVA.

Figure reproduced from Giovannoni et al. 1999 with permission (Giovannoni, van Schalkwyk et al., 1999)

A year later Homann et al. examined whether the BRAIN TEST© could discriminate PD and HC subjects and predict the clinical severity of PD (Homann, Suppan et al., 2000). 154 PwPD *on* were assessed and 73 were matched for age and sex with HC. All four parameters measures in the MA PD hands were significantly correlated with UPDRS parts 2 and 3, HY and Schwab & England rating scales (Figure 15). These results suggest a possible role in monitoring PD as there is a *trend* for reduced FT frequency

with clinically more advanced disease, but Figure 15 demonstrates there is a wide range of KS for each disease stage or UPDRS score.

Figure 15 Correlation of BRAIN test with clinical rating scales



Legend: Scatterplots of the correlation between Kinesia score (equivalent to tapping frequency) and PD rating scales: (A) UPDRS part III motor examination (UPDRS-ME), (B) Hoehn & Yahr scale, (C) UPDRS part II activities of daily living (UPDRS-ADL), and (D) Schwab & England Scale. *Reproduced with permission from Homann et al, 2000 (Homann, Suppan et al., 2000)*

This study was the first to show that quantification of FT could be used for diagnostic classification with 85% of subjects correctly classified when the means of each parameter score were used. The main limitations of the study

were not addressing other confounding factors such as topographical motor deficit, presence of tremor or LID, cognitive function and mood, and using only one rater for the clinical scales.

In 2001 Pal et al. correlated quantification of FT with nigrostriatal degeneration as assessed by 6-FD PET. Eighty-six PwPD *off* and 136 HC performed FT for 10 seconds with each hand separately on a computerised drum machine interfaced with MIDI. PD tapping rates were significantly lower than HC and were strongly associated with disease duration, PPB, disease severity, and reductions in contralateral 6-FD uptake (but not with advanced age). FT frequency measures had sensitivity/ specificity of 0.79/ 0.70 for discriminating PD from HC. (Pal, Lee et al., 2001).

In 2004 Kandori et al. used index finger and thumb EM sensors to classify 20 PD and 18 HC FT data sets according to HY stage PD FT had reduced speed and amplitude in PwPD compared to controls and these measures correlated with HY stages. The authors also reported another curious finding in HCs: in young HC the amplitude remained constant but velocity reduced over the recording period whereas in older HC the speed remained constant and amplitude varied considerably (Kandori, Yokoe et al., 2004). The relationship between amplitude and speed will be discussed further in Chapter 4.

In 2005 Taylor Tavares et al. used a musical keyboard and MIDI to collect rapid alternating FT movement data of the index and middle fingers in PwPD before and after dopaminergic medications and/or DBS (Figure 11). The authors focussed on new methods of assessing rhythm using SD of velocity and coefficient of variance (CV = SD/mean) for interval and duration as they proposed that the 'signature' of PD was the "*temporal aspect of repetitive movements"* (Taylor Tavares, Jefferis et al., 2005). Thirty-three PwPD were assessed pre-operatively whilst *on* and *off* medication. Seventeen were assessed approximately 10 months after DBS and in three states: on meds/on DBS, off meds/on DBS and off meds/off DBS. Each hand was tested separately for 30 seconds of rapid tapping.

In the pre-surgery *off* group the two variables that correlated best with UPDRS part 3 score were measures of rhythm and velocity. A combination of velocity, interval (i.e. inverse of tap frequency) and rhythm best predicted total modified UPDRS 3 score (r = 0.7) and also the bradykinesia sub-scores (r = 0.63-0.67). When *on* there was clinical improvement of the UPDRS score by 43% but DBS improved it by 72%. Kinematic analysis showed that after medication was taken velocity, interval, and some (but not all) rhythm measurements significantly improved. DBS improved velocity by a similar magnitude to medications but improved rhythm much more. In other words medication improved velocity but DBS was better at improving measures of velocity and rhythm.

This study was pivotal in demonstrating that objective measurements of movement identifies differential improvements in motor control after DBS compared to medications and these were not detected by clinical UPDRS scores.

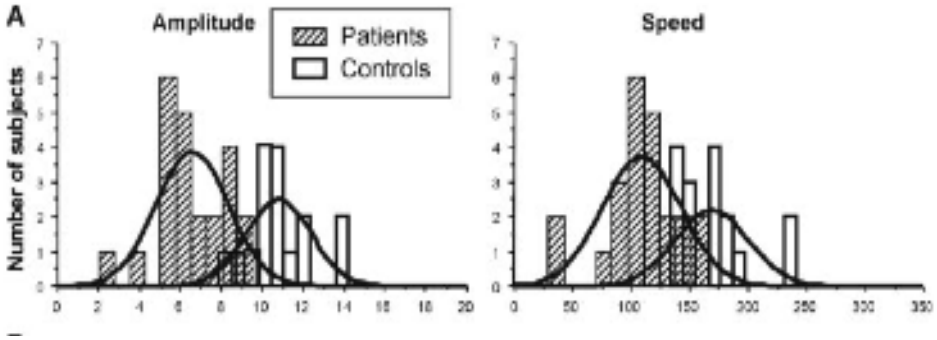
In 2005 Jobbagy et al. used an OKS to assess piano playing FT movements in 32 HCs and ten PwPD (Jobbagy, Harcos et al., 2005). The subjects placed their hand on a table with fingers separated and were asked to tap with their little, ring, middle and then index fingers in turn for 30 seconds. They found that a composite score of rhythmicity, amplitude and frequency was lower in PD than HC and tended to be lower in more advanced HY stages. Most subjects performed the task several times on the same day and this revealed that their scores generally improved for the first three trials and then plateaued, suggesting a learning effect and prompting the authors to recommend that only scores obtained after three practice trials should be used to improve accuracy.

Espay et al. compared three objective measures of bradykinesia to clinical rating scale scores in 24 PwPD and 16 HCs. The MA hand of patients, and the dominant hands of HCs were assessed in *on* and *off* states performing the PPB, tapping two buttons 30 cm apart over 30 seconds, and a 15 second

rapid FT task whilst wearing EM tracking sensors (Espay, Beaton et al., 2009).

In the *off* state the total UPDRS 3 scores were inversely correlated with amplitude, but not speed, and levodopa normalized speed to a greater extent than amplitude. The authors hypothesized that speed and amplitude in PD may therefore have different pathophysiological processes. This study demonstrated that the range of speed and amplitude overlapped considerably between patients and HCs (Figure 16) and this makes discrimination of diagnostic groups based on linear statistical analysis difficult. It is important to note that this study excluded PwPD with moderate tremor, perhaps suggesting that the group largely comprised the PIGD subtype of PD.

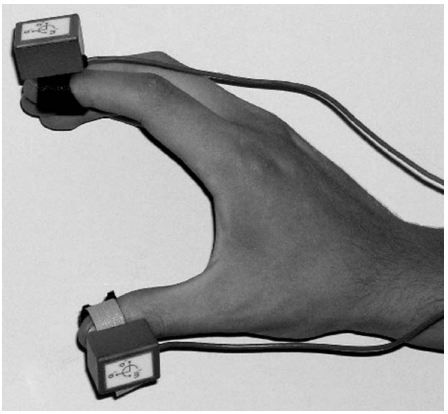
Figure 16 Finger tapping speed and amplitude in patients and controls



Legend: Histograms of amplitude and speed measurements for PD patients *off* (shaded bars) and controls (white bars). *Figure reproduced with permission from Espay et al.2009 (Espay, Beaton et al., 2009)*

In 2011 Espay et al. used KinetiSense (Great Lakes Neuro Technologies Inc.) movement sensors (Figure 17) to determine whether four components of bradykinesia in PD – slowness, reduced amplitude, dysrhythmia and fatigue – were associated with differential response to dopaminergic medication (Espay, Giuffrida et al., 2011).

Figure 17 KinetiSense motion sensors



Legend: Lightweight KinetiSense motion sensors containing three accelerometers and three gyroscopes were placed on the subject's index finger and thumb but kinematic data was only extracted from the gyroscopes. *Image reproduced from Espay et al. 2011 with permission (Espay, Giuffrida et al., 2011).*

Eighty-five PwPD wore the sensors on their MA hand and were assessed *on* and *off* for 15 seconds whilst performing FT, hand grasping and pronation. The assessments were video-recorded and four neurologists clinically rated speed, amplitude, rhythm and fatigue using the Modified Bradykinesia Rating Scale (MBRS). The clinical scores were averaged across three tasks and four clinicians to minimise variability.

Amplitude impairments were more severe and more prevalent than speed or rhythm impairments in all tasks whilst *off*. The speed component improved in the clinical assessments, and this was the predominant effect by quantitative measures also. Rhythm and amplitude slightly improved quantitatively but these changes were not detected by the MBRS. These findings were largely in line with Taylor-Taveres' et al.'s findings but with the benefit of including information on amplitude too.

In 2012 Ling et al. used an OKS to quantify FT in 15 PwP, 9 PSP and 16 HCs in order to evaluate whether PSP exhibits the same bradykinesia as PD

(Ling, Massey et al., 2012). Three sets of 15- second recordings were performed with each hand and PwPD were tested *off* and *on*. In contrast to previous kinematic studies that measured components of movement over the whole recording period, in this study the amplitude, duration, and speed of *each single tap* cycle was measured. Using this detailed approach the progressive change in performance (slope of linear regression line of variable against cycle number) could be calculated in addition to regularity (coefficient of variation) and mean performance of each component.

The most significant FT differences between PD *off* and HC were reduced speed and greater variability between tap cycles. Amplitude was numerically smaller ($p = 0.1$) and cycle duration longer ($p = 0.062$) but with borderline significance. These findings contrast with Espay's study that reported amplitude as the most impaired component of PD (Espay, Beaton et al., 2009, Espay, Giuffrida et al., 2011). There are two possible explanations for this discrepancy: firstly, amplitude was measured in different ways – Espay measured linear separation between digit sensors whereas Ling measured degrees of angular separation that may better account for different sized hands. Secondly Ling combined the scores of both hands together whereas Espay used only the MA hands scores and it is possible that Ling's measurments have become 'diluted' by the LA hand's data. Interestingly the authors found that the amplitude of FT movement in PSP was roughly half of those in PD and the amplitude did not decrement at all with repetitive movements, suggesting that PSP and PD have different component profiles of bradykinesia.

Only this study and Yokoe et al. (Yokoe, Okuno et al., 2009) have examined peak velocity measures *within* the tap cycle. This showed that when *off*, PD subjects' peak opening velocity was less than HC ($p = 0.03$) but there was no difference in peak closing velocity. After levodopa tap frequency, mean speed, and rhythm all significantly improved but decrement, fatigue and amplitude did not, lending support to Espay et al's findings (Espay, Giuffrida et al., 2011). Akin to Espay's findings, the total UPDRS part 3 scores in Ling et al.'s study were more strongly correlated

with amplitude than rhythm or speed and there was no significant correlation with SE measures.

2.3.9 Summary of bradykinesia

Bradykinesia is a complex clinical phenomenon and the underlying pathophysiology remains incompletely understood. Some aspects of bradykinesia are not DA responsive suggesting that the current BG functional models do not fully explain the underlying pathophysiology. However, largely based on such models the current understanding of bradykinesia in PD may be summarised as follows:

SN cells degenerate in the midbrain leading to reduced DA input to the BG. This results in less activation of the direct pathway and relative increased activation of the indirect pathway; the net result is an increase in the inhibitory output from the BG, namely the GPi. This results in less excitation of the cortex, in particular the midline SMA that is important for planning and preparing movements. Consequently there is an underscaling of internally generated movement programmes and insufficient recruitment of muscle force during initiation of movement. PD patients' movements undershoot the target and this results in reduced amplitude and speed of actions. Medial cortical areas are more active during internally generated movements whilst lateral areas are associated with externally cued movements. The BG preferentially access the medial rather than lateral motor cortical areas but this may be overcome if an external stimulus is provided. Hence PwPD may recruit additional circuits during movement to compensate for the primary BG defect and the clinical presentation of bradykinesia is probably a mixture of the primary deficit and these compensatory processes.

The gold standard diagnostic and monitoring tools for grading bradykinesia rely on subjective interpretation of the multiple movement components during dynamic assessments and are vulnerable to inter-rater variability. The clinical grades of bradykinesia severity are coarse scales and require a

composite assimilation of the separable components that means that differential component responses to new therapies may go unnoticed.

Modern technology has facilitated the development of small, lightweight, cheap and reliable movement sensors. They have revealed that the components of bradykinesia respond differently to pharmacological and surgical therapies for PD and clinical rating scales are insensitive to such changes. There remains a need for a small portable device that could objectively measure bradykinesia in the clinical setting to aid accurate diagnosis and monitoring of PD.

Chapter 3 Methodology

3.1 Patients and controls

3.1.1 Recruitment

The aims and objectives of the FT study have been described in Chapter 1.2. Patients with clinically definite PD attending neurology outpatient clinics at Leeds Teaching Hospitals NHS Trust (LTHT) were invited to take part in the study via a letter and information sheet sent in the post. The patient was telephoned one week later and asked if they were agreeable to participate in the study. They were also asked if they might be able to bring along their spouse or friend to act as a HC for the study but it was stressed that there was no obligation to do so.

An appointment was arranged by telephone, and confirmed in writing, for the research participant to attend the outpatient department at one of two LTHT hospitals - Leeds General Infirmary or Wharfedale Hospital, Otley. The appointment was scheduled for a date that was at least a week later in order to allow the patient an opportunity to withdraw from the study if they changed their mind.

3.1.2 Consent

Forty-nine PD patients and 41 HC were recruited. All subjects provided informed written consent and were assessed between August 2009 and October 2010. Approval was obtained from the National Research Ethics Service (reference 08/H0903/36) and the Medicines and Healthcare Products Regulatory Agency. Medication was not altered for the study and all patients were assessed whilst *on*. All but three of the patients were taking dopaminergic drugs and the mean levodopa equivalent daily dose (LEDD) was calculated using standard conversion factors (Tomlinson, Stowe et al., 2010). Seven patients had dyskinesia and 29 had resting tremor evident on the study day. None of the HCs were taking dopaminergic or anticholinergic drugs. Two HCs had a postural tremor and one had generalised epilepsy but the other controls had no history of neurological conditions.

3.1.3 Independent validation set

A validation sample of 13 patients diagnosed with clinically definite PD and nine age-matched HCs were recruited from a study on PD at the University of California-San Francisco (UCSF). Ethical approval was obtained from the UCSF and Veterans Affairs Medical Center Committees on Human Research (reference 11-06926) and all subjects provided written consent. Subjects were assessed between March 2012 and November 2012. The same procedures were followed for the UCSF sample and additionally they were assessed whilst in *on* and *off* (12 hours without dopaminergic medication) states.

3.2 Apparatus

3.2.1 Polhemus Patriot EM tracking sensor system

Assessments were performed in a standard hospital clinic room with the subject sat in a non-swivel high backed chair facing the long edge of a table. A laptop computer and Polhemus Patriot EM tracking sensor system (Polhemus, Inc., Vermont USA) were placed on the table and connected (Figures 18 and 19). The researcher sat by the short edge of the table and the laptop screen was adjusted so it could only be viewed by the researcher. This meant that participants did not receive any visual feedback from the movement recording traces displayed on the screen.

Figure 18 Ariel schematic view of the apparatus set up

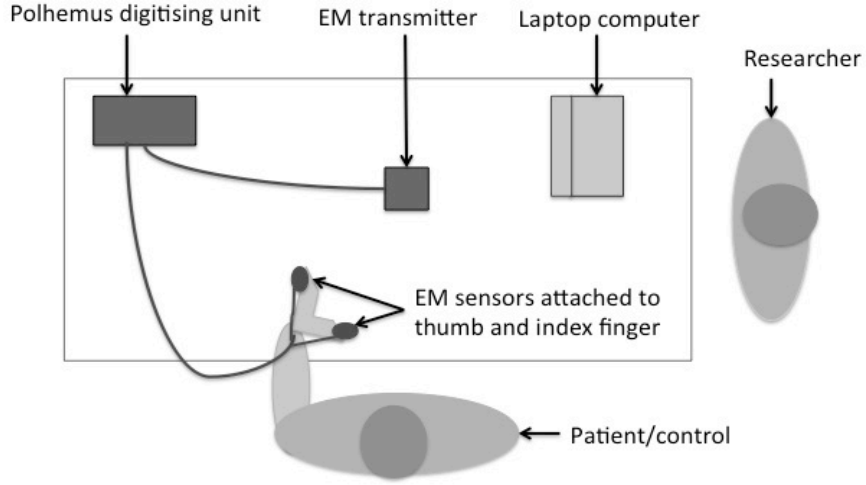
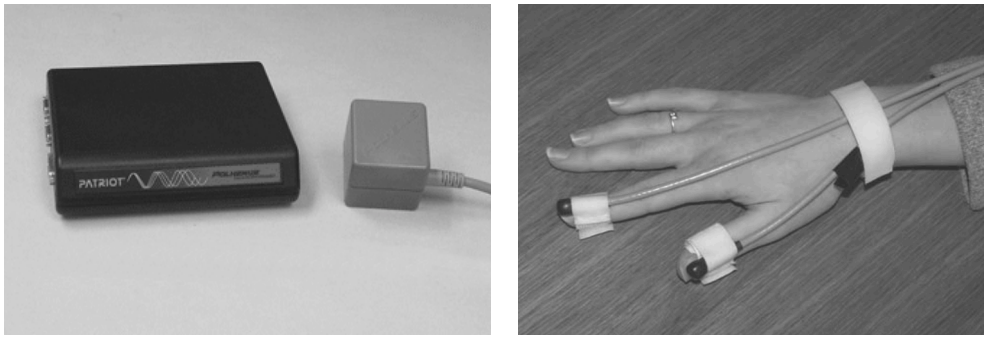


Figure 19 Polhemus electromagnetic tracking system



(a) **(b)**

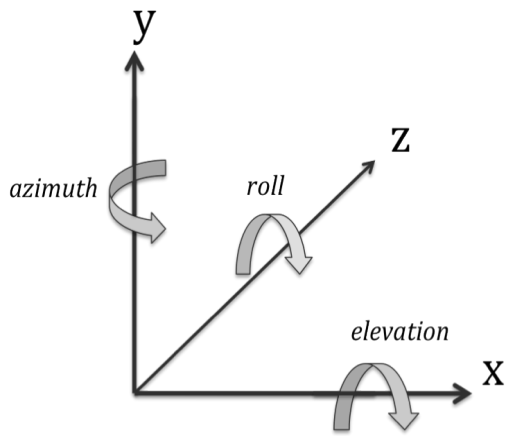
Legend: (a) Polhemus Patriot systems electronics unit (left) and magnetic transmitter (right) **(b)** Polhemus Patriot EM sensors attached to the thumb and index finger with Velcro straps.

The Polhemus device comprises a systems electronic unit (SEU), a magnetic transmitter and two EM tracking sensors. The SEU and magnetic transmitter were placed on the table and the two tracking sensors were worn on the research participant's hand (Figure 19).

In each sensor and the magnetic transmitter there are three EM coils arranged on orthogonal axes and encased in plastic. When an AC passes through the transmitter a pulse of a magnetic field is sequentially emitted from each coil in turn. The strength of the magnetic field detected by each coil inside the sensor enables its position and orientation *relative to the transmitter* to be accurately measured with six degrees of freedom (three positional and three orientation) in 3D space (Figure 20).

The SEU samples the positional and orientation data from each sensor 60 times per second thereby effectively measuring each sensor's movement in real-time. The SEU then transfers the movement time series data to the laptop for subsequent offline analysis.

Figure 20 Sensor movements measured with six degrees of freedom



Legend: Each sensor's movements are measured in 3D space using three positional coordinates (x, y, z) and three orientation coordinates (azimuth, roll, elevation).

3.2.2 Electromagnetic tracking sensors

The EM sensors are ideally suited to collecting FT kinematic data as they are small ($1cm^3$) and lightweight (2 grams) so any dampening or distortion of movements is minimised. The small size also makes them easily portable and comfortable for subjects to wear. Secondly the EM sensors are very sensitive to subtle changes in movement; when they are within a 30 cm range from the transmitter they have a sensitivity of 0.01 mm for position and 0.004 degrees for orientation (PATRIOT^{TM}). Thirdly they were used to quantify PD FT in a previous study (Espay *et al.*, 2009). The Polhemus Patriot systems costs approximately £2680 (quote from website in July 2012 (PATRIOT^{TM}).) and whilst this is more expensive than most accelerometer or gyroscope devices it is less than half the cost of OKSs.

The main disadvantage of EM sensors is their vulnerability to interference from magnetic and ferrous metal objects and electronic devices. These can distort the EM field and reduce accuracy of measurements. This means that it is important to keep computers, telephones and large metal objects such as filing cabinets at least a metre away from the sensors. The second limitation is that the accuracy of the measurements will begin to decay if the sensors

are moved too far away from the magnetic transmitter. This distance depends on the strength of the magnetic field and for the Polhemus system the manufacturer recommends that the sensors are kept within an operating distance of 1.5 m from the transmitter (PATRIOT^{TM}).

3.2.3 Positioning of the sensors

Each subject wore two EM sensors – one each over the nail beds of the thumb and index finger (Figure 19b). By placing the EM sensors distally any movements of the digits would be magnified but the effect of hand size on the kinematic measures would also be more apparent i.e. subjects with longer digits, would have an advantage for producing movement with greater amplitude and speed. Calculations to adjust for the effect of hand size on FT kinematics are described in section 3.5.1.

Velcro was used to secure the EM sensors in place because it resulted in a reliable secure fit and could easily be adjusted for different sized digits. A strip of Velcro was used around the digits and wrist (Figure 19b). A similar FT study used finger-stalls to attach EM sensors and accelerometers to their subjects' digits (Yokoe, Okuno et al., 2009) but Velcro was chosen as it was considered cheaper and equally strong.

The main disadvantage of using Velcro is that the subjects required the researcher's assistance to put the sensors on. Earlier work by Edgar demonstrated that it was feasible to incorporate the EM sensors into a glove so participants could put them on themselves (Edgar, 2007). However this method was not used as a series of gloves to accommodate different hand sizes would probably be required and it is likely that the gloves would interfere with normal sensorimotor integration due to the fingertips being covered.

3.3 Assessment procedure

3.3.1 Collection of demographic and clinical details

The researcher commenced the study assessment for each participant by summarising the purpose of the research, checking that they had read the

information sheet and offering the opportunity to discuss any queries or concerns that may have arisen. The participant was then asked to complete the consent form. The following demographic and clinical details were collected from each participant: name, age, gender, hand dominance, history of neurological condition, co-morbidities and a list of medications. Additional clinical details were collected from the patients: disease duration (years since diagnosis) and MA side. If the participant was unsure of the details for any of these categories the researcher sought written consent to consult their medical notes in order to obtain the information.

3.3.2 Physical and cognitive assessment

A brief physical examination was performed to assess for tremor, rigidity and postural instability in order to allocate a HY stage and to confirm the patient's MA side. The participant then completed the Montreal Cognitive Assessment (MoCA).

3.3.3 Finger tapping assessment

The EM sensors were attached and the participant was asked to sit upright with their back resting against the chair, then to raise their dominant hand so it was approximately at the level of their shoulder and the upper limb was unsupported. They were instructed to repeatedly tap their dominant index finger and thumb together with movements "as fast and as big as possible for 30 seconds until asked to stop". The researcher demonstrated ten FTs by moving her finger and thumb roughly perpendicular to the ground and with wide amplitude, fast and rhythmic movements, but these movements were not continued during the recording of the participants' movements.

The participants were not given a practice period and the first attempt at FT for each hand was recorded. The assessment was only repeated if a technical error occurred such as the movement data not registering on the laptop screen, or the sensor slipping off the nail bed. If a motor arrest occurred during the recording the researcher verbally reminded the participant that

they should continue trying to tap for the full 30 seconds but no 'trick'[16] techniques were used and the recording was not repeated. The researcher was an MDS-UPDRS certified clinician and she graded the FT performance on a scale of zero to four immediately after each recording period using item 3.4 of the MDS-UPDRS (Goetz, Tilley et al., 2008); see Chapter 2.2.7. The sensors were then transferred to the non-dominant hand and the assessment repeated.

3.3.3.1 Why was a finger tapping task used to evaluate bradykinesia?

The FT test is ideally suited for evaluating movements in PD patients and controls for several reasons. Firstly it is a familiar and established clinical test with a validated clinical grading scale available - item 3.4 on MDS-UPDRS (Goetz, Tilley et al., 2008). If a device is to be developed that objectively measures bradykinesia in clinical practice or research trials it is important that test procedures are simple and quick. By selecting the FT test, collection can occur *during* standard clinical assessments rather than additional assessments being required. Secondly it is more sensitive to PD bradykinesia than other MDS-UPDRS upper limb bradykinesia items (Agostino, Curra et al., 2003) and thirdly, clinical and kinematic assessments of the FT task have been shown to correlate well with other measures of bradykinesia (Agostino, Curra et al., 2003, Goetz, Tilley et al., 2008). A final advantage of using FT to evaluate PD bradykinesia is that it is a compact test requiring only a small room for the assessment and, in contrast to gait analysis it can be performed sitting down thereby not reducing EM sensor accuracy by moving away from the magnetic transmitter, or adding confounding of the effects of balance on the results.

3.3.3.2 Why is it important to assess both hands but separately?

There are several benefits for assessing the FT task in both hands but separately. It has been shown that when PwPD perform FT using both hands simultaneously the performance of the MA hand improves and the performance of the least affected (LA) hand deteriorates (Kishore, Espay et

[16] Trick techniques include counting or using visual cues to reduce the duration of a motor arrest

al., 2007). It is important to assess both hands though rather than just the MA side because the degree of asymmetry of FT performance may in itself be a helpful discriminating feature for PD. The LA side data is likely to provide a more challenging test for the device too. Finally Haxmaa et al. proposed that PwPD and HY stage 1 disease (i.e. unilateral parkinsonism) the 'non- affected' side's performance is a useful model for "*preclinical*" PD (Haaxma, Bloem et al., 2010).

3.3.3.3 Choosing the duration of finger tapping assessment

For purposes of diagnosis, the UKBBDC does not specify how bradykinesia should be determined clinically in terms of which clinical tests to use or how many repetitions to assess. For monitoring, the MDS-UPDRS instructs subjects to perform ten consecutive FTs but concern has been raised that this may not be long enough for the SE to manifest (Ling, Massey et al., 2012).

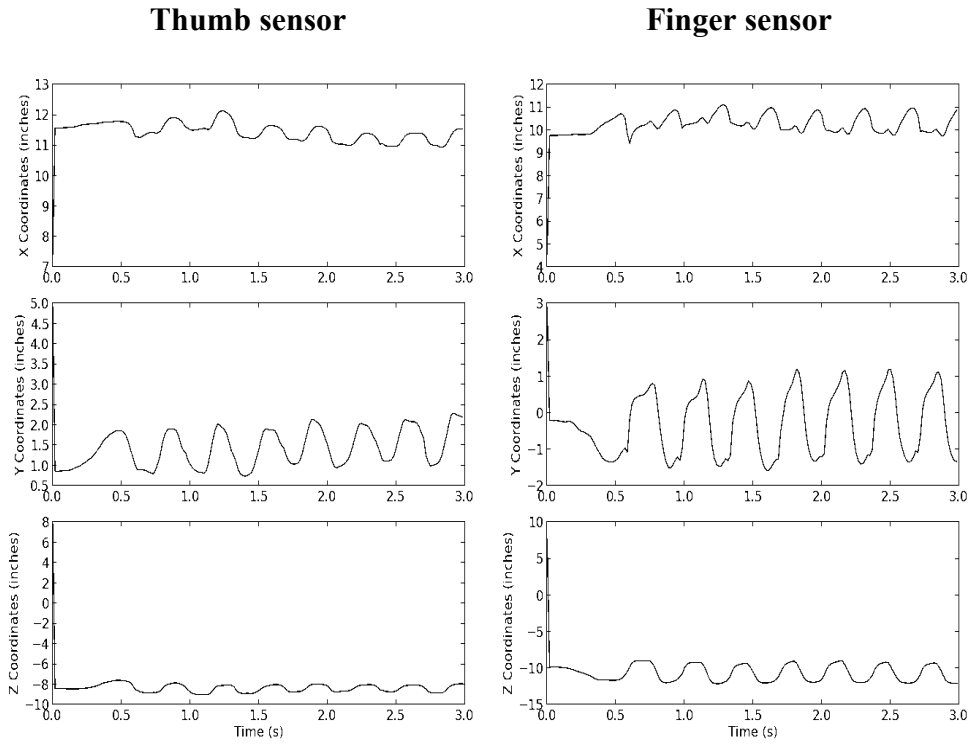
Previous kinematic studies of PD have used a range of FT durations including 5 seconds (Agostino, Curra et al., 2003), 15 seconds (Espay, Beaton et al., 2009, Espay, Giuffrida et al., 2011, Heldman, Giuffrida et al., 2011, Ling, Massey et al., 2012), 30 seconds (Kandori, Yokoe et al., 2004), and 60 seconds (Yokoe, Okuno et al., 2009). Thirty seconds was chosen for the present FT study protocol as this was felt to be long enough to allow the SE to manifest but probably not long enough to result in physiological fatigue. It also provided the opportunity for further analysis of shorter segments of data.

3.4 Pre-processing of movement data

Only the positional movement data was used in this study. The difference between the x, y and z coordinates in one sensor and the corresponding x, y and z coordinates in the second sensor gave the separation distance between the index finger nail and the thumb nail (Figure 21).

For each 1/60th second time point the relative distance between the sensors was calculated by subtracting the x, y, z coordinates of the index finger sensor from the x, y, z coordinates of the thumb sensor. The Euclidean distance D, or overall positional separation, between index finger and thumb was then calculated as: $D = \sqrt{(x^2 + y^2 + z^2)}$ where x, y and z are the coordinate distances of the index finger relative to the thumb.

Figure 21 Raw positional data in the thumb and index finger sensors

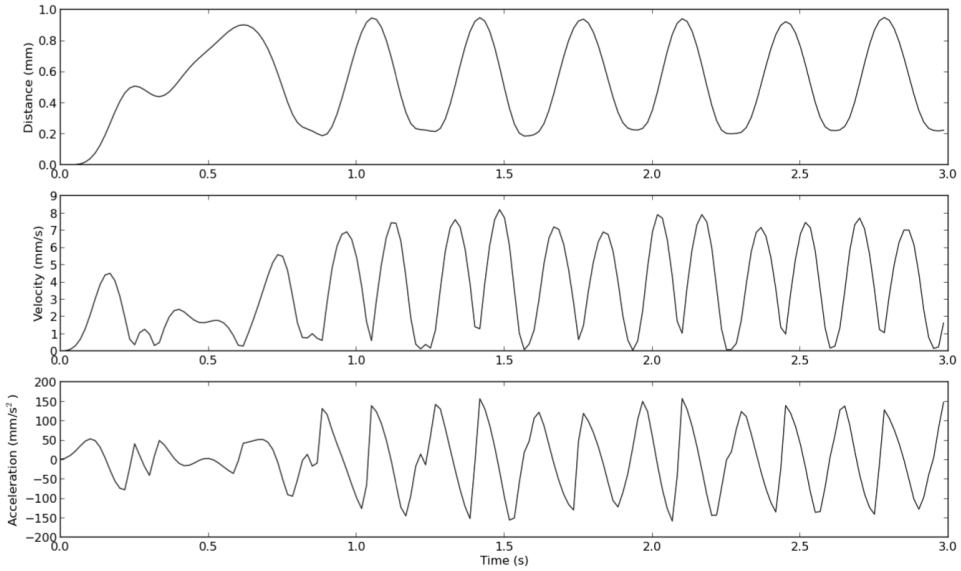


Legend: An example of the raw positional data collected from the thumb (left column of charts) and index finger (right column of charts) sensors of a participant. From top to bottom each sensor's time series positional data are shown for the x-, y- and z -coordinates. *Data series produced by Stuart Lacy.*

This generated a sequence of digit separations over time (Figure 22). For each subject the separation time series data, D, was differentiated to produce the speed time series data (*dD/*t) and then differentiated again to give the

acceleration time series data (dD^2/t). In other words the first and second derivatives of Euclidean distance (or relative separation in 3D space) were calculated in order to produce the speed and acceleration data respectively.

Figure 22 Example of finger tapping kinematic data



Legend: Kinematic data from seven FTs are shown. The x-axes denote time, measured in seconds, since the assessment recording started. The top chart shows the relative distance between the two sensors in mm. The middle chart shows the corresponding speed data, in mm/s. The bottom chart shows the corresponding acceleration data, in mm/s^2. *Data series produced by Stuart Lacy.*

Two different approaches to analysing the movement data were used in this study - separable component measures of bradykinesia and EA induced classifiers as outlined below.

The data was pre-processed to remove noise and this process was slightly different for each approach to analysis. For the separable component analysis the x, y, z coordinate raw data was passed through a Low Pass 5Hz Butterworth filter to remove any high frequency noise elements. For the EA analysis noise was first removed by down-sampling the acceleration data by

a factor of two. This means that only every other data point was used, resulting in 30 data points per second rather than 60. The acceleration data was then truncated to one standard deviation around the mean and scaled uniformly to the interval 0,1 in order to remove noise and information about absolute amplitude (Figure 23).

3.5 Calculation of separable movement components

For the separable component analyses of the FT movement data the separation, speed and acceleration time series data were analysed by purpose-written SciPy (Jones, Oliphant et al., 2001) and R (R Development Core Team, 2008) scripts. The principles of the calculations used within the scripts are outlined in this section and in Figures 24 and 25. In summary two types of calculations were made – the first for components within an individual tap cycle and the second for measuring the average, the variability and the trend of these measurements over a defined duration of consecutive FTs.

3.5.1 Measuring components within each finger tap

Defining individual tap cycles

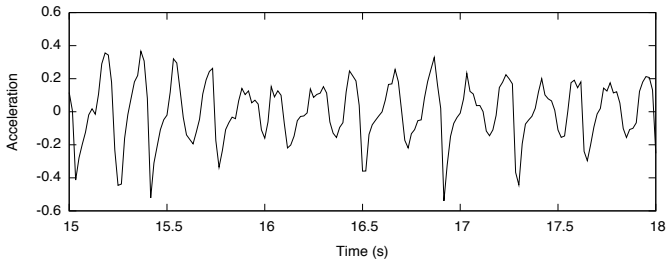
Each FT cycle comprises an opening phase and a closing phase. The opening phase begins once the digits begin to separate from opposition and finishes when the digits are maximally separated; the closing phase begins once the digits move towards one another after the point of maximal separation and finishes when the digits are opposed again (Figure 24).

The cycle of opening and closing movements is repeated successively during the assessment period to produce a sequence of FTs. Each 30-second recording of sensor separation values was segmented into separate tap cycles that were defined by the period between two successive minimal separation points i.e. two consecutive oppositions (Figure 24). This allowed movement component measurements to be calculated for *every individual* tap cycle performed by each subject.

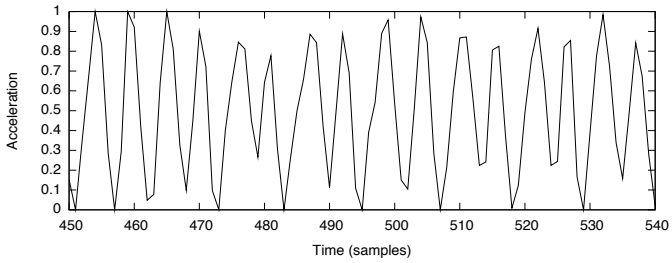
Figure 23 Pre-processing raw movement data to remove noise

Patient data

(a) Raw data

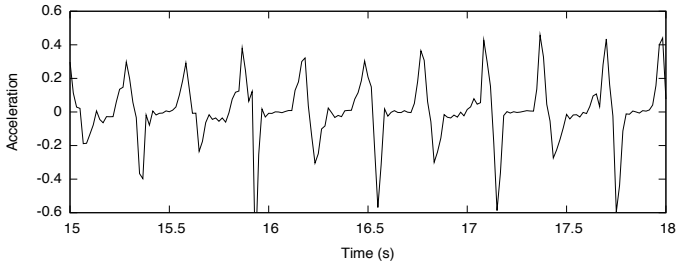


(b) Pre-processed data

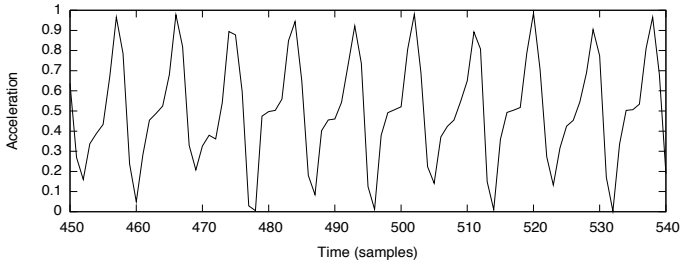


Control data

(c) Raw data

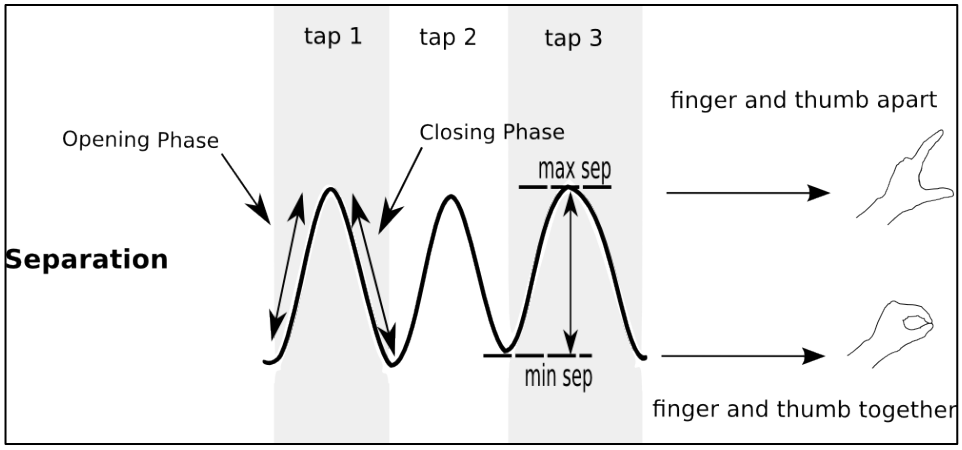


(d) Pre-processed data



Legend: The raw acceleration data from a patient (a) and a control (c) was pre-processed by using alternate data points, truncating these to one SD about the mean and then scaling the time series data uniformly between zero and one. This resulted in acceleration data with less noise and no information on the absolute amplitude (b, d). *Figure produced by Mic Lones*

Figure 24 Finger tapping cycles



Legend: Positional separation data recorded from a control demonstrating the opening and closing phases of three consecutive FT cycles. Max sep, maximum separation; min sep, minimum separation.
Illustration adapted from an original produced by Stuart Lacy.

Normalisation of data

In order to minimise the impact of different hand sizes between subjects the separation data were normalised. This was done using two methods. Firstly, as the minimal separation of the two EM sensors, when the digits are opposed, depends on the antero-posterior dimension or 'thickness' of the participant's finger and thumb (as the sensors were positioned over the nail bed, or posterior aspect, of each digit) the minimum separation distance recorded for each subject was subtracted from all of their other separation data points. This gave a series of separation measurements for each subject's hand *relative* to their minimal separation.

Secondly, as the maximal separation of the digits depends on the length of the digits all of the separation data points were divided by the maximum separation achieved during the 30 second recording. This resulted in a measure of normalised amplitude from zero to one representing the relative distance between the index finger and thumb.

Amplitude

The maximum amplitude of each FT cycle was calculated as the largest amplitude achieved during each single tap cycle (Figure 25). The mean amplitude of each tap cycle was calculated by dividing the sum of all the separation data points by the total number of data points.

Speed

The normalised amplitude data were differentiated to calculate the corresponding velocity profile and the absolute values of velocity taken in order to obtain the speed data. The following measurements of speed were made for each individual tap cycle (Figure 25):

Mean speed = sum of all speed data points/ number of data points
Maximum opening speed (OS) = maximum speed during opening phase
Maximum closing speed (CS) = maximum speed during closing phase
Maximum speed = whichever is greater of OS and CS

Periodicity

It was recognised that subjects may exhibit different *patterns* of tapping that may not be captured with measures of amplitude and speed e.g. a subject with small *and* slow FT movements may have the same amplitude score as a subjects with small but fast FT movements. To capture the relationship between these components a variable called 'periodicity' was calculated for each tap cycle as follows:

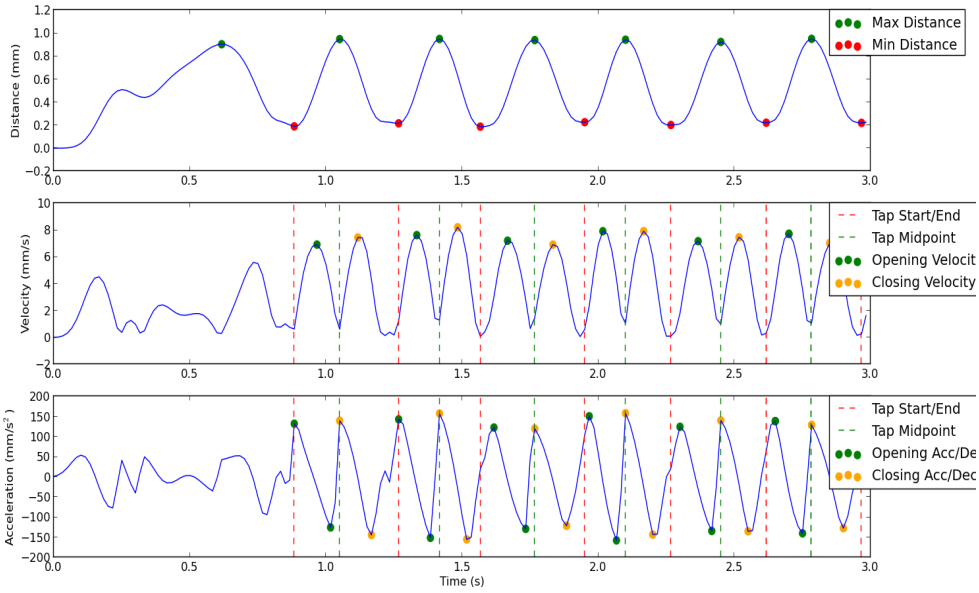
Periodicity = maximum amplitude x maximum speed

Duration

The time period between two consecutive amplitude minima defined the tap cycle duration (Dur). Other movement components of the tap cycle could then be calculated as follows:

% Dur in opening phase = (time in opening phase/ Dur) x 100
% Dur in closing phase = (time in closing phase/Dur) x 100

Figure 25 Calculating movement components from kinematic data



Legend: From top to bottom FT data is shown for distance (separation), speed and acceleration time series. The computer program overlays the coloured dots and dashes to define tap cycle parameters and enable calculation of tap cycle components. Note that two speed maxima occur during each tap cycle corresponding to the opening and closing phases. *Data series produced by Stuart Lacy.*

Halts

Halts were measured by calculating the percentage of the tap cycle duration spent at 'zero' (< 5% of the maximum) speed:

% Dur with 'zero' speed = (time < 5% max speed/ Dur) x100

Acceleration

The acceleration series was derived from the amplitude data by taking the second derivative. For each tap cycle the following components were calculated:

Opening acceleration = maximum acceleration point during opening phase
Opening deceleration = maximum deceleration point during opening phase

Closing acceleration = maximum acceleration point during closing phase
Closing deceleration = maximum deceleration point during closing phase

A measure of zero acceleration, which may be considered a measure of initiation delay, or akinesia, was also calculated as follows:

% Dur with 'zero' acceleration = (time <5% mean acceleration/Dur) x 100

3.5.2 Measuring movement components over tapping sequences

A number of descriptive statistics were used to convey the central tendency and spread of data of each movement component over multiple FTs: mean and median, standard deviation (SD), coefficient of variation (COV), and the regression line gradient. This enabled the average, the variability and the trend, respectively, of FT movements to be measured over defined periods.

Mean

Taking the mean value of a particular movement component gives an average value, or measure of central tendency, for that component over a defined duration. Mean values were calculated for the following individual tap cycle components: maximum amplitude, mean amplitude, maximum speed, mean speed, maximum OV, maximum CV, periodicity and % zero speed. For each component the calculation was:

Mean x = sum of x during n tap cycles / n tap cycles,

where x is the movement component and n is the number of FTs during a defined period.

Median

The median value of a movement component is the middle point that separates the top half of the data from the bottom half of the data. Therefore the median also gives an average value of all FTs over a defined duration but it is less affected by outlier data that the mean. Median values were calculated for the same components as outlined in the mean section above using the following calculation:

Median x = middle value of n tap cycles,

where x is the movement component and n is the number of taps during a defined period.

Standard deviation

The SD reflects how the data is distributed about the mean value, with larger values indicating that the data is more widely distributed. It is calculated by:

SD = √(variance),

where variance = (sum of each data point – mean)^2 / sample size

Coefficient of Variation

The COV reflects how much a movement component measure varies over a defined period. It may be considered a measure of how rhythmic the repetitive FT movements are and has been used in this manner in a number of kinematic FT studies previously (Espay, Giuffrida et al., 2011, Ling, Massey et al., 2012, Taylor Tavares, Jefferis et al., 2005). High COV values imply less rhythmic movements than small COV values. COV was calculated over defined FT durations as follows:

COV = SD/ mean

COV amplitude = SD max amplitude/ mean amplitude

COV speed = SD max speed/ mean speed

COV duration = SD duration/ mean duration

Regression line gradient

To calculate the trend of speed and amplitude over a series of repetitive finger taps the maximum speed and maximum amplitude for each consecutive tap cycle was linearly regressed against the number of tap cycles according to the model:

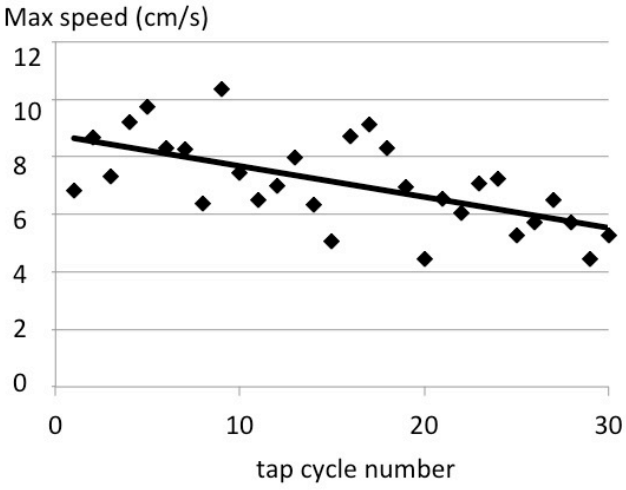
Maximum speed (or amplitude) = y axis intercept + β x + error,

where β is the slope of the regression line and x is the tap cycle number.

A negative slope is a measure of the SE as it indicates that the overall trend of a particular movement component measure is decrementing and has been used in a number of previous studies (Iansek *et al.*, 2006, Chee *et al.*, 2009, Ling *et al.*, 2012). A zero or positive slope indicates that the speed (or amplitude) is not decrementing and hence not exhibiting the SE (Figure 26).

Figure 26 Calculation of finger tapping decrements

(a) Patient P11



β = - 2.1

Decrementing speed

(b) Control C21

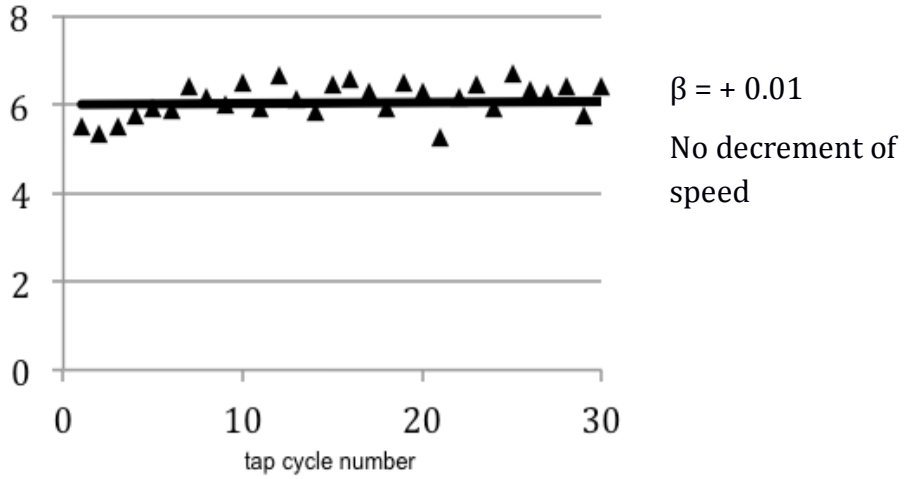


Figure 26: The maximum speed of each FT cycle (y-axis) is plotted against the corresponding tap cycle number (x-axis) for **(a)** a patient, P11 and **(b)** a control, C21. A regression line of best fit is applied to the data points and the gradient of the line gives a measure of the trend of speed measurements during the first 30 taps. These charts show that P11 exhibits decrementing FT speed (negative gradient) but C21 does not (flat/ positive gradient).

3.6 Evolutionary algorithm analysis

3.6.1 What are evolutionary algorithms?

The second approach to analysing the movement data was using EAs and these are a type of computer program that have been designed to find optimal solutions to problems. An 'algorithm' is a list of sequential instructions traditionally specified by the programmer but EAs differ as they are formed through a process based on the Darwinian theory of biological evolution[17] and involve activities that replicate selection, inheritance, reproduction and mutation (Larson, 2009). In short EAs solve problems through a repetitive process of producing vast numbers of possible solutions, searching through these to select the best ones, and then recombining these to get potentially even better solutions. In other words the EAs involve repeatedly optimising different solutions to a problem until the best one possible is produced.

[17] states that organisms and species develop through the natural selection of small inherited variations that increase the individuals ability to compete, survive and reproduce.

3.6.2 How does an evolutionary algorithm solve a problem?

Training, validation and test data

There are different types of EA, and in this study two different types were used (see section 3.6.5) but the principles outlined in this section are largely common to all. The main steps of the EA process can be demonstrated by considering a simple hypothetical problem that is adapted from Larson (Larson, 2009) such as 'what lists of four numbers will equal 100 when summed together?'

In order for the EA to solve this problem the data (e.g. 1,000 sets of random four-string numbers) must first be split into training, validation and test subsets. The 'training' data (e.g. 600 of the 1000 sets of four-string numbers) is used by the EA to develop equations that can potentially solve the problem. The 'answer' (e.g. the sum of each set of four-string numbers) is revealed to the EA alongside the training data so it can 'learn' through a process of selection, reproduction and evolution (see below) what components of equations will best map the data to the correct answer.

The accuracy of these equations are subsequently evaluated on previously unseen data (i.e. the other 400 sets of 4-string numbers) known as the 'non-training' data. The non-training data is split into validation and test sets. The validation set data (e.g. 200 sets of 4 string numbers with the sum of each concealed) is processed through the equations to find out which equations give accurate results beyond the training set e.g. how many of the 4-string numbers that were predicted to have a sum of 100 actually did.

The high performing equations selected through the validation process are then assessed on a 'test' set of data (the remaining 200 sets of previously unseen 4-string numbers) in the same manner and it is this final step that gives a measure of the accuracy of the best solutions, or equations, to the problem.

Optimisation through selection, reproduction and mutation

The EA finds the best solutions to solve a particular problem by a process called optimisation. This means that elements from the solutions that give the best results are selected then recombined in some manner to produce new solutions, some of which should give even better results than the previous solutions. This process is repeated numerous times and tends to result in better and better solutions.

The first step of the EA is to randomly propose possible solutions to the problem. So in the hypothetical example of trying to find an equation that can predict which four numbers will sum 100 the computer generates a random set of potential solutions such as:

solution 1 = [95,0,0,0],
solution 2 = [90,1,4,0],
solution 3 = [80,10,0,0],
solution 4 = [20,20,20,10] etc.

In EAs each solution is called an 'individual' and all the proposed individuals grouped together are called a 'population'. The next step is to assess how good each individual is, also known as the 'fitness' of the individual. The 'fitness function' is the procedure used to evaluate individuals so in this case it is a measure of how near to 100 (the optimum) the sum of each solution's string of numbers is. This can be calculated because for the training set data the 'answer' is revealed – i.e. solution 1 and 2 have more 'fitness' (each of their sums is 95) than solution 3 (sum = 90) which in turn has more fitness than solution 4 (sum = 70).

The next step of the EA process is to select the fittest individuals from the population to be the 'parents' of the next population of 'children' in order to 'evolve' even better individuals. There are different ways of defining which individuals are selected as parents depending on the particular problem being solved but, in this hypothetical example, it might be only individuals that have a sum within the range of 95-100.

The first generation parent individuals can undergo ‘crossover’ to recombine them in some manner to produce new child individuals in the second generation, analogous to breeding and reproduction. In order to keep some ‘genetic diversity’ (i.e. enable individuals to evolve that will generalise beyond the training data population) a random selection of lesser performing individuals are also selected as parents. This means that each child individual will be a combination of the parents’ individuals and likely to be different to all the other child individuals in that generation.

Usually each child individual is produced by combining 50% of the mother’s data with 50% of the father’s data; for example:

Father individual = solution 1 = [95, 0, 0, 0]
Mother individual = solution 2 = [90, 1, 4, 0]
Child individual = 50% father [95, 0] and 50% mother [4, 0] = [90, 5, 4, 0]

In this particular example the child individual is ‘fitter’ than either of the parents because it’s sum is nearer to 100. The parent individuals from the first generation are recombined until enough children individuals are reproduced so the second generation is the same size as the first generation. This means that if the first generation had 100 individuals and 20 of these were selected as parents, they would need to produce 80 children individuals so that the second generation had a total of 100 individuals. This usually means that each parent needs to cross-over with multiple individuals (and hence are not monogamous!). The second ‘generation’ would therefore be a combination of the 20 fittest parent individuals from the first generation and the 80 new children individuals. The final step of the evolution cycle is for a small randomly determined portion of the new population individuals to undergo mutation. This increases genetic diversity and hence increases the chance of the individual generalising to data beyond the training set.

This process of fitness evaluation, selection and reproduction is repeated numerous times. Each repetition of the cycle leads to a generation that contains some individuals that are even fitter than the best ones in the

previous generation. In other words the individuals 'evolve' to be fitter, or better performing, the more times the cycle is repeated. The limits for this evolution may be defined based on the number of generations, time limits or until a minimum fitness is achieved e.g. in the hypothetical example until the sum equals 100. However as these methods lead to a gradual loss of genetic diversity there is also a limit to how long the individuals can continue to improve; this is called the 'convergence' of an algorithm and in simple terms means that the next generation of children individuals are no better than the parent individuals from where they derived.

Finally, it is important to note that each time the EA process is run again from the beginning there will be different individuals produced because of differences in terms of initial random population, parent cross-over patterns, random selection of low-fitness parents and random mutations introduced.

3.6.3 Why were evolutionary algorithms used in this study?

EAs were used in this study to develop, or 'induce' equations that could predict whether FT movement data was from a PD patient or a control. In other words the fittest individual was an equation known as a 'classifier' that could best match the kinematic data to the correct diagnostic group.

Classifiers are mapping techniques that find the best 'route' from a set of data (e.g. FT data) to a group label (e.g. PD vs. HC) and this makes them particularly helpful for classifying data. Previous studies have demonstrated that component measures of FT movements such as speed and amplitude have overlapping ranges in PD and controls (Espay, Beaton et al., 2009, Ling, Massey et al., 2012) so linear statistical methods are limited in their capacity to classify because threshold levels lead to considerable numbers of false positive or negative results.

A further advantage of using EAs to generate classifiers is that they have the potential to discover novel clinical information. This is because EAs do not use any assumptions about what features of the data are likely to give the best solutions and hence they search broadly and can 'find' discriminating

features without bias or constraints. In contrast standard methods of comparing PD and HC movement data involve measuring focussed components of the data that have been observed clinically, such as reduced amplitude in PD. As there remains an incomplete understanding of the pathophysiology of PD bradykinesia, and how it manifests in FT, the former approach may be beneficial. The evolved classifier equations can then be examined to understand what the most discriminating features of the data are.

3.6.4 How did evolutionary algorithms discriminate PD from HC data?

The acceleration data series were uniformly partitioned into training, validation and test sets in a ratio of 4:1:1 (Table 8). Two thirds of the data was used for training the EAs to develop classifiers and one third of the data was used as non-training data to assess how well the classifiers generalised to other unseen data. The non-training data was subdivided into validation and test data sets and the validation set was used to identify the best performing classifier and the tests set is used to give an unbiased measure of its ability to discriminate between PD and HC.

Table 8 Division of finger tapping data sets for EA analysis

	Training	**Validation**	**Test**
Patients	66	16	16
Controls	56	14	12

Legend: Each participant had two acceleration data sets (left and right hand) and these were kept together within the same test sets (training vs. validation vs. test) but analysed separately.

The training FT data sets were presented with the individual diagnoses revealed as PD = +1 and HC = -1. The EA used a repetitive selection, reproduction and mutation processes for 100 generations until a classifier

was developed that maximised the area under the ROC curve (AUC) for discriminating between PD and HC.

The whole EA process was repeated 50 times so 50 different classifiers were produced. In order to identify the classifiers with the best diagnostic potential (that generalise to subjects beyond the training set) all 50 classifiers were re-evaluated on the 'validation' set - PD and HC data not previously seen by the EA and with the diagnosis concealed. The best classifiers selected by this process were then evaluated on another subset of unseen PD and HC data – the test set. This gives an unbiased estimate of their diagnostic ability and is a measure of the classification accuracy. The ability of the selected classifiers to generalise beyond the training data set was further tested by assessing the ensemble classifier on an independent sample of FT data collected from 13 PD patients and nine HC at UCSF.

3.6.5 Two distinct approaches to classifier development

In this study two different evolutionary approaches to classifier development were used. These were based on a novel type of Cartesian genetic programming (CGP) (Smith, Leggett et al., 2005) and artificial biochemical networks (ABNs) (Smith and Timmis, 2008; Lones, Smith et al., 2013). These are both advanced algorithmic paradigms inspired by conventional genetics and biochemical pathways respectively.

The CGP method uses a sliding window of 0.6-second duration to assess the local patterns of movement in the FT data to induce a classifier. The overall output of the classifier represents the mean occurrence of the movement patterns over the whole assessment period. ABNs are a form of classifier induction based on the chemical reactions seen in cells – hence enzymatic reactions occur and the concentrations of various chemicals varies over time in relation to these reactions, and ABNs thus analyse changes over time within a data set. Whilst the two evolutionary approaches to classifier development used in this study are quite distinct, the resultant classifiers may be considered somewhat complementary as CGP looks at local features of the data within a FT cycle and ABNs respond to global features across a

number of FTs. The best classifier produced by CGP was then combined with the best classifier produced by ABNs to produce an 'ensemble classifier.'

A similar type of standard generational EA was used to evolve both types of classifier: a standard generational EA with an initial population size of 200 individuals, evolved over 100 generations, with children individuals generated from parents using uniform crossover, and a random mutation rate of 6%.

3.6.6 Examining the classifiers to reveal discriminating features

An advantage of using CGP classifiers is that they can be explored relatively easily to reveal what features of the data were used to construct them. This was done by taking twelve data windows (0.6 seconds each) from six PD and six HC that had the strongest CGP classifier results (closest to +1 for PD and -1 for HC) and then examining these to find out what features best discriminated PD from HC. Specifically, the CGP classifier was expressed in order to reveal which acceleration data points from training data were used in its formation and these were then overlayed onto the corresponding sections of the tap cycle in the strongest scoring windows to focus the search for the most distinguishing features.

3.7 Statistical Analysis

Statistical analyses were performed using IBM Statistical Package for the Social Sciences release 19 (Chicago SPSS Inc.), SciPy (Jones *et al.*, 2001) and R (R Team, 2013). A p-value ≤ 0.05 denoted statistical significance. Demographic and clinical data were compared using independent t-tests and applying Levene's test for equality of variance. The Spearman Rank Correlation Coefficient was used to assess the relationship between demographic, clinical and kinematic variables.

D'Agostino's K-squared test, a measure of departure from a normal distribution, showed that all movement component measures in the patient group, and most in the control group, were positively skewed (D'Agostino,

1971, D'Agostino and Pearson, 1973). Transformations with logarithm and square of the results did not achieve sufficient normalisation of the data so non-parametric tests were used. Movement component measures were compared between PD and HC groups using the Mann-Whitney U-test, and compared between PD patients *on* and *off* using the Wilcoxon Signed Rank test. The Friedman non-parametric ANOVA and Wilcoxon Signed-Rank *post hoc* tests were used to assess if the SE results were associated with length of tapping sequence; p-values were adjusted using the Bonferroni Correction for four repeated measures. The Kendall Tau-b Rank Correlation Coefficient was used to assess the degree of independence between HY stages and MDS-UPDRS grades and decrementing amplitude and speed values.

Logistic regression was used to form composite variables comprising the separable component measures. The classification accuracy of the separable component analysis and the EA induced classifier was measured using AUC of ROC curves[18] with the null hypothesis denoted by AUC of 0.5.

[18] See Chapter 2 for more details on ROC curves

Chapter 4

Analysis of the separable component measures of bradykinesia: Results and Discussion

The clinical details of the participants and the kinematic analyses results from the FT study are presented in this chapter. Firstly the clinical and demographic details of the patient and control groups will be outlined. Secondly the issue of corrupt data is discussed and the clinical details of the groups presented again for when only the non-corrupt, or approved, data sets are included. The kinematic data results will then follow with a specific focus on measuring the separable components of bradykinesia (as defined by UPDRS and UKBBDC) in the FT kinematic data, namely amplitude, speed, rhythm, frequency, halts, decrementing speed and decrementing amplitude.

The separable component measures will then be evaluated in two main ways – firstly to assess how accurately each component, individually or when combined into a 'bradykinesia composite' model classifies the data sets into the correct diagnostic group and secondly to assess how well the component measures correlate with disease progression. These two broad approaches to analysis are important for informing the development of a device to aid PD diagnosis and monitoring respectively.

4.1 Clinical Results

4.1.1 Demographic and clinical details of participants

The demographic and clinical details of the subjects recruited from LTHT and UCSF are summarised in Table 9. In the LTHT cohort the patient group was significantly older and had a greater proportion of men than in the control group. There was no significant difference between the UCSF groups in terms of age or gender.

Table 9 Demographic and clinical details of participants

LEEDS	**Patients N=49**	**Controls N=41**	***p***
Gender, M : F	31 : 18	15 : 26	< 0.001
Age, years	67.2 ± 8.7	63.7 ± 9.6	0.01
Dominant hand, R : L	41 : 8	32 : 9	0.49
MoCA[1]	26.1 ± 2.6	28.3 ± 1.4	< 0.001
HY stage	2.4 ± 0.7 *(1 - 4)*	NA	-
PD duration, years	5.9 ± 3.9 *(0.5 - 18)*	NA	-
LEDD, mg	723 ± 435 *(140 - 2080)*	NA	-

SAN FRANCISCO	**Patients N=13**	**Controls N=9**	***p***
Gender, M : F	9 : 4	4 : 5	0.24
Age, years	67.2 ± 7.0	70.2 ± 4.3	0.27
Dominant hand, R : L	13 : 0	8 : 1	0.22
MoCA	27.6 ± 1.3	27.7 ± 1.3	0.94
HY stage	1.7 ± 0.7 *(1 - 3)*	NA	-
PD duration, years	4.5 ± 3.5 *(0.5 - 12)*	NA	-
LEDD, mg	540 ± 321 *(75 – 1188)*	NA	-

Legend: All values are presented as mean ± one standard deviation (*range*) except for gender and handedness ratios. Abbreviations: M, male; F, female; R, right hand; L, left hand, NA, not applicable. [1]The first four patients and three controls in Leeds did not undertake MoCA as it was introduced into the protocol after their assessments had been completed.

Most participants were right-handed with no difference in the proportion of left-handedness between the patient and control groups at either site. The LTHT and UCSF patients had similar age and gender distributions (ps >

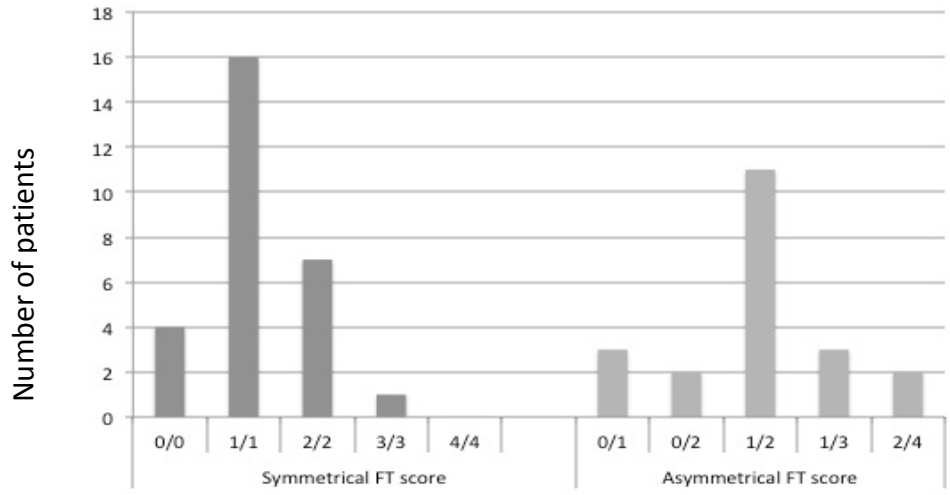
0.1) but the LTHT patient group had longer disease duration ($p < 0.05$) and more advanced disease stage ($p < 0.05$) than UCSF patients.

4.1.2 Clinical grades of finger tapping assessments

Each subject received two MDS-UPDRS bradykinesia grades (one for each hand) ranging from zero to four for their FT assessments. All controls' FT assessments were graded zero, denoting there was no clinical evidence of bradykinesia.

For the LTHT patients the grades were normal in both hands for 8% (n = 4), and the rest had a degree of bradykinesia in at least one hand as follows: 39% (19) slight, 41% (20) mild, 8% (4) moderate and 4% (2) severe. Fifty seven percent (28) of the PwPD had symmetrical clinical bradykinesia grades. Of this group 14% (4) were graded as normal, 57% (16) slight, 25% (7) mild, and 4% (1) moderate bilaterally. Forty three percent (21) of patients had asymmetric clinical bradykinesia grades, with 77% (14) of these having one point difference and 33% (7) two points difference between hands (Figure 27).

Figure 27 MDS-UPDRS finger tapping grades for Leeds patients



Legend. Distribution and asymmetry of FT clinical grades for LTHT patients. The different combinations of grades for each pair of hands is shown on x-axis i.e. 0/0 denotes grade zero in both hands and 0/1 denotes grade zero in one hand and grade one in the other.

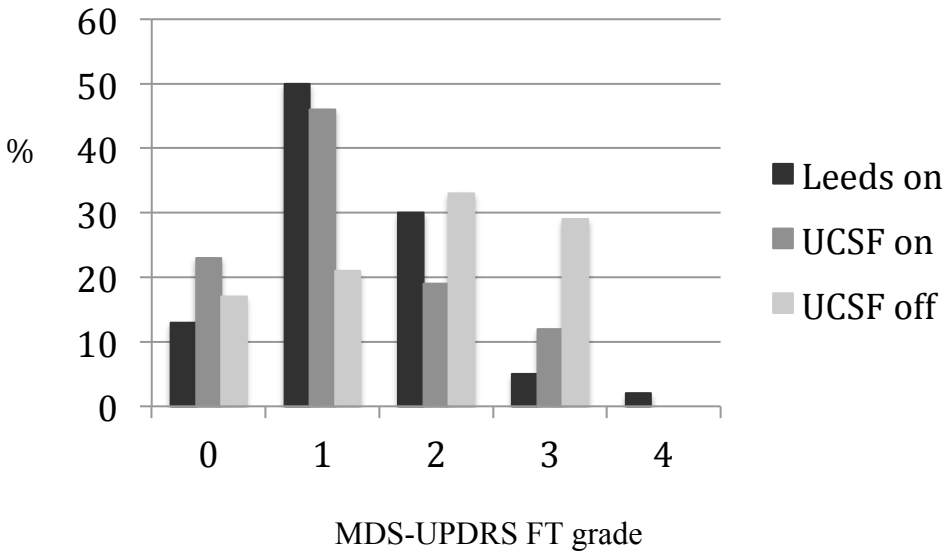
The distribution of patient clinical FT grades was skewed with 63% of LTHT patients, and 69% of UCSF patients *on* being allocated grade zero or one (Table 10). In UCSF patients *off* there was a greater proportion of higher clinical grades with 63% of the assessments receiving MDS-UPDRS grades two or three. None of the UCSF patient assessments, and only two out of 98 Leeds assessments received a grade four (Figure 28). In summary the majority of the patients *on* had only slight bradykinesia and the distribution of grades was similar in the Leeds and UCSF patient groups.

Table 10 Distribution of patient clinical bradykinesia grades

		MDS-UPDRS FT grade				
Patient group	**FT assessments, n**	**0**	**1**	**2**	**3**	**4**
Leeds *on*	98	13	49	29	5	2
UCSF *on*	26	6	12	5	3	0
UCSF *off*	24	4	5	8	7	0

Legend: One UCSF patient was not tested in the *off* state so there are only 24 UCSF *off* assessments.

Figure 28 Distribution of clinical grades for LTHT and UCSF patients



Legend: Percentage of FT assessments from Leeds and UCSF allocated each MDS-UPDRS clinical grade.

4.2 Corrupted kinematic data

4.2.1 Reasons for corrupted data

For the majority of assessments the FT movements were reliably recorded as continuous kinematic data, with the separation of the two EM sensors measured in three translational planes (x, y and z). Unfortunately some of assessments contained sections of 'corrupted' kinematic data. This means that the movement data recorded had EM interference or 'noise' within it (Figure 29).

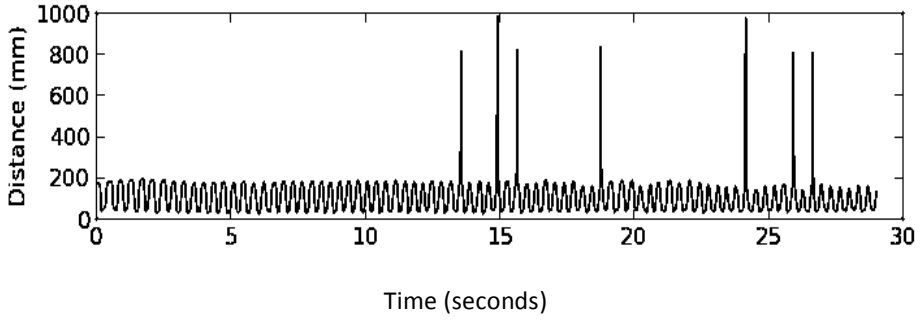
The two likely explanations for corrupted kinematic data are interference from other electrical equipment (e.g. telephones and computers) nearby and misplacement of the EM tracking sensors in relation to the EM transmitter. Displacement of the two sensors are calculated based on the EM signals that the transmitter box receives from the sensors in a 180^{o} arc anterior to the front aspect of the box. This means that if the sensors move into the hemisphere *posterior* to the transmitter the EM signals from the sensors are recorded as occurring in the opposite direction. Furthermore, as the sensors cross the hemisphere boundary there is a momentary loss of reception and this artefact results in a break in the data recording and additional corruption.

Various procedures were tried to 'de-corrupt' the data such as inverting the polarity of the corrupted points, and interpolating between the last known accurate data point and the corrupted data points but these methods were unsuccessful as there were too many data points to be able to accurately interpolate in most cases.

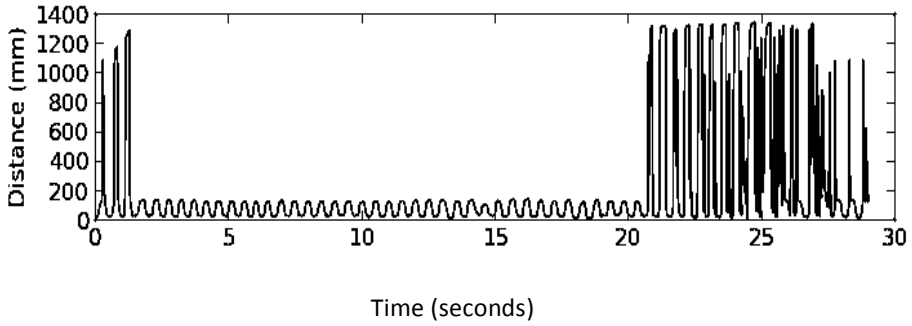
Figure 29 clearly demonstrates that the corrupted data points would lead to erroneous results when measuring the separable components of bradykinesia. The corrupted data sets were therefore excluded from the separable component (Chapter 4) and SE analyses (Chapter 6).

Figure 29 Examples of corrupted data recordings

(a) Control C11



(b) Control C42



Legend: Examples of corrupted separation data are shown for control subjects (a) C11 and (b) C42. Approximately 5% of the data in C11's recording is corrupted: after the first 14 seconds there are then seven sections of corrupted data, shown by large vertical spikes in the data. Approximately 30% of the data in C42's recording is corrupted with corrupted data points from zero to two seconds and then from approximately 20 seconds until the end of the recording.

For the EA analyses (Chapter 5) all kinematic data sets were included. This analytic method searches for patterns of movement that are over–represented in the patient data compared to control data in order to form classifiers that can predict the diagnostic group. As the sections of corrupted data were generally only a small section of the whole kinematic data collected and as the patient and control group data was affected to roughly the same degree, it was hypothesised that the corrupt section of data would

have negligible effect on the diagnostic accuracy of this method; this hypothesis was tested and shown to be correct (section 5.1.4).

4.2.2 Prevalence of corrupted data

Eleven percent of the LTHT data sets (21 out of 180 recorded assessments), and 10% of the UCSF data sets (7 out of 68) had sections of corrupted data. Of the 21 corrupted LTHT data sets, eleven were from patients (4:7 dominant: non-dominant hand) and ten were from controls (4:6 dominant: non-dominant hand). Of the seven UCSF corrupted data sets, two were from controls (both non-dominant hands) and five from patients (two *off*, three *on*, 4:1 dominant: non-dominant hand). No participant had both recorded data sets corrupted.

4.2.3 Demographic and clinical details when corrupted data excluded

The revised demographic and clinical details of the approved data sets that were used for the separable component measures and SE analyses, *with the corrupted data sets excluded* are outlined in Tables 11 and 12.

Table 11 Demographic and clinical details of approved LTHT data sets

		Patients N = 49	**Controls N = 41**	***p***
Total assessments	N	98	82	
Approved data sets	N (%)	87 (89%)	72 (88%)	0.84
Gender	M : F	57 : 30	23 : 49	< 0.001
Age, years	Mean ± SD	68.4 ± 8.4	63.1± 9.5	< 0.001
Dominant hand	R : L	74 : 13	56 : 16	0.2
MoCA score	Mean ± SD	25.9 ± 2.7	28.3 ± 1.4	< 0.001

In the LTHT patient group the mean FT grade for the approved data sets was 1.32 ± 0.84 distributed as: 11 grade zero, 45 grade one, 25 grade two, four grade three and two grade four. The approved LTHT patient and control groups was even less closely matched for age; when all data sets were included the mean ages (rounded to the nearest integer) from the

patient and control groups respectively were 67 and 64 years ($p < 0.01$) but with the corrupt data sets excluded these were 68 and 63 years ($p < 0.001$). Table 12 shows that a similar proportion of data sets from each site was corrupted and that the approved patient and control data sets from LTHT and UCSF were similar in terms of age, gender, FT grade and PD disease duration. This is an important point because it means that the UCSF data is a good validation group to assess how well the LTHT results generalise to data collected independently from similar subjects.

Table 12 LTHT and UCSF groups when corrupt data excluded

PATIENTS	**Leeds N = 49**	**UCSF N = 13**	***p***
Approved data, n/total	87/98, 89%	23/26, 88%	0.96
Gender, M : F	57 : 30	17 : 6	0.45
Age, years	68.4 ± 8.5	67.3 ± 6.4	0.54
Dominant hand, R : L	74 : 13	23 : 0	0.07
MoCA	25.9 ± 2.7	27.6 ± 1.3	<0.001
H&Y stage	2.4 ± 0.7	1.7 ± 0.7	<0.001
MDS-UPDRS FT grade	1.32 ± 0.84	1.26 ± 0.96	0.78
PD duration, years	5.6 ± 3.5	4.4 ± 3.6	0.17

CONTROLS	**Leeds N = 41**	**UCSF N = 9**	***p***
Approved data, n/total	72/81, 88%	16/18, 89%	0.89
Gender, M : F	23 : 49	7 : 9	0.37
Age, years	63.1 ± 9.5	70.5 ± 4.4	0.003
Dominant hand, R : L	56 : 16	14 : 2	0.51
MoCA	26.3 ± 2.5	27.6 ± 1.2	0.07

Legend: The clinical and demographic details of the groups comprising approved data sets i.e. when corrupted data sets are excluded. Values are presented as mean ± one standard deviation except for gender and handedness ratios. The UCSF patient data presented is for the assessments performed in the *on* state.

4.2.4 Correlation of demographic and clinical variables

The Spearman correlation coefficient, r_s is a non-parametric measure of the strength and direction of association between two variables. It was calculated for the demographic and clinical variables of the approved LTHT data. Control and patient data was analysed separately and the patient results are presented in Table 13. In the control group there was a trend towards older age being correlated with lower MoCA scores (r_s -0.21, p = 0.081) but there were no other significant correlations between the variables.

Table 13 Correlogram of clinical and demographic variables in LTHT patients

	Gender	Age	Handed	Dom hand	HY	Duration	MoCA	LEDD	UPDRS
Gender		.22*	-.10	.03	.05	-.15	-.09	-.17	.006
Age	.22*		-.02	.04	.35*	.04	-.13	.26*	.06
Handed	-.10	-.03		-.02	.18	.32*	-.09	.16	.02
Dom hand	.03	.04	-.02		.002	.05	-.005	-.02	.10
HY	.05	.35*	.18	.002		.35*	-.24*	.33*	.20*
Duration	-.15	.04	.32*	.05	.35*		-.10	.24*	-.03
MoCA	-.09	-.13	-.09	-.005	-.24*	-.10		.09	-.05
LEDD	-.17	.26*	.16	-.02	.33*	.24*	.09		.05
UPDRS	.006	.06	.02	.10	.20*	-.03	-.05	.05	

Legend: Abbreviations: Handed, side of dominant hand i.e. right vs. left; Dom hand, dominant hand vs. non-dominant hand; HY, Hoehn and Yahr disease stage; Duration, PD disease duration; MoCA, Montreal Cognitive Assessment; LEDD, levodopa equivalent daily dose; UPDRS, Unified Parkinson's Disease Rating Scale finger tapping score; * significant at $p < 0.05$.

Table 13 shows that several clinical and demographic variables were significantly associated in the patient group. Age and gender were correlated (r_s 0.22, p = 0.04) reflecting that the majority of patients were male and they tended to be older than female patients. Age was also correlated to HY stage (r_s 0.35, p = 0.001) and LEDD (r_s 0.26, p = 0.01) suggesting that the older patients tended to have more advanced disease and be taking higher doses of medications. Handedness was correlated only with disease duration (r_s 0.32, p = 0.002) presumably reflecting that most patients were right handed and that the few left handers in the study tended to have a shorter than average disease duration.

HY stage was the variable with the greatest number of significant correlations: in this study the patients with a higher HY stage (more advanced disease) tended to be older (p = 0.001), have a longer disease duration (p = 0.001), a lower cognitive score (p = 0.04), be taking higher doses of dopaminergic drugs (p = 0.002), and have greater severity of clinical bradykinesia (p = 0.06).

Summary of section 4.1 and 4.2 results

The results so far have shown that approximately 10% of data sets were corrupted. When these were excluded (for the separable component measures and the SE analyses) the remaining approved UCSF patient and control group data were matched for age and gender. However this was not the case for the LTHT subjects with mean age for approved patient and control group data being 68 and 63 years old. Importantly the LTHT patient and control groups were matched with the respective UCSF groups for age, gender, FT grade and disease duration though, which makes the UCSF groups particularly useful for validating the Leeds results.

4.3 Classification accuracy of individual separable component measures

Two different approaches to data analysis were used. The first approach was to measure the separable components of clinically defined bradykinesia (amplitude, speed, rhythm, halts and decrements) using the normalised FT separation data and then assess how well these component measures,

individually or as composite models, classified the patient and control data into the correct diagnostic group. The diagnostic classification accuracy results for individual component measures and for bradykinesia composite models are presented in sections 4.3 and 4.4 respectively. The second approach, presented in Chapter 5, examined how well the ensemble classifier, developed through EA analysis of the FT normalised acceleration data, could discriminate patient and control data.

4.3.1 Classifying Leeds PD and healthy control data

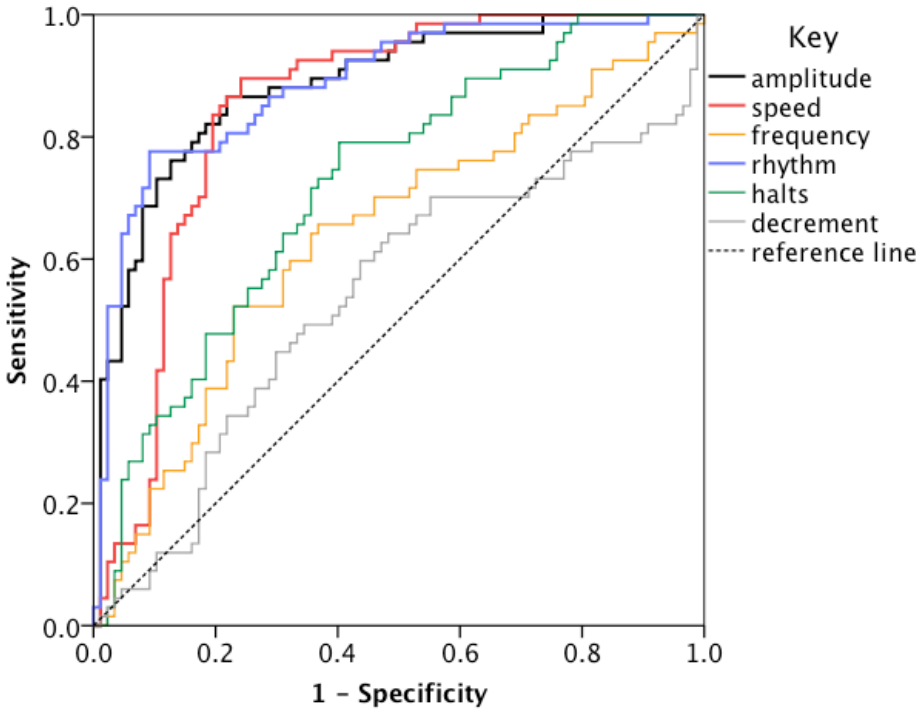
The results presented are based on measuring the separable components of bradykinesia in 87 Leeds patient FT data sets and 72 Leeds HC FT data sets, thus using only the approved data. AUC was used to compare diagnostic accuracies i.e. how well each component measure of bradykinesia classified data into the correct diagnostic group. The sensitivity/specificity of each component to discriminate PD from HC data is presented at the threshold of equal trade off. The following separable movement components were chosen for analysis as they represent measures of the clinical components of bradykinesia described in the UKBBDC and MDS-UPDRS respectively:

- Tapping frequency
- Speed - mean and maximum
- Amplitude – mean and maximum
- Rhythm – Coefficient of variation (COV) for speed and amplitude
- Halts - % of tap cycle at 'zero' speed (<5% maximum speed)
- Decrementing speed
- Decrementing amplitude

For the speed, amplitude and halts measurements both mean and median values were initially calculated and are presented in full in Table 14. Subsequently only the mean values are presented in the rest of this chapter as they were consistently, albeit marginally, more discriminative than the median measurements.

The results in Figure 30 and Table 14 demonstrate that the most discriminating component measures were maximum amplitude, rhythm (both COV speed and COV amplitude), mean amplitude and maximum speed with AUCs 0.88, 0.88, 0.86 and 0.84 respectively. Mean speed and halts discriminated the groups moderately well, with AUCs of 0.79 and 0.72, but tapping frequency, decrementing amplitude and decrementing speed discriminated the subjects poorly.

Figure 30 ROC curves for separable component measures of bradykinesia



Legend: The ability of each bradykinesia separable component measure to discriminate PD from HC FT data is summarised. For clarity only the best scoring measure of each component is presented. Amplitude is the mean of the tap cycles' maximum amplitude, speed is the mean of the tap cycles' maximum speed, frequency is tapping frequency, rhythm is COV amplitude, halts is the mean of the tap cycles' percentage at zero speed, decrement is amplitude decrement. The reference line denotes an AUC of 0.5 or a discrimination that is no better than chance i.e. 50% sensitivity and 50% specificity.

Table 14 Diagnostic accuracy of individual separable components measures of bradykinesia

Component measure of FT		AUC	95% CI	*P*
Tapping frequency		0.62	0.54 – 0.71	0.007
Mean speed	mean	0.79	0.72 – 0.86	2.6×10^{-10}
	median	0.79	0.72 - 0.86	3.7×10^{-10}
Maximum speed	mean	0.84	0.78 – 0.91	1.1×10^{-13}
	median	0.84	0.78 - 0.90	1.9×10^{-13}
Mean amplitude	mean	0.86	0.80 - 0.92	3.8×10^{-15}
	median	0.86	0.80 – 0.92	8.2×10^{-15}
Maximum amplitude	mean	0.88	0.82 - 0.93	4.0×10^{-16}
	median	0.87	0.81 – 0.93	1.3×10^{-15}
Rhythm	COV amplitude	0.88	0.82 – 0.93	4.8×10^{-16}
	COV speed	0.88	0.82 – 0.93	3.9×10^{-16}
Halts	% zero speed mean	0.72	0.64 – 0.80	1.6×10^{-6}
	% zero speed median	0.70	0.62 – 0.78	1.34×10^{-5}
Decrementing amplitude		0.57	0.48 – 0.66	0.14
Decrementing speed		0.54	0.44-0.63	0.41

Abbreviations: AUC, area under ROC curve; CI, confidence interval; COV, coefficient of variation – calculated by the SD/ mean.

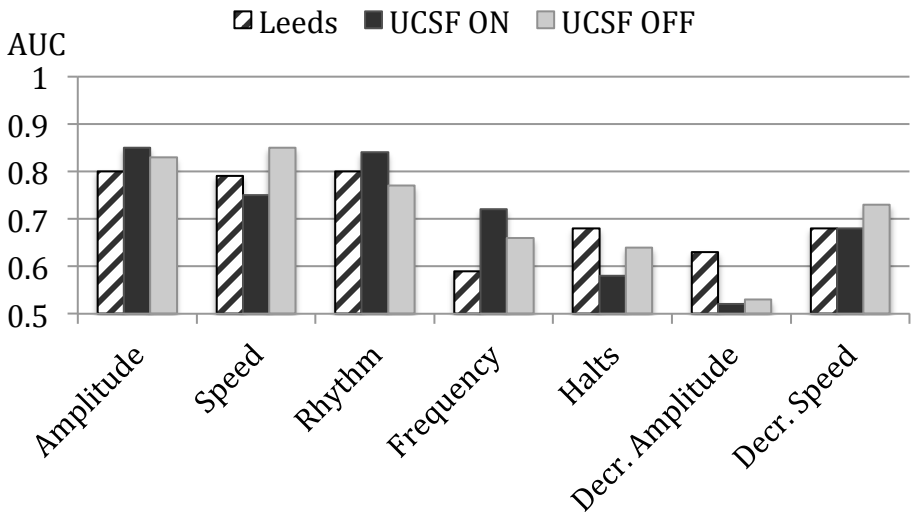
4.3.2 Validating classification accuracy on San Francisco data

In order to test how well these findings generalise to an independent data set, the separable components were measured in the UCSF FT data too. The mean FT grades ± 1SD in the LTHT and UCSF patient data were 1.32 ± 0.84 and 1.26 ± 0.96 ($p = 0.78$) respectively. There were very few UCSF controls and this makes AUC a less reliable measure of classification

accuracy because each misclassified case will have a disproportionately large effect on the overall results. To get around this problem the classification accuracy of the separable component measures was assessed on the validation set by using the UCSF patient data (n = 23) and a control group comprising the UCSF (n = 16) *and* the LTHT controls (n = 72) combined. This method also enabled direct comparison of the AUC results for the LTHT and UCSF patients, because they could both be compared to the *same set of combined control data*.

Using this method the component measure results in the UCSF *on* patient data showed a very similar trend of results to the LTHT patient data with amplitude, rhythm and speed being the most discriminating components (AUCs 0.85, 0.84 and 0.75 respectively); see Figure 31 and Table 15. In contrast to the LTHT patients, tapping frequency (AUC 0.72) and decrementing speed (AUC 0.68) classified the UCSF patients better than halts (AUC 0.58).

Figure 31 Classification accuracy of separable component measures in LTHT and UCSF



Legend: The classification accuracy of each separable component measure to discriminate patient data (LTHT, UCSF ON and UCSF OFF) from 'all controls' (LTHT and UCSF data sets combined) is presented as AUC.

The results suggest that all the separable component measure results from LTHT patient data (except for the tap frequency) generalise to the UCSF *on* patient data with the same rank order (except for tap frequency) and roughly the same magnitude of classification (i.e. similar AUCs) observed. In descending order of classification accuracy the separable component measures is summarised below:

Leeds *on*:	**UCSF *on*:**
Amplitude	Amplitude
Rhythm	Rhythm
Speed	Speed
Decrementing speed	Frequency
Halts	Decrementing speed
Decrementing amplitude	Halts
Frequency	Decrementing amplitude

Regarding different clinical states, Table 15 shows that measures of speed and decrementing speed are more discriminatory when the patients are in the *off* state. The most discriminatory component measures, in descending order for each clinical state are outlined below:

***on* clinical state**: amplitude > rhythm > speed

***off* clinical state**: speed > amplitude > rhythm

This finding is discussed further in section 4.7 but, in brief, is consistent with studies that showed dopaminergic drugs (that promote the *on* state) disproportionately improve the speed component of bradykinesia and have less effect on improving amplitude and rhythm (Espay, Giuffrida et al., 2011).

Table 15 Classification accuracies of separable component measures in LTHT and UCSF

	LTHT *on*	UCSF *on*	UCSF *off*
Amplitude	**0.81**	**0.85**	0.83
Speed	0.79	0.75	**0.85**
Rhythm	0.80	0.84	0.77
Frequency	0.59	0.72	0.66
Halts	0.68	0.58	0.64
Decrement amplitude	0.63	0.52	0.53
Decrement speed	0.68	0.68	0.73

Legend: Comparisons of AUC were made for approved LTHT patient data (n = 87) vs. all controls (approved LTHT and UCSF control data combined; n = 88), for UCSF *on* patient data (n = 23) vs. all control data (n = 88), and for UCSF *off* patient data (n = 22) vs. all control data (n = 88). The most discriminatory component AUC in each group is highlighted in bold.

4.3.3 Examining how PD patient movements differ from controls

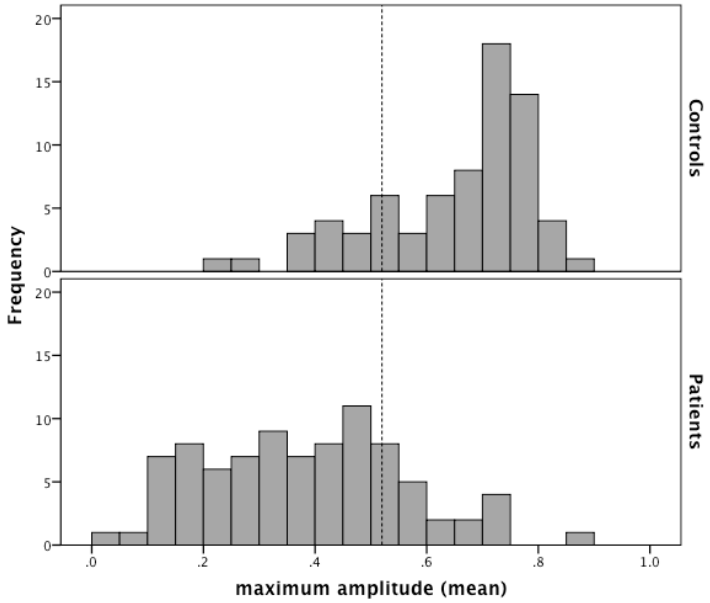
In order to better understand what the AUC results signify clinically the distribution of the separable component measures within each group and between groups was examined. Histograms of the component data are presented for the LTHT PD and HC groups with ROC threshold levels, that give maximal sensitivity/specificity with equal trade off, marked as a vertical dashed line; see Figures 32 (a-k).

Figure 32 (a-k) Comparing separable component measures of bradykinesia in Leeds patients and controls

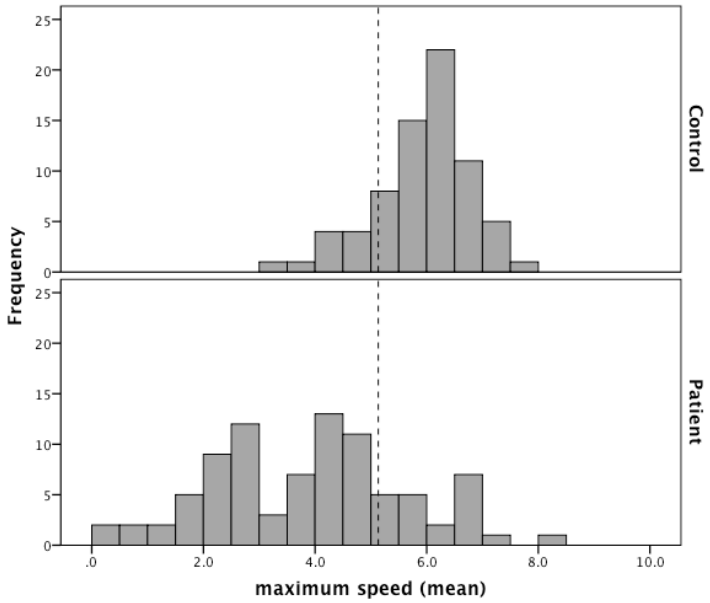
4.3.3.a Maximum amplitude

For each data set the mean of the normalised maximum amplitude of each tap cycle during the 30 second recording is presented. The histograms show that PD subjects tend to have a smaller maximal separation of the finger and

thumb than HCs. In other words the finger and thumb do not 'open up' as much between each tap. This component measure has a sensitivity/specificity of 0.81/0.79.



4.3.3.b Maximum speed

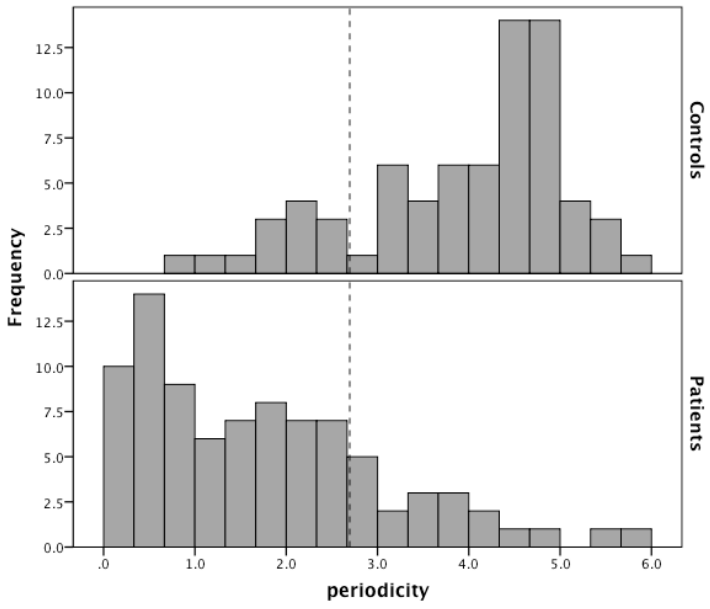


The histograms of maximum speed measurements (mean of maximum speed for each tap cycle) confirm that PwPD tend to have a slower maximum speed than controls. This component has a sensitivity/specificity

of 0.79/0.78 for discriminating the groups. There is considerable overlap between the groups though and it is noteworthy that some PwPD are tapping as fast as the fastest controls. It is possible that these patients have a fast speed but small amplitude of movement though and to take this into account the product of the mean maximum amplitude x mean maximum speed is also presented.

4.3.3.c Maximum speed x maximum amplitude

By multiplying the maximum speed and maximum amplitude of each tap cycle together, to obtain 'periodicity' the potential reciprocal relationship of each component is taken into account. This gives an interesting distribution with each group skewed but in opposite directions. This suggests that patients tend to tap with smaller *and* slower movements than controls. However the maximal sensitivity/specificity is not significantly better (at 0.82/0.79) than each individual component measure as the range of values still overlap considerably.



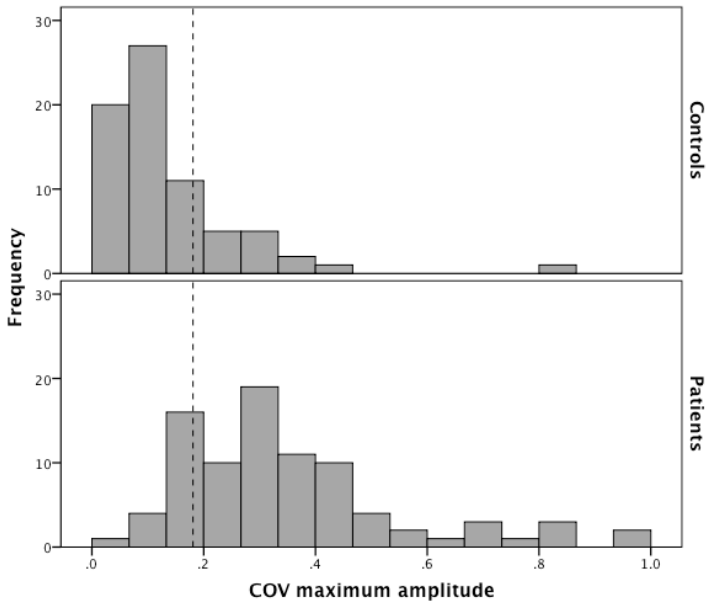
Rhythm

There are two measurements for rhythm, namely COV amplitude and COV speed, measuring how much variability in the maximum amplitude and speed of the tap cycles respectively during each 30-second assessment

period. Higher values of COV denote more variation between tap cycles and hence less rhythmic FT movements.

4.3.3.d COV maximum amplitude

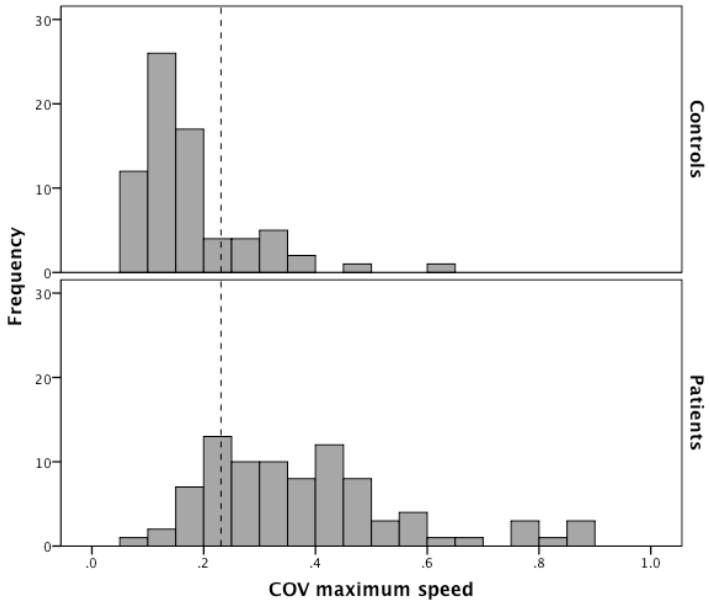
The COV maximum amplitude histogram shows that in the control group the vast majority of data sets have zero (n = 19) or minimal, 0.1 (n = 27) variation from the mean amplitude score, implying that the maximum amplitude of each tap stays fairly uniform throughout the 30 second tapping sequence. This finding contrasts with the PD group results where nearly all the PwPD exhibited a degree of variability in the maximum amplitude of tap cycles. Only one PD patient FT assessment had zero, and three had minimal (0.1), variation in maximum amplitude. A threshold value of 0.18 had sensitivity/specificity of 0.85/0.78 for discriminating the groups.



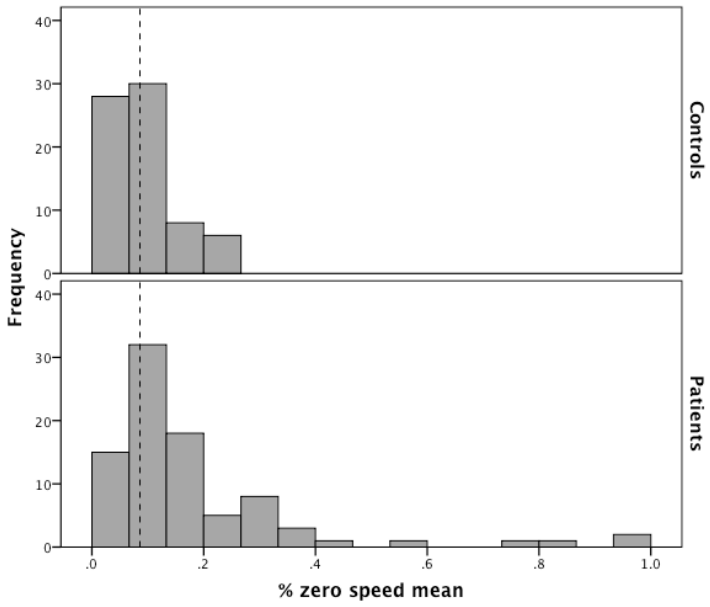
4.3.3.e COV maximum speed

The COV maximum speed distributions in the PD and HC groups have a similar pattern to COV maximum amplitude. In the PD group the vast majority of assessments show that the maximum speed per tap cycle varies considerably over the 30-second assessment whereas the majority of the control assessments showed only minimal variability. At the threshold COV

maximum speed of 0.23 this measurement of rhythm has a sensitivity/specificity of 0.80/0.78.



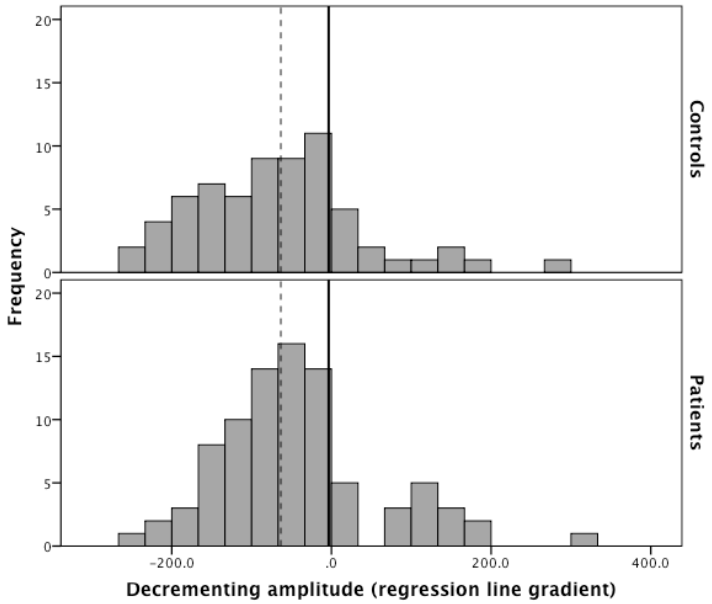
4.3.3.f Halts



Very few of the assessments in either group had any significant halts or hesitations. Even in the PD group less than 1% of the duration of the FT assessment is spent at such low speeds as to be considered a halt. Fifteen perecent of the PD group had no halts compared to 34% of the controls. However halts did not discriminate the groups particularly well and a

threshold of equal trade off, at 0.086%, gave a sensitivity/specificity of 0.69/0.64.

4.3.3.g Decrementing amplitude

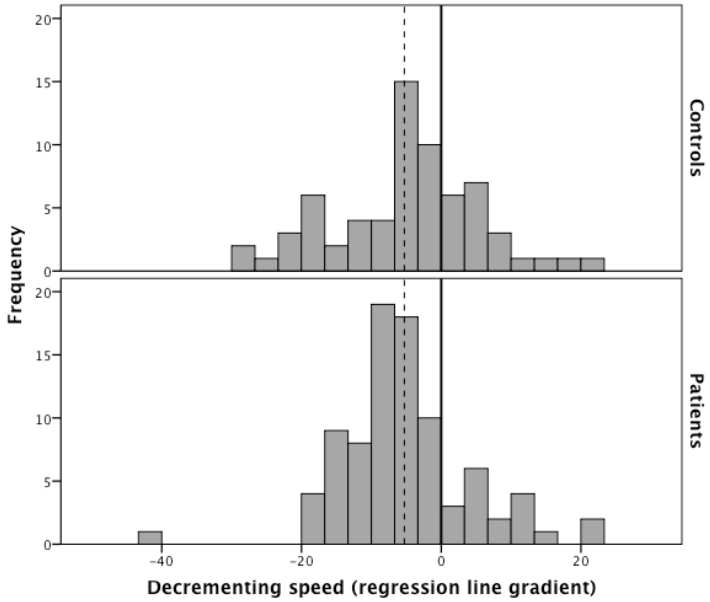


The solid vertical line on the histograms 4.3vii and 4.3.viii denote a zero regression line gradient. This is equivalent to a neutral trend for how the maximum amplitude (or maximum speed in 4.3.viii) of each tap cycle varies over the 30 second tapping assessment. Scores to the left of the line denote decrementing amplitude whereas scores to the right denote incrementing amplitude. In both patient and control groups the maximum amplitude of each tap cycle tends to decrement during the FT task. Decrementing amplitude poorly discriminates the groups with a low sensitivity/specificity value of 0.55/0.54. This is a surprising result because decrementing amplitude is considered a characteristic feature of PD.

4.3.3.h Decrementing speed

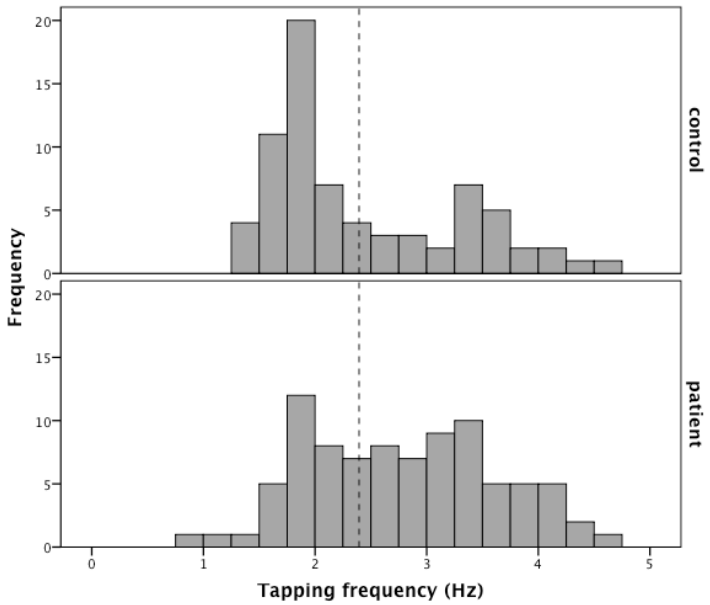
The histogram shows that the majority of PwPD and HCs slow down during the FT assessment period. As a result decrementing speed poorly differentiates the groups, with a sensitivity/specificity of only 0.58/0.56. It is noteworthy that a small number of the controls slow down *more* than patients and it is possible that this may be due to a floor effect in the PD

group i.e. PwPD start off with slower movements so they have a limited range of speeds to further slow down to before they stop altogether. Decrementing amplitude and decrementing speed have been analysed in further detail in Chapter 6 including adjusting the gradients for mean amplitude and mean speed.



4.3.3.i Tapping frequency

All subjects (except one PD subject who had a tap frequency of 0.8Hz) had a FT frequency in the range of 1 to 4.6 Hz. The most common tap frequency in both groups was 1.8 Hz. There was a wide range of results within each group but a tendency for controls to tap at a lower frequency than patients – 65% of controls tapped at a frequency of less than 2.5 Hz whereas 62% of patients tapped at a frequency greater than 2.5Hz. At a threshold of 2.4 Hz the sensitivity/specificity was 0.64/0.61which suggests it is fairly poor at discriminating the groups.



Summary of section 4.3.3 results

These results show that the FT movements of PwPD differ from those of HCs primarily in terms of having smaller amplitude, slower speed, and less rhythmicity. To a lesser extent they also differ in that they have more brief halts and there is a tendency to have a higher tap frequency. Decrementing amplitude and decrementing speed did not discriminate the patient and control group data well for two reasons – there was a wide range of results *within* the groups, with some decrementing and some incrementing, but also there was an overall similar pattern between the groups with the majority of PwPD *and* HCs exhibiting a decrementing during the assessment period.

4.3.4 Does classification accuracy improve using data from one hand?

PwPD typically have asymmetric clinical signs: the side of the body that is first symptomatic tends to remain more severely affected throughout the course of the disease even when both sides of the body exhibit signs of parkinsonism. Hence PwPD have a MA and a LA side. Also some earlier studies have shown that the non-dominant (NonDom) hand of both PD and HCs tends to tap more slowly and less rhythmically than the dominant (Dom) hand (Dunnewold, Jacobi et al., 1997, Muir, Jones et al., 1995). In view of these observations the data was further analysed to see if more

discriminative results could be achieved by comparing separable component measures from the just the MA PD hand to those in the Dom HC hand. Also the results were analysed to look at just the LA hand as this may be considered a fairer assessment of the diagnostic abilities of the device – i.e. assessing only very subtle PD signs, or even asymptomatic signs (in the case HY stage one PD denoting unilateral disease) when the LA side has a clinically normal bradykinesia grade.

4.3.4.1 Most affected and least affected PD hand data

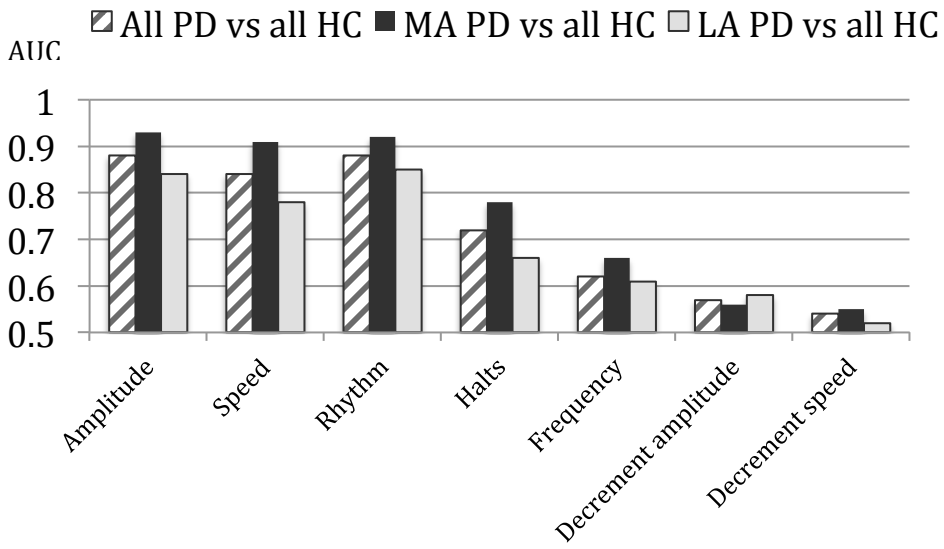
The MA and LA Leeds patient group data had mean FT grades of 1.57 ± 0.94 and 1.05 ± 0.63 respectively. Figure 33 summarises the AUC scores when the separable component measures for the following combinations of hand data were compared:

- All PD and all HC,
- MA PD and all HC, and
- LA PD and all HC.

Figure 33 shows that the following component measures are highly discriminating for PD, regardless of which sub-group of data is compared: maximum amplitude, maximum speed and rhythm. These component measures generally have AUC's scores in the range of 0.80 - 0.92, which denote very good classification of the data.

A similar pattern of results is seen for the component measures regardless of which of the three combinations of hand data is used in the ROC AUC analysis: halts discriminates PD from HC moderately well (AUC 0.66 – 0.78), tap frequency is less discriminatory (AUC 0.61 - 0.66) and decrementing amplitude and decrementing speed discriminate very poorly (AUC 0.51 – 0.58).

Figure 33 Classification accuracy of component measures of bradykinesia when patient data is limited to the most affected hand



Legend: The ability of each movement component measure (x-axis) to classify the PD and HC data is expressed as AUC, area under a ROC curve (y-axis). Results are demonstrated for three different sets of data discrimination indicated by column shading: striped = discriminating all PD data from all HC data, dark grey = discriminating MA PD hand data from all HC data, light grey = discriminating LA PD hand data from all HC.

Secondly Figure 33 shows that the AUC scores are highest when only the MA hand PD data is used, rather than data from both hands of the PwPD. This is not surprising as by definition bradykinesia, assessed clinically, should be more severe in the MA hand of PwPD and hence it would be expected that the components of bradykinesia, measured kinematically, would also have a greater deviation from normal than when all PD data was used. In other words only the 'easiest' PD data sets with the most severe grades of bradykinesia need to be discriminated from HC and the overall result will not be impaired by the clinically more difficult (i.e. LA, or more subtle bradykinesia) cases.

Although this may be clinically obvious, these results demonstrate that for purposes of developing a device to provide a diagnostic prediction the most

discriminative results may be obtained by using just the MA (or in the cases of very early PD – the *only* affected hand) kinematic results rather than including data from both hands.

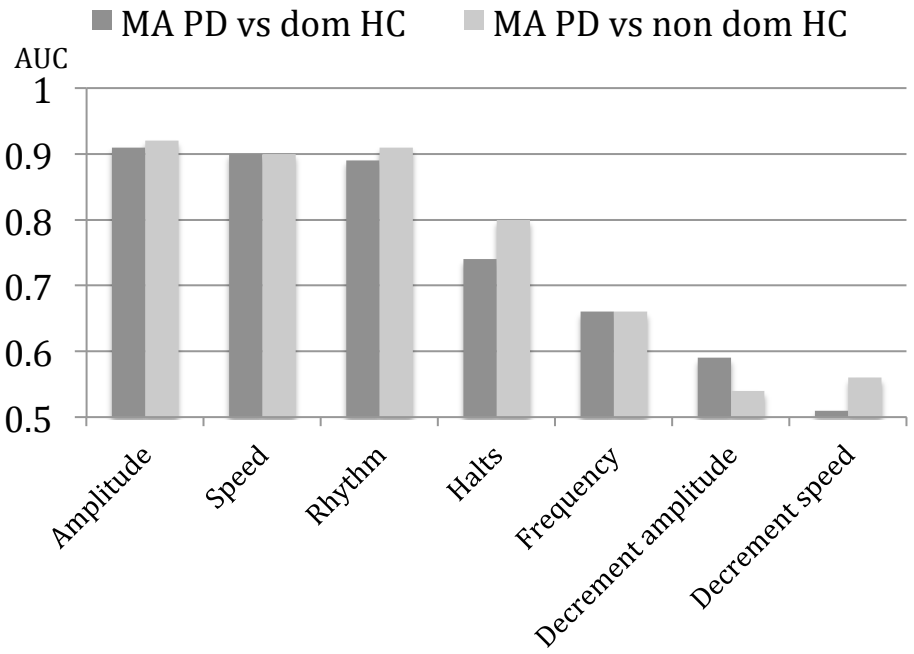
It could be argued that the LA data results are a truer reflection of how the system would perform as a diagnostic predictive device though. In the Leeds cohort only patients with a clinically definite diagnosis of PD have been included i.e. all patients met the gold standard diagnostic criteria, as set by UKBBDC. In a 'real life' clinical situation the subjects being assessed by the device are more likely to be clinically indeterminate and have only subtle bradykinesia, or even no bradykinesia at all, on clinical assessments though. It is likely that they would have kinematic data that is more difficult to discriminate from control data so it is reassuring that the LA hand data AUC remains moderately good in the range of 0.78 – 0.85 for amplitude, speed and rhythm. Approximately half of the MA data sets were the dominant side (23/45 51%) so further analyses including just the dominant hand PD data were not undertaken.

4.3.4.2 Dominant and non-dominant hand control data

The effect of just including dominant hand data from controls in the analysis was examined. As some previous studies suggest that healthy adults tap more slowly and less rhythmically with their non-dominant hand it is possible that classification accuracy could be improved by only including the 'best' performing hand data from controls.

Figure 34 shows that the AUC scores did not significantly differ when just the dominant hand (or just the non-dominant hand) HC data was used though. Hence better discrimination of the groups is unlikely to be achieved by using only a database of dominant hand HC data to compare with the test patient data.

Figure 34 Classification accuracy of each component measures of bradykinesia when control data is limited to only the dominant hand data



Legend: The ability of each movement component (x-axis) to discriminate the PD and HC subjects is expressed as the AUC, area under a ROC curve (y-axis). Results are demonstrated for two different types of data discrimination analysis indicated by column shading: dark grey = discriminating MA PD data from dominant HC data; light grey = discriminating MA PD from non dominant HC.

4.3.5. Correlation of the separable components of finger tapping

The Spearman correlation coefficient, r_s was calculated for the FT variables in the Leeds PD and HC data separately. Table 16 shows that in controls the strongest correlations were for rhythm measurements: COV amplitude and COV speed were strongly correlated ($r_s + 0.78$). Both rhythm measurements were also strongly and inversely correlated with maximum amplitude, suggesting that controls with large FT amplitudes have the least variation in rhythm, possibly because large amplitudes are maximum amplitudes for that individual so there is a ceiling effect (i.e. the amplitudes cannot further increase because they are already maximal) and COV amplitude is minimal, and consequently the COV speed also. Frequency and halts were also

strongly correlated (r_s - 0.65) suggesting that controls' FT with less halts have higher tap frequencies, as might be intuitively expected.

Table 16 Correlogram of the separable component measures in Leeds healthy control finger tapping data

	Maximum amplitude	Maximum speed	Rhythm COV amplitude	Rhythm COV speed	Halts	Frequency	Decrement amplitude	Decrement speed
Maximum amplitude		0.45[1]	-0.76[1]	-0.51[1]	0.21	-0.69[1]	0.03	0.06
Maximum speed	0.45[1]		-0.33[1]	-0.28[1]	-0.14	-0.06	0.09	0.06
Rhythm COV amplitude	-0.76[1]	-0.33[1]		0.78[1]	0.04	0.46[1]	0.18	-0.13
Rhythm COV speed	-0.51[1]	-0.28[1]	0.78[1]		0.26[2]	0.21	-0.06	-0.13
Halts	0.21	-0.14	0.04	0.26[2]		-0.65[1]	-0.14	-0.10
Frequency	-0.69[1]	-0.06	0.46[1]	0.21	-0.65[1]		-0.13	-0.13
Decrement amplitude	0.03	0.09	0.18	-0.06	-0.14	-0.13		0.45[1]
Decrement speed	0.06	0.06	-0.13	-0.13	-0.10	-0.13	0.45[1]	

Key: Grey shading and correlation coefficient

> 0.9 | 0.80-0.89 | 0.70-0.79 | 0.60-0.69 | 0.50-0.59

Legend: Spearman correlation coefficients are presented for eight component measures of FT using data obtained from both hands of Leeds controls. The coefficients marked with [1] were significantly correlated at the $p < 0.01$ level (2-tailed) and those denoted by [2] were significant at $p < 0.05$ level (2-tailed).

Table 17 Correlogram of the separable component measures in Leeds PD patient finger tapping data

	Maximum amplitude	Maximum speed	Rhythm COV amplitude	Rhythm COV speed	Halts	Frequency	Decrement amplitude	Decrement speed
Maximum amplitude		0.87[1]	-0.81[1]	-0.66[1]	-0.63[1]	-0.27[1]	0.31[1]	0.24[1]
Maximum speed	0.87[1]		-0.77[1]	-0.75[1]	-0.82[1]	-0.15	0.21[1]	0.28[1]
Rhythm COV amplitude	-0.81[1]	-0.77[1]		0.87[1]	0.69[1]	0.05	-0.25[2]	-0.27[2]
Rhythm COV speed	-0.66[1]	-0.75[1]	0.87[1]		0.78[1]	-0.20	-0.21	-0.20
Halts	-0.63[1]	-0.82[1]	0.69[1]	0.78[1]		-0.42[1]	-0.09	-0.18
Frequency	-0.27[2]	-0.15	0.05	-0.20	-0.42[1]		-0.24[2]	-0.17
Decrement amplitude	0.31[1]	0.21[2]	-0.25[2]	-0.21	-0.09	-0.24[2]		0.85[1]
Decrement speed	0.24[1]	0.28[1]	-0.27[2]	-0.20	-0.18	0.17	0.85[1]	

Key: Grey shading and correlation coefficient

> 0.9 | 0.80-0.89 | 0.70-0.79 | 0.60-0.69 | 0.50-0.59

Legend: Spearman correlation coefficients are presented for eight component measures of FT using data obtained from both hands of Leeds patients. The coefficients marked with [1] were significantly correlated at the $p < 0.01$ level (2-tailed) and those denoted by [2] were significant at $p < 0.05$ level (2-tailed).

It is striking how different the patient and control correlograms are (Tables 16 and 17) Firstly the patient correlogram contains generally much more shading than the control correlogram both in terms of the *number* of boxes and also the *depth* of shading. This suggests there is a stronger correlation between the component measures in patient data, and is consistent with the fact that these variables are all separable measures of the same obligatory pathological clinical sign, bradykinesia.

The shading is localised to the top left portion of the patient correlogram denoting that these components – amplitude, speed, rhythm and halts - are all strongly associated with each other. It is noteworthy that measures of tapping frequency and the decrementing variables are not strongly correlated to any of the other component measures in the patient data though (except the decrementing measures with each other, and possibly frequency and halts). This raises the question again of whether these components are part of the same clinical phenomenon, bradykinesia. On the other hand there is a significant correlation between amplitude, and speed, and the decrementing variables for the patient data but not for the control data suggesting that patients who have bigger or faster movements are more likely to exhibit a decrement in amplitude and speed respectively, and hinting at a floor effect being present.

There are a number of specific differences for how the FT variables correlate in the patient and control data. For example, amplitude and speed are more strongly correlated in the patient data (r_s + 0.87 PD vs. + 0.45 HC). This suggests that PwPD tend to tap small *and* slow (or big and fast) whereas controls exhibit more variability in how these variables associate so other combinations such as small and fast movements, or big and slow movements are likely too. Another difference is that in the PD group halts are more strongly correlated with amplitude (r_s - 0.63 PD vs. + 0.21 HC), speed (r_s - 0.82 PD vs. - 0.14 HC) and rhythm (COV amplitude r_s + 0.69 PD vs. + 0.04 HC). This suggests that patients who have larger, faster or more rhythmic movements (i.e. less impaired) also have less halts (i.e. less

impaired) and this supports the premise that these are all component measures of the same clinical phenomenon.

Likewise there is a difference between groups in terms of how tapping frequency and amplitude are associated: these variables have a much stronger association in the control group (r_s - 0.69) compared to patients (r_s - 0.27). The control result is fairly intuitive – if a HC has a larger amplitude of tap their fingers have to travel further so if the speed is kept constant (and it is only moderately correlated to amplitude in the control group) the tap frequency will tend to be lower because the fingers have a longer 'path' to travel during each tap cycle, so less taps are performed over the assessment period. Perhaps this association has been lost in the patient data because there are so many other impaired separable components in the patient FT movements that interfere with this relationship.

Summary of section 4.3 results

PD FT movements are characterised by smaller, slower, less rhythmic and more interrupted movements than in controls. Of all the component measures amplitude, rhythm and speed have the highest classification accuracies with AUCs in range of 0.84 - 0.88. The rank order varied depending on clinical state with amplitude most discriminatory in the *on* state and speed more discriminatory in the *off* state. The LTHT *on* data was validated on the UCSF *on* data. Halts and tapping frequency had moderate classification accuracy but the SE measures did not discriminate the data any better than chance. The classification accuracy of all component measures was improved to a maximum AUC of 0.92 (for amplitude) by using only MA PD data that corresponded to a mean UPDRS FT grade of 1.57 ± 0.94. There was no improvement in AUC by limiting HC data to the dominant hand. When LA PD data was used, corresponding to a UPDRS FT grade of 1.05 ± 0.63, the classification accuracy was still good with AUC 0.84 (for amplitude). The component measures were more strongly correlated in the PD data than the control data supporting the premise that these are measures of the same clinical phenomenon.

4.4. Classification accuracy of bradykinesia composite models

When a skilled clinician evaluates whether a patient demonstrates bradykinesia during the FT task they need to integrate analyses of the separable components. Furthermore the clinician is likely to be using other demographic information such as the age and gender of the patient, and other clinical signs such as altered gait to formulate a clinical diagnosis. Therefore, it is important to examine whether the classification accuracy of a device that objectively measures bradykinesia can be improved by using a composite score of the separable component measures and whether the inclusion of basic demographic details may further improve this.

Logistic regression analysis can be used to form an equation that combines several individual measures into an equation that predicts a dichotomous outcome. Thus bradykinesia composite models (equations) were produced through logistic regression analysis of the separable component measures in order to predict whether each data set came from a patient or control. The composite model summarises the relative 'weighting' or importance of each kinematic component measure in combination with a constant to produce the best discrimination of the PD and HC data.

4.4.1 Classifying Leeds patient and healthy control data

4.4.1.1 Composite model comprising the four most discriminating components

The first approach was to use just the four most discriminating component measures in the regression model – amplitude, speed, COV amplitude and COV speed. The strong correlations between each of these measures meant there was considerable multicollinearity. This is not a problem when using the model to predict diagnostic group but does limit its use for exploring how the individual variables affect the outcome. When data from both hands of the Leeds subjects was used, the following regression model was produced:

PD predicted if (4.2 -9.4 amplitude -0.04 speed -9.6 COV amplitude +12.8 COV speed) > 0.5

In other words the FT data is predicted to come from a patient if the solution to the equation above is greater than 0.5. This model accounted for 61% of the FT data variability (NagelKerke R^2) between the groups, with a classification accuracy of 0.90 AUC (95% CI 0.85 - 0.95; $p = 4.5 \times 10^{-18}$) or 85% (Table18). All FT components were significant variables (< 0.05) except for maximum speed (p = 0.9).

Table 18 Classification accuracy of amplitude, speed and rhythm composite model

		Predicted		
		Control	Patient	**% correct**
Observed	Control	59	13	81.9
	Patient	12	75	86.2

4.4.1.2 Composite comprising all finger tapping component measures

When all the FT components were included in the logistic regression model the classification accuracy improved marginally to an AUC of 0.93 (95% CI 0.89 – 0.97, $p = 2.3 \times 10^{-20}$) or 87.7% ; Nagelkerke R^2 0.69 (Table 19). The regression model was:

PD predicted if (15.4 -2.5 tap frequency -28.3 amplitude +1.2 speed -13.0 COV amplitude + 15.8 COV speed -7.5 halts + 0.01 decrementing amplitude -0.05 decrementing speed) > 0.5

All variables were significant (p < 0.05) except for decrementing speed (p = 0.21). When this was removed from the logistic regression analysis there was a slight reduction in overall classification accuracy to AUC 0.91

(87.0%), with an additional control misclassified as a patient (i.e. a false positive).

Table 19 Classification accuracy of composite model comprising all separable measures of bradykinesia

		Predicted		
		Control	Patient	**% correct**
Observed	Control	57	10	85.1
	Patient	9	78	89.7

4.4.1.3 Composite comprising all FT component measures plus age and gender

An automated diagnostic device may combine objective measures of movements with demographic information in order to improve diagnostic classification. The operator of the device could input the demographic details of the subject and this information could be combined with the results of the kinematic data analysis in a mathematical model to provide an overall diagnostic prediction of whether the test data is likely to be normal or indicative of PD. When demographic information on age and gender was included in the model the overall classification accuracy improved to an AUC 0.94 (95% CI 0.90 - 0.97, $p = 3.4 \times 10^{-21}$) or 88.3 % (R^2 0.73); see Table 20. Inclusion of hand variable data, left vs. right and dominant vs. non-dominant, were not significant predictors ($p > 0.1$) and did not improve the classification accuracy any further. The combined kinematic and demographic regression model was:

PD predicted if (10.5 -2.7 tap frequency -29.2 amplitude +1.4 speed -13.6 COV amplitude +17.4 COV speed -8.3 halts +0.01 decrementing amplitude -0.05 decrementing speed +0.06 age +1.4 gender) > 0.5,

where age is measured in years and male gender is 1 and female gender is 2. All variables were significant ($p < 0.05$) except the decrementing speed ($p =$

0.3). The three different composite model results are summarised in Table...21.

Table 20 Classification accuracy of bradykinesia composite model comprising all separable component measures plus age and gender

		Predicted		
		Control	Patient	**% correct**
Observed	Control	58	9	86.6
	Patient	9	78	89.7

Table 21 Classification accuracies of bradykinesia composite models

Composite model	**Correctly classified**		**Total, n**	
	Controls	Patients	**(%)**	**AUC**
Amplitude, speed, rhythm	59	75	134 (85)	0.90
All FT variables	57	78	135 (87.7)	0.93
All FT variables+ age + gender	58	78	136 (88.3)	0.94

4.4.2 Validating classification accuracy on San Francisco data

It is important to validate predictive models on an independent cohort in order to test whether the model generalises to subjects whose data was not used to develop the equation. The classification accuracies of the three models were tested on the UCSF PD patient and HC data sets (Table 22) and this showed that the overall classification accuracies were inferior in the UCSF group. This suggests that the models may not generalise so well to independent samples but it is noteworthy that the UCSF cohort is small so any misclassified subjects have a large effect on the overall accuracy. The optimal model (all variables, age + gender) has acceptable classification accuracy in the UCSF group suggesting that this model *may* generalise beyond the Leeds cohort but it is striking that all of the models tend to misclassify the controls in particular. It is reassuring that the clinical state

(*on* vs. *off*) does not seem to significantly impact on the classification accuracy though.

Table 22 Classification accuracy of the Leeds bradykinesia composite models' when applied to the San Francisco data

	Classification accuracy %				
Composite models	**Leeds**	**UCSF**	UCSF HC	UCSF PD *on*	UCSF PD *off*
Amplitude, speed, rhythm	**85**	**73**	33	87	93
All FT components	**88**	**71**	33	91	83
All components, age, gender	**88**	**81**	55	91	96

Legend: The classification accuracy results are rounded to the nearest integer. The first two columns of results represent the overall classification accuracy at each site. The last three columns of results gives the individual breakdown of the UCSF results for each cohort of data.

4.4.3. Assessing composites in models of newly diagnosed PD

When developing a device to aid clinical diagnosis of PD it would be ideal to test the diagnostic predictions in a clinically indeterminate group of patients and follow them up longitudinally to see if the device's earlier predictions were accurate. This is not feasible with the current subjects' data as all the patients had clinically definite PD and the study was cross-sectional. Nevertheless it is possible to use subsets of the current data to test the diagnostic accuracy by examining just those PwPD with clinically 'normal', or 'slight bradykinesia' FT grades. Using patient data that is graded zero is using data that has clinically undetectable bradykinesia (i.e. the unaffected hand in HY stage one (unilateral) disease, or undetectable in the *on* state due to treatment) and this has previously been used as a proxy for 'clinically indeterminate' patient data (Haaxma, Bloem et al., 2010). Alternatively using data from patients who have grade one, or slight bradykinesia, may also be considered another way of obtaining representative data of early, or possibly clinically indeterminate, PD.

Classification accuracy when only grade zero patient data is used

There were 11 sets of approved Leeds patient data that had been clinically graded 'normal' or grade zero. When these were combined with the 72 HC data sets (all graded zero clinically) the second composite comprising all FT variables had an AUC of 0.77 (95% CI 0.60 - 0.94, p = 0.005). This was equivalent to 85% (59/72) and 62% (8/11) of HC and PD respectively being correctly classified. The results were not significantly altered when the third composite comprising 'all variables plus age and gender' with an AUC of 0.77 (95% CI 0.62 – 0.92, p = 0.004).

Classification accuracy when only grade one patient data is used

Grade one denotes 'slight bradykinesia' and 45 sets of approved patient data had this clinical grade. The composite, comprising all FT variables had AUC 0.89 (95% CI 83 - 95, p < 0.0001) and this was equivalent to 85% (59/72) of HC, and 80% (36/45), of PD being correctly classified. When the third composite, comprising all FT component measures plus age and gender was used, the AUC improved to 0.94 (95% CI 0.89 - 98, p < 0.0001) which was equivalent to 86% (61/72) of HC, and 84% (38/45) being correctly classified.

Summary of section 4.4 results

The composite bradykinesia models had higher classification accuracy than the individual component measures. A composite comprising all FT variables and the best performing individual component measure, maximum amplitude had AUCs of 0.93 and 0.88 respectively. The classification accuracy of the composite models were similar for UCSF patients *on* and *off* but did not generalise well to the whole UCSF sample as there were many misclassified controls. The bradykinesia composite had good classification accuracy for discriminating slight PD bradykinesia FT data from HC data (AUC 0.89) and moderate accuracy for discriminating clinically 'normal' PD FT data from HC data (AUC 0.77).

4.5 Correlation of bradykinesia measures to demographics and PD progression

The results so far have been focussed on using component measures of bradykinesia to discriminate PD from HC data as this information may inform the development of a diagnostic device. In addition it would be useful to know whether such a device could *monitor* PD progression as this would open up the potential for a device to aid epidemiological studies and assess the effects of therapeutic interventions.

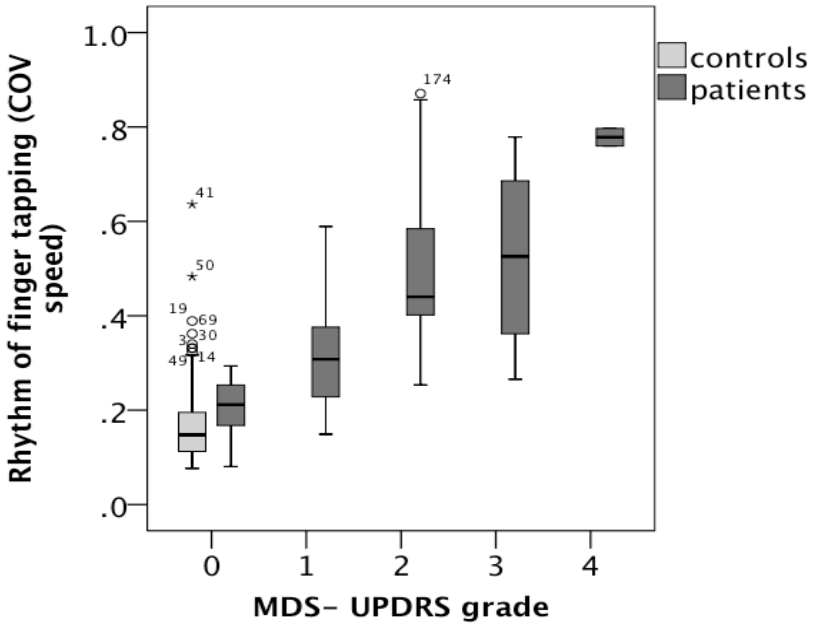
Ideally a longitudinal study should be performed to address these questions, with repeated measures of the same subjects over time. However some preliminary data from this cross –sectional study may be obtained by correlating the clinical measures of disease progression with the component measures of bradykinesia. This was done for the individual measures and the composite model (of all FT components) using the Leeds data.

4.5.1 Individual component measures of bradykinesia and disease progression

In the patient group the UPDRS FT grade correlated significantly with the following individual component measures, in descending rank order: COV speed (r_s + 0.67), speed (r_s - 0.66), halts (r_s + 0.63), COV amplitude (r_s + 0.63), amplitude (r_s - 0.58), all ps < 0.001.

Figures 35-38 show there is a trend between increasing severity of UPDRS FT grade and movements that are less rhythmic, smaller, slower and have more halts. They also show that grade zero, or 'normal' FT movements in PwPD are not truly 'normal' as there is a tendency for these patients to have smaller, slower and less rhythmic movements than HCs. When the group means for these kinematic measures were compared between grade zero patients and HCs only amplitude showed a significant difference (p = 0.04).

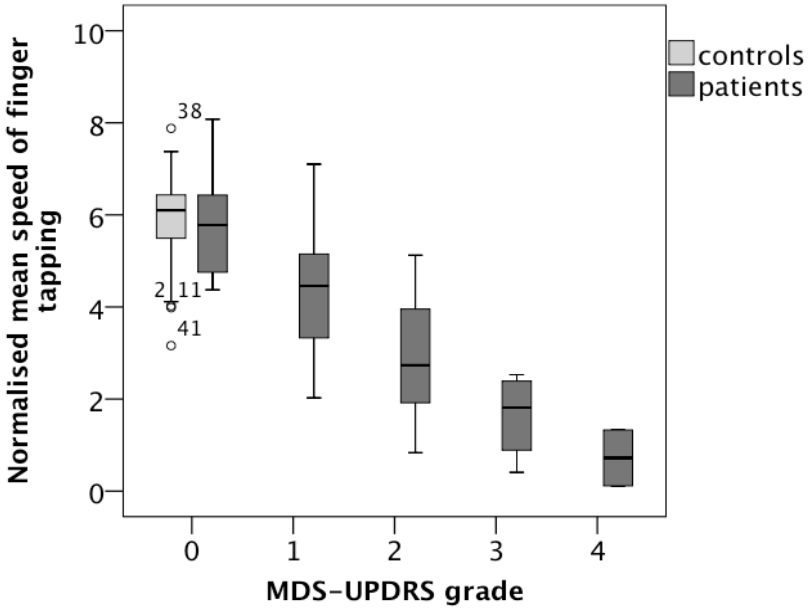
Figure 35 Rhythm of finger tapping for each UPDRS grade



Legend: Box and Whisker plots for Leeds approved data shows that higher UPDRS FT grades are associated with less rhythmic movements, $r_s + 0.67$. Rhythm is measured by COV of maximum speed i.e. higher COV values denote less rhythmic movements. Values more than 2 or 3 SDs from group mean are denoted by circles and crosses respectively.

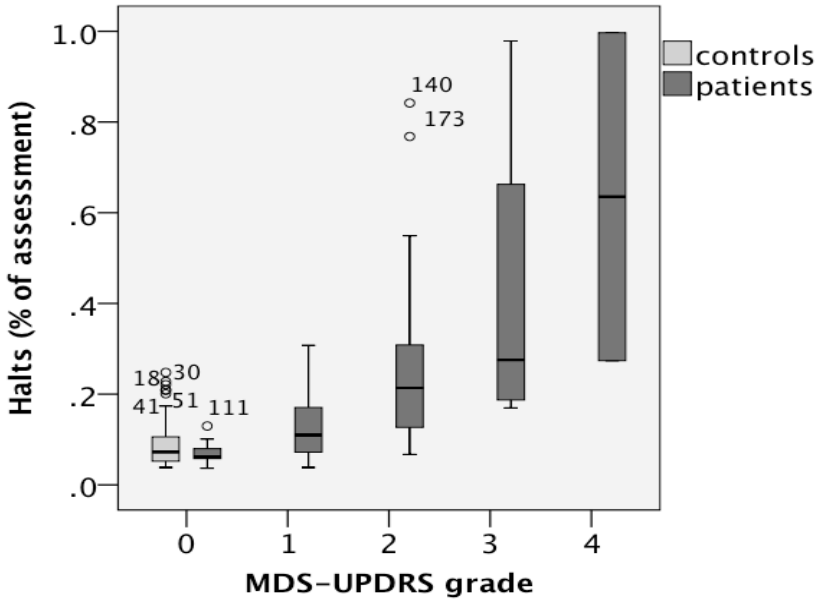
The UPDRS clinical FT grade showed a borderline significant association with tap frequency (r_s - 0.21, p = 0.047) but none with decrementing speed and decrementing amplitude measurements (ps > 0.11). MoCA was only correlated with rhythm measurements (r_s= - 0.30 COV speed; p = 0.009). HY stage, disease duration, and LEDD were not significantly correlated with any of the individual component measures.

Figure 36 Speed of finger tapping for each UPDRS grade



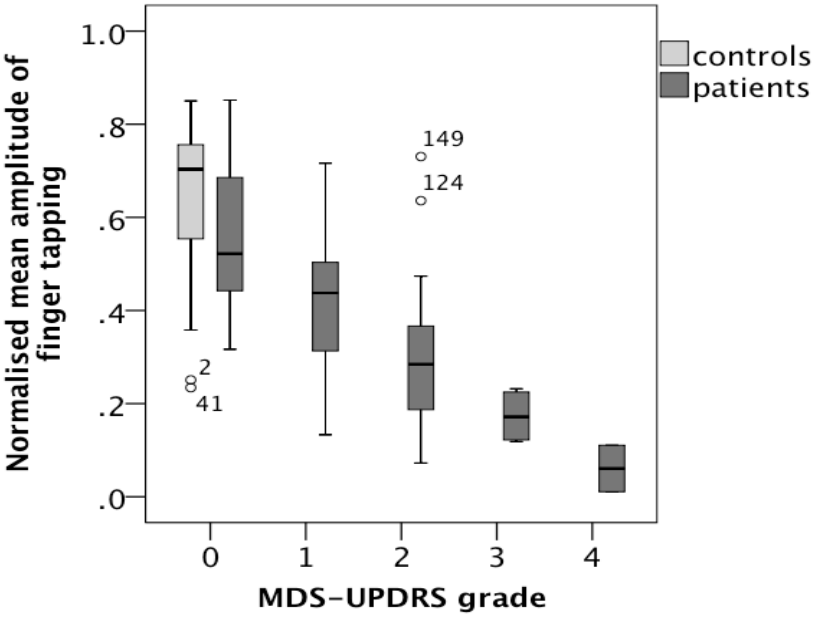
Legend: Box and Whisker plots for Leeds data show how higher UPDRS FT grades are associated with slower movements, r_s -0.66. Also mean speed for grade zero PD data is slower than HC (all grade zero) data. Values more than 2 SDs from mean are denoted by circles.

Figure 37 Halts in finger tapping for each UPDRS grade



Legend: Box and Whisker plot show how higher UPDRS FT grades are associated with more halts, r_s 0.63. Circles denote values more than 2 SDs from mean.

Figure 38 Amplitude of finger tapping for each UPDRS grade

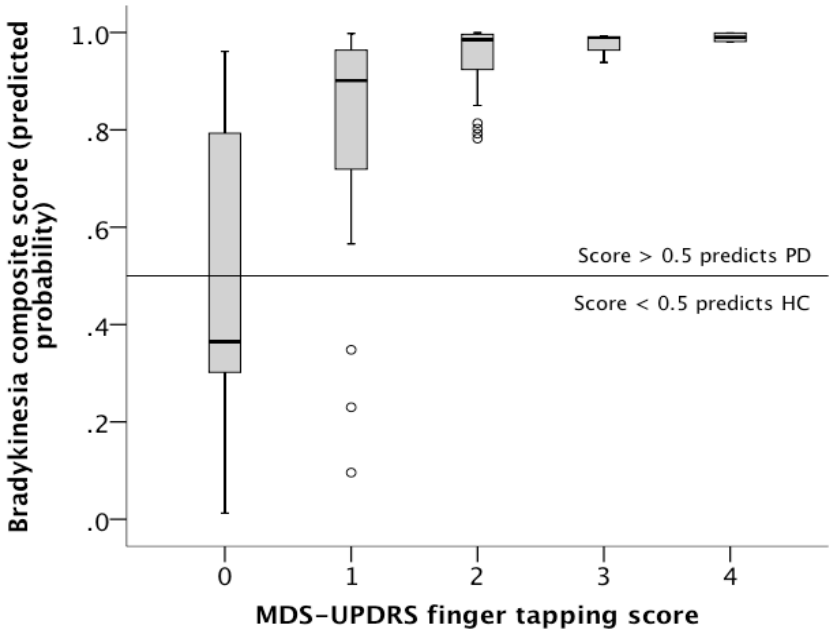


Legend: Box and Whisker plot show how higher UPDRS grades are associated with smaller amplitude of FT; r_s -0.58. Also mean amplitude for grade zero PD data is smaller than the HCs (all grade zero) data; p = 0.04. Circles denote values more than 2 SDs from mean.

4.5.2 Composite bradykinesia measures and disease progression

There was a significant correlation between the bradykinesia composite (of all components) model score and UPDRS FT grade (r_s + 0.60, p < 0.0001) with higher composite scores associated with more severe clinical grades of bradykinesia (Figure 39). This suggests that the composite may have the ability to not only classify FT data into PD or HC diagnostic groups but also to provide information on *the degree* of clinical bradykinesia present. There was a significant difference (p = 0.002) in the composite score for patients with UPDRS FT grade zero (0.51 ± 0.33) and all HC data (0.23 ± 0.26). Figure 39 also demonstrates that the misclassified PD data sets (below the 0.5 reference line) are predominantly those that had grade zero FT grades.

Figure 39 Bradykinesia composite model scores of Leeds PD patients for each UPDRS grade



Legend: The bradykinesia composite scores (logistic regression equation predicted probability) are presented as box plots for each MDS-UPDRS FT grade. The centre line of each box represents the median score, the upper and outer limits of the box, the 75% and 25% quartiles and the error lines ± 2 SD from the mean. The reference line at 0.5 is the cut-off point for the bradykinesia composite model predicting whether the data came from a patient (score > 0.5) or a control (score < 0.5). Therefore patient data below this line were misclassified as controls. The circles are data with composite scores > 2 SD from the mean.

When the clinical and objective scores for each Leeds patient data set were plotted the following regression line of best fit (least squares) applied:

Bradykinesia composite score = 0.142 x UPDRS grade + 0.635

However because of the wide spread of bradykinesia composite model scores for grade zero data, and the small numbers of data sets that were allocated grades three and four, the regression line model does not fit the

data particularly well (R^2 0.27) despite UPDRS FT score being a significant predictor of the bradykinesia composite score ($p < 0.0001$).

The bradykinesia composite score was also inversely correlated with the MoCA score (r_s - 0.46, $p < 0.0001$) suggesting that cognitive impairment is associated with more severe objective measures of bradykinesia. None of the other markers of disease progression such as HY stage (r_s + 0.05), PD disease duration (r_s - 0.09) or LEDD (r_s - 0.20) significantly correlated with the bradykinesia composite score.

4.5.3 Component measures of bradykinesia and age, gender and dominant hand

Age was not correlated with the bradykinesia composite score in either the control group (r_s - 0.02) or the patient group (r_s 0.09). Age was also not significantly correlated with any of the individual separable component measures of bradykinesia in either group. There was no association between gender and any of the component measures in the patient group. None of the component measures, in either patient or control group data, were significantly correlated to hand dominance.

4.6 Summary of Chapter 4 results

The results obtained from analysing the separable component measures of bradykinesia in the FT data are summarised:

- FT performed by PwPD was characterised by movements with smaller amplitude, slower speed, less rhythm and more halts than in controls. The three measures that best discriminated PD and HC group data were amplitude, rhythm and speed and these results were validated on the independent data collected in UCSF.
- The discriminatory value of each of the top three separable components varied depending on the clinical state of the patients. In the *on* state, amplitude was the most discriminating component followed by rhythm then speed, but in the *off* state, speed and amplitude were most discriminating (speed non significantly greater)

followed by rhythm. This suggests that speed improves more than either of the other components in response to dopaminergic drugs.

- Halts discriminated the groups moderately well as PwPD exhibited more halts than controls.
- The tapping frequency tends to be higher in PD than controls but the range of frequencies varied considerably within the groups so this component was a fairly poor discriminator of group data.
- The SE measures, decrementing amplitude and decrementing speed, did not discriminate the Leeds PwPD from HCs at all. However the decrementing speed results were not confirmed in the UCSF data so this requires further analysis (see Chapter 6)
- The most discriminating results, with AUCs > 0.9 were obtained when PD data from just the MA hand (mean FT grade 1.57) was used, but discrimination was still moderately good, with AUCs 0.76 - 0.85, when only the LA PD hand data was used (mean FT grade 1.05).
- Using just dominant hand data from controls did not improve diagnostic classification.
- The individual component measures were more strongly inter-correlated in the PD group than controls. Amplitude, speed, rhythm and halts were particularly closely correlated. SE measures did not correlate well with the other component measures of bradykinesia.
- Combining the top four most discriminatory separable components together (amplitude, speed and two measures of rhythm) into a composite model improved the classification accuracy marginally from an AUC of 0.88 (for the best individual component, amplitude) to an AUC of 0.90. Combining all the FT components together into a composite improved the classification accuracy to AUC 0.93
- The bradykinesia composite models developed using Leeds data had similar accuracy in the UCSF patient data whether tested in the *on* or *off* states. However they misclassified many UCSF control data sets and it remains unclear whether the composite models generalise well to independent data.

- Patients with the same subjective clinical grade as controls in the *on* state (grade zero or 'no bradykinesia') had significantly different objective measures of amplitude and the bradyinesia composite score.
- The composite models differentiated grade zero PD data from grade zero HC data moderately well with AUC 0.77, and differentiated grade one patient data (clinically 'slight' bradykinesia) from grade zero control data very well with AUC 0.89.
- The clinical grade of bradykinesia was significantly correlated with measures of rhythm, speed, halts and amplitude and also with the bradykinesia composite model. More severe clinical grades of bradykinesia were associated with less rhythmic, slower and smaller movements with more halts.
- Cognitive impairment was correlated to the composite bradykinesia score and rhythm but advancing age, HY disease stage, PD disease duration and LEDD were not associated with any objective measures of bradykinesia.

In summary these results highlight the potential for objective measures of bradykinesia using FT data to aid diagnosis and monitoring of PD. The objective measures correlate well with the clinical bradykinesia grade and this is important for monitoring. Additionally there remains good classification accuracy when bradykinesia is clinically 'slight', and the amplitude measures in controls and PD patients are significantly different even when no bradykinesia was clinically apparent; these results suggest that the device may be able to provide diagnostic predictions to support clinical evaluation.

4.7 Discussion of separable component analyses

This chapter has focussed on objectively measuring the clinically defined components of bradykinesia in PD and HC FT data. Two main types of analysis have been undertaken. Firstly, classification accuracy has been evaluated by assessing how well each component measure of bradykinesia differentiates PD and HC kinematic data. These results have implications

for the development of a device to support clinical diagnosis. Secondly, how the components of bradykinesia correlate with one another and with the subjective clinical grades of bradykinesia has been explored. These results are more relevant to how a device that measures bradykinesia could potentially monitor the progression of PD and response to intervention.

4.7.1 Discussion in context of developing a diagnostic device

Amplitude, speed and rhythm are the most discriminating components

PD patients' FT movements were characterised by smaller amplitude, slower speed, less rhythm and more halts than controls. In the *on* state the mean of the maximum amplitude of each FT cycle was the most discriminating component, followed by measures of rhythm then speed. In the *off* state the mean of the maximum speed of each FT cycle and the mean of the maximum amplitude of each FT cycle were the most discriminating component measures (not significantly different), followed by rhythm. These objective measures support the inclusion of amplitude, speed, rhythm and halts within the gold standard clinical definitions of bradykinesia, as defined by UKBBDC and MDS-UPDRS (Gibb and Lees, 1989, Goetz, Tilley et al., 2008). However no difference in SE measures was found between PD and HC group data.

Only four studies have previously quantified the clinically defined components of bradykinesia during the FT task. Espay at al.'s pilot study in 2009 (Espay, Beaton et al., 2009) assessed the MA hands of 23 PwPD and the dominant hand of 16 HC who each performed FT for 15 seconds. The same movement sensors as the present study (Polhemus EM tracking system) were used and attached to the index finger and thumb. Patients were assessed in a defined *off* state, 12 hours after their last dose of PD medication, and in an *on* state one hour after their first morning dose of levodopa. The kinematic scores of HC were used to define the amplitude and speed of PD movements into 'normal' (< 1 SD below HC group mean score), 'slow speed' or 'low amplitude' (1 - 3 SDs below HC mean) or 'very slow', or 'very low amplitude' (> 3 SDs below HC mean).

The authors found that in PwPD *off,* amplitude was more impaired than speed with most having 'slow speed' and either 'low amplitude' or 'very low amplitude.' In the *on* state speed normalised more than amplitude and most patients *on* had 'normal speed' but 'low amplitude'. Half of PwPD *on* had normal speed but only four had normal amplitude. These results lend support to the present study's finding that amplitude was the most discriminating component measure for patients in the *on* state in both the LTHT and UCSF cohorts. However Espay et al.'s results for the *off* state contrast with the present study that found speed was numerically, but not significantly, more discriminating than amplitude. Nevertheless in both studies the patient numbers assessed *off* were quite small (n = 23 MA hands in Espay, n = 24 both hands in current study) and these may make conclusions less robust. Furthermore the methods used to define speed and amplitude in Espay's study using a three point ordinal scale defined by the degree of deviation from the control group mean scores makes it difficult to directly compare the results with the present study where a continuous scale of objective measurements and AUC was used.

Espay et al. published a second study in 2011 quantifying the separable components of bradykinesia using objective and subjective assessments of FT (Espay, Giuffrida et al., 2011). The MA hand of 85 PwPD was assessed *on* and *off* during 15-second periods of FT, hand opening/closing and then pronation/supination movements. There was no control group. Four clinicians graded each video-recorded assessment using UPDRS and the Modified Bradykinesia Rating scale (MBRS). The MBRS requires three separate grades, on a scale of zero to four, to be allocated for the speed, the amplitude and the rhythm of movements. Each subject wore KinetiSense (Figure 17) movement sensors containing accelerometers and gyroscopes on their index finger and thumb and the kinematic measures of amplitude, speed, rhythm and decrements were compared to the clinical grades.

Clinically, in the *off* state, the MBRS amplitude grade was more impaired in terms of prevalence and severity than either the speed or rhythm grades but

in the *on* state only the MBRS speed grades showed significant improvement. The quantitative measures, obtained through the movement sensors, showed that in the *on* state speed improved the most, then amplitude and rhythm, and SE measures did not improve at all. These results highlighted several important points. Firstly the UPDRS bradykinesia grade (a composite grade of all the components of bradykinesia) is more strongly influenced by the degree of reduced amplitude observed clinically than any other component. Perhaps this explains why amplitude correlated so strongly to the UPDRS FT grade in the current study. Secondly when patients switched into the *on* state, speed was the *only* MBRS component to improve whereas the movement sensors not only detected a large improvement in speed but also significant improvements in rhythm and amplitude too. This suggests that any changes in rhythm or amplitude may go unnoticed when using either the UPDRS or the MBRS and suggests that clinical scales are less sensitive to changes in movement characteristics than objective measures. Extrapolating from this point it may be that the current gold standard clinical diagnostic criteria are also less sensitive to early changes in the components of bradykinesia than objective measures. The differential response of the components of bradyknesia is an interesting point not only when considering the underlying pathophysiology of bradykinesia, but also because it has important practical implications for protocols used to collect kinematic data that may aid diagnosis prediction or for monitoring progression of disease.

There are several aspects of the Espay et al. 2011 study that need to be considered when trying to compare the results to the present study though. Firstly the overall UPDRS grade of bradykinesia for the FT task in PwPD is not presented in the paper. Instead the authors averaged the amplitude, speed and rhythm MBRS grades across the three bradykinesia tasks and the four raters into a composite clinical grade and then compared these composites to the individual quantitative measures. This may be an important methodological flaw as PD patients are known to be disproportionately impaired on the FT task compared to the hand opening and arm pronation-supination tasks (Agostino, Berardelli et al., 1998) so

combining the clinical grades runs the risk of disguising any differential clinical response within each task. The lack of a control group means that the focus is limited to how PwPD's movements change in the *off* and *on* states and we cannot be sure about which component was most impaired quantitatively in either state when compared to HC. Also it is noteworthy that Espay et al.'s patients were clinically different to those in the present study as they excluded all patients with more than grade one or 'slight' tremor and those with cognition test scores < 27/30. This means they have selected a subset of PwPD that may not be representative of the population with PD. Furthermore the details of the mean age, HY stage and disease duration are not stated.

In 2009 Yokoe et al. quantified 60-second periods of FT in 16 PwPD *on* and 32 HC. Each subject wore a movement sensor system comprising tri-axial accelerometers and touch sensors attached to the index finger and thumb. The authors showed that the biggest difference between PD and HC group data were measures of amplitude and speed. Rhythm measurements were less discriminating and tapping frequency measures did not discriminate the groups at all. They used principal component analysis to find out which component measures accounted for the greatest variability between the PD and HC data. The first component accounted for 40% of variability and comprised measures of maximum speed, maximum amplitude, total amplitude (i.e. cumulative amplitude of all tap cycles over 60 seconds) and rhythm. The second component, accounting for 27% of variability, comprised tapping frequency and a measure of halts. The authors also used logistic regression to see how well each individual measure classified data into patient and control groups. They found that the most discriminating component was the mean of maximum opening velocity (with a misclassification rate of 16%) followed by total amplitude (19%) and rhythm (21%).

Yokoe et al.'s finding that speed was more discriminatory than amplitude in the *on* state contrasted with the present study and with Espay's two studies. However it is worth noting that they specifically measured the maximum

opening speed rather than the overall speed of FT and this component measure is further explored in Chapter 5. Otherwise the results are broadly similar to the present study that demonstrated misclassification rates of 20%, 22% and 23% for amplitude, rhythm and speed respectively. However direct comparisons of results between the present study and Yokoe et al. are not entirely straightforward as different methods for calculating each component were used, the tapping sequence length was twice as long in Yokoe's study, in neither study were the patient and control groups age matched and Yokoe et al. selected a sub-group of PwPD without marked action tremor, severe dyskinesia or cognitive impairment.

In 2012 Ling et al. measured FT movements in 15 PwPD, 9 PSP patients and 16 HCs using an OKS in order to examine whether the components of bradykinesia differed between the groups (Ling, Massey et al., 2012). The participants performed three sets of 15-second periods of FT with each hand separately and PwPD were assessed *on* and *off*. The authors found that in comparison to HC the characteristic FT pattern of PwPD *off* consisted of significantly reduced speed and impaired rhythm. However whilst amplitude and SE measures were numerically more impaired in PwPD than HCs these components did not reach statistical significance. The amplitude measurements contrast with the present study but otherwise the results are broadly similar.

Ling et al.'s study highlighted that different parkinsonian conditions may have different bradykinesia profiles with the components impaired to different degrees. For example the PSP patients had significantly smaller amplitudes of movements with significantly less decrement compared to PwPD and this observation hints at the exciting potential that quantifying FT movements may not only predict whether the movements are bradykinetic but may also differentiate PD from other parkinsonian conditions. If this were the case such a device would have a major advantage over DaT scans that cannot reliably differentiate the various neurodegenerative parkinsonian conditions such as PD, PSP, CBD and MSA.

Halts and tapping frequency are moderately discriminatory

The present study found that halts and tapping frequency discriminated patient and control data moderately well. Out of the FT studies described above, only Yokoe et al. measured halts and reported classification accuracy as approximately 70%. Indirect evidence from keyboard tapping studies suggests that objective measures of halts may have the potential to be more discriminatory though; for example Homann et al. found that the cumulative time over a 60 second assessment period for keeping a key depressed > 17msec (which corresponds to the keyboard repeat rate) was greater in PD and differentiated PD and HC better than objective measurements of rhythm (Homann, Suppan et al., 2000).

Regarding tapping frequency the present study found that patients had a tendency to have higher frequencies, which may at first seem counter-intuitive for a condition that is underpinned by slowed movements. However as the PD patient movements had very small amplitudes the distance moved by the fingers in each tap cycle was less and hence for the same speed or even slightly slower speeds a higher tap frequency could still be achieved. Yokoe et al. also found tapping frequency to be moderately discriminatory with a classification accuracy of 67% but with PwPD having a slightly lower frequency than HC (Yokoe, Okuno et al., 2009). Ling et al. found that tapping frequency was marginally lower in PwPD *off* (2.8 Hz) than controls (3.3 Hz) and that tap frequency increased in the *on* state (3.1 Hz) driven largely by improvements in speed and then rhythm. Tapping frequency was highly discriminatory in keyboard tapping studies with a tendency for PwPD to have a lower frequency (Dunnewold, Jacobi et al., 1997, Homann, Suppan et al., 2000, Muir, Jones et al., 1995, Pal, Lee et al., 2001, Shimoyama, Ninchoji et al., 1990). It is difficult to compare FT measurements with keyboard tapping measurements though as it is not possible to measure amplitude with the latter and the tapping frequency is constrained by distance between the keys, making frequency largely a measure of speed.

Bradykinesia composite models improve classification accuracy

This study showed that when the individual component measures of bradykinesia were combined into composite models the classification accuracy improved from 80% for the best performing individual component (amplitude) to 85% for a composite comprising amplitude, speed and rhythm components, and to 88% for a composite comprising all FT component measures. The improved classification accuracy with composite measures has been reported just once before (Taylor Tavares, Jefferis et al., 2005). Although these composites accurately classified the UCSF patients, there were many false positive misclassifications amongst the UCSF controls giving overall classification accuracies of 71% for the composites comprising all FT components. Reassuringly the composite's classification accuracy was upheld for PwPD in either *on* or *off* states. Examining the composite equation shows that the most important measures are rhythm, amplitude and speed with greater relative weight given to these components.

Classification accuracy in models of clinically indeterminate PD

When only LA PD hand data was used (mean MDS-UPDRS grade 1.05) the AUC remained moderately good ranging between 0.78 – 0.85 for amplitude, speed and rhythm. When the grade zero PD data sets (n =11) were compared to grade zero control data (n = 72) a significant difference between the mean amplitude scores remained and overall classification accuracy for the bradykinesia composite comprising all component measures was AUC 0.77. This was equivalent to 85% and 62% of HC and PD data respectively being correctly classified whereas when all data was included, 89% of HC and 85% of PD data sets were correctly classified. The bradykinesia composite differentiated grade one, or 'slight bradykinesia' PD data (n = 45) from HC data (n = 72) with AUC of 0.89.

Section 4.5.2 showed that PD data sets misclassified by the bradykinesia composite tended to be either those that had been graded as 'normal' or as 'slight bradykineisa' (n = 3) but with kinematic values more than 2 SD below the grade one group median value. In other words misclassified PD data sets are either those with no bradykinesia clinically apparent or those

with unusually low kinematic results, and the former suggests that further assessments in clinically indeterminate cases of PD, or those with very mild signs need to be undertaken in order to assess the true diagnostic potential of objectively measuring a clinical sign that is not yet clinically apparent, or only very mildly so.

Summary of classification accuracy of component measures of bradykinesia

The present study has shown that objective measures of the clinically defined components of bradykinesia discriminate PD patient and control data with different accuracies but amplitude, speed and rhythm are consistently the most discriminating measures. Halts and tapping frequency are less discriminatory and SE measurements did not discriminate the groups at all. Even when the FT performance had been graded 'normal' the mean measures of amplitude remained significantly different in the PD data compared to controls.

In the Leeds patients *on*, reduced amplitude is the most discriminating individual component of bradykinesia, followed by rhythm and this is probably because the speed component has been 'normalised' by the dopaminergic drugs (Espay, Giuffrida et al., 2011). These results are broadly supported by the previous four studies that quantified PD FT with speed improving much more than amplitude or rhythm when the patient is *on*. When the UCSF patients were in the *off* state speed and amplitude were roughly equivalent as the most discriminatory components but these results require further validation as some previous studies support this (Ling, Massey et al., 2012) whilst others suggest that amplitude remains more discriminatory in either clinical state (Espay, Beaton et al., 2009, Espay, Giuffrida et al., 2011).

Clinical and technical implications of diagnostic classification results

These kinematic measure results may inform clinical practice as they suggest that clinicians should focus on the rhythm, amplitude and speed of movements when considering whether a new patient may have PD. Heldman et al. showed that most clinicians focus on amplitude when allocating a UPDRS bradykinesia grade (Heldman, Giuffrida et al., 2011) but it is not known if this is also the case when formulating a diagnosis of PD. The importance of accurately evaluating bradykinesia was highlighted by Bajaj et al. (Bajaj, Gontu et al., 2010) who showed that the most common reason for misdiagnosing PD from other tremulous movement disorders was clinical misinterpretation of bradykinesia.

From a technical aspect, if one was to develop a device that can predict diagnosis of PD based on measuring the separable components of bradykinesia, these results suggest that a composite measure of all the components of bradykinesia, with particular emphasis on amplitude, speed and rhythm, would increase the classification accuracy beyond the level of any individual component measure. It is also encouraging that in the models of clinically indeterminate PD (grade zero UPDRS or 'normal' FT movements) there remained a significant difference between the objective measures of amplitude despite no difference in the clinical grade.

From an investigative tool aspect the present study and the four published studies quantifying FT suggest dissociation between the components of bradykinesia in PD, in both the *off* state and in response to intervention. It has been shown that there is a differential response of the components to drugs (Espay, Beaton et al., 2009, Espay, Giuffrida et al., 2011, Ling, Massey et al., 2012) and other interventions such as simultaneous movements (Kishore, Espay et al., 2007) and neurosurgery (Kimber, Tsai et al., 1999). For example Kishore et al. showed that amplitude, but not speed, of movements improves when PD patients perform bimanual tasks (Kishore, Espay et al., 2007). Kimber at el. showed that after pallidotomy surgery PW wPD exhibited an improvements in movement amplitude but not in speed or rhythm (Kimber, Tsai et al., 1999). These suggest that there may be

different underlying pathological processes for each component and this is discussed further in Chapter 5.

4.7.2 Discussion in context of developing a monitoring device

The study methodology was not specifically designed to evaluate how well the device could monitor disease progression but useful information could still be gleaned on how the component measures may reflect the degree of bradykinesia and other measures of disease progression.

Correlation of kinematic data and subjective clinical grades

This study found that the MDS-UPDRS FT grade was significantly associated, in descending order of correlation, with rhythm, speed, amplitude and halts. More severe clinical bradykinesia was associated with less rhythmic, smaller, slower and more interrupted movements measured kinematically. These findings are similar to Yokoe et al.'s where, in descending order of correlation, lower speeds, smaller amplitudes and less rhythmicity of FT movements were associated with more severe clinical grades of bradykinesia in PwPD *on* (Yokoe, Okuno et al., 2009). These results suggest a role for objective measurements of bradykinesia components to monitor clinical changes.

In contrast to improved classification accuracy with the bradykinesia composite models, the strength of correlation was no better for the composites compared to when only rhythm or speed were used. There are no previous studies comparing bradykinesia composites to clinical grades of bradykinesia.

Other measures of disease progression such as HY stage, disease duration and LEDD were not associated with any of the individual component measures of bradykinesia. Advancing age was also not associated with any of the component measures in either patients or controls, suggesting that subclinical bradykinesia, or generalised slowing does not occur simply with ageing.

In Espay's 2009 study the total UPDRS motor examination score correlated with FT amplitude measurements when patients were *off* but there was no correlation with either speed or amplitude measures in the *on* state (Espay, Beaton et al., 2009). This suggests that the objective measures are more sensitive that the UPDRS and perhaps this is not surprising as each UPDRS bradykinesia assessment had a composite grade comprising all the separate components of movement, and all of these composite grades are then summed together to obtain the overall motor score. Espay et al. also found that HY stages were not correlated with the speed and amplitude kinematic measures. This is an important point because it suggests that clinical scales of clinical severity or disease stage are too crude to correlate with the detailed objective measurements. It could be argued that this is related to the fact that clinical scales are rating features other than bradykinesia (i.e. overall mobility/ clinical signs of PD) but in the case of UPDRS this is not the case and at least some correlation with the kinematic scores would be expected. This lack of correlation suggests the UPDRS scale is not sensitive enough to reflect the objective changes in the movement components.

The Espay 2011 study suggests that even the MBRS, a detailed clinical scale specific for bradykinesia, is not sensitive enough to detect the changes in amplitude and rhythm, with the MBRS grades not significantly changing between *on* and *off* states, despite evidence of improvement on the kinematic measures (Espay, Giuffrida et al., 2011). In other words the subjective assessments were not as sensitive as the objective assessments.

Adjustment for demographic details of test subjects

When developing a device that could potentially diagnose or monitor PD it would be useful to know whether the accuracy of the device would be improved by adjusting the results for the demographic details of the test subjects. This was briefly explored in the present study.

Accuracy not improved by using only healthy control dominant hand

The results showed that accuracy was not improved by using only the dominant hand data from controls, suggesting that there was little difference

in the component measures between HC hands. Ling et al. found similarly that HC had longer tap cycle durations in the non-dominant hand but no other performance parameters differed between hands (Ling, Massey et al., 2012). Espay and Yokoe did not present the data for the hands separately. The results from studies assessing touch pad tapping are mixed: Muir found tapping frequency in the non dominant hand of controls was 7% slower than dominant hand but the results were not significant (Muir, Jones et al., 1995), Dunnewold found both controls and PD patient tapped significantly slower with the non dominant hand (Dunnewold, Jacobi et al., 1997) but Kandori found no difference at all between the hands (Kandori, Yokoe et al., 2004).

4.7.3 Strengths and limitations of the study

The present study is the largest study in terms of number of participants to quantify the clinical FT task in PwPD *on* and *off* and compare the results to HCs. Choosing the FT task as an evaluation of bradykinesia has the advantage that this test is already used clinically and is familiar to clinicians. Testing both hands separately was important as it has been shown that bimanual tasks in PD tend to improve the movements of the MA hand but degrade the movements of the LA hand (Kishore, Espay et al., 2007).

The findings that amplitude, speed and rhythm are disproportionately more impaired in PD than other clinically defined components of bradykinesia are broadly supported by previous studies. However it is the first study to validate the results on an independent data set. In addition this study provides a detailed evaluation of how each component measure varies within, and between, the patient and control groups. It is the second study to combine the component measures into a bradykinesia composite model and this improved the classification accuracy. It is also one of only two studies to highlight that there is no significant difference between SE measurements in PD and HC groups. Methodologically a broad spectrum of PwPD have been recruited with wide ranging symptoms, signs and stages of the disease – this means the test has not been limited to those with intact cognition or minimal tremor or dyskinesia. The predominance of patients with only

slight bradykinesia offers credibility to the results on diagnostic classification.

Several limitations of the study are acknowledged though and most of these are related to the methodology. Firstly, only one task has been quantified and it is not possible to correlate the bradykinesia findings to other impairments in PD, or to clinical subtypes of PD, because the whole UPDRS motor examination was not assessed. Although Agostino et al. showed that FT is a more discriminatory task for PD than hand opening/closing or forearm pronation/supination (Agostino, Curra et al., 2003) and Ling et al. showed that deficits in FT were mirrored by the same deficits in handwriting performed by PwPD (Ling, Massey et al., 2012), we cannot be absolutely sure that the component measures impairments in FT movements reflect bradykinesia generally in PD.

Secondly, the task was only assessed once and it is possible that the movements vary considerably within individuals over repeated assessments i.e. the test retest reliability was not assessed and this would be an important next step when developing the device. Furthermore the protocol required all subjects to start the assessments with their dominant side and this may have led to a degree of motor learning by the time the non-dominant side was assessed and distorted the results according to hand dominance. In support of this concern, O'Sullivan showed that there was a significant increase in tap frequency on the second trial in ten PD patients who performed three trials of tapping with the same hand (O'Sullivan, Said et al., 1998). A further potential area of bias was having only one clinician to grade the UPDRS severity who was not blinded to the diagnosis.

Thirdly, for a true evaluation of a new diagnostic device, it should be tested on a group of clinically indeterminate patients and compared to the current gold standards of diagnosis based on clinical assessments with or without DaT scan. Using clinically definite PwPD limits any claims about diagnostic accuracy. Nevertheless the results are still an important first step in proving the principle that separable component analysis of FT data shows

potential for diagnostic classification and grading of bradykinesia severity. Also, as with all FT studies thus far, the gold standard has been clinical diagnosis rather than pathological diagnosis, and it is possible that some of the HC group have sub-clinical PD and some of PD group have another parkinsonian condition. Finally, approximately 10% of the collected data was corrupted and this not only led to considerable loss of data, but also when the corrupted data sets were excluded the groups were far less matched for age and gender.

4.7.4 Conclusions

Objective measures of the separable components of bradykinesia can be used to discriminate FT data from PwPD and HCs. PD bradykinesia is characterised by movements of reduced amplitude, slower speed, less rhythm and more halts. Decrementing amplitude and speed were no different between PD and HC groups. The response of each component to levodopa is dissociable with reduced amplitude being the most discriminative component of bradykinesia in PwPD *on* but the results are less clear for the *off* state. Even when there is no bradykinesia apparent clinically, objective measures of amplitude are significantly different between the grade zero PD and HC groups; this suggests potential for using such measures as a diagnostic tool. When a bradykinesia composite model was formed from all the component measures, classification accuracy improved beyond that of any single component. Amplitude, speed, rhythm and halts measurements are also associated with the clinical *severity* of bradykinesia and this suggests a role for monitoring PD. In summary objective measures of bradykinesia show potential to discriminate PD from HC subjects and to monitor the progression of PD and the response to interventions.

Chapter 5

Evaluating bradykinesia with evolutionary algorithms: Results and Discussion

5.0 Introduction

This chapter will focus on the classification accuracy of the ensemble classifier for discriminating FT data from PwPD and HCs. The ensemble classifier was developed using EA analyses of the data and the classification results are compared to the best performing separable component measures of bradykinesia from Chapter 4. The classifier expression is examined in order to better understand the behaviour of the classifier and to investigate the characteristics of bradykinesia. The discussion section summarises the main findings, outlines how these results may inform the development of a device to aid clinical diagnosis and monitoring of PD and reviews how the results fit in with the current understanding of the pathophysiology of bradykinesia.

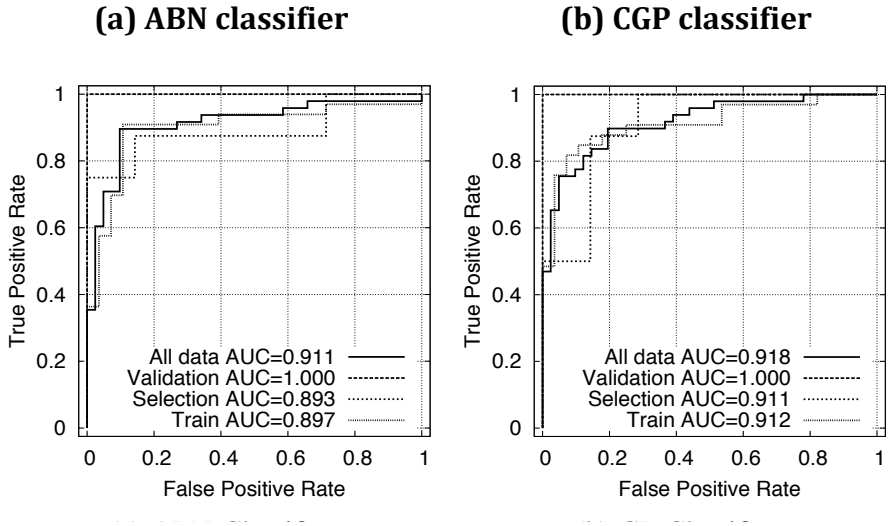
5.1 Ensemble classifier accuracy

EA analysis of the FT data led to the production of a number of classifiers that discriminated patient from control data very well. FT data from each 30 second recording was processed by the classifier to produce a solution in the range - 1 to +1. A solution greater than zero meant that the classifier predicted the data set came from a patient and a solution of less than zero meant the classifier predicted that the data came from a control. The diagnostic accuracy of the classifiers was assessed using AUC.

Two different types of classifier were developed through EA analysis techniques – an Artificial Biochemical Network (ABN) classifier and a Cartesian Genetic Programming (CGP) classifier; see Chapter 3. The ABN classifier was developed using EAs to examine overall patterns of FT movement during the 30 second FT assessment period and using the differentiating features of these gross patterns of movement to form the equation; the ABN classifier thus evaluates movements on a macroscopic

scale. In contrast, the CGP classifier was developed by focussed evaluation of 0.6-second windows of FT data and using the subtle differences of movement over each individual FT cycle as the defining components of the equation; hence the CGP classifier can be considered to evaluate movements on a microscopic scale. The classification accuracy of the best performing ABN and CGP classifiers are summarised in Figure 40 with 'Selection' (or test) AUC values of 0.893 and 0.911 for ABN and CGP classifiers respectively.

Figure 40 Classification accuracy of the most discriminative ABN and CGP classifiers



Legend: ROC curves summarising classification accuracy of best performing (a) ABN and (b) CGP classifiers.
ROC curves produced by Dr Mic Lones

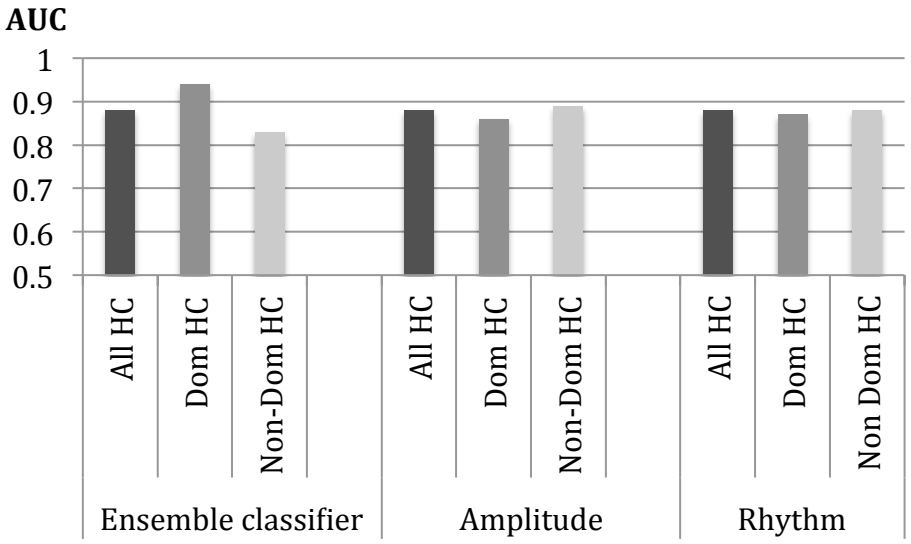
The behaviour of the ABN and CGP classifiers may be considered complementary and their outputs were normalised and then averaged in order to create an 'ensemble' classifier that maximised the ability to discriminate PD and HC data based on both 'microscopic' *and* 'macroscopic' evaluation of FT movements. It will be shown that the classification accuracy improves when the classifiers are combined into an ensemble classifier in this manner.

5.1.1 Accuracy improves with dominant hand control data

In order to assess the ensemble classifier accuracy and compare the results to the separable component analysis results, the approved data sets were used (see Table 12 in Chapter 4). When all LTHT approved data sets were used (87 PD, 72 HC) the ensemble classifier had good classification accuracy with an AUC of 0.88 (95% CI 0.83 - 0.94, $p < 0.0001$). At equal trade-off the sensitivity/ specificity was 0.84/0.82 and this corresponded to 77 out of 87 PD, and 54 out of 72 HC data sets being correctly classified. These results are similar to the best performing individual separable component measures of bradykinesia, amplitude and rhythm (AUCs 0.88).

Figure 41 demonstrates the ensemble classification accuracy when sub-sets of HC hand data are used and compares this to the equivalent results when the most discriminating separable components were used to classify subjects.

Figure 41 Comparison of ensemble classifier accuracy when control data sets limited to dominant or non-dominant hand



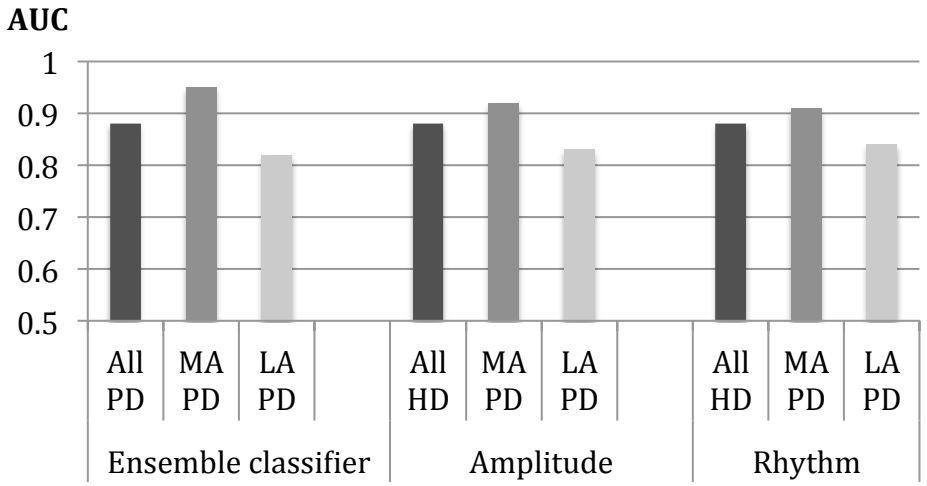
Legend: Classification accuracy for the ensemble classifier and the best two performing separable components, amplitude and rhythm, are presented. The AUCs are shown when all PD data is combined with either all HC data, or all PD data is combined to a subset of dominant (Dom) hand HC data, or non-dominant (Non Dom) hand HC data.

Figure 41 shows that when HC data sets are limited to just the dominant hand kinematic data (rather than using both hands data) the accuracy of the ensemble classifier improves from an AUC of 0.88 to 0.94 (95% CI 0.90 - 0.98, $p < 0.0001$) but there is no corresponding improvement in the separable component classification accuracy results with AUCs 0.86 and 0.87 for amplitude and rhythm respectively. This suggests that the ensemble classifier includes a feature of FT movements that is over-represented in dominant hand HC data. The fact that the ensemble classifier seems less likely to misclassify the dominant hand data sets is an informative point when considering how to further develop a diagnostic device using EA induced classifiers: it may be that even better results could be obtained if the algorithms are trained, and tested, using solely dominant hand control data.

5.1.2 Accuracy improves when most affected hand PD data is used

Figure 42 shows that when PD data is limited to the MA hand, classification accuracies are similar for the ensemble classifier and the separable component measures of bradykinesia.

Figure 42 Comparison of ensemble classifier's classification accuracy when patient data limited to the most affected hand



Legend: The classification accuracy of the ensemble classifier and the separable component measures improve when only the MA PD hand data is used, and deteriorate when only the LA PD hand data is used.

5.1.3 Comparison of classifier and separable component accuracy

The classification accuracies of the ensemble classifier, the individual separable component measures, and the bradykinesia composite using various combinations of the approved data sets is summarised in Table 23.

Table 23 Comparison of classification accuracy for ensemble classifier and separable component measures

Test data	Ensemble classifier	Individual component (amplitude)	Composite of all FT components
	AUC (95% CI)	AUC (95% CI)	AUC (95% CI)
All PD, all HC	0.88 (0.83 - 0.94)	0.88 (0.82 - 0.93)	0.93 (0.90 - 0.97)
MA PD, all HC	0.95 (0.91- 0.98)	0.91 (0.87 - 0.97)	0.96 (0.94 - 0.99)
All PD, Dom HC	0.94 (0.90 – 0.98)	0.86 (0.79 - 0.93)	0.93 (0.88 - 0.98)
MA PD, Dom HC	**0.99** (0.98 – 1.00)	0.90 (0.84 – 0.97)	0.96 (0.92 - 0.99)

Legend: The classification accuracy for three different methods of analysing the FT data is presented. MA, most affected hand; Dom, dominant hand.

The first row of Table 23 shows that when all data sets are used the composite model (comprising all the separate component measures of bradykinesia) classifies the data sets more accurately (AUC 0.93) than the other two methods (AUCS 0.88). The second row shows that for all methods the AUC improves when patient data is limited to the MA hand. The third row results (when compared to the first row results) demonstrate that only the ensemble classifier has improved classification accuracy when the control data is limited to the dominant hand data. Therefore the ensemble classifier is better than the other two methods at discriminating MA PD data from dominant hand HC data with an AUC of 0.99 (0.98 -1.00,

$p < 0.0001$). The best performing individual component, amplitude, had an AUC of 0.90 (0.84 - 0.99, $p < 0.0001$) and this was significantly inferior to the ensemble classifier accuracy ($p = 0.042$). The composite model has an AUC of 0.96 (0.92 – 0.99, $p < 0.0001$) with a trend towards being inferior to the ensemble classifier accuracy ($p = 0.071$).

Only the ensemble classifier showed a significant improvement in classification accuracy when the data was limited to MA PD and Dom HC compared to using all data sets (0.88 improved to 0.99; $p < 0.05$) whereas the other two methods showed a slight numerical increase in AUC (0.88 improved to 0.91, and 0.93 improved to 0.96) when the data was limited to certain sub-sets of data but this was not a significant improvement from the baseline AUC when all data sets are included ($p > 0.09$). This suggests that the best classification accuracy of the device would be obtained if the ensemble classifier was used to predict diagnostic classification and 'query PD' test data were obtained from just the MA hand and then compared to dominant hand control data.

However, it could be argued that assessment of the classification accuracy of the composite model is flawed as this model was produced through logistic regression analysis to provide dichotomy of all patient vs. all control data, and not MA hand PD vs. dominant HC data. The comparison of the composite model and the ensemble classifier accuracies may thus be biased in favour of the ensemble classifier, especially if the evolutionary algorithms preferentially used kinematic features from the HC dominant hand to induce the classifier.

In view of this an additional composite model was produced using all the separable components of FT as before but limiting the regression analysis to dichotomising MA PD data and dominant hand HC data. This new model, termed 'MA PD vs. Dom HC composite' was expressed as:

MA PD data is predicted if (- 26.4 + 4.2 frequency + 45.8 amplitude - 1.8 speed + 20.1 COV amplitude - 20.0 COV speed +12.2 halts - 0.01 decrementing amplitude + 0.16 decrementing speed) > 0.50

Table 24 shows that this model gave an overall classification accuracy of 91.4% ($R^2 = 0.79$). This corresponds to an AUC of 0.97 (0.94 - 0.99, $p < 0.0001$) that was numerically inferior to the classification accuracy of the ensemble classifier at AUC 0.99 and showed a statistical trend towards being significant ($p = 0.081$). This suggests that the ensemble classifier more accurately classifies MA PD kinematic data from dominant hand HC data than either of the composite models comprising separable component measures of bradykinesia.

Table 24 Classification accuracy of composite bradykinesia model developed to discriminate MA PD data from dominant hand HC data.

		Predicted		
		MA PD	Dom HC	**% correct**
Observed	MA PD	43	3	93.5
	Dom HC	4	31	88.6

5.1.4 Kinematic data corruption and classifier accuracy

Any corruption in the kinematic data would cause erroneous separable component analysis results because these measures are calculated by taking an average, or trend, of the components in each individual tap performed over the 30-second recording period. The necessary exclusion of the corrupt data sets is a major weakness of the separable component measures technique. For this research study the loss of data has not been too troublesome as there were still enough remaining approved data sets for useful analysis to be undertaken. However, if the device were to be used as an objective test of bradykinesia in a clinical setting any corruption within the data sets would probably require the patient to undergo a repeat test and

it is likely that a test with a 10% repeat test rate would be considered too unreliable to be clinically useful.

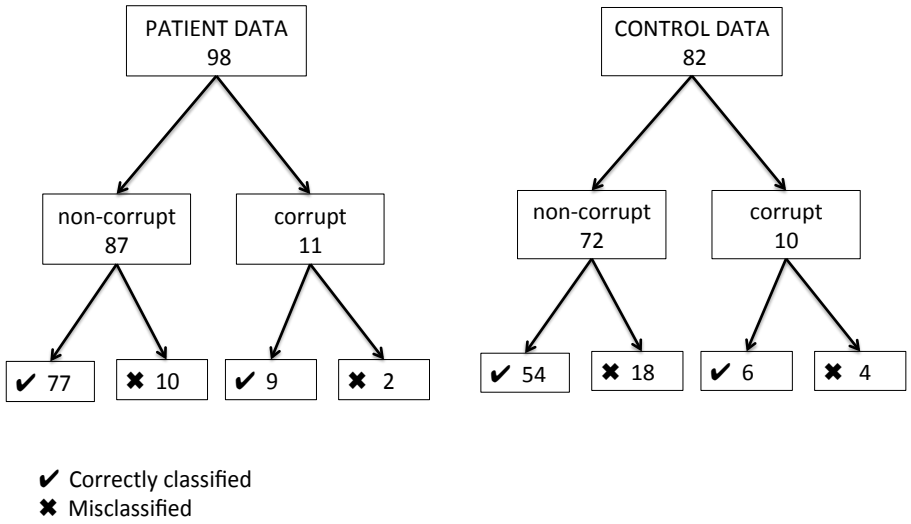
In contrast, it is hypothesised that the EA induced classifiers may not be affected by FT data corruption. This is because EAs 'search' for patterns of movement data that are over-represented in one group's data (e.g. patients) compared to another group (e.g. controls) and then use these kinematic features to form a classifier equation that best 'splits' the data predictions into the correct diagnostic groups. If the corrupt data occurs to the same extent in the PD and HC data, and especially if it does not occur very frequently compared to all other intact FT data points within a recording, it is likely that the corrupt data points will not affect the overall classifier equation. In other words the EA uses features that discriminate the groups from one another and as the corrupt data is a feature of both patient *and* control data it is unlikely to be a key discriminatory feature included within the classifier equation.

At the time of the training data being used by the EA to induce classifiers it had not been recognised that some of the kinematic data was corrupt. Hence the classifiers were inadvertently developed without excluding corrupt data from the training and test data sets. This initial oversight allows for evaluation of how the EA classifiers perform on all data sets (including corrupt data) in addition to how they perform on the approved data. If there is little difference between the classification results when corrupt data sets are included this highlights a distinct advantage of the ensemble classifier over the separable component analyses approach.

Figure 43 summarises the accuracy of the ensemble classifier for corrupt and approved data sets. It shows that nine out of 11 (82%) of the corrupt, and 77 out of 87 (89%) of the approved data sets from PwPD were correctly classified with no association between correct classification and the presence of corruption ($X^2 = 0.41$, $p = 0.52$). Likewise for HC data six out of 10 (60%) of the corrupt, and 54 out of 72 (75%) of the approved data sets were correctly classified ($X^2 = 1.0$, $p = 0.32$).

When *all* data sets were included the ensemble classifier correctly classified 88% (86 / 98) PD sets and 73% (60 / 82) control sets. The classification accuracy was not significantly different when *the corrupt data sets were excluded* with 89% (77 / 87) PD approved data sets and 75% (54 / 72) control approved data sets correctly classified (p's > 0.1).

Figure 43: Accuracy of ensemble classifier in corrupt and approved data sets



It is noteworthy that the ensemble classifier was more accurate at classifying dominant hand HC data (4/41 data sets misclassified) than non-dominant hand HC data (18/41 data sets misclassified). In view of this, and the fact that the best ensemble classification results were obtained for the MA PD and dominant hand HC data, a further analysis was undertaken to be absolutely sure that the superior results of the EA classifier were not simply because these sub-groups of data lacked any corrupt data sets.

This revealed that three of the 49 MA PD data sets contained corruption and one of these was misclassified; this misclassification rate was not significantly different to the remaining 44 approved MA PD data sets where two sets were misclassified (p = 0.18). Four of the dominant hand HC data sets were corrupted but none of these were misclassified and four out of the

37 remaining approved dominant hand HC data sets were misclassified. No association between classification accuracy of dominant hand HC data sets and the presence of corruption within the data was found ($p = 0.67$). In summary these results show that data sets with corruption are no more likely to be misclassified by the EA ensemble classifier than approved data sets.

5.1.5 Validation of classifier on San Francisco data

Using all data

There were 26 data sets collected in UCSF for PwPD *on* but one assessment did not record. Twenty-three out of 25 UCSF PD data sets were correctly classified but only six out of 18 UCSF HC data sets. The classification accuracy for patients *on* was similar for LTHT, 88% (86/98) and UCSF, 92% (23/25), $p = 0.73$. However the classification accuracy for control data in LTHT, 73% (60/82) and UCSF, 33% (6/18) significantly differed, $p < 0.0001$.

Using most affected PD hand and dominant hand control data

The ensemble classifier correctly classified the MA hand data for all 13 PwPD and the dominant hand data for six out of nine HCs assessed in UCSF. These results show that the ensemble classifier (developed using Leeds data) generalises to independently collected data from patients' MA hands with similar classification accuracy results at the two centres ($p = 0.48$). There was a trend towards the classifier misclassifying more of the dominant HC data from UCSF than LTHT ($p = 0.09$) but it is difficult to draw firm conclusions from this due to the small size of the control validation data set. Whether the inferior classification results in UCSF controls is due to the classifier not generalising to other data sets, or whether it could be a calibration problem remains to be determined.

Using *on* and *off* clinical state data

Twenty-six data sets from UCSF patients in the *off* state were collected and 25 in the *on* state. In the *off* state, 96% (25/26) of data sets were correctly classified and in the *on* state 92% (23/25) of data sets. All 13 patients MA

hand data sets were correctly classified regardless of whether the patients were assessed in an *on* or *off* clinical state. These results suggest that the ensemble classifier is able to correctly classify patients regardless of their clinical state. This is a very encouraging result when considering how such a device could aid clinical diagnosis – it suggests flexibility in test protocol is possible allowing subjects with possible early PD to be tested in an *off* or *on* state.

5.2 Elucidating the behaviour of the classifier

5.2.1 Examining the classifier expression

It has been shown that the ensemble classifier discriminates PD and HC data more accurately than the separable component measures. This suggests that the classifier is using features of the FT data other than solely measures of the defined components of bradykinesia. In other words the EAs have searched for the most discriminating features of the FT data in an unbiased manner and the classifier has not been constrained by the current clinical definition of bradykinesia. Classifiers induced by EA can be examined in order to discover what features of the movement data were used in their formation. This means that in addition to being used for diagnostic classification, the classifier can be used to investigate what are the most discriminating features of PD movements. This exciting aspect of classifiers potentially enables new features of bradykinetic movements to be found.

Methods of examining classifiers

The behaviour of the ABN classifier is considered more difficult to understand than the CGP classifier so this section focuses on examination of the latter. There are different ways to explore a classifier and one fairly straightforward method involves examining what differs between the kinematic data sets with ‘strong patient’ and ‘strong control’ solutions i.e. patient data that was scored very close to +1 and control data that scored very close to -1. The *differences* between the two groups of data are likely to be the key discriminating kinematic features used as components in the classifier equation.

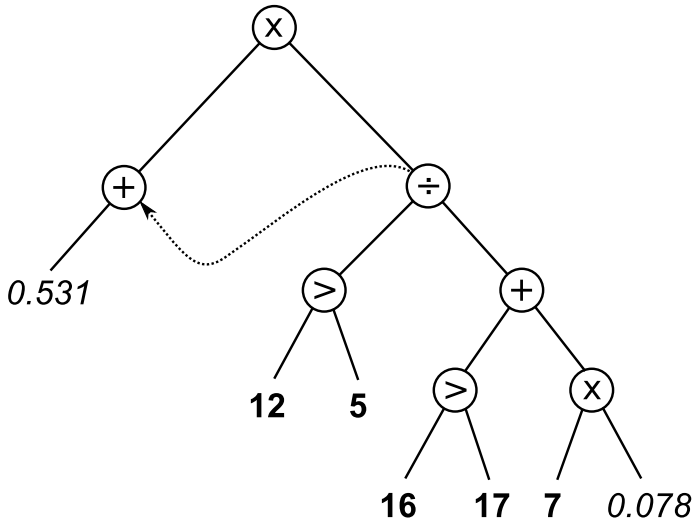
Therefore the top six PD and top six HC windows of data (each of 0.6 seconds duration) that had the strongest GP classifier results (closest to +1 for PD and -1 for HC) were visually inspected for the kinematic features that best discriminated the PD and HC data.

To focus this inspection the CGP classifier equation was expressed in order to reveal which acceleration training data points, or 'window offsets', were used to construct it (see below for further explanation of window offsets). These window offsets were overlaid onto corresponding sections of the tap cycle in the strongest scoring PD and HC windows to highlight the distinguishing features between the groups. Quantitative evaluation of the diagnostic accuracy of the kinematic features highlighted by visual inspection could then be undertaken using ROC analysis.

Expression of genetic programming classifier

The most discriminative CGP classifier is expressed in Figure 44. This is called a 'parse tree representation' and is read upwards, starting from the terminals or 'leaves' at the bottom of the tree and reaching the classifier solution at the top.

Figure 44 Expression of CGP classifier

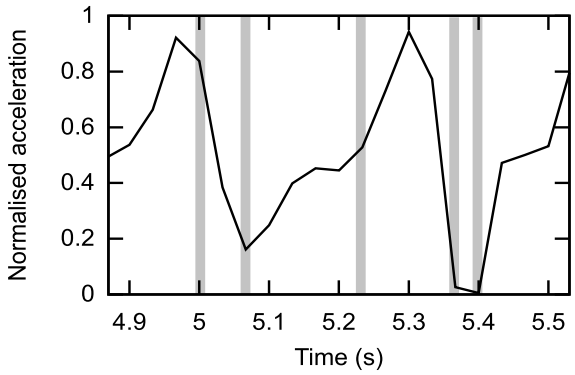


Expression produced by Dr Mic Lones

At each ‘node’, denoted by a number or a mathematical symbol within a circle, the expression is evaluated and then these node sub–results form part of the solution at the next node. The numbers in italic are constants and the bold numbers (12, 5, 16, 17, 7) denote the ‘window offsets’. Window offsets are the data points in the normalised acceleration data (out of the 20 data points in each 0.6 second window) that enable the best classification of the data when included in the expression. The window offsets are demonstrated in relation to the kinematic data in Figure 45.

The dashed line in Figure 44 indicates that the result evaluated at the divide sign node is used twice in the classifier expression – once as an input to the plus sign node (at the left side of the tree i.e. to be added to 0.531) and once as an input to the final multiplication node.

Figure 45 Window offsets used in CGP classifier



Legend: The classifier uses normalised acceleration data points in the expression. For a 0.6 second window there are 20 data points because the movement data was recorded 60 times per second (60 Hz) but down-sampled by a factor of two to remove noise. The offset windows used in the expression are marked as vertical grey lines. The normalised acceleration data point marked by the first vertical line (at 5 seconds time) is the ‘5’ window offset, the second line (at approximately 5.06 seconds time) is the ‘7’ window offset, the third line (at 5.24 seconds) is the ‘12’ window offset and so on.

The steps for calculating the solution to the CGP classifier expression in Figure 45 For each 0.6 window of data are as follows:

Step 1: Data point at window offset 7 is multiplied by 0.078

Step 2: Whichever of window offset data points 16 and 17 is greater is then added to the solution from Step 1.

Step 3: Whichever of window offset data points 5 and 12 is greater is then divided by the solution from Step 2.

Step 4: The solution from Step 3 is added to 0.531

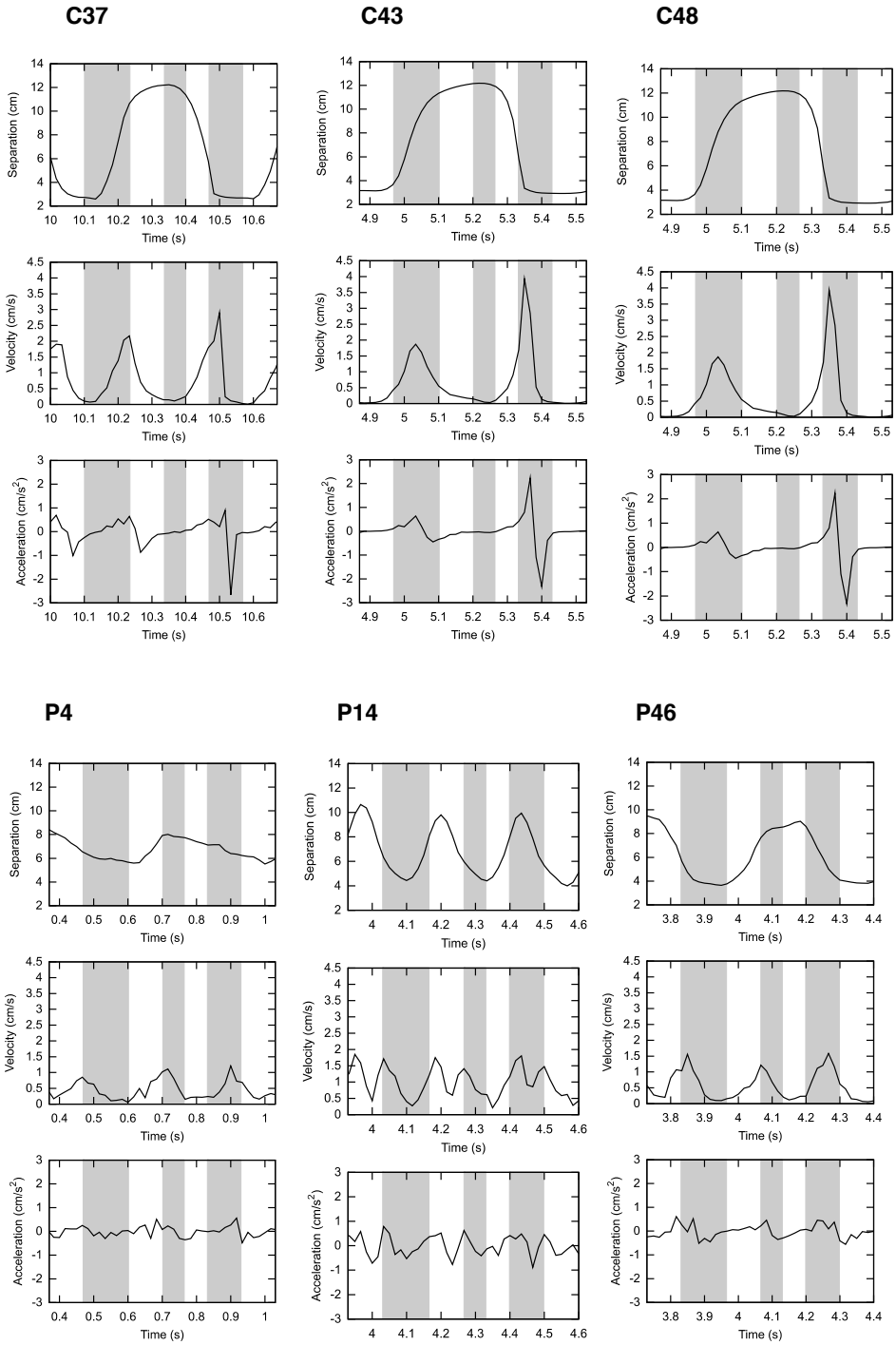
Step 5: The solution from Step 4 is multiplied by the solution from Step 3 to obtain the classifier solution for this individual 0.6 seconds window of data.

If the solution is greater than zero (up to a maximum of +1) the window of data is classified as PD and if the solution is less than zero (to a minimum of -1) the window of data is classified as HC. The overall classification of the 30-second period of FT was obtained by taking the mean classifier solution score for *all* the 0.6 second windows i.e. 50 x 0.6 second window classifier scores divided by 50.

Comparing PD and HC kinematic profiles

Three PD and three HC data windows with the strongest GP classifier results are shown in Figure 46 and the shaded areas mark the window offsets in the acceleration data and their corresponding relationship to finger separation and speed profiles, when the acceleration data is integrated. Visual inspection revealed that the shaded areas were consistently positioned over three points of the tap cycle in control acceleration data: the opening acceleration and deceleration peaks, the zero acceleration period corresponding to maximum digit separation, and the closing acceleration and deceleration peaks.

Figure 46 Comparison of kinematic profiles of patient and control data windows that had the strongest CGP classifier solutions



Legend: Separation, velocity and acceleration data windows (of 0.6 second duration) that had the strongest GP classifier solutions are presented for three HCs, C37, C43,C48, and three PwPD, P4, P14, P46 are presented. Shaded areas highlight the most significant acceleration data points that were used in classifier formation.

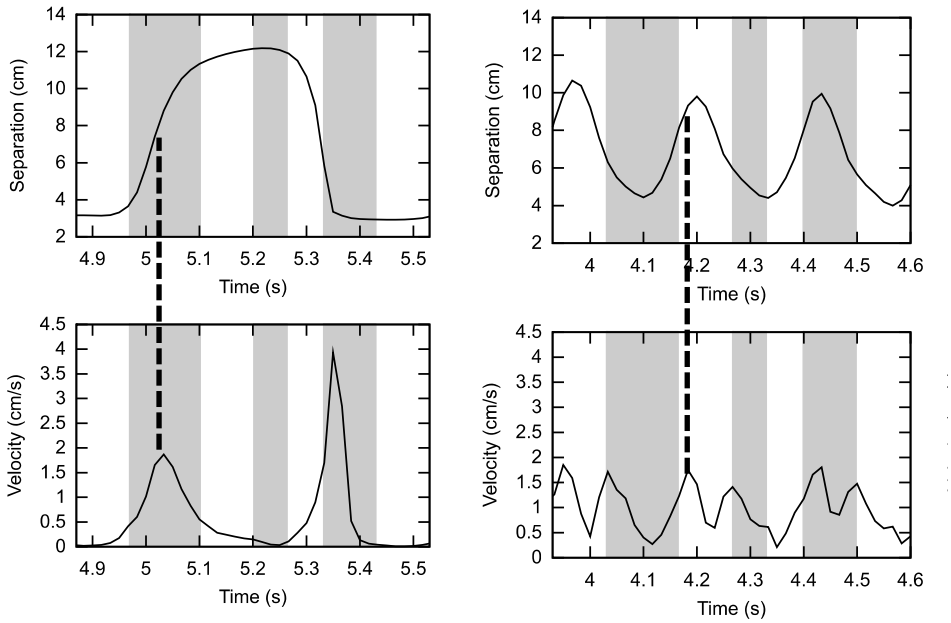
Visual inspection of the 12 data windows with the strongest CGP scores showed that HC data was characterised by a large sharp peak of acceleration then deceleration in the opening phase of the tap, with a similar pattern, of an even greater magnitude in the closing phase of the tap. In between these peaks at the point of maximum (and also minimum) digit separation there was almost zero acceleration. This acceleration profile corresponded to a brief sharp peak of speed in the opening and closing phases, with the closing phase speed consistently greater, and a smooth separation profile characterised by rapid opening, then a pause at maximum aperture followed by rapid closing and a further pause at opposition.

In contrast the PD kinematic data was characterised by a jerky acceleration profile throughout the tap cycle with the magnitude of the peaks of acceleration and deceleration smaller and more dispersed, and no obvious difference in the opening and closing phases. The peak opening speed was similar to the peak closing speed and both were slower than in HCs. The shape of the speed peaks was more dispersed and the maximum speed occurred later in the tap cycle phase i.e. in PD the peak opening speed occurred as the fingers approached maximum separation whereas in HCs the peak speed occurred earlier when the fingers were approximately 50% separated (Figure 47).

These observations of the data, and correlation to the GP parse tree expression, prompted further quantification of the maximum acceleration, maximum deceleration and maximum speed of the opening and closing phases of the tap cycle. The mean values of these variables over the 30 second FT period was calculated then the classification accuracy assessed on MA PD and dominant hand HC LTHT data using AUC.

Figure 47 Relationship between peak opening speed and maximum separation of finger tapping

(i) Control C43 **(ii) Patient P14**



Legend: The timing of the peak opening speed in relation to the opening phase of separation highlighted with dashed vertical lines. (i) In the example of control data (C43) the peak opening speed occurs in the first half of the opening phase of the tap cycle when the digits are open approximately 50% of the maximum separation. (ii) In the example of patient data (P14) the peak opening speed occurs in the second half of the opening phase of the tap cycle when the digits are almost at their maximum separation.

The most discriminating separable component measures so far have been the mean of the maximum amplitude, mean of the maximum speed and rhythm with AUCs of 0.90, 0.89 and 0.90 respectively when assessed in LTHT MA PD *on* and dominant hand HC data. Table 25 shows that the opening phase speed was numerically more discriminatory (AUC 0.91) than the closing phase speed (AUC 0.87) measurements, although there was no significant difference statistically between these ($p = 0.43$) or the overall measures of speed, amplitude and rhythm ($p = 0.87$).

Table 25 Classification accuracy of opening and closing phases of finger tap cycle

FT component	AUC	95% CI	*p*
Opening phase			
Acceleration	0.79	0.70 - 0.89	1×10^{-6}
Deceleration	0.78	0.68 - 0.898	1×10^{-5}
Speed	**0.91**	0.85 - 0.97	1×10^{-10}
Closing phase			
Acceleration	0.70	0.59 - 0.81	0.02
Deceleration	0.81	0.71 - 0.90	2×10^{-6}
Speed	0.87	0.79 - 0.95	1×10^{-8}

Abbreviations: AUC, area under ROC curve; CI, confidence interval

The CGP classifier expression (Figure 44) suggests that the relative sizes, or ratios, of certain aspects of the tap cycle are important in classifier formation i.e. whichever is the greater of window offsets 12 and 5 (that possibly correspond to the maximum separation and opening acceleration in Figure 47) is divided by whichever is greater of window offsets 16 and 17 (that seem to correspond to the closing acceleration/deceleration in Figure 47). In view of this several ratios of opening and closing acceleration and deceleration measurements were also calculated.

Table 26 shows that the ratio values are generally no better than the individual component measures of the opening and closing phase though. Nevertheless, plots of the ratio of closing acceleration to closing deceleration showed that PwPD tended to have a closing acceleration that was larger than the closing deceleration (ratio > 1) whereas HCs tended to have a closing deceleration that was greater than the closing acceleration (ratio < 1). Also plots of the ratio of maximum separation acceleration to the maximum opening acceleration showed that the ratio was > 1 in patients but < 1 in controls. In other words in controls there was more acceleration in the opening phase of the cycle than when the fingers were maximally separated (and had very little acceleration) but this pattern was lost in PwPD who had

bursts of acceleration even when the fingers were stationary in maximum separation. Indeed visual inspection of the high scoring CGP windows of PD and HC data showed that patients seem to move from opening to closing in a sinusoidal manner whereas controls have defined bursts of acceleration and deceleration followed by zero acceleration (at maximum and minimum opposition). Possible explanations for these results are discussed in section 5.6.

Table 26 Classification accuracy of acceleration ratio measurements

Ratio calculated	**AUC**	**95% CI**	***p***
Op accel : Cl accel	0.67	0.52 - 076	0.03
Op decel : Cl decel	0.65	0.53 - 0.77	0.02
Op accel : Op decel	0.52	0.35 - 0.61	0.75
Cl accel : Cl decel	**0.77**	0.64 - 0.87	< 0.0001
Max sep accel : Min sep accel	0.57	0.43 – 0.70	0.31
Max sep accel : Op accel	**0.75**	0.64 – 0.86	< 0.0001

Abbreviations: Op, opening phase; Cl, closing phase; accel, maximum acceleration; decel, maximum deceleration; max sep, maximum digit separation (i.e. fingers open); min sep, minimum separation (i.e. fingers opposed).

In view of these observations the percentage duration of each tap cycle that was spent in the opening phase, the closing phase, in opposition, in acceleration, in deceleration and with zero acceleration were calculated (Table 27). These results show that percentage opposition was a fairly good discriminator with AUC 0.79 (p = 0.001). Further analyses showed that PwPD spent a greater part of the FT cycle, (mean % ± SD) 12.1% ± 11.6 with the fingers opposed than HCs, 6.0% ± 4.1, p = 0.02. The excess opposition in patients was not associated with less acceleration though; in fact PwPD had a smaller percentage of the tap cycle with zero acceleration (41% in patients vs. 45% in controls) and these results are discussed further in section 5.6.

Table 27 Percentage of each tap cycle spent in opening, closing and opposition phase

FT component	AUC	95% CI	p
% opposition	**0.79**	0.61 – 0.83	0.001
% opening phase	0.62	0.55 – 0.74	0.06
% closing phase	0.62	0.50 – 0.74	0.06
% acceleration	0.66	0.54 – 0.78	0.02
% deceleration	0.66	0.54 - 0.78	0.02
% zero acceleration	0.69	0.55 – 0.72	0.01

Visual inspection showed that PwPD had several small accelerations in each opening and closing phase but HCs tended to have just one. Thus the number of accelerations and number of decelerations in each of the opening and closing phases of the tap cycle were calculated but these measures were poor discriminators with AUCs in range of 0.51- 0.61 (all *p*s > 0.09).

These results suggest that in addition to the EA classifier being useful for diagnostic prediction it may also have a role in investigating the characteristics of bradykinetic movements. When the classifier expression is examined the key acceleration data points used in its formation are revealed. When these are overlaid on high scoring PD and HC kinematic data further exploration of potentially important components of bradykinesia becomes more focussed. This method has highlighted the acceleration profiles of PD depart further from HCs than the separation and speed profiles. The PD separation and speed profiles have smaller amplitude and slower speeds than HC but the shape of the kinematic data waveform is fairly similar to PD. In contrast when the acceleration profiles are examined they are totally different between PD and HC.

5.2.2 Examining the misclassified data sets

Examining the misclassified data sets may also provide useful information about what movement features best characterise PD bradykinesia. Comparison of the misclassified data sets with the correctly classified data sets, for each group of subjects, is another method of evaluating the key components of the EA classifier.

The ensemble classifier resulted in one out of 46 MA PD, and four out of 35 dominant hand HC, LTHT data sets being misclassified. The 'MA PD vs. Dom HC composite' model led to three out of 46 PD MA data sets and four out of 35 dominant hand HC data sets being misclassified. The misclassified subjects for each method of analysis are listed below, with the prefixes 'P' and 'C' denoting patient and control data respectively, and the subjects misclassified by both methods are highlighted in bold font:

Ensemble classifier analysis:	**P17**
	C3 C18 C34 C47
Bradykinesia composite analysis:	**P17** P23 P37
	C3 C18 C9 C30

Misclassified patients

The MA data set from P17 was the only PD data set to be misclassified by the EA classifier. It was also misclassified by the bradykinesia composite. This patient was a 59 year old man with HY stage 3 PD who had been diagnosed six years earlier. He was treated with Stalevo (levodopa-based drug) and an apomorphine pump with LEDD of 756 mg. Both hands' FT performances *on* were graded zero on MDS-UPDRS and he scored 26/30 on MoCA. The data set did not contain any corruption and the patient did not have any tremor or dyskinesia. Two further MA PD data sets were misclassified by the bradykinesia composite: P23 and P37. Table 28 summarises the details of the three MA PD data sets that were misclassified in comparison to the rest of the correctly classified MA PD data sets.

Table 28 Comparison of clinical and kinematic details of correctly classified and misclassified MA PD data sets

	Correctly classified MA PD data[1]	Misclassified MA PD data sets		
		P17	P23	P37
EA classifier score[2]	+ 0.39 ± 0.21	- 0.34*	+ 0.32	+ 0.20
Composite score[3]	0.13 ± 0.20	0.99*	0.51*	0.77*
Age, years	68.3 ± 8.4	59	64	75
Disease duration, yrs	6.01 ± 4.0	6	5	5
HY stage	2.44 ± 0.69	3	2	3
FT score	1.59 ± 0.95	0	1	1
LEDD, mg	695 ± 445	756	1030	1148
Amplitude	0.33 ± 0.16	0.85*	0.30	0.50
(raw, mm)	(43.1 ± 36.5)	(114.8)	(43.8)	(64.6)
Speed	3.45 ± 1.62	6.71	3.29	6.55
(raw, mm/sec)	(692 ± 548)	(1366)	(748)	(1057)
Rhythm (COV speed)	0.41 ± 0.18	0.08*	0.59	0.17
Halts (%)	0.22 ± 0.23	0.06	0.12	0.04
Frequency (Hz)	2.77 ± 0.81	1.87	3.4	4.1
Amplitude decrement	- 43.1 ± 98.4	- 72.9	- 103.8	- 98.1
Speed decrement	- 5.21 ± 9.72	- 19.9	- 12.6	- 17.8

Results are mean ± 1 SD. [1]This group comprises all MA PD data sets that were correctly classified by EA and component analysis methods. [2]EA classifier score ranges from -1 to + 1 with scores > zero predicting PD and scores < zero predicting control; [3] Bradykinesia composite score ranges from zero to one with scores < 0.5 predicting PD; [*]significantly ($p < 0.05$) different to correctly classified patient data

Regarding the patient data sets misclassified by the bradykinesia composite P23 was a 64 year-old man with HY stage 2 PD who had been diagnosed five years earlier. His FT performance was graded one on MDS-UPDRS, his MoCA score was 29/30, and he was treated with Stalevo, ropinirole and selegiline with a LEDD of 1030 mg. There was no tremor or LID clinically

and no corruption in the data. P37 was a 75 year old man who had been diagnosed with PD 5 years earlier, had HY stage 3 disease, UPDRS FT grade one and MoCA score of 30/30. He was treated with co-careldopa, entacapone and amantadine, with a LEDD 1148 mg. Table 28 shows that the age, disease duration, FT grade, LEDD and HY stage were not significantly different between the correctly classified and misclassified MA PD data sets (all $p > 0.1$). P17 had larger raw amplitudes ($p = 0.06$), and less variation in rhythm ($p = 0.07$), than the correctly classified MA PD data but otherwise had similar component measurements. There was no difference between the kinematic measures in P23 and P37 compared to the correctly classified PD data.

Misclassified controls

Dominant hand control data from C3, a 66 year old man, and C18, an 82 year old woman, were misclassified by the classifier *and* the composite model. C34, a 35 year old woman, and C47, a 68 year old man, were misclassified by the EA classifier but correctly classified by the composite. Conversely C9 and C30 were misclassified by the bradykinesia composite but not by EA classifier. C9 was from a 69 year old man and C30 a 75 year old man. All misclassified HCs were right handed. The kinematic data of the misclassified dominant hand HC data sets were compared to the dominant hand HC data sets that were correctly classified by both methods in Table 29. This showed that C3, C9 and C30 had smaller raw amplitudes ($ps < 0.01$) and C3, C47 and C30 had slower raw speeds ($p < 0.05$). C18, C34, C9 and C30 had less rhythmic FT than the rest of the control group ($p < 0.02$) and C18, C9 and C30 had more halts ($p < 0.01$). C3 and C47 had significantly higher tap frequency than rest of control group ($p < 0.02$). C47 had a greater degree of amplitude decrement ($p = 0.03$) but there was no difference between speed decrements in the misclassified data and the correctly classified data.

Table 29 Comparison of correctly classified and misclassified dominant hand control data

	Correctly classified control data	Misclassified					
		EA & composite		EA		Composite	
		C3	C18	C34	C47	C9	C30
EA score[1]	- 0.4 ± 0.2	+ 0.07	+ 0.01	+ 0.2	+ 0.05	- 0.002	- 0.7
Amplitude	0.3 ± 0.2	0.25	0.76	0.42	0.39	0.59	0.23
(raw)	(92 ± 19)	(26*)	(80.9)	(59.9)	(38.9)	(86.6)	(27.8)
Speed	6.0 ± 0.7	4.0	5.9	3.9	4.9	5.2	3.2
(raw)	(1546± 400)	(517*)	(882)	(876)	(749*)	(1233)	(483*)
Rhythm	0.2 ± 0.07	0.26	0.32*	0.33*	0.14	0.32*	0.64*
Halts	0.09 ± 0.05	0.07	0.23*	0.07	0.05	0.25*	0.21*
Frequency	2.3 ± 0.7	4.6*	1.4	3.3	4.0*	1.6	3.3
Ampl Dec.	-82.9 ± 69	- 50.0	+18.1	+ 1.5	- 243*	- 22.7	- 88.4
Speed Dec.	-6.8 ± 10.3	- 4.9	+ 3.1	+ 6.8	- 12.3	- 0.98	- 10.8

[1]EA classifier score ranges from -1 to + 1 with scores > zero predicting PD and < zero predicting HC; *significantly ($p < 0.05$) different to correctly classified HC data; rhythm is COV maximum speed. Dec. is decrement.

Summary of section 5.2.2 results

In summary closer examination of the misclassified data confirms that amplitude, speed, rhythm and halts are the most discriminatory measurements of FT used by both methods of classification. Control data with smaller amplitude, slower speed, less rhythm or more halts are more likely to be misclassified and this is not surprising as these features are characteristic of PD movements. On the whole, the decrement measurements did not differ between correctly classified and misclassified HC data sets. MA hand data from P23 was borderline for misclassification with a composite prediction of 0.51 (score > 0.50 predict control) but it is not entirely clear why the bradykinesia composite model misclassified P37 as there was no significant difference in any of the component measures of this data set when compared to the correctly classified data group. It is noteworthy that the data sets misclassified by the EA tended to have only 'weakly PD' scores i.e. scores very near to zero (+ 0.01, + 0.05, + 0.07 and

+ 0.15) suggesting that the classifier 'recognised' that the kinematic profiles of these data sets was not entirely typical for patients either.

Non-dominant hand HC data

Dominant HC data was much more likely to be correctly classified by the ensemble classifier than non-dominant hand data. This suggests that there are features in the dominant hand data that are over-represented in HCs, relative to PwPD, and are used as components of the classifier equation. This was explored further by comparing the mean EA and component scores for dominant and non-dominant hands of HCs to find out what these differentiating components might be.

This showed that in the control group the classifier scores (mean ± 1 SD) of dominant and non-dominant hands were - 0.31 ± 0.25 and - 0.01 ± 0.32 respectively, $p = 0.003$ (where negative scores predict the data came from a control). The patient scores did not significantly differ between dominant and non-dominant hands ($p = 0.14$). When the separable component measures of bradykinesia were compared between HC dominant and non-dominant hands there was no difference for amplitude, speed, halts, rhythm, tapping frequency, decrements, opening and closing speed/acceleration or any of the other components listed in Tables 25-27. Only raw speed showed a trend ($p = 0.057$) towards being different in the dominant and non-dominant hands of controls and this was in an unexpected manner with the dominant hand having a slower mean speed of FT than the non-dominant hand (1423 mm/sec vs. 1650mm/sec). In summary it remains unclear why HC dominant hand data is more likely to be correctly classified from this analysis.

5.3 Discriminating clinically slight bradykinesia from normal movement

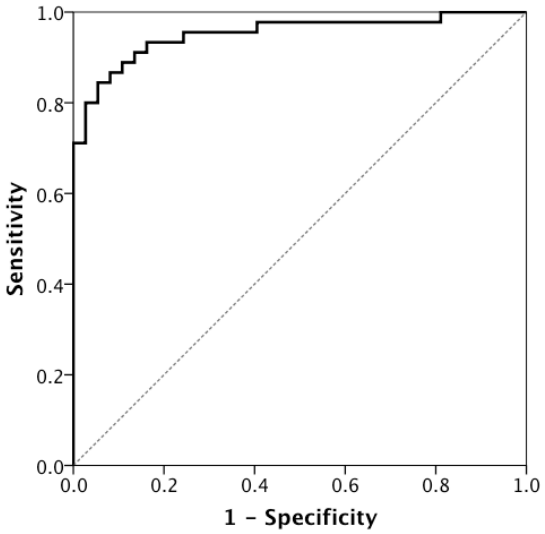
So far it has been shown that the optimal diagnostic accuracies for discriminating MA PD (mean UPDRS FT grade 1.57 ± 0.94) from dominant HC data (all had UPDRS FT grade zero) were 0.99 and 0.97 for the EA ensemble classifier and the bradykinesia composite respectively. However

these results are for discriminating a heterogenous group of PwPD from controls. Whilst most of the patients had only clinically mild bradykinesia (Figure 28) some had more severe grades so the classification results are likely to be an over-estimate of how well the device would discriminate kinematic data from patients with mild bradykinesia in clinically indeterminate cases. Assessing how well the device differentiates a sub-group of PD patients with clinically slight bradykinesia (MDS-UPDRS grade one) from HCs better reflects how the device might perform as a diagnostic device, albeit with the caveats that all the patients still had clinically definite PD and were on treatment.

Classification accuracy of ensemble classifier for slight bradykinesia

Forty-five FT assessments performed by LTHT patients had 'slight bradykinesia' and the EA ensemble classifier (that was trained on the whole data set) was assessed to see how well it could discriminate these patient data sets from the dominant hand HC data; see Figure 48.

Figure 48 Accuracy of ensemble classifier for differentiating 'slight' bradykinesia from control data



Legend: ROC curve with AUC 0.959 (CI 0.907- 0.997 $p < 0.0001$) for discriminating grade one MDS-UPDRS PD and grade zero HC dominant hand data. The reference level of a test with 50% sensitivity and 50% specificity is shown as a dashed line.

The AUC of 0.959 denotes excellent classification accuracy and is equivalent to 41 out of 45 (91%) of grade one PD data sets, and 38 out of 41 (93%) HC data sets being accurately classified at the threshold point of equal trade off.

Validation of slight bradykinesia classification accuracy on UCSF data

Fourteen out of 16 (88%) grade one PD FT data sets, and five out of nine (56%) of control dominant hand data sets from UCSF were correctly classified by the ensemble classifier. This suggests that the EA ensemble classifier generalises beyond the Leeds cohort to an independent patient data set, but not so for control data. However in both groups, but especially the control group, there are only small numbers of subjects and further validation on larger independent sets is required.

The patient data that was included in the training and test sets for the ensemble classifier was from patients in an *on* state. When the UCSF grade one data was split according to clinical state it was encouraging that the classifier seems to generalise beyond the *on* clinical state as nine out of 11 (82%) UCSF *on*, and five out of five (100%) UCSF *off* data sets were correctly classified (Table 30). This suggests that the clinical status of the patient probably does not interfere with the diagnostic accuracy of the EA classifier. If this were the case it would allow for a more flexible use of the test i.e. patients may be tested whether on or off medications.

Classification accuracy of bradykinesia composite model for slight bradykinesia

When classifying grade one PD data from dominant hand HC data the bradykinesia composite model comprising all the FT variables (and developed using all PD data and dominant hand HC data had an AUC of 0.930 (95% CI 0.87 - 0.98; $p < 0.0001$) which was numerically slightly less accurate than the ensemble classifier ($p = 0.7$); see Table 31.

Table 30 Classification accuracy of ensemble classifier for UCSF 'slight' bradykinesia data

UCSF patient code and hand		Clinical state	
		ON	OFF
10975	Dominant	✔	-
	Non-dominant	✖	-
14713	Dominant	✔	-
15105	Dominant	✔	-
	Non-dominant	✖	-
15180	Non-dominant	✔	✔
15235	Non-dominant	✔	-
7662	Dominant	✔	✔
8560	Dominant	✔	-
14312	Dominant	-	✔
15431	Dominant	✔	-
15431	Non-dominant	-	✔
15557	Dominant	✔	✔

Legend: All the UCSF PD data sets from FT assessments that were clinically graded as slight bradykinesia (grade one MDS-UPDRS) are presented with ticks and crosses denoting that the ensemble classifier correctly, or incorrectly, classified the data respectively. A dash denotes that the corresponding *on/off* data received a grade other than 'slight' and hence these are not included in the analysis.

Table 31 Accuracy of bradykinesia composite model and ensemble classifier for discriminating slight bradykinesia from control data

	AUC (95% CI)	Correctly classified data PD	HC
Ensemble classifier	0.959	91 %	93 %
	(0.92 – 1.0)	41/ 45	38/ 41
Bradykinesia composite	0.930	93 %	86 %
	(0.87 – 0.98)	42/ 45	32/ 37

Legend: The PD group comprised all data sets allocated grade one FT scores and the HC group comprised all dominant hand data. The denominators are different for the HC data because four of the dominant hand data sets were corrupted so could not be used for the composite analysis.

5.4 Comparing kinematic data with demographic and clinical variables

It would be useful to know how the individual component measures of bradykinesia, the bradykinesia composite model, and the EA ensemble classifier scores vary according to demographic and clinical variables. This information may aid the development of a diagnostic device for two main reasons; firstly so that in addition to making a prediction about whether the data is from a patient or a control, a prediction of disease stage/duration and FT grade could also potentially be made, and secondly so that the information about the association between the variables and the objective measures could be taken into account (i.e. adjust the scores) to further improve diagnostic classification if appropriate.

The results for the separable components, and the composite models were presented in section 4.5. In summary this showed that HY stage, disease duration, and LEDD were not significantly correlated with any of the

individual component measures or the composite model. All the individual components of bradykinesia except the decrement measurements correlated with the UPDRS FT grade but only amplitude showed a significant difference between patients and controls with grade zero UPDRS.

Ensemble classifier correlations

Spearman's correlation coefficient, r_s, was used to examine the association between EA ensemble classifier scores (of approved data) and gender, age, MoCA, dominant vs. non-dominant hand, and handedness of LTHT patients and controls separately. PD disease duration, HY stage, LEDD, and UPDRS FT grade were also correlated in the patient data only.

Ensemble classifier score correlations in the control group

The ensemble classifier score was significantly associated with dominant hand (r_s - 0.32; $p = 0.005$) in the control group but not with any of the other variables (all $ps > 0.25$). This suggests that dominant hand data tends to receive higher EA scores than the non-dominant hand data and is in line with the previous classification results that showed the dominant hand HC data is less likely to be misclassified.

Ensemble classifier score correlations in the patient group

Table 32 shows that for patients, the only variable that significantly correlated with the ensemble classifier score was UPDRS FT grade (r_s + 0.43, $p < 0.0001$) and all other variables had a p value > 0.37.

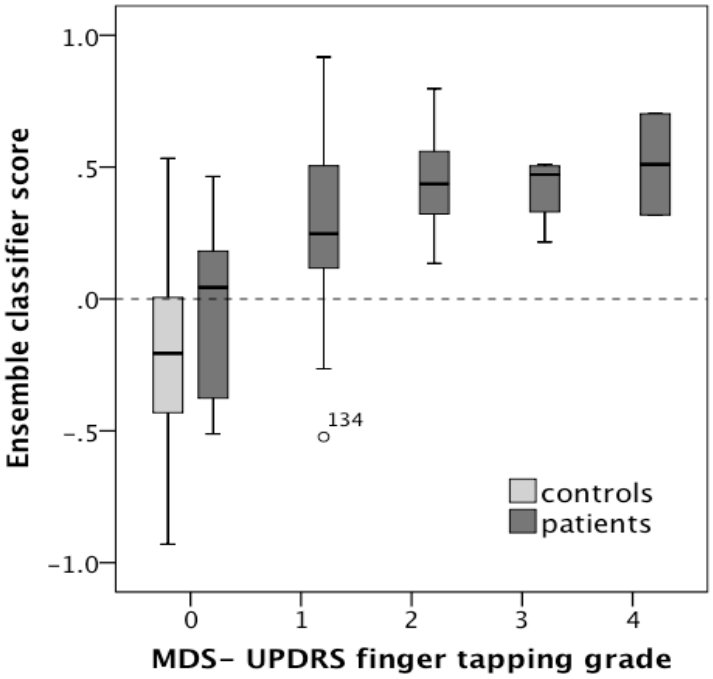
Table 32 Correlogram of ensemble classifier scores and demographic and clinical variables in Leeds patients

	Gender	Age	Handedness	HY	Duration	MoCA	LEDD	UPDRS	EA score
Gender		.22*	-.10	.05	-.15	-.09	-.17	.006	.04
Age	.22*		-.02	.35*	.04	-.13	.26*	.06	-.04
Handedness	-.10	-.03		.18	.32*	-.09	.16	.02	-.04
HY	.05	.35*	.18		.35*	-.24*	.33*	.20*	-.05
Duration	-.15	.04	.32*	.35*		-.10	.24*	-.03	-.03
MoCA	-.09	-.13	-.09	-.24*	-.10		.09	-.05	-.10
LEDD	-.17	.26*	.16	.33*	.24*	.09		.05	-.08
UPDRS	.006	.06	.02	.20*	-.03	-.05	.05		.43*
EA	.04	-.04	-.04	-.05	-.03	-.10	-.08	.43*	

Legend: *significant at <0.05. Abbreviations: Dom hand, dominant hand; HY, Hoehn and Yahr stage; Duration, PD duration since diagnosis; MoCA, Montreal Cognitive Assessment; LEDD, levodopa equivalent daily dose; UPDRS, Unified Parkinson's Disease Rating Scale finger tapping score; EA, evolutionary algorithm induced ensemble classifier score. NB the dominant vs. dominant hand variable has been removed from the table as it did not correlate significantly with any of the other variables.

Figure 49 demonstrates that there is a trend for patients with higher clinical grades of bradykinesia to have a higher ensemble classifier scores. This is an important point because it suggests that the classifier score not only predicts whether data came from a PwPD or HC but also reflects the clinical severity of bradykinesia. This is an expected finding when one considers how the CGP classifier expression was developed i.e. by searching for kinematic features that discriminate PD from HC. Hence patients with more severe bradykinesia would intuitively have a greater degree of such discriminating features.

Figure 49 Ensemble classifier scores for each UPDRS finger tapping grade



Legend: All PD and dominant HC data sets are presented. The horizontal dashed line at zero marks the decision threshold of the classifier solution. If the classifier score is > zero the data will be classified as PD and if the score is < zero it will be classified as HC. This figure shows that the majority of misclassified patients had a UPDRS FT grades of zero. It also shows that classifier solutions for grade zero PwPD are significantly higher than for grade zero controls.

Mean EA scores for HC dominant hand data were significantly lower at -0.31 ± 0.25 than those for patients with UPDRS grade zero at - 0.04 ± 0.35, p = 0.007. This was a similar to the difference between the bradykinesia composite scores in grade zero PD (0.51± 0.33) and dominant hand HC data sets (0.23 ± 0.27), p = 0.002. Amplitude was the only individual component to show a significant difference between the PD and HC group data with grade zero, p = 0.04.

5.5 Discussion

5.5.1 Summary of classification accuracy results

The classification accuracy improved when the ABN and CGP classifiers were combined into an ensemble classifier. The likely explanation is that both the subtle movement abnormalities over each tap cycle (focused on by the CGP) and the overall patterns of movement during the 30 second recording (focused on by the ABN) were being evaluated. In contrast to the separable component and bradykinesia composite measures, the ensemble classifier accuracy improved when HC data was limited to just the dominant hand (AUC 0.88 to 0.94). This suggests that movement features specific to controls are over-represented in the dominant hand data and are used to form the classifier. However the analyses undertaken in sections 5.1-5.3 did not reveal what these are. It may be that further exploration of the ABN classifier expression would reveal such features but this is beyond the scope of the thesis. The behaviour of the ABN classifier is considered more difficult to understand but Lones et al. demonstrated that the ABN classifier seems to respond to an association between amplitude and frequency with small amplitude and low frequency movements more likely to be classified as patient data (Lones, Smith et al. 2013).

Both the ensemble classifier and bradykinesia composite model discriminated PD and HC data better when PD data was limited to the MA hand. This is expected because the MA PD data group had more severe bradykinesia (UPDRS grade 1.57 ± 0.94) than the PD group comprising data from both hands (1.32 ± 0.84). The best bradykinesia composite model and the ensemble classifier discriminated PD MA and HC dominant hand

data with AUCs of 0.97 and 0.99 respectively (p = 0.081). This equates to three misclassified patients (out of a total of 46) and four misclassified controls (out of a total of 37) with the composite model and one misclassified patient and four misclassified controls with the ensemble classifier.

Control data sets with smaller amplitude, slower speed, less rhythm or more halts were more likely to be misclassified and this is probably because these features are characteristic of PD bradykinesia. On the whole the decrement measurements did not differ between correctly classified and misclassified data sets. It is noteworthy that control data sets misclassified by the ensemble classifier tended to have only 'weak PD' scores i.e. scores much nearer to zero than one (+0.01, +0.05, +0.07 and + 0.15) suggesting that the classifier 'recognised' the kinematic profiles of these data sets were not entirely typical for PD either.

Section 5.1.4 showed that data corruption does not affect the ensemble classifier's accuracy. The overall prediction of diagnostic group is based on a combination of the average score of the taps (CGP classifier) and general trends in the data (ABN classifier) over a 30 second period, so even if some taps are misclassified due to corruption this will have minimal effect on the overall result assuming that the rest of the recording is correctly classified. Hence the presence of corruption does not necessitate exclusion of the whole data set from analysis and this is a major advantage over the separable component analysis technique.

The EA classifier was validated on UCSF MA PD and dominant hand HC data; this showed that classification accuracies were similar between the centres (p = 0.48 for patients, p = 0.09 for controls). This implies that the classifiers developed using LTHT data generalise beyond this population and can accurately classify independently collected data too. This result is paramount to the development of a device that aims to aid clinical diagnosis and these results lend strong support to the use of EA induced classifiers for discriminating patient and control kinematic data.

Additionally the ensemble classifier correctly classified all UCSF PD MA data sets in both *on* and *off* states. When all UCSF PD data sets were included 96% (25 / 26) of PD *off*, and 92% (23 / 25) of PD *on* data sets were correctly classified. These results are very encouraging and suggest that a device employing EA classifiers can predict diagnostic group regardless of the patient's clinical status. This would enable flexibility to the investigation protocol as patients could be tested without having to alter their medication regimen.

Taken together these results suggest that diagnostic accuracy of the FT device would be optimised of patient test data is limited to the MA hand and then compared to a database of dominant HC hand data using classifiers induced through EA analysis. Specifically evaluating the most subtle clinical grade of bradykinsia the ensemble classifier accurately discriminated HC FT movements from UPDRS grade one PD FT movements with an AUC of 0.959. This was equivalent to four (out of 45) PD data sets and three (out of 41) HC data sets being misclassified. These results suggest potential for the device to aid early clinical diagnosis as it was able to identify PD data even when the clinical signs are very subtle.

5.5.2 Correlation of objective and clinical measures of bradykinesia

Both approaches to objectively measuring bradykinesia - separable component measures of bradykinesia and EA induced classifiers - were significantly correlated with the UPDRS FT grade. This shows that the objective measures of FT not only classify the data but also reflect the clinical severity of bradykinesia. All the individual component measures of bradykinesia (except decrement measurements) correlated with the UPDRS FT grade. Only amplitude showed a significant difference between HCs and PwPD *on* and UPDRS grade zero though, $p = 0.04$. The bradykinesia composite model scores and the EA classifier scores were correlated with the UPDRS FT grades ($r_s + 0.60$ and $r_s + 0.43$ respectively, both p's < 0.0001) and there was a significant difference in the scores between grade zero patients and controls ($p = 0.002$ for composite model, $p = 0.007$ for EA classifier).

None of the objective measures of bradykinesia were significantly correlated with age, HY stage, disease duration or LEDD. MoCA was inversely correlated with the bradykinesia composite score, and the individual rhythm measures suggesting that cognitive problems are associated with more bradykinesia and less rhythm. There was no association between MoCA score and ensemble classifier score and arguably this may be considered advantageous because classification based on EA analysis may not need to be adjusted according to the cognitive score of the test subject.

5.5.3 Understanding the pathophysiology of bradykinesia

There remains an incomplete understanding of the pathophysiology of bradykinesia in PD but the present study results are now discussed in the context of the medical literature. The EA classifier was used as a tool to investigate the most discriminating features of PD movements. The acceleration profiles were focussed upon as normalised acceleration data had been used to induce the ensemble classifier. Examination of the high scoring PD and HC data windows showed that controls had a characteristic pattern of a sharp burst of acceleration in the opening phase of the FT cycle followed by almost no acceleration when the fingers were maximally separated and were momentarily static. This pattern was lost in the PD data which had many small bursts of acceleration throughout the FT cycle and even when the fingers were stationary at the point of maximum separation. This qualitative observation was confirmed by plotting the ratio of maximum opening acceleration to maximum separation acceleration and finding that controls had a ratio > 1 but patients had a ratio of < 1. Examination of the closing phase of the tap cycle showed that controls had a sharp peak of acceleration and then deceleration, with the magnitude of the acceleration peak being smaller than the deceleration peak (ratio < 1). In contrast PwPD had a smaller magnitude and more dispersed pattern of acceleration and deceleration in the closing phase with the magnitude of each peak fairly similar (ratio of approximately 1). This suggests that after the point of maximum separation controls accelerate their finger and thumb

rapidly in the closing phase with a subsequent sudden large deceleration as the finger 'crashes into' or opposes the thumb at high speed. This contrasts with the pattern seen in PwPD where there is a gradual 'drifting' of the digits towards opposition, driven by lots of small acceleration bursts that 'nudge' the digits towards their target.

These characteristic findings on the high scoring kinematic profiles identified by the ensemble classifier are consistent with previous EMG studies that have demonstrated abnormal muscle activity in PD. In healthy subjects there is a characteristic tri-phasic EMG pattern seen with single rapid movements such as finger extension: first there is a large agonist muscle burst to accelerate the finger to its peak speed, followed by a smaller antagonist burst to slow it down as it approaches its target, and this is followed by a final smaller second agonist burst to bring the finger to is final position without any oscillation (Hallett, Shahani et al., 1975). In PD it has been shown that there is gross departure from this normal tri-phasic EMG pattern with several small agonist bursts occurring before a lower peak speed is reached and then de-synchronisation of the antagonist bursts, leading to excessive periods of co-contraction of agonist and antagonist muscles and a jerky acceleration profile (Hallett and Khoshbin, 1980). Dopaminergic medication only partly normalises the acceleration and speed profiles in PD (Vaillancourt, Prodoehl et al., 2004, Vaillancourt, Prodoehl et al., 2006) so these features highlighted in the present study may be even more discriminatory when patients are tested in the *off* state; indeed there was a suggestion of this in the small UCSF PD *off* sample.

Further exploration of the CGP classifier's behaviour showed that the percentage duration of each tap cycle spent with the digits opposed (minimum separation) discriminated the PD and HC data fairly well (AUC 0.72, $p = 0.001$). Patients spent a greater part of the whole tap cycle in opposition (12.1 % ± 11.6) than HCs did (6.0 % ± 4.1), $p = 0.02$. This probably reflects of a form of akinesia where there is difficulty switching from one task (flexion of the fingers during the closing phase) to another task (extension of the fingers during the opening phase). This 'switching

time' akinesia, or pause between two different movements, was described by Homann et al. in a keyboard tapping study to differentiate it from the more typical 'initiation' akinesia of PD (Homann, Suppan et al., 2000). Homann et al. measured how long each key was depressed and concluded that the prolonged duration of key depression between alternating key taps was due to "*an inability to execute automatically learned motor plans.*" Similarly Beradelli et al. found that when PwPD drew geometric shapes they tended to pause for longer periods than controls at the vertices and the authors speculated that this was due to PD patients having difficulty running two motor programs concurrently (Berardelli, Accornero et al., 1986). Sheridan et al's study using surface EMG recordings in six PwPD showed that reaction time has a premotor and motor phase. The premotor phase before any EMG activity was recorded was prolonged in PD but the motor phase was almost the same in PD as controls (Sheridan, Flowers et al., 1987).

The increased opposition period demonstrated in the present study was not associated with more periods of zero acceleration; in fact patients had a smaller percentage of the FT cycle with zero acceleration than controls (41% in patients vs. 45% in controls) and even when there was no net movement according to the separation profile there were many small bursts of acceleration and deceleration being recorded. These results taken together in the context of the literature suggest that the 'wrong' motor programs are being planned or selected for the desired movement and the subsequent small acceleration bursts are not adequate to move the fingers out of the opposition position so there is increased akinesia.

This study found that PD FT movements are characterised by a sinusoidal separation profile of gradual opening and closing movements. These 'drifting' movements of the digits are associated with small and dispersed peaks of speed during the opening and closing phases and are driven by multiple small bursts of acceleration that persist throughout the FT cycle. This contrasts sharply with HC speed and acceleration data profiles where there are brief sharp spikes of speed at the beginning of each opening and

closing phase with each spike driven by a defined coordinated burst of acceleration and deceleration. In between each opening or closing movement there is a period with no acceleration when the digits are at maximum or minimum separation. This cycle of brief large acceleration/deceleration followed by zero acceleration then another brief acceleration/deceleration repeats as the FT cycle proceeds resulting in a coordinated pattern of quick short opening phase, a brief pause at maximum separation then a quick short closing phase followed by another brief pause at minimum separation. In other words, whilst the shape of the separation data waveform in PD may appear fairly similar to HCs, albeit with smaller magnitudes, the speed and especially the acceleration data that are driving this separation profile differ markedly between patients and controls. These movement characteristics fit with one particular theory of bradykinesia outlined by Mark Hallett in 2011 (Hallett, 2011). He describes how in PD there is:

"*a loss of motor energy – movements are not given the full motor command that they require*"

and that

"*larger movements should be faster, but patients tend to have the same velocity for all movements. This requires more time to accomplish the movement."*

(Hallett, 2011).

The smaller bursts of EMG activity (Hallett and Khoshbin, 1980) result in 'underscaled' acceleration and speed which lead to a reduction in amplitude of movement. It has been shown that PwPD have the ability to make faster or bigger movements in response to cueing (Griffin, Greenlaw et al., 2011, Oliveira, Gurd et al., 1997), attention (Oliveira, Gurd et al., 1997) or emotional excitement (Bonanni, Thomas et al., 2010). However under normal circumstances when there is no visual cue or specified target the motor program selected by PwPD results in muscle acceleration that is too

small in magnitude for the task. This means the movements are not only slower but also smaller – both due to inadequate first burst of agonist muscle contraction. However if there is a target this may be reached by adding on several extra small acceleration bursts of muscle activity that repeatedly 'nudge' the limb along towards the desired position. It is likely that the FT task results in a combination of these two patterns of movement – there is no set target for the opening phase so the maximum speed and amplitude of digit separation are both reduced whereas there *is* a target in the closing phase (i.e. digit opposition) so additional acceleration bursts are added on until the digits are fully opposed.

This may partly explain why the opening phase speed was the most discriminatory measure of the FT cycle in the present study – as there is no target for the opening phase the fingers separate largely dependent on the first agonist muscle burst and there is not the 'benefit' of the additional acceleration bursts that are necessary in the closing phase to bring the digits to their target of complete opposition. Only two previous studies have measured the opening and closing phases of FT in PD but both found that the opening speed was disproportionately slower than the closing speed when compared to HCs (Ling, Massey et al., 2012, Yokoe, Okuno et al., 2009).

Whilst it is clear that PD EMG patterns are abnormal it remains less clear why the wrong motor program that produces this EMG pattern is selected in the first place. 'Internally driven' movements are much more impaired in PD than 'externally driven' movements; in other words movements that are initiated by the individual (internal) are slower and smaller than when the movements are made in response to cueing or with conscious attention (external) (Oliveira, Gurd et al., 1997). For example Oliveira et al. demonstrated that PwPD have much less micrographia if they are reminded to write bigger or are given visual targets (Oliveira, Gurd et al., 1997) and conversely if their attention is drawn away from the task there is a further reduction in amplitude of movements (Oliveira, Gurd et al., 1998). Functional imaging studies have shown that in PD there is reduced SMA

activity and overactivity in the lateral premotor regions (Ikeda, Luders et al., 1992, Jahanshahi, Jenkins et al., 1995, Jenkins, Fernandez et al., 1992). The SMA is important for automatic movements whereas the premotor cortex is particularly responsive to external cueing. This suggests that motor control is being moved to the premotor areas to compensate for the underactive SMA. The downside of this compensation is that movements become less automatic and more reliant on external cues. This theory fits with the observation that PD movements are more impaired if the individual has to attend to two motor tasks (Benecke, Rothwell et al., 1986), or a motor and a cognitive task (Morris, Iansek et al., 1996, Oliveira, Gurd et al., 1998). The exact reason for the deficient SMA activity still remains indeterminate but is likely to be due to reduced dopaminergic stimulation leading to relative increased excitation of the indirect BG pathways and hence increased inhibitory activity from the BG projections via the thalamus to the SMA.

It seems that an increase in beta frequency (10-35 Hz) neuronal oscillations (detected by EEG) throughout the BG and cortex are also important in the pathophysiology of bradykinesia. These have been detected during human DBS surgery and also in animal models (Delong, 1990) but their role in bradykinesia remains incompletely understood. It is hypothesised that if there is synchronous oscillatory activity in a large population of neurons there may be a breakdown in the ability of individual neurons to process and relay specific information and thus to effectively control complex movements (Wichmann and Dostrovsky, 2011) which may lead to an underscaling of the motor program (Hallett, 2011).

5.5.4 Strengths and limitations of the study

The strengths and limitations of the methodology have been discussed in section 4.7.3. In addition there are some specific points to highlight about the EA classifier analysis. Firstly this study demonstrated that combining classifiers evolved based on different patterns of data enables both characteristics of individual tap cycles (akinesia, amplitude etc.) and of patterns of movement over time (changes in amplitude and tapping frequency etc.) to be assessed and hence the classification accuracy

improves. Few studies have used EA classifiers to analyse movement data in PD and these have tended to be have smaller numbers of subjects and only test patients *on*. Smith et al. used two different types of EA based on immune system interactions and CGP to classify drawing movements in 12 PD and ten HCs (Smith, Gaughan et al., 2007, Smith and Timmis, 2008) and Tsanas et al. used 'machine learning algorithms' to classify 33 PD patients and 10 controls using voice recording data (Tsanas, Little et al., 2012). In both studies the authors reported 99% accuracy of the classifiers induced but the results were not validated on independent data and the small sample size makes it difficult to assess how well the classifiers might generalise. The current study has demonstrated the potential benefits of using EA classifiers to discriminate PD and HC data and validated the results on independently collected patient data assessed in both *on* and *off* states. There has also been a detailed evaluation of how the classification accuracy varies according to dominant hand, corrupt data and severity of clinical bradykinesia.

The 'black box' behaviour of the CGP classifier could to an extent be examined to better understand what features of PD bradykinesia were most discriminatory but there still remains limited understanding of the classifier functions, especially the ABN. Nevertheless with better understanding of classifiers there remains the possibility of finding novel features of bradykinesia and these might lead to a more focussed clinical examination and even better understanding of the underlying pathophysiology.

Chapter 6
Investigating the sequence effect: Results and Discussion

6.1 Definition of the sequence effect

The SE describes a phenomenon whereby the amplitude and speed of repetitive voluntary movements progressively decrement and this clinical feature is included within both the diagnostic and monitoring definitions of PD bradykinesia (Gibb and Lees, 1988)(Goetz, Tilley et al., 2008). The gold standard definitions for the purposes of diagnosis and monitoring differ slightly though: for diagnosis the SE is a defining feature when ascertaining whether bradykinesia is present or not, whereas the *grade* of bradykinesia allocated during monitoring is a composite of all the components of bradykinesia and it is difficult to gauge how much the SE contributes to this grade.

The SE, which has also been called 'decrementing' (Bajaj, Gontu et al., 2010) and 'fatiguing' (Espay, Giuffrida et al., 2011), is considered to be a specific clinical attribute of PD (Abdo, van de Warrenburg et al., 2010, Berardelli, Rothwell et al., 2001, Evarts, Teravainen et al., 1981, Hallett and Khoshbin, 1980, Ling, Massey et al., 2012, Marsden, 1984). The SE has been shown to occur in PwPD when the same movement is repeated (Chee, Murphy et al., 2009, Espay, Giuffrida et al., 2011, Iansek, Huxham et al., 2006, Ling, Massey et al., 2012), when a series of different movements are made sequentially (Agostino, Berardelli et al., 1992a, Agostino, Berardelli et al., 1994) and during well-learned movements such as walking (Chee, Murphy et al., 2009, Iansek, Huxham et al., 2006), handwriting (Ling, Massey et al., 2012), ocular pursuit (Lekwuwa, Barnes et al., 1999) and grasping (Bennett, Marchetti et al., 1995). Whether the SE is present in *all* PwPD is less clear though.

The SE has important clinical implications – it is the only component of bradykinesia not to improve with levodopa (Espay, Giuffrida et al., 2011, Iansek, Huxham et al., 2006, Kang, Wasaka et al., 2010), dopamine agonists (Espay, Giuffrida et al., 2011, Iansek, Huxham et al., 2006) or transcranial

magnetic stimulation (Kang, Wasaka et al., 2010) and is strongly associated with freezing of gait (FOG) (Chee, Murphy et al., 2009, Iansek, Huxham et al., 2006)

6.2 Why investigate the sequence effect?

Despite numerous studies the pathophysiology of the SE remains poorly understood and it is unclear whether it is definitely specific to PD. This may be in part because the studies examining the SE have tended to be small, with most including no more than ten PD subjects (Agostino, Berardelli et al., 1994, Benecke, Rothwell et al., 1987, Bennett, Marchetti et al., 1995, Connor and Abbs, 1991, Iansek, Huxham et al., 2006, Lekwuwa, Barnes et al., 1999, Longstaff, Mahant et al., 2003, Plotnik, Flash et al., 1998), some have not had a control group (Espay, Giuffrida et al., 2011, Iansek, Huxham et al., 2006, Kang, Wasaka et al., 2010) and some have assessed only the dominant hand (Agostino, Berardelli et al., 1998, Agostino, Curra et al., 2003, Kang, Wasaka et al., 2010). The studies have also differed markedly in the length of sequence assessed and Ling *et al*. highlighted the important point that it is still unclear how many repetitive movements are needed for the SE to manifest (Ling, Massey et al., 2012).

The studies also used different methods of quantification and varied in terms of how the SE was defined with some assessing decrementing speed (Agostino, Berardelli et al., 1992b, Agostino, Berardelli et al., 1994, Benecke, Rothwell et al., 1987, Bennett, Marchetti et al., 1995, Berardelli, Accornero et al., 1986, Kang, Wasaka et al., 2010, Plotnik, Flash et al., 1998), some decrementing amplitude (Chee, Murphy et al., 2009, Iansek, Huxham et al., 2006, Longstaff, Mahant et al., 2003) and others decrementing speed *and* amplitude (Agostino, Curra et al., 2003, Connor and Abbs, 1991, Espay, Giuffrida et al., 2011, Lekwuwa, Barnes et al., 1999, Ling, Massey et al., 2012)). Such variability of methodology makes it more difficult to compare the results.

The fundamental question of whether the SE is a defining characteristic of PD bradykinesia thus remains unanswered and there is considerable

evidence to suggest it may not be. Firstly, some studies found that the SE did not manifest in people with PD any more than in HCs (Connor and Abbs, 1991, Longstaff, Mahant et al., 2003). Secondly, several studies showed that the SE was limited to only a subset of PwPD (Bennett, Marchetti et al., 1995, Chee, Murphy et al., 2009, Kang, Wasaka et al., 2010, Wagle Shukla, Ounpraseuth et al., 2012). Thirdly there is evidence that the SE occurs in other movement disorders too with Michell et al. demonstrating that approximately half of Huntington's disease patients exhibited the SE during hand tapping (Michell, Goodman et al., 2008).

Conversely, there is evidence that the SE may be common in healthy adults too. Only one study specifically highlighted this observation (Plotnik, Flash et al., 1998) but examination of the control group results in some of the other studies that focussed on the SE in PD provides further supportive evidence (Chee, Murphy et al., 2009, Ling, Massey et al., 2012). For example Chee et al. showed that under certain gait conditions the HC group had decrementing stride length (amplitude) that was comparable to the PD *off* group without FOG (Chee, Murphy et al., 2009). Taken together these results raise the possibility that the SE may be a physiological phenomenon rather than a specific feature of PD bradykinesia.

It is of the utmost importance to clarify whether the SE is a defining characteristic of PD bradykinesia. The diagnosis of PD is based on clinical ascertainment of gold standard criteria defined bradykinesia and an inaccurate definition of bradykinesia may thus lead to mis-diagnosis. The aims of this section were to quantitatively evaluate the SE in order to examine whether the SE is a feature specific to PD bradykinesia and whether it occurs any more frequently or more severely in PD patients than in HCs.

6.3 Clinical and demographic details of participants

The methods used for measuring the components of bradykinesia using the movement sensor data have been described in section 3.5. All approved data sets from the 49 LTHT patients, 13 UCSF patients and nine UCSF controls were used in the analysis. In order to match the groups for age only data sets from 38 of the 41 LTHT HCs were used. The demographic and clinical details of the groups are summarised in Table 33.

Table 33 Demographic and clinical details for patient and age-matched controls included in sequence effect analysis

	Approved data sets		
	PD N = 62	**Controls N = 47**	***p***
Age, years	67.8 ± 8.3	65.6 ± 7.6	0.16
Gender, M : F	41: 21	18 : 29	0.004
Hand dominance, R : L	54 : 8	38 : 9	0.37
MoCA	26.4 ± 2.5	28.3 ± 1.4	<0.0001
HY stage	2.30 ± 0.72 (1 - 4)	NA	
PD duration, years	5.51±3.82 (0.5-18)	NA	
LEDD, mg	696.6 ± 418.4	NA	

Legend: Values stated as mean ±1 SD (range). Abbreviations: M, male; F, female; R, right; L, left; MoCA, Montreal Cognitive Assessment (maximum score is 30 with lower scores indicating more impairment); HY, Hoehn and Yahr stage; LEDD, levodopa equivalent daily dose calculated using standard conversion factors (Tomlinson, Stowe et al., 2010); NA, not applicable.

The groups were closely matched for age and hand dominance but not for gender with significantly more men in the patient group. Some kinematic studies have shown that men tend to perform tapping movements more quickly than women (Homann, Suppan et al., 2000, Pal, Lee et al., 2001)

but others have not found any difference (Giovannoni, van Schalkwyk et al., 1999). No association between gender and the SE has been previously reported. Nevertheless the association of gender and the SE results was examined as outlined later.

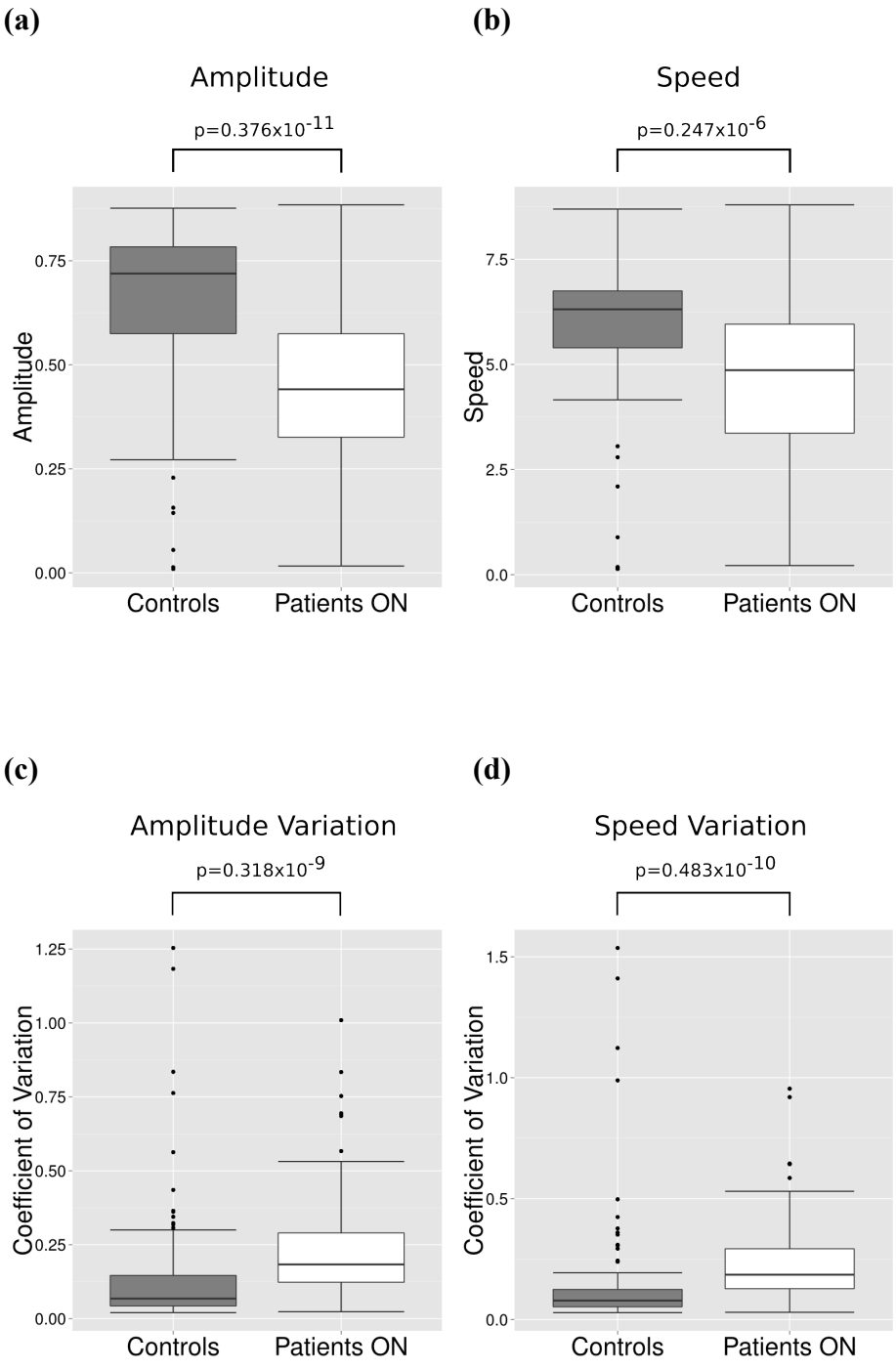
Two MDS-UPDRS FT grades were allocated for each subject giving a total of 124 grades for PwPD and 94 for HC. The grades were distributed in the patient groups as follows: 15% (19) grade zero, 51% (63) grade 1, 27% (33) grade 2, 6% (7) grade 3 and 1% (2) grade 4. The mean FT grade for the PD group was 1.28. The mean FT grade in the LTHT PD group (1.32 ± 0.84) was not significantly different to the UCSF patients *on* (1.26 ± 0.96), $p = 0.78$. All controls had a clinical FT grade of zero.

6.4 Does the sequence effect manifest over ten finger tap cycles?

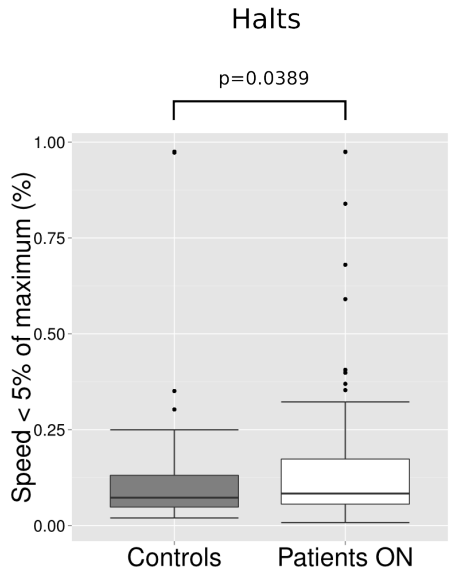
Figures 50 a-e show that in comparison to HCs the PD group's FT performance had significantly smaller amplitude, $p < 0.001$, slower speed, $p < 0.001$, more variation of amplitude, $p < 0.001$, more variation of speed, $p < 0.001$, and more halts, $p = 0.039$. In other words the FT movements in the PD group were slower, smaller, less rhythmic and interrupted by more halts and these findings are consistent with the MDS-UPDRS, and UKBBDC definitions of bradykinesia.

However Figures 50 f and 50 g show that there was no significant difference in decrementing amplitude, $p = 0.20$, or decrementing speed, $p = 0.78$ between the patient and control groups.

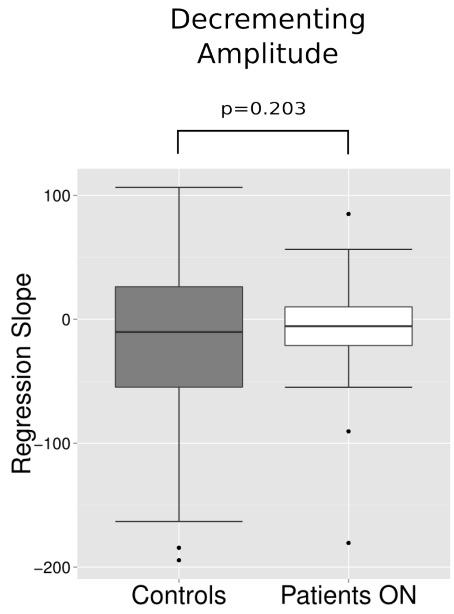
Figure 50 Measuring separable components of bradykinesia over first ten finger tap cycles



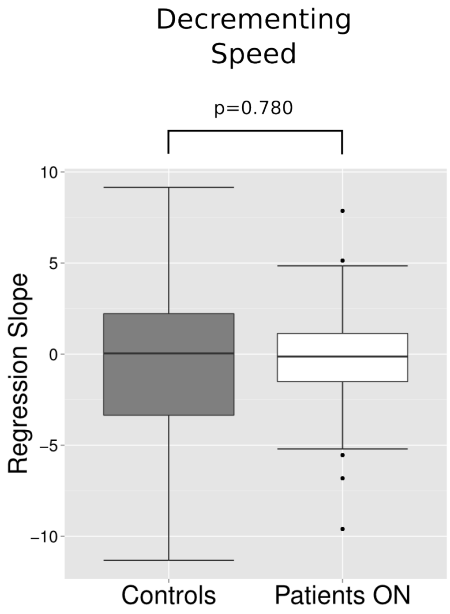
(e)



(f)



(g)



Legend: Box and Whisker plots summarising group results for the clinically defined component measures of bradykinesia during the first ten FT cycles. The centre line is the group median, the upper and lower limits of the box are the 75th and 25th percentile respectively and the error lines show the range (±2 SD from the mean). All measurements were significantly different between the PD and HC groups (a-e) except for decrementing amplitude (f) and decrementing speed (g). *Figures produced by Stuart Lacy*

6.5 Sequence effect may not be a defining characteristic of bradykinesia

In the PD and HC groups respectively, the median (± 1 SD) gradients for decrementing amplitude were - 5.61 (29.64) and - 10.29 (57.57), p = 0.203, and for decrementing speed were - 0.133 (2.39) and + 0.041 (4.20), p = 0.780 (Figures 50 f and g). The distribution of results within each group demonstrates that not all patients had a negative gradient (i.e. decrementing trend) and conversely that some HCs did have a negative gradient. In fact the proportion of each group exhibiting the SE was very similar: 61% and 52% of the PD FT data, and 59% and 49% of the HC FT data, had negative gradients for amplitude and speed respectively (p = 0.9). Further analyses were undertaken to explore possible explanations for these findings.

Sub-sets of PD data based on gender, age, hand dominance, MA side, MDS-UPDRS clinical grade, disease duration and HY stage were compared but the regression gradients for amplitude and speed did not significantly differ between groups for any of these variables (Table 34). Likewise the amplitude and speed gradients did not differ significantly for any of the variables when comparisons of the HC data were made based on gender, age and hand dominance.

The possibility of the patient group results being due to a floor effect (i.e. the mean amplitude and speed was initially lower in the PD group than in HCs so there was a smaller range of values over which they could decrement) was evaluated by adjusting the gradients for the mean amplitude and speed calculated during each ten tap recording but no significant difference between the groups was found. These results suggest that the lack of difference in SE measurements between PD and HC groups is unlikely to be related to any of these variables.

Table 34 Sequence effect measurement in Parkinson's disease group

Variable	Amplitude decrement		Speed decrement	
	Median± SD	*p*	Median ± SD	*p*
Male	- 6.32 ± 23.7	0.83	- 0.44 ± 2.3	0.87
Female	- 4.40 ± 39.3		0.19 ± 2.7	
Dominant hand	- 5.62 ± 24.1	0.89	- 0.08 ± 2.3	0.44
Non-dominant hand	- 5.49 ± 35.5		- 0.44 ± 2.5	
MA hand	- 6.22 ± 33.4	0.39	- 0.44 ± 2.5	0.22
LA hand	- 1.77 ± 24.7		0.23 ± 2.3	
Right hand	- 5.56 ± 24.4	0.68	0.02 ± 2.3	0.23
Left hand	- 5.65 ± 34.6		- 0.53 ± 2.5	
Age	0.03[1]	0.76	0.01[1]	0.88
Disease duration	- 0.02[1]	0.86	0.11[1]	0.26
HY stage	0.046[1]	0.52	0.05[1]	0.50
MDS-UPDRS grade	- 0.043[1]	0.56	- 0.09[1]	0.23

Legend: Comparisons of the speed and amplitude gradient coefficients were made between sub-sets of the PD data based on demographic and clinical variables but none reached statistical significance. All values are presented as the mean ± 1 SD except for age, disease duration, HY stage and MDS-UPDRS score that are presented as [1]Spearman's correlation coefficients.

6.6 Effect of dopaminergic drugs on the sequence effect

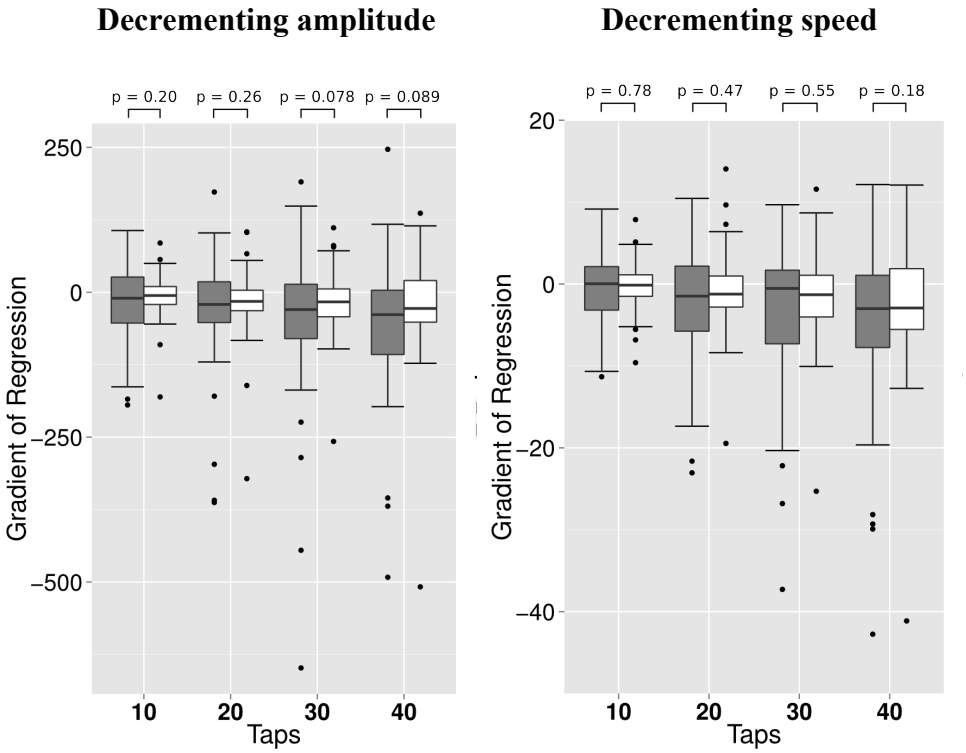
Previous studies have shown that the separable components of bradykinesia respond differently to dopaminergic drugs: speed improves the most, amplitude and rhythm improve by a lesser degree and the SE does not show any significant response (Espay, Giuffrida et al., 2011, Iansek, Huxham et al., 2006, Kang, Wasaka et al., 2010, Ling, Massey et al., 2012). Nevertheless the data from the sub-group of UCSF patients (n = 13) who were tested in *off* and *on* states was examined more closely to see whether the lack of difference in SE measures between the HC and PD groups could be due to the patient group being assessed whilst *on*. The clinical bradykinesia grades for the UCSF patients were 2.1 ± 0.79 *off* and 1.26 ± 0.96 *on*, p = 0.001 but there was no significant difference between the mean gradients in the *on* and *off* states for amplitude (- 8.31 *on* vs. - 13.07 *off*, p = 0.39) or speed (- 0.51 *on* vs. - 0.71 *off*, p = 0.14). Likewise when the gradient coefficients for the subgroup of PD patients in the *off* state were compared to the HC group there remained no significant difference for decrementing amplitude, p = 0.91, or decrementing speed, p = 0.16. This suggests that the *on-off* clinical state is unlikely to be an explanation for the lack of difference in SE measures between HC and PD groups.

6.7 Measuring the sequence effect over longer tapping sequences

A sequence length of ten taps was chosen for the primary analysis because this replicates the MDS-UPDRS motor examination instructions for assessment of FT and all other MDS-UPDRS bradykinesia items. Sequence length specified is not specified in the UKBBDC. It is possible that ten taps are not enough for decrements to manifest so the association between sequence length and decrementing amplitude and speed was investigated by comparing gradient coefficients from longer tapping sequences to those from the ten taps sequences. Figure 51 shows that with longer sequence lengths the median SE gradients for PD and HC groups tend to be more negative. In the patient group each of the longer sequences (20, 30, and 40) was associated with greater decrements in speed and amplitude than when assessed over just ten taps, all p values < 0.05. Likewise in HCs the longer FT sequences were associated with greater decrements in speed and greater

decrements in amplitude (with the exception of the 20 taps sequence; p = 0.27) than the ten tap sequence. As the same association between length of sequence and decrement occurred with approximately the same magnitude in both groups the net result was no significant net difference between PD and HCs for any of the sequence lengths. There was a trend for sequences of 30 taps (p = 0.078) and 40 taps (p= 0.089) to show a difference in amplitude decrement between PD and HC groups but this was in an unexpected manner with the HC group having a *greater* median decrement. In summary longer sequence lengths are associated with more decrement of amplitude and speed but this is no greater in PD patients than age-matched controls.

Figure 51 Sequence effect measurements over longer finger tapping sequences



Legend: The gradient coefficients for HCs (dark grey) and patients (white) groups are compared for four different lengths of FT sequence: 10, 20, 30 and 40 taps. Group data (presented using Box and Whisker plots - see Figure 50 for definitions) demonstrate that there is no significant difference in SE measurements for PD and HC groups at any of the sequence lengths.

6.8 Discussion

6.8.1 Summary of findings

The kinematic data from the first ten FTs has been used to measure the clinical components of bradykinesia, as described by the UKBBDC and MDS-UPDRS in PwPD and age-matched HCs. The results confirm that repetitive FTs performed by PwPD are characterised by slower, smaller and less rhythmic movements than HCs and these findings are consistent with the current clinical definitions of bradykinesia. However the SE measurements do not differ significantly between PD and HC groups. Over a sequence of ten FTs, approximately 50% of patients exhibited decrementing amplitude, and 60% exhibited decrementing speed but the proportions were similar in the HC group. Likewise the *magnitude* of the decrements was not significantly different between PwPD and HCs.

These results call into question whether the SE should be included in the gold standard clinical definitions of bradykinesia. This is an important finding because the diagnosis and monitoring of PD focuses on the clinical assessment of bradykinesia and it is essential that the definition of this complex clinical sign is precise in order to optimise diagnostic accuracy, accurate monitoring and the quality of clinical research.

6.8.2 Previous quantitative sequence effect studies

The previous studies that have quantified the SE are summarised in Table 35. Although the present study results may be considered controversial, detailed inspection of some of the previous studies provides evidence to support the conclusions regarding the SE reached here.

Several earlier studies demonstrated that not all PwPD exhibit the SE (Agostino, Berardelli et al., 1992a, Bajaj, Wang et al., 2012, Chee, Murphy et al., 2009, Connor and Abbs, 1991, Iansek, Huxham et al., 2006, Kang, Wasaka et al., 2010, Wagle Shukla, Ounpraseuth et al., 2012). For example Wagle Shukla et al. showed that progressive micrographia (i.e. decrementing amplitude) was present in only 50% of PwPD (Wagle Shukla, Ounpraseuth et al., 2012).

Table 35 Summary of previous quantitative sequence effect studies in Parkinson's disease

Study	Task	Decrement measured	PD n *on/off*	HC n
Berardelli 1986	Drawing	Speed	12	10
Benecke 1987	Hand and arm movements	Speed	8 *off* 2 *on*	9
Connor 1991	Speaking	Speed Amplitude	6 *on*	6
Agostino 1992	Drawing	Speed	14 *on*	13 9HD 7DT
Agostino 1994	Drawing	Speed	8 *off*	8
Bennett 1995	Drinking	Speed	9 *on*	9
Plotnik 1998	Aiming	Speed	9 *off*	7
Lekwuwa 1999	Ocular pursuit	Speed Amplitude	7 *off*	7
Longstaff 2003	Drawing	Amplitude	10 *off*	12
Agostino 2003	FT Hand opening	Speed Amplitude	11 *off*	9
Iansek 2006	Walking	Amplitude	10 *on off*	0
Chee 2009	Walking	Speed Amplitude	26 *off*	10
Kang 2010	PPB	Speed	11 *on off*	0

Espay 2011	FT Hand opening	Speed Amplitude	85 *on off*	0
Bajaj 2012	Handwriting	Amplitude	27 *on*	39 SWEDD
Wagle Shukla 2012	Handwriting	Amplitude	68 *on*	?[1]
Ling 2012	FT Handwriting	Speed Amplitude	15 *on off*	16 9 PSP

Abbreviations: SE, sequence effect; PD, Parkinson's disease; HC, healthy controls; n, number; y, years; HD, Huntington's disease; DT, Dystonia; PPB, Purdue Pegboard Test; SWEDD, subjects without evidence of dopaminergic deficit; PSP, progressive supranuclear palsy. [1]The number of controls is not stated in the paper.

Additionally many controls in previous studies were manifesting the SE (Agostino, Berardelli et al., 1992a, Agostino, Berardelli et al., 1994, Bajaj, Wang et al., 2012, Chee, Murphy et al., 2009, D'Agostino, 1971, Ling, Massey et al., 2012, Michell, Goodman et al., 2008) and yet, perhaps because the focus had been on the results of the PD group, this had rarely been commented on. For example in Chee et al.'s study of stride length, the amplitude gradient coefficients were negative in the HC group at preferred and 100% stride length conditions (with the mean values ± SD reported as -0.28 ±0.36 and -0.30 ±0.13 respectively) suggesting that decrementing amplitude was occurring in some controls (Chee, Murphy et al., 2009). Also Ling *et al.'s* study demonstrated a clear lack of SE in PSP patients relative to PwPD, but it is less clear whether there was a definite relative difference in SE measurements between the PD group and HCs. The mean ±SD values for PD and HCs respectively were -0.12 ± 0.12 and -0.2 ± 0.21 for decrementing amplitude, and -1.52 ± 0.81 and -1.71 ± 1.59 for decrementing speed. These slopes were not significantly different. After mean amplitude and speed were included as covariates, the decrementing

amplitude reached borderline significance (p = 0.046) and decrementing speed showed a trend towards significance (p = 0.070) but these borderline results did not clarify whether the SE is more reliably observed in PD than in HCs (Ling, Massey et al., 2012).

Previous studies that measured the SE objectively also have a number of limitations that should be considered when evaluating the respective conclusions. The majority assessed only small numbers of subjects with 13 of the 17 studies summarised in Table 35 having 15 or fewer PwPD and the mean number of HC per study was just 7.5. The current study is a much larger age-matched controlled kinematic study of the SE in PD. The choice of statistical tests also needs to be considered as kinematic studies have repeatedly demonstrated that the separable component measurements of repetitive voluntary movements tend to be positively skewed or have other non-normal distributions (Connor and Abbs, 1991, Giovannoni, van Schalkwyk et al., 1999, Homann, Suppan et al., 2000, Michell, Goodman et al., 2008, Taylor Tavares, Jefferis et al., 2005). Two of the SE studies have explicitly tested for normality and confirmed this finding (Connor and Abbs, 1991, Michell, Goodman et al., 2008) but the majority of the other studies outlined in Table 35 have used parametric tests without any statement about the normality of data distribution (Agostino, Berardelli et al., 1992a, Agostino, Curra et al., 2003, Benecke, Rothwell et al., 1987, Bennett, Marchetti et al., 1995, Berardelli, Accornero et al., 1986, Chee, Murphy et al., 2009, Iansek, Huxham et al., 2006, Ling, Massey et al., 2012). Given the non-normality of the movement parameters in our study, we selected non-parametric tests to interpret group differences, without assumptions about the particular parametric distribution.

It is difficult to compare the results of the previous studies as the definition of SE, patients' clinical state, and the length of sequence used has varied considerably. For example only four of the previous studies have measured both decrementing amplitude and decrementing speed (Connor and Abbs, 1991, Espay, Giuffrida et al., 2011, Lekwuwa, Barnes et al., 1999, Ling, Massey et al., 2012). Some patients have been assessed only when *on*

(Bajaj, Wang et al., 2012, Bennett, Marchetti et al., 1995, Wagle Shukla, Ounpraseuth et al., 2012), others only when *off* (Agostino, Curra et al., 2003, Lekwuwa, Barnes et al., 1999) and some have been part of a mixed group that were assessed when some were *on* and others were *off* (Benecke, Rothwell et al., 1987). Only two previous studies measured decrementing amplitude and decrementing speed in subjects who were assessed in both the *on* and *off* states (Espay, Giuffrida et al., 2011, Ling, Massey et al., 2012) and only one of these also included an age-matched control group (Ling, Massey et al., 2012). Different sequence lengths have also been assessed ranging from five seconds (Agostino, Curra et al., 2003) to 15 seconds (Espay, Giuffrida et al., 2011, Ling, Massey et al., 2012) for FT. Such variation in terms of definition, clinical state and length of sequence combined with small numbers of subjects in each study makes it difficult to build up a clear consensus about the conclusions. This study addressed these limitations by evaluating all central components of bradykinesia in a large sample on medications, including a sample tested off medications, and reporting results at sequence lengths of 10, 20, 30 and 40 taps in both PD patients and age-matched HCs.

6.8.3 Limitations

Several limitations of the present study are recognised though. Only one task was used to measure the SE and it would be useful to measure other sequential movements such as walking, foot tapping and hand opening/closing to see if any of the results generalise to other actions. This may be particularly important for over-learned movements such as walking where a 'closed loop' model of movement control is more likely than in FT which it could be argued is a new skill, especially for controls. Only one sequence of FT was assessed for each hand and this could be another potential weakness as the magnitude of the SE has been shown to vary to a certain degree within individuals between trials (Iansek, Huxham et al., 2006). It is noteworthy that the sub group of patients tested o*ff* in addition to *on* was fairly small so these particular results may be less robust. Also it could be argued that an overnight washout period for the *off* state is not long enough and there could still be effects of dopaminergic medications on the

SE or other component measures of bradykinesia. Ideally one would need to test drug naïve PD patients in order to fully evaluate what differentiates PD movements from HCs, but this method also has potential limitations in that patients with more advanced disease or more severe grades of bradykinesia are less likely to be represented in such a group. Finally it is conceivable that some of the HC subjects may develop PD and clinical follow up is necessary to determine unequivocally whether SE occurs under normal conditions.

6.8.4 Conclusions

This study has demonstrated that decrements of speed and amplitude during repetitive FT movements occur as frequently in age-matched HCs as they do in PD and the magnitude of the decrements is similar between the groups. This result suggests that the SE may not be a defining feature of PD. The explanation for this observation remains speculative. Plotnik *et al*. showed that healthy elderly adults exhibit delays when switching from one movement to another so perhaps the physiological aging process is contributing to progressive slowness or decrementing amplitude (Plotnik, Flash et al., 1998). However in the current study, advancing age was not significantly correlated with either decrementing amplitude or decrementing speed in patients or controls. It is intriguing that the SE has not been explicitly reported in controls before. This could simply be because we have not looked for it as FT, and other repetitive movement tasks, are tests focussed on assessing people with suspected PD or related disorders. Furthermore, even if healthy adults were assessed with a FT task the SE may still be clinically more apparent in PD because the baseline amplitude and speed are already reduced and hence any further reductions are visually easy to recognise i.e. a slow movement that becomes even slower is easier to detect than a ‘very’ fast movement that slows to ‘quite’ fast. Alternatively perhaps there has been so much focus on detailing the complexities of PD bradykinesia that it has been overlooked that some of the defined components of bradykinesia are also common to normal motor control. Similarly when there has been evidence of the SE in controls it may have been interpreted as ‘physiological fatigue’ (Ling, Massey et al., 2012).

Finally these results provide an explanation for the consistent finding that the SE is the only clinical component of bradykinesia not to respond to levodopa or transcranial magnetic stimulation (Espay, Giuffrida et al., 2011, Kang, Wasaka et al., 2010, Ling, Massey et al., 2012): it is 'unresponsive' because it is not part of pathological bradykinesia, but rather a physiological phenomenon superimposed on pathologically small and slow movements.

The clinical ramifications of inaccurately defining bradykinesia are very important. Bradykinesia is the obligatory motor feature of PD and accurate clinical interpretation of this sign is critical for diagnosis. It is not difficult to see how a healthy subject with generally slow movements, perhaps secondary to musculoskeletal problems or pain, who then exhibits decrements may be wrongly diagnosed with PD. Could the inclusion of the SE as a core feature of bradykinesia be contributing to the considerable misdiagnosis rates of PD? These results suggest that removing SE from the definition of bradykinesia may improve the accuracy of diagnosis and the clinical monitoring of PD patients.

Chapter 7
Summary, conclusions and future research

7.1 Summary of the research and original contributions

This thesis examines bradykinesia, the obligatory motor sign of PD, through detailed quantitative analysis of FT movements. It evaluates how objective measurements of FT may be used to predict diagnostic classification, reflect the clinical severity of bradykinesia, and investigate the characteristic kinematic features of bradykinesia. The FT movements of 49 PwPD and 41 HCs in Leeds were recorded using small and lightweight EM tracking sensors attached to the finger and thumb. Two different methods were used to analyse the movement data: firstly standard statistical methods to quantify the clinically defined components of bradykinesia and secondly a modern computer science technique called EAs were used to induce mathematical expressions called classifiers. An independent validation group comprising FT data from 13 PwPD and nine HC in UCSF was used to assess how well the results from each method of analysis generalised beyond the Leeds sample.

The thesis provides several original contributions to the current literature on PD bradykinesia. It is the first study to compare kinematic data analysis using EAs with standard statistical measures of bradykinesia in PwPD and HCs, and this approach has demonstrated the superior classification accuracy of EAs. Three previous studies have quantified PD FT movements using statistical measures, but some of the results were conflicting and only two of the studies included HCs (Espay, Giuffrida et al., 2011, Ling, Massey et al., 2012, Yokoe, Okuno et al., 2009). EAs have been used to analyse PD movement and vocal data in just a few studies so far (Smith, Gaughan et al., 2007, Smith and Timmis, 2008, Tsanas, Little et al., 2012) and the number of participants in each has been very small. This is the largest study to use EAs to analyse PD movement data and it is the only one to evaluate FT movements. It is the only study using EA analysis of PD data to validate the EA induced classifiers on an independent data set. It is also the first study to correlate the classifier results with demographic variables

and clinical measures of disease progression. This showed that classifiers induced through EAs for predicting group membership also reflect the clinical severity of bradykinesia. These results suggest that a device employing EA analysis of FT movement data has potential to predict diagnosis and also to monitor progression or response of bradykinesia in PD. A final significant original contribution is the finding that a component of PD bradykinesia, known as the SE, occurred as frequently, and with similar magnitude, in age-matched HC as it did in PwPD. This result challenges the gold standard diagnostic definition of PD bradykinesia and suggests that the SE is a physiological phenomena rather than a defining characteristic of PD bradykinesia.

7.2 Why was bradykinesia researched?

PD is characterised by a clinical syndrome comprising several different abnormal movements including bradykinesia, tremor, rigidity and postural instability. The fundamental motor feature of PD is bradykinesia though and this is the only abnormal movement that is obligatory for diagnosis. Bradykinesia is also a focus of PD clinical monitoring because it is common to all patients, results in functional disability and typically reflects disease progression and also the response to intervention. However bradykinesia is a complicated clinical sign that comprises several components, or separate sub-categories of abnormal movement, and it may be difficult to evaluate all of these accurately using visual inspection alone. It has been shown that even clinicians expert in movement disorders frequently misinterpret whether bradykinesia is present or not (Bajaj, Gontu et al., 2010) and the bradykinesia items of the gold standard clinical rating scale, the MDS-UPDRS, have the lowest inter-rater reliability of all items (Camicioli, Grossmann et al., 2001, Henderson, Kennard et al., 1991, Martinezmartin, Gilnagel et al., 1994).

Imprecise clinical ascertainment of the *presence* of bradykinesia has important potential ramifications as it may contribute to the mis-diagnosis rates of PD, which are as high as 20% amongst consultant neurologists with

a specialist interest in movement disorders (Hughes, Daniel et al., 1992, Rajput, Rozdilsky et al., 1991). Imprecise clinical interpretation of the *severity* of bradykinesia may lead to inaccurate monitoring of disease progression and response to therapeutic interventions. A method that allows objective assessment of bradykinesia could therefore potentially aid clinical diagnosis and monitoring of PD. It could also be used as a tool to investigate the characteristic features of bradykinesia in PD and other movement disorders.

7.3 What aspects of bradykinesia were researched?

This research has focussed on quantitative evaluation of bradykinesia in PD patients and HC using two different data analysis methods: standard statistical measures of the clinically defined components of bradykinesia and classifiers induced through novel EAs. The accuracy of each method to classify FT movement data into the correct diagnostic group was evaluated. The movement features that best discriminated PD data from HC data were examined in order to define the characteristic components of bradykinesia. This research has also explored how objective measurements of bradykinesia correlate with clinical grades of bradykinesia severity and other clinical markers of disease progression. These two approaches, classification and correlation, may inform the development of a non-invasive device that could potentially aid clinical diagnosis and monitoring respectively. Finally the application of using FT movement analysis as an investigative tool was explored with a specific focus on how the SE varies between PwPD and HC.

7.4 How was bradykinesia researched?

Bradykinesia was researched by evaluating objective measurements of FT movements. EM tracking sensors were attached to the index finger and thumb of 49 PwPD and 41 HCs to measure their FT movements over a 30 second assessment period. FT data was also collected from 13 PwPD and nine HC in UCSF so that the LTHT results could be validated on an independent sample.

The FT test was chosen as the method for evaluating bradykinesia for several reasons: it is an established clinical method of examining for bradykinesia, it already has a validated rating scale (Goetz, Tilley et al., 2008), it is quick to perform, correlates well with other bradykinetic movements in PD (Agostino, Berardelli et al., 1998, Ling, Massey et al., 2012), is easy to assess in a standard clinic room and it is not dependent on a subject's mobility or balance. EM sensors were chosen because they are small, lightweight, non-invasive, fairly cheap and can precisely measure movements in 3D space. Each hand's FT performance was assessed separately because simultaneous hand movements tend to lead to an improvement in the performance of the MA PD hand, and a deterioration in the LA hand (Kishore, Espay et al., 2007).

PD patients with a range of clinical severity were included in order to assess how accurately the device classified data with different degrees of bradykinesia. However there was a predominance of patients with only slight or mild clinical bradykinesia and this was important for beginning to gauge how well the device might perform as a diagnostic device in clinically indeterminate cases. The objective measures of FT were also correlated to a number of demographic and clinical measures in order to assess not only whether the device could detect the presence of bradykinesia (used for classification or diagnostic prediction) but also whether the measurements could reflect the *severity* of bradykinesia and other markers of disease progression. These assessments are useful when considering how the device could be developed into a tool for objectively monitoring PD. Finally the FT data measurements obtained from PwPD and HCs were compared to investigate what are the defining components of PD bradykinesia and whether these kinematic features match the current clinical definitions.

7.5 Conclusions

The conclusions of the research will be summarised under three main sub-headings in order to highlight how the results from objective measurements

of bradykinesia could inform the development of a device that aids diagnosis, monitoring and investigation of PD.

7.5.1 Diagnostic classification

7.5.1.1 Individual component measures of bradykinesia

The current gold standard definitions of bradykinesia (UKBBDC and MDS-UPDRS) describes bradykinesia as comprising several clinical components that may be summarised as follows: reduced amplitude, slowed speed, impaired rhythm, increased halts, and decrementing speed and decrementing amplitude. The results in Chapter 4 confirmed that FT movements performed by PwPD had smaller amplitude, slower speed, less rhythm and more halts than those performed by HCs. The results in Chapters 4 and 6 showed that measurements of the SE, decrementing speed and amplitude, did not significantly differ between the PD and HC groups though. The Chapter 5 results, obtained by examining the components of the EA induced classifier expression and the composite models combining the separable components supported these findings by demonstrating that amplitude, speed, rhythm and halts were the most discriminatory measurements of FT used by both methods of data classification.

Regarding classification accuracy, Chapter 4 demonstrated that amplitude, rhythm and speed were the clinically defined components of bradykinesia that best discriminated PD from HC FT data, each with an AUC of 0.88. Halts were moderately discriminatory with AUC of 0.72 but the SE measurements did not classify PD and HC group data at all well (AUCs 0.51 and 0.58). These findings based on the LTHT subject data were largely validated in the UCSF group. Amplitude was the single most discriminatory component in PwPD *on* whereas speed was the most discriminatory when *off*. These findings are in line with previous studies that demonstrated that dopaminergic drugs (that switch PD patients from an *off* to an *on* clinical state) improve the speed component of bradykinesia much more than the amplitude and rhythm components (Espay, Giuffrida et al., 2011). The best classification accuracy for any individual component measure was obtained by limiting the PD group data to the MA hand and this resulted in an AUC

of 0.91 for amplitude. The MA hand PD group data had a UPDRS FT grade of 1.57 ± 0.94. The classification results for the individual component measures were not improved by limiting the HC data to the dominant or non-dominant hand though.

The separable component measures of bradykinesia had much stronger correlations in the patient data than in the control group data and this supports the premise that these components are measures of the same clinical pathological phenomenon. The exception to this were the SE measurements as they did not significantly correlate with any of the other component measures of bradykinesia in the patient group, again raising the question whether the SE is part of PD bradykinesia and this was further explored in Chapter 6 and these results are outlined below in Section 7.5.3.

Only four other studies have quantified all the clinically defined components of bradykinesia in PD FT data (Espay, Beaton et al., 2009, Espay, Giuffrida et al., 2011, Ling, Massey et al., 2012, Yokoe, Okuno et al., 2009) and these all used standard statistical measures. The results from these earlier studies largely support the present study findings regarding the most discriminating components of bradykinesia and have been discussed in detail in Chapter 4. The present study has the greatest number of participants (n = 90) and whilst Espay et al's 2011 study had more patients (n = 85) there was no control group. Espay et al.'s 2009 study and the studies by Ling et al. and Yokoe et al. did include HC groups but their numbers of patients were smaller than the current study (n = 23, n = 15 and n = 16 respectively). This is the first quantitative study of PD FT to include a validation group. It is the second study, after Ling et al's to quantify PD FT in each hand separately as the other studies either assessed only one hand (Espay, Beaton et al., 2009, Espay, Giuffrida et al., 2011) or combined the data from both hands together (Yokoe, Okuno et al., 2009).

7.5.1.2 Composite models of bradykinesia

Chapter 4 demonstrated that when all the individual component measures of bradykinesia were combined using logistic regression into a composite

model the classification accuracy improved to an AUC of 0.93 for all data and to an AUC of 0.96 when PD data was limited to the MA hand. The bradykinesia composite model was not validated on the UCSF data though as it misclassified two-thirds of control data. There are no previous PD FT studies that formed a composite bradykinesia model to compare these results to.

7.5.1.3 Evolutionary algorithm induced classifier

Chapter 5 showed that the EA induced ensemble classifier had an accuracy of 0.88 AUC when all data was included, but this improved to 0.97 when PD data was limited to the MA hand. Both the ensemble classifier and the bradykinesia composite model had better classification accuracy when PD data was limited to the MA hand and this is expected because the MA PD group data had greater clinical severity of bradykinesia (UPDRS grade 1.57 ± 0.94) than the PD group comprising data from both hands (UPDRS grade 1.32 ± 0.84).

The ensemble classifier accuracy improved from an AUC of 0.88 to 0.94 when HC data was limited to just the dominant hand which suggests that movement features specific to controls are over-represented in the dominant hand. However the detailed analysis undertaken in sections 5.1-5.3 did not reveal why this was so.

The EA classifier and the bradykinesia composite model discriminated PD MA and HC dominant hand data with AUCs of 0.99 and 0.97 respectively. This equates to the EA classifier misclassifying one patient (out of 46 data sets) and four controls (out of 37 data sets) whereas the composite model misclassified three patients and four controls. There was a trend towards the EA classifier being significantly more accurate than the composite model ($p = 0.081$). Furthermore, in contrast to composite model the ensemble classifier results for MA hand and HC dominant hand *were* validated in the UCSF data.

Control data with smaller amplitude, slower speed, less rhythm or more halts were more likely to be misclassified by the EA classifier as these features are characteristic of PD movements. In contrast the SE measurements did not differ between correctly classified and misclassified HC data sets further prompting the question whether the SE is specific to PD or not.

These results lend strong support to the use of EA induced classifiers as the preferred method of data analysis when developing a device to provide objective measurements of bradykinesia. The EA induced classifiers had marginally greater classification accuracy than standard statistical measures but importantly they also could generalise beyond the population data they were developed from to accurately classify independently collected data too. In addition the ensemble classifier was validated on the UCSF patient data collected in both *on* and *off* states, suggesting that it can predict diagnostic group membership of data regardless of the patients' clinical status. This is a very encouraging result when considering development of the device to aid clinical diagnosis – it suggests that flexibility in the test procedure protocol is feasible so that subjects suspected with early PD may be tested without having to withhold, or delay the initiation of, dopaminergic medications.

7.5.1.4 Models of clinically indeterminate or newly diagnosed PD

As all of the patients in the study had clinically definite PD it is acknowledged that the classification results will over-estimate how the device is likely to perform as a diagnostic tool. In order to gain some insight into how objective measures of bradykinesia could be used to predict diagnosis in clinically indeterminate, or very early stage PD, the data from just those patient data sets that had been allocated MDS-UPDRS grade zero (i.e. clinically 'normal') or grade one (i.e. 'slight bradykinesia') were analysed separately and the results presented in Chapter 5.3.

Eleven PD FT assessments were allocated MDS-UPDRS grade zero, either because of the effects of treatment (as LTHT patients were tested *on*) or

because the assessments were from the non-affected side in HY stage one PD patients. Using the sub-set of data from grade zero PD FT data sets (n =11) and grade zero dominant hand HC FT data sets (n = 37) the bradykinesia composite model and the EA classifier had moderate classification accuracy with AUCs of 0.77 and 0.72 respectively. This corresponded to eight of the eleven grade zero PD data sets being correctly classified by the composite model and seven by the EA. Although the numbers are small this remains an encouraging result as it suggests that the device is more sensitive than clinical assessments – i.e. none of the assessments were clinically considered to be bradykinetic and yet both methods of analysis detected enough bradykinesia in the kinematic data in the majority of the patient assessments to classify them as PD rather than HC.

Forty-five of the PD data sets were clinically graded as slight bradykinesia (MDS-UPDRS grade one) and the composite model discriminated these very well from HC dominant hand data sets with an AUC of 0.93. This was equivalent to 42/45 (93%) of PD, and 32/37 (86%) HC data sets being correctly classified. The EA classifier had excellent classification accuracy with AUC of 0.96, corresponding to 41/45 (91%) PD, and 38/41 (93%) HC dominant hand data sets being correctly classified. As the 95% CI were wide there was no statistically significant difference between these results. These figures are for the threshold of equal trade off and clearly the threshold could be altered depending on the application of the device i.e. whether it was being used for screening in epidemiological studies, when sensitivity may be paramount, or for supporting diagnosis when specificity, and minimal false positives may be more important. A longitudinal study of truly clinically indeterminate patients is clearly a better method of assessing the value of using FT data analysis for diagnostic prediction, but the results here are encouraging.

7.5.2 Monitoring

The objective measures of bradykinesia were correlated to bradykinesia clinical *severity* as measured by the MDS-UPDRS FT grade. The individual

component measures of amplitude, speed, rhythm and halts (r_s range = 0.58 to 0.67), the composite bradykinesia model (r_s = 0.60) and the EA classifier score (r_s = 0.43) were all significantly correlated with MDS-UPDRS FT grade. These results indicate that methods employing quantification of bradykinesia could have a role in objectively monitoring the clinical progression of PD, and response to intervention, as they not only reflect the *presence* of bradykinesia but also *the degree of clinical severity*.

The results also suggest that objective measures of bradykinesia are more sensitive than clinical grading of severity. For example the mean EA classifier scores for the dominant hand of controls were significantly lower (-0.31 ± 0.25) than those from PD hands graded zero on the MDS-UPDRS (-0.04 ± 0.35), p = 0.007, despite the fact that bradykinesia was not clinically apparent in either group. Similarly the bradykinesia composite model scores were significantly different between MDS-UPDRS grade zero PD (0.51 ± 0.33) and dominant hand HC data sets (0.23 ± 0.27), p = 0.002. One explanation for these findings are that the MDS-UPDRS scale is too coarse to detect the subtle differences detected by the objective measures i.e. the MDS-UPDRS FT scale has just five grades (zero to four inclusive) of bradykinesia severity so it is likely that subjects within each grade allocation will have a *range* of severity. If this is the case it suggests that subtle, or individual component, response to new therapeutic interventions, may go unnoticed when assessed in trials relying on clinical grades as an outcome measure, but the device could potentially detect such responses. This theory could be tested further by examining the relative changes in clinical grades and objective measures of bradykinesia before and after dopaminergic drugs are given; this will be undertaken in the proposed future study entitled 'Using a non-invasive novel device to analyse bradykinesia and tremor in different movement disorders' that is discussed further in section 7.7.2.

The objective measures of bradykinesia were correlated with other demographic and clinical variables that reflected disease progression. This showed that none of the individual component measures or the bradykinesia

composite model or the EA classifier scores correlated with age, HY stage, disease duration or LEDD. The association of cognitive impairment with bradykinesia severity requires further exploration. The MoCA score inversely correlated strongly with rhythm ($r_s = -0.30$, $p = 0.009$) and the bradykinesia composite model score ($r_s = -0.46$, $p < 0.0001$), suggesting that cognitive impairment is associated with more impaired rhythm and a greater severity of bradykinesia. This is an interesting area of research to develop in the future as it raises the possibility that objective measures of bradykinesia could have a role in screening for cognitive impairment. The neuro-psychology literature on pre-hension (reaching and grasping) movements lends some support to this hypothesis and the proposed research study entitled 'A novel diagnostic device for the objective assessment of Parkinson's disease with and without dementia' aims to explore the association between cognitive and motor assessments in more detail and is outlined in Section 7.7.4.

7.5.3 Investigation

The FT data was examined to investigate the kinematic profile of FT in PD and whether the SE is a characteristic component of PD bradykinesia.

7.5.3.1 Kinematic features of finger tapping in PD

EAs search widely for solutions without any prior assumptions. The classifiers induced by EAs to distinguish PD from HC movement data thus may include novel discriminating features of bradykinesia i.e. their formation is not confined to using the components of bradykinesia defined clinically by UKBBDC and UPDRS. Furthermore it is possible to 'open the black box' and examine the classifier expression in order to find out what components were used to differentiate patient from control movement data. This offers an exciting method of investigating how the movements in PD differ from HC without being constrained by the current clinical definitions of bradykinesia. Some classifiers are easier to examine than others and in this study only the CGP classifier part of the ensemble classifier was examined in detail with the results discussed in Chapter 5 (section 5.5).

When the classifier parse tree expression was examined the important acceleration data points, or window offsets, used in its formation were revealed. Overlaying these data points on PD and HC kinematic profiles that received the highest classifier scores (i.e. 'highly PD' or 'highly normal') enabled a more focussed exploration of the potentially important components of bradykinesia. This method has highlighted that opening speed is numerically more discriminatory than any of the clinically defined components of bradykinesia with an AUC of 0.91. This is the second study to find this - opening speed was the most discriminatory feature of PD FT in Yokoe et al.'s study (Yokoe, Okuno et al., 2009) but the other three studies that quantified FT in PD did not specify the most discriminating movement component (Espay, Beaton et al., 2009, Espay, Giuffrida et al., 2011, Ling, Massey et al., 2012).

Overlaying the classifier expression offset windows on high scoring kinematic profiles also showed that the acceleration profiles of PD depart further from those of HC than the separation and speed profiles do. The PD separation and speed profiles have smaller amplitude and slower speeds than HC but the shape of the kinematic data waveform is fairly similar. In contrast when the acceleration profiles are examined they are totally different between PD and HC: controls had a characteristic pattern of a sharp bursts of acceleration then deceleration in the opening and closing phases of the FT cycle with the periods in between, when the fingers were either maximally separated or opposed, having almost no acceleration at all. This pattern was lost in the PD data with many small bursts of acceleration and deceleration occurring throughout the FT cycle, even when the fingers were stationary at the point of maximum or minimum separation. These profiles are consistent with EMG studies that have demonstrated the normal triphasic EMG pattern of single rapid movements is lost in PD and replaced by ill-coordinated and inefficient bursts of acceleration (Hallett and Khoshbin, 1980).

Examination of the CGP classifier's expression also pointed towards the opposition period of the FT cycle being important in distinguishing PD from

HC data and the percentage duration of each tap cycle spent with the digits opposed was found to discriminate the PD and HC data fairly well with an AUC of 0.72. Patients and controls spent approximately12% and 6% respectively of the whole tap cycle with their digits opposed, p = 0.02. Counter-intuitively the PD data also had a smaller percentage of the tap cycle spent in zero acceleration though – i.e. despite PwPD spending more of the FT cycle with the digits stationary (during opposition) there were fewer periods without any acceleration. Small acceleration and deceleration bursts were recorded in the PD data even when there was no net movement of the digits. These results suggest that in PD the 'wrong' motor programs are being selected for the desired movement and hence the muscle acceleration bursts that are too small to move the fingers out of the opposition position. This means that additional acceleration bursts are required so there is a prolonged pause, or "*switching akinesia"* (Homann, Suppan et al., 2000) between the opening and closing phases as the alternate motor programs are executed. Similar conclusions have been reached in studies assessing PD movements through drawing (Berardelli, Accornero et al., 1986), computer keyboard taps (Homann, Suppan et al., 2000) and EMG recordings (Sheridan, Flowers et al., 1987) but this is the first detailed study to quantify akinesia between the flexion and extension movements of FT in PD.

7.5.3.2 Sequence effect

It was noted in Chapter 4 that SE component measures did not correlate well with the other component measures of bradykinesia. They also did not discriminate the diagnostic group data well as PD *and* HC subjects had a tendency to decrement during the 30-second assessment period. Furthermore when the bradykinesia composite models in Chapters 4 and 5 were examined it was clear that the SE measurements made minimal contribution to the logistic regression expression. These results prompted the question of whether the SE was definitely a specific component of PD bradykinesia or whether it may simply be a physiological phenomena.

In Chapter 6 the kinematic FT data was examined more carefully in order to investigate the SE. Firstly the issue of whether ten taps, as specified by the gold standard monitoring scale - MDS-UPDRS, are a long enough sequence for all the components of bradykinesia to manifest was explored. This showed that nearly all of the clinically defined components of bradykinesia – reduced amplitude, slowed speed, less rhythm and more halts - did manifest during the first ten taps made by PwPD when compared to the first ten taps made by HCs. However there was no net difference in the SE measurements *between* the groups over this period. Possible reasons for this were explored further: firstly there was no association between the SE results and clinical/demographic variables in the groups. Secondly the SE results did not significantly change when patients switched from *off* to *on* clinical states, despite the clinical UPDRS grades improving from 2.10 ± 0.79 to 1.26 ± 0.96, $p = 0.001$. Thirdly longer sequences, of up to 40 taps did not lead to a significant net difference in frequency or severity of decrement between the PD and HC groups.

In summary no explanation related to the methodology or the demographics of the groups for the lack of difference in the SE measurements between the PD and HC groups could be found in this cohort, leading to the conclusion that the SE may not be a defining component of PD bradykinesia. This finding is controversial as it questions the gold standard UKBBDC of bradykinesia on which the clinical diagnosis if PD is based upon. A detailed literature review is presented in Chapter 6 that provides further evidence from previous studies to support the current study conclusions.

7.6 Limitations of thesis

The limitations of the study methodology have been discussed in detail in Chapter 4 and the main limitations are highlighted again here with the potential implications of these discussed in the wider context of the thesis.

When considering how the device might be developed into a diagnostic tool, the main limitation of the methodology was a cross-sectional study design using patients with clinically definite PD. By definition all the patients will

have had bradykinesia that is clinically apparent, at least when they are off treatment, so diagnostic prediction results will be an overestimate of how the device is likely to perform in clinically indeterminate, or very early PD subjects. Assessing patients who were *on* may have made the objective diagnostic prediction more difficult though as many patients had no obvious bradykinesia clinically and there were 11 grade zero MDS-UPDRS FT performances in the LTHT PD group. Arguably the current study was a proof of concept study, or an essential first step to test the hypothesis that diagnostic group classification can be accurately predicted using a new form of movement data analysis, before commencing more extensive studies. Whilst the results are encouraging the next crucial step would be to perform a longitudinal study comprising patients with possible early PD and compare the diagnostic predictions based on objective measurements of their movements with standard clinical methods of diagnosis over time to see whether the device adds any value to the clinical predictions.

Also it would be useful to objectively measure other movements in PD in order to clarify whether FT is the best test to use for diagnostic prediction. Agostino et al. reported that FT was disproportionately more impaired in PD than other tests of upper limb bradykinesia (Agostino, Curra et al., 2003) but it would be useful to measure a range of movements such as walking, foot tapping, or handwriting to confirm that FT is indeed the most sensitive test to use in a diagnostic test protocol. As the whole UPDRS motor examination was not performed in this study it was impossible to evaluate how FT measurements correlated with clinical bradykinesia assessments involving other body parts or with other movement abnormalities in PD. It was also not possible to assess whether bradykinesia that is localised clinically to just the legs would have a normal FT test or not.

A further limitation was the lack of detailed evaluation to confirm that subjects were in the correct diagnostic group. For example very few of the PD patients have had $I^{(123)}$- SPECT scans to confirm dopaminergic neuronal degeneration (and these were not used as an inclusion criteria) and many were only recently diagnosed with PD. This means that it is possible there

were some people with atypical parkinsonism (vascular, MSA etc.) inadvertently included in the PD patient group. Conversely the controls did not undergo a detailed examination to evaluate whether they had any parkinsonism and there may be some controls that had signs of early PD but had not been formally diagnosed yet.

Most of the thesis has focussed on the characteristic features of bradykinesia, how these can be used to discriminate PD from HC FT movement data, and hence how such features could be used to aid diagnosis. The evaluation of the movement data to aid development of a monitoring device was much more limited. This was largely because the LTHT subjects only underwent assessments that were brief and at one point in time, and also because the UCSF patient cohort tested in *on* and *off* states was small. The simple correlations of the EA classifier, and bradykinesia composite scores with the clinical grades of bradykinesia lend some support to the usefulness of objective measurements of movements as a method of monitoring but further studies over a longer period of time, assessing patients in *on* and off states, are needed to better understand this potential application. A study assessing PwPD over several hours has begun recruitment and is described further in section 7.7.2 below.

Finally a technical limitation of the study was the high prevalence of corruption in the movement data. Approximately 10% of the recordings had corrupt sections and these data sets were excluded from the separable component, and the SE analyses in Chapters 4 and 6 respectively. It would have been useful to have a signal generated from the laptop at the time of data collection to alert the researcher that corruption had occurred. This would allow for the assessment to be repeated again immediately rather than incurring loss of data at the analysis stage. Subsequent assessment of the test protocol has revealed that the accuracy of the Polhemus system deteriorates if the EM sensors are moved more than 20cm away from the magnetic source. Also some of the corruption was occurring because the EM sensors were crossing over into the posterior hemisphere of the magnetic source field and thus results in the sensor position being

erroneously ‘reversed’. A warning signal to alert the researcher that the data recorded contains corrupt section has now been incorporated into the software for the future studies described below (Section 7.7) that use the Polhemus system. Therefore it is likely that with better positioning of the equipment and the inclusion of the warning signal the corruption of data will be less of an issue with future research studies.

7.7 Future research

This study has demonstrated that objective evaluation of FT can be used to accurately discriminate PD from HC movement data. This opens up several areas of research for further development. The proposed studies are discussed in detail below but a brief overview is provided herein. Firstly for the device to be developed as a diagnostic test it is important to assess the classification accuracy in clinically indeterminate subjects, or those with possible early PD. The device’s prediction of diagnostic group should be compared to the current clinical methods of diagnosis and then the patients followed up over time to see how the initial objective and subjective diagnoses correlated to the final clinical diagnosis. Secondly, and related to this, it would be useful to see how well the device can differentiate movement disorders that look similar to PD (eg. other parkinsonian or tremulous conditions) and by doing so elucidate what kinematic features differentiate these conditions. This potentially would reveal an advantage of movement data analysis over DaTscan imaging that cannot reliably discriminate the various neurodegenerative parkinsonian conditions.

Thirdly in order to develop a device to provide objective monitoring of PD it would be necessary to assess whether EA classifiers could be induced that accurately predict abnormalities of movement over longer periods of time. In particular it would be important to assess whether classifiers could predict whether PwPD were in an *off* state, an *on* state or, an *on* state with dyskinesia present as this would effectively be a step towards developing a 24 hour PD monitor. Finally there is evidence from the current study, and also the prehension literature, that objective measurements of movement may be applied to screen for cognitive deficits and this hypothesis will be

tested in the fourth proposed study. These four study propositions, and their stage of development are discussed in greater detail in sections 7.7.1 to 7.7.4.

7.7.1 Diagnostic tool and pre-clinical screening

The results of this research support the principle that quantifying FT can be used as a tool to discriminate normal from bradykinetic movements, even when the severity of bradykinesia is clinically very subtle. However this study is only the first step towards developing a non-invasive device that could be used to aid diagnosis of PD. All the patients in this study had clinically definite PD so, despite attempts at modelling early PD with data from just those patients with MDS-UPDRS grade zero or one bradykinesia, there has not been a true evaluation of how the device performs as a 'diagnostic' test.

In order to do so the next necessary step would be to perform a longitudinal study of patients with possible bradykinesia but an indeterminate diagnosis and compare the device's prediction of the diagnosis with the current clinical methods for making a diagnosis. Over several years the clinical diagnosis tends to become clearer as more clinical signs develop and the response to levodopa is apparent. The final clinical diagnosis (possibly supported by $I^{(123)}$-SPECT imaging) would be considered the gold standard to compare the initial objective (device) and subjective (clinical) diagnostic predictions to. It would be prudent to also compare the original clinical diagnosis to the initial device prediction in order to evaluate whether the device was able to add any additional value to the clinical method in predicting the final clinical diagnosis i.e. calculating the positive and negative likelihood ratios.

This type of study could be extended even further 'backwards' to the earlier pre-motor stage of PD when the process of neuro-degeneration has already begun but the clinical signs of parkinsonism have not fully manifest. It is recognised that approximately 50% of dopaminergic neurons have died by the time motor signs of parkinsonism develop and this phase of 'silent' neurodegeneration probably occurs for five years preceding the onset of

definite clinical signs (Fearnley and Lees, 1991). It may be particularly important to detect patients in the pre-motor stage if a neuro-protective drug becomes available that can slow down the rate of further neurodegeneration.

REM sleep behaviour disorder (RBD) patients are an ideal 'at risk' group to study for incipient parkinsonism because they have a 50% risk of developing a neurodegenerative parkinsonian disorder (usually PD, MSA or LBD) within 12 years of being diagnosed with RBD (Postuma, Gagnon et al., 2009). Postuma et al. conducted a seven-year longitudinal study of 78 RBD and age-matched controls, assessing them annually using UPDRS and a number of quantitative motor tasks. During the study eleven patients developed PD and 9 LBD. The clinical and quantitative motor scores were plotted over time and then regression analyses used to calculate when the 20 parkinsonian patients' scores had begun to significantly deviate from normal. This showed that a very simple quantitative test of bradykinesia (number of hand taps in 1 minute) had greater sensitivity for detecting early signs of parkinsonism than the detailed UPDRS examination - the hand tapping task and UPDRS total scores deviated from those of HCs approximately 8 years and 4 years, respectively, before clinical diagnosis and the AUCs three years before clinical diagnosis were 0.81 for hand tapping and 0.76 for the total UPDRS.

The current study has shown that EA induced classifiers tend to be more accurate than standard statistical measures for predicting diagnostic classification of slight bradykinesia (grade one UPDRS) and this suggests that there may be potential to improve diagnostic prediction to an even earlier pre-motor stage by combining quantification of bradykinesia with EA classifier analysis. Detailed protocols for these two longitudinal studies of clinically indeterminate, or 'at risk' patients have not been devised yet.

7.7.2 Differentiating bradykinetic or tremulous movement disorders

There are several different conditions that manifest clinically with bradykinesia including PD, MSA, PSP, CBD, DLB, DIP, VP, AD, HD, and SCA (Chapter 2). Some of these conditions, especially MSA, PSP and CBD,

can initially be very difficult to discriminate clinically from PD. The various pathological bradykinetic conditions in their early stages may also be difficult to differentiate from the movement changes associated with normal physiological ageing, particularly if only a few of the components of bradykinesia (i.e. just small amplitude or just slowing) are present (Collier et *al.*, 2011, Buchman *et al.* 2012).

Current imaging techniques do not reliably differentiate these conditions as the vast majority can have abnormal DaTscans (Brooks, 2012) and, initially at least, the structural MRI brain scans tend to be normal or show non-specific changes. The phenomenology of bradykinesia in these movement disorders remains incompletely understood so careful clinical examination of the components of bradykinesia is unlikely to elucidate the diagnosis either. This means that clinicians often rely on a 'watchful waiting' approach to guide their clinical diagnosis; in other words with the passage of time (months or years) additional signs of 'atypical' parkinsonism or cognitive changes emerge which make the diagnosis clearer. However this method is imperfect and the final clinical diagnoses are often not supported by the pathological findings, even when they are made by clinicians expert in movement disorders: Hughes et al. for example showed that approximately 20% of the cases referred from the National Hospital of Neurology and Neurosurgery's movement disorders clinic in London with an ante-mortem clinical diagnosis of a neurodegenerative parkinsonian conditions (PD, MSA, PSP etc) had their diagnosis revised at post-mortem (Hughes, Daniel et al., 2002).

In view of the limitations of current clinical and imaging methods a non-invasive method to objectively differentiate the bradykinetic movements disorders earlier is required. This is important because the response to medications and prognosis vary considerably between the conditions. A future area of research would be to develop the FT study to objectively measure clinical tasks performed by patients with a wide range of bradykinetic conditions. The discriminating kinematic features could inform

the development of a diagnostic device capable of differentiating different bradykinetic conditions and also the clinical examination of future patients.

There have been very few studies that have objectively measured the separable components of bradykinesia across different movement disorders. Ling et al. compared the kinematic features of FT in 16 PwPD patients, nine PSP patients and 16 HCs and reported that PSP bradykinesia, compared to PD bradykinesia, had much smaller amplitudes and no SE (Ling, Massey et al., 2012). There are no studies comparing FT in MSA patients and PD patients but one abstract (main paper in Russian) reports that the frequency of EMG bursts in MSA patients' limbs (n = 18) were higher than those in PD (n = 21). (Levin, Khutorskaya et al., 2003). Also gait analysis in 15 HC, 15 HD and 15 PD patients has shown that HD bradykinesia is characterised by reduced speed and rhythm but largely normal amplitude (Delval, Krystkowiak et al., 2006). Taken together these studies suggest potential for objective measures of movements to differentiate the various parkinsonian conditions if they have characteristic bradykinesia component profiles.

Similarly it may be difficult to discriminate patients with an isolated tremor using just clinical assessments. For example a PD tremor will typically have an asymmetric distribution and be present at rest but the tremor of MSA often mimics this. There are also atypical PD tremors that predominantly occur with posture or action and these may resemble ET or DT. It could be argued that most non-parkinsonian tremors could be reliably differentiated from PD as the DaTscan will remain normal, but this method has several drawbacks including expense at £800 per scan, test procedure time (approximately 3-5 hours), limited worldwide access and an ionising radiation dose. Moreover DaTscans will not differentiate the neuro-degenerative parkinsonian tremulous disorders such as PD, MSA and CBD.

Therefore a non-invasive simple objective method for differentiating the movement disorders characterised by bradykinesia or tremor would be an important development in aiding early clinical diagnosis and guiding appropriate management. A research study protocol has been developed

between LTHT and University of York in collaboration with Monash Medical Centre in Melbourne, Australia entitled 'Using a non-invasive novel device to analyse bradykinesia and tremor in different movement disorders' and ethics approval is currently awaited (National Research Ethics Service (NRES). The aims of this study are to gain a better understanding of the discriminating features of bradykineia and tremor in a number of movement disorders and in normal ageing, and to assess how accurately EA classifiers predict diagnosis based on analysis of the movement data.

The plan is to recruit 60 PD, 30 ET, 30 HD, 30 DT, 30 PSP, 30 MSA, 30 DLB and 30 DIP patients and 50 HCs. The subjects will each be assessed (in *on* and *off* states for those on dopaminergic medications) using the same EM tracking equipment as in the present study. The sensors will be attached to the finger and thumb of the dominant hand whilst the subject performs the following bradykinesia and tremor assessments from the UPDRS: 10 FTs, 10 hand opening- closing, 10 pronation-supination, 30 seconds of rest tremor, 30 seconds of postural tremor and 10 finger-nose repetitions. The sensors will then be moved to the non-dominant hand and the assessments repeated. The assessments will be video-recorded and each task graded (using the MDS-UPDRS) by two clinicians blinded to the current clinical working diagnosis. Some patients will have had a DaTscan performed as part of their clinical investigation (i.e. not specifically for this study) and the kinematic data collected for these patients will also be compared to the results of their DaT scans (graded 0-4 where zero is normal and 4 is severe striatal neuronal degeneration) (Benamer *et al.*, 2000, Benamer *et al.*, 2000) in addition to the diagnoses and grades of clinical severity.

Data will be analysed using EAs and standard statistical measures of the separable components of bradykinesia and tremor such as frequency, amplitude, speed etc. The EAs will enable development of classifiers to discriminate the data from separate diagnostic groups and analysis of these classifiers, and the separable component measures will provide information on the differentiating features. The strength of using EAs for this study is

the potential to learn more about bradykinesia without *a priori* assumptions about the characteristics of bradykinesia and tremor in each diagnostic group.

7.7.3 Monitoring dyskinesia in Parkinson's disease

A further exciting area of research development is to use movement sensors in combination with EA classifiers to monitor the severity of involuntary movements, called dyskinesia, or LID in PD patients. LID are a troublesome motor complication that commonly occur in PwPD who are treated with dopaminergic drugs, with cumulative rates of 40% and 90% in PD patients who take levodopa for more than four years and 10 years respectively (Ahlskog and Muenter, 2001); see Chapter 1. The presence of LID limits the ability to optimise the PD treatment regimen as a fine balance has to be made between giving enough dopaminergic drugs to alleviate the symptoms of PD (such as bradykinesia, tremor and rigidity) but not so much that LID is worsened. Over time this balance becomes more precarious and eventually a compromise has to be made where the dose of medication is somewhat sub-therapeutic for the PD symptoms in an attempt to restrict the time spent with troublesome LID.

The management of patients with LID typically involves altering the timing and doses of dopaminergic drugs in an attempt to provide a more steady level of dopaminergic stimulation to the brain. This is typically quite a complicated process though as every patient is individual in terms of what drug dose triggers LID, at what time of day their LID occurs and the topographical distribution of LID. Furthermore patients may also report variability of the severity and character of their symptoms from day to day. Often PwPD have LID occurring tens of times per day, or almost constantly but with varying degrees of severity and this taken together with the fact that many take ten or more dopaminergic drug doses over the course of 24 hours, makes it extremely difficult to decide exactly which doses of medication need to be altered.

In order to better guide changes the management of LID by altering drug regimens various methods of monitoring LID have been tried. The most common method is reviewing the patient every few months in the outpatient clinic, asking what times they think LID occurs and then reducing the drug dose that precedes any episodes of troublesome LID. Other methods for monitoring LID have been tried but they all have limitations. For example symptom diaries are onerous for patients to complete and generally lack accuracy (Golbe and Pae, 1988), questionnaires or scales based on the patient's perspective are prone to bias with the mood and cognition impacting on responses, and clinician rated assessment scales tend to be time consuming to administer, require training, only provide a snapshot of the clinical signs at the point in time that they are completed and may necessitate a period of expensive inpatient monitoring if administered over more than a few hours.

A device that could accurately record the presence and severity of LID would revolutionise the management of PD. It would also benefit the pharmaceutical industry that require accurate methods to assess new anti-dyskinesia drugs as well as testing new anti-parkinsonian drugs for dyskinetic AEs. Various devices that objectively measure LID have been assessed but none of these have been incorporated into mainstream clinical practice yet. One of the main limitations of previous devices was the fact that the frequency of voluntary movements (especially walking) overlapped considerably with the frequencies of LID so standard statistical methods were imprecise at discriminating these movements (Hoff, van den Plas et al., 2001, Keijsers, Horstink et al., 2000, Manson, Brown et al., 2000). More recently there have been promising results for discriminating different grades of LID and tremor during a range of voluntary actions using dynamic neural networks that are similar to EAs (Roy, Cole et al., 2013).

A proof of concept study protocol entitled 'A novel device for the objective assessment of levodopa-induced dyskinesia in Parkinson's disease' has been developed to discriminate different clinical severities of LID from normal movements in PD patients. Ethical approval was granted (NRES reference:

11/NW/0541) and the study was added to the NIHR Clinical Research Network portfolio (Portfolio ID 11762). Ten patients have been assessed on the neurology day ward for six hours whilst wearing movement sensors that comprise three accelerometers and three gyroscopes. The patients were continuously video-recorded and sections of the video were subsequently graded, or marked up, by two clinicians independently using the Unified Dyskinesia Rating Scale and MDS-UPDRS. The movement data linked to the marked up sections of the video were then compared to the clinical grades of LID to assess the accuracy of the classifiers. The preliminary results show that EA classifiers can discriminate moderate and severe (grades 3 and 4) LID from normal movements (grade zero) with AUC of 0.90. However the classification accuracy for discriminating slight or mild LID (grades 1 and 2) from normal movements was less with AUCs of 0.68 and 0.8 respectively. In view of this an amendment has been recently been granted (March 2014) to assess 30 more patients over a two-hour period each in order to improve the accuracy of the EA classifiers.

7.7.4 Movement analysis and cognition

A fascinating area of research that builds upon the current FT study findings is to examine the association between abnormalities of movement and cognition. Although the methods for assessing cognition in the current FT study were not particularly detailed (MoCA), there was a signal that impaired rhythm was associated with cognitive impairment. In view of this, and also based on a literature review of the sensory-motor integration in PD and dementia, a protocol called ‘A novel diagnostic device for the objective assessment of Parkinson's disease with and without dementia has been devised. The aim of this study is to find out what movement features are associated with PD dementia (PDD) by comparing prehension (reach and grasp) movements and detailed cognitive assessments in PwPD, PDD patients and in HCs. The results may inform the development of a new test that predicts, or at least screens for, cognitive impairment in PD accurately, quickly and more cheaply than current methods of imaging and detailed cognitive tests. The study data may also enable further understanding of bradykinesia, and other motor deficits in PD.

The theory that this study's hypothesis is based upon may be summarised as follows: it is well recognised that non-demented PD patients have deficits in executive aspects of visuospatial processing and in PDD patients these deficits are even more pronounced. The reaching phase of prehension requires information about where the object is positioned in 3D space and this depends on the function of the dorsal occipito-parietal, or 'where', visual stream. The grasping phase of prehension requires visual information about the size, shape and orientation of an object to be converted into appropriate patterns of finger and wrist movement and this depends on the ventral occipito-temporal, or 'what', visual stream. Prehension is an overlearned complex movement that develops in childhood so there is no need for new skill acquisition (which is impaired in PD). This means that prehension is a useful paradigm to investigate motor control deficits and visuospatial processing – both of which are known to be impaired in PD and PDD.

Therefore the hypothesis of the study states that analysis of prehension movement data will allow prediction about the cognitive state of PD patients (demented or non-demented). The control group is included in order to compare PD cognitive deficits to age-matched controls and also to compare prehension (developmental) and FT (non-developmental) movement data for predicting diagnostic group (PD vs. HC).

Ethical approval has been obtained (NRES reference 10/H1308/5) and testing began in Leeds in March 2014 with the aim to recruit 30 HCs and 60 patients with PD (approximately 30 with PDD and 30 without PDD). Collaborative study assessments have also begun in UCSF and Dubai. The participants will perform a number of motor tasks including prehension, drawing and FT whilst wearing special Lycra gloves called Computer Data Gloves that contain various movement sensors embedded in them. The movement data recorded by the gloves will be analysed by EAs to provide classifiers for diagnostic prediction and also separable component analysis to examine how the abnormalities of movement in PD correlate with cognitive deficits. The research participants will also undergo standard

motor assessments, including UPDRS, cognitive tests including MoCA, Benson figure copy and delayed recall test, Benton line of orientation tasks and Trails A and B, and the Geriatric depression scale to screen for pseudo-dementia. In order to make the diagnosis of 'probable PD dementia' using the Emre 2007 criteria (Emre, Aarsland et al., 2007) an informant interview comprising the neuropsychiatric inventory questionnaire and the Clinical Dementia Rating Scale Informant interview will also be conducted.

7.8 Final summary

This thesis has examined PD bradykinesia in detail. Quantification of the FT task in PD patients and HCs assessed in Leeds has been used as a method to objectively evaluate bradykinesia. A modern computer science technique called EAs have been employed to analyse movement data recorded with EM tracking sensors and these results compared to standard statistical measures of the components of bradykinesia. A small sample of data collected independently in UCSF has been used to validate the results. The results have been discussed in the context of the current medical literature with the potential for developing a device to aid diagnosis, monitoring and investigation of PD highlighted.

The FT study results have shown that EA induced classifiers tend to have better diagnostic accuracy than methods based on measuring the components of bradykinesia. Only the EA classifier results generalised to the validation data set. With both methods of analysis, quantification of FT using was associated with the clinical severity of bradykinesia. There was an association between cognitive impairment and reduced rhythm of FT movements but physiological ageing was not associated with any of the component measures of bradykinesia. The results have confirmed that PD patients have slower, smaller and less rhythmic movements than HC but challenge whether the SE should be included within the gold standard diagnostic definition of bradykinesia as objective SE measurements of did not differ between PwPD and age-matched HCs.

A number of research studies are currently underway, or planned, to develop the ideas outlined in this thesis and there is the potential that objective measurements of movements may be a future method for aiding the diagnosis of PD, screening for PD in epidemiological studies, discriminating PD from other movement disorders, monitoring motor fluctuations over the course of the day and investigating the characteristics of PD motor and cognitive deficits.

References

Abdo, W. F., et al. (2010). The clinical approach to movement disorders. *Nature Reviews Neurology* 6(1) 29-37.

Agostino, R., et al. (1992). Sequential arm movements in patients with Parkinson's disease, Huntington's disease and dystonia. *Brain* 115(5) 1481-1495.

Agostino, R., et al. (1994). Analysis of repetitive and norepetitive sequential arm movements in patients with Parkinson's disease. *Movement Disorders* 9(3) 311-314.

Agostino, R., et al. (1998). Clinical impairment of sequential finger movements in Parkinson's disease. *Movement Disorders* 13(3) 418-421.

Agostino, R., et al. (2003). Impairment of individual finger movements in Parkinson's disease. *Movement Disorders* 18(5) 560-565.

Ahlskog, J. E. and Muenter, M. D. (2001). Frequency of levodopa-related dyskinesias and motor fluctuations as estimated from the cumulative literature. *Movement Disorders* 16(3) 448-458.

Alberts, J. L., Tresilian, J. R. and Stelmach, G. E. (1998). The co-ordination and phasing of a bilateral prehension task. The influence of Parkinson's disease. *Brain* 121(4) 725-742.

Altman, D. G. and Bland, J. M. (1994a). Diagnostic tests 2: Predictive values. *British Medical Journal* 309(6947) 102.

Altman, D. G. and Bland, J. M. (1994b). Diagnostic tests 3: Receiver operating characteristic plots. *British Medical Journal* 309(6948) 188.

Altman, D. G. and Bland, J. M. (1994c). Diagnostic tests 1: Sensitivity and specificity. *British Medical Journal* 308(6943) 1552.

Amanzio, M., et al. (2010). Impaired awareness of movement disorders in Parkinson's disease. *Brain and Cognition* 72(3) 337-346.

Antonini, A., et al. (1998). Differential diagnosis of parkinsonism with F-18 fluorodeoxyglucose and PET. *Movement Disorders* 13(2) 268-274.

Antonini A, et al. (2013). Validation of the Italian version of the Movement Disorders Society- Unified Parkinson's Disease Rating Scale. *Neurological Sciences* 23(5) 683-687.

Bain, P. G., et al. (1994). A study of hereditary essential tremor. *Brain* 117(4) 805-824.

Bajaj, N. P., et al. (2010). Accuracy of clinical diagnosis in tremulous parkinsonian patients: a blinded video study. *Journal of Neurology Neurosurgery and Psychiatry* 81(11) 1223-1228.

Bajaj, N. P., et al. (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(11) 2335-2340.

Bajaj, N.P., Hauser, R.A. and Grachev, I.D. (2013) Clinical utility of dopamine transporter single photon emission CT (DaT-SPECT) with (123I) ioflupane in diagnosis of parkinsonian syndromes. *Journal of Neurolpgy Neurosurgery and Psychiatry* 84(11) 1288-1295.

Banaskiewicz, K., et al. (2009). Spiral drawing time as a measure of bradykinesia. *Neurologia I Neurochirurgia Polska* 43(1) 16-21.

Baroni, A., et al. (1984). Human ballistic arm abduction movements – effects of L-dopa treatment in Parkinson's disease. *Neurology* 34(7) 868-876.

Behnke, S., et al. (2005). Differentiation of Parkinson's disease and atypical parkinsonian syndromes by transcranial ultrasound. *Journal of Neurology Neurosurgery and Psychiatry* 76(3) 423-425.

Benamer, H. T., et al. (2003). Prospective study of presynaptic dopaminergic imaging in patients with mild parkinsonism and tremor disorders: part 1. Baseline and 3-month observations. *Movement Disorders* 18(9) 977-984.

Benecke, R., et al. (1986). Performance of simultaneos movements in patients with Parkinson's disease. *Brain* 109 739-757.

Benecke, R., et al. (1987). Disturbance of sequential movements in paients with Parkinson's disease. *Brain* 110 361-379.

Bennett, D. A., et al. (1996). Prevalence of Parkinsonian signs and associated mortality in a community population of older people. *New England Journal of Medicine* 334(2) 71-76.

Bennett, D. A., et al. (1997). Metric properties of nurses' ratings of parkinsonian signs with a modified Unified Parkinson's Disease Rating Scale. *Neurology* 49(6) 1580-1587.

Bennett, K. M., et al. (1995). The drinking action of Parkinsons disease subjects. *Brain* 118 959-970.

Berardelli, A., et al. (1986). Fast complex arm movements in Parkinson's disease. *Journal of Neurology Neurosurgery and Psychiatry* 49(10) 1146-1149.

Berardelli, A., et al. (2001). Pathophysiology of bradykinesia in Parkinson's disease. *Brain* 124 2131-2146.

Berg, D. (2011). Hyperechogenicity of the substantia nigra: pitfalls in assessment and specificity for Parkinson's disease. *Journal of Neural Transmission* 118(3).

Berg, D., Siefker, C. and Becker, G. (2001). Echogenicity of the substantia nigra in Parkinsons disease and its relation to clinical findings. *Journal of Neurology* 248(8) 648-9.

Bernheim, H., et al. (1973). Brain dopamine and syndromes of Parkinson and Huntington. Clinical, morphological and neurochemical correlations. *Journal of the Neurological Sciences* 20(4) 415-55.

Biglan, K. M., et al. (2009). Motor Abnormalities in Premanifest Persons with Huntington's Disease: The PREDICT-HD Study. *Movement Disorders* 24(12) 1763-1772.

Bonanni, L., et al. (2010). Protracted benefit from paradoxical kinesia in typical and atypical parkinsonisms. *Neurological Sciences* 31(6) 751-756.

Braak, H., et al. (2003). Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiology of Aging* 24(2) 197-211.

Brooks, D. J. (2012). Parkinson's disease: diagnosis. *Parkinsonism & Related Disorders* 18 Suppl 1.

Brooks, D. J., et al. (1992). Isolated tremor and disruption of the nigrostrtaital dopaminergic system: an F-18 dopa PET study. *Neurology* 42(8) 1554-1560.

Caligiuri, M. P., et al. (2006). Quantitative measurement of handwriting in the assessment of drug-induced parkinsonism. *Human Movement Science* 25(4-5) 510-522.

Camicioli, R., et al. (2001). Discriminating mild parkinsonism: Methods for epidemiological research. *Movement Disorders* 16(1) 33-40.

Carlsson, A. (1959). The occurrence, distribution and physiological role of catecholamines in the nervous system. *Pharmacological reviews* 11(2) 490-493.

Chaudhuri, K. R., et al. (2005). Long duration asymmetrical postural tremor is likely to predict development of Parkinson's disease and not essential tremor: clinical follow up study of 13 cases. *Journal of Neurology Neurosurgery and Psychiatry* 76(1) 115-117.

Chee, R., et al. (2009). Gait freezing in Parkinsons disease and the stride length sequence effect interaction. *Brain* 132(8) 2151-2160.

Connor, N. P. and Abbs, J. H. (1991). Task-dependent variations in Parkinsonian motor impairements. *Brain* 114(1) 321-332.

D'Agostino, R. B. (1971). An Omnibus Test of Normality for Moderate and large size samples. *Biometrika* 58(2) 341-348.

Dafotakis, M., et al. (2008). The impact of subthalamic deep brain stimulation on bradykinesia of proximal and distal upper limb muscles in Parkinson's disease. *Journal of Neurology* 255(3) 429-437.

Daniel, S. E., de Debruin, V. M. and Lees, A. J. (1995). The clinic and pathological spectrum of Steele-Richardson-Olszewski syndrome (progressive supranuclear palsy): a reappraisal. *Brain* 118(3) 759-70.

de Lau, L. M. and Breteler M. M. (2006). Epidemiology of Parkinson's disease. *Lancet Neurology* 5(6) 525-535.

Delong, M. R. (1990). Primate models of movement disorders of basal ganglia origin. *Trends in Neurosciences* 13(7) 281-285.

Delval, A., et al. (2006). Role of hypokinesia and bradykinesia in gait disturbances in Huntington's disease: a biomechanical study. *Journal of Neurology* 253(1) 73-80.

Department of Health (2005). *The National Service Framework for long term conditions.* March 2005.

deRijk, M. C., et al. (1997). Prevalence of parkinsonism and Parkinson's disease in Europe: The EUROPARKINSON collaborative study. *Journal of Neurology Neurosurgery and Psychiatry* 62(1) 10-5.

Dick, J. P., et al. (1984). The corticomotorneuron connection is normal in Parkinson's disease. *Nature* 310(5976) 407-409.

Dick, J. P., et al. (1989). The Bereitshaftspotential is abnormal in Parkinson's disease. *Brain* 112(1) 233-244.

Diem-Zangerl, A., et al. (2009). Mortality in Parkinson's Disease: A 20-Year Follow-Up Study. *Movement Disorders* 24(6) 819-825.

Dodel, R. C., Berger, K. and Oertel, W. H. (2001). Health-related quality of life and healthcare utilisation in patients with Parkinson's disease - Impact of motor fluctuations and dyskinesias. *Pharmacoeconomics* 19(10) 1013-1038.

Dodel, R. C., et al. (1997). Cost of Parkinson's disease. A retrospective 3-month analysis of direct costs. *Nervenarzt* 68(12) 978-984.

Drew, M. R., et al. (2007). Transient overexpression of striatal D-2 receptors impairs operant motivation and interval timing. *Journal of Neuroscience* 27(29) 7731-7739.

Dunnewold, R. J., Jacobi, C. E. and van Hilten, J. J. (1997). Quantitative assessment of bradykinesia in patients with Parkinson's disease. *Journal of Neuroscience Methods* 74(1) 107-112.

Edgar, J. A. (2007). The Application of Evolutionary Algorithms towards the Diagnosis of Parkinson's Disease. MEng thesis. *University of York.*

Edwards M., Quinn N., and Bhatia, K.P. (2008). Parkinsons disease and other movements disorders. *Oxford University Press.*

Emara, T. H., et al. (2010). Repetitive transcranial magnetic stimulation at 1Hz and 5Hz produces sustained improvement in motor function and disability after ischaemic stroke. *European Journal of Neurology* 17(9) 1203-1209.

Emre, M., et al. (2007). Clinical diagnostic criteria for dementia associated with Parkinson's disease. *Movement Disorders* 22(12) 1689-1707.

Espay, A. J., et al. (2009). Impairments of Speed and Amplitude of Movement in Parkinson's Disease: A Pilot Study. *Movement Disorders* 24(7) 1001-1008.

Espay, A. J., et al. (2011). Differential response of speed, amplitude, and rhythm to dopaminergic medications in Parkinson's disease. *Movement Disorders* 26(14) 2504-2508.

Esper, C. D. and Factor, S. A. (2008). Failure of recognition of drug-induced parkinsonism in the elderly. *Movement Disorders* 23(3) 401-404.

Evarts, E. V., Teravainen, H. and Calne, D. B. (1981). Reaction time in Parkinson's disease. *Brain* 104(1) 167-186.

Fahn, S., et al. (2004). Levodopa and the progression of Parkinson's disease. *New England Journal of Medicine* 351(24) 2498-508.

Fearnley, J. M. and Lees A. J. (1991). Aging and Parkinson's disease – substantia nigra regional selectivity. *Brain* 114(5) 2283-2301.

Fellows, S. J., Noth, J. and Schwarz, M. (1998). Precision grip and Parkinson's disease. *Brain* 121(9) 1771-1784.

Filippi, L., et al. (2006). I-123-FP-ClT in progressive supranuclear palsy and in Parkinson's disease: a SPECT serniquantitative study. *Nuclear Medicine Communications* 27(4) 381-386.

Findley, L., et al. (2003). Direct economic impact of Parkinson's disease: a research survey in the United Kingdom. *Movement disorders* 18(10) 1139-1145.

Gasparoli, E., et al. (2002). Clinical predictors in Parkinson's disease. *Neurological Sciences* 23 S77-S78.

Gasser, T. (1998). Genetics of Parkinson's disease. *Annals of Neurology* 44(3) S53-S57.

Gerschlager, W., et al. (1999). Bilateral subthalamic nucleus stimulation improves frontal cortex function in Parkinson's disease: an electrophysiological study of the contingent negative variation. *Brain* 122(12) 2365-2373.

Gibb, W. R. and Lees, A. J. (1988). The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry* 51(6) 745-752.

Gibb, W. R. G. and Lees, A. J. (1989). The significance of the Lewy body in the diagnosis of Idiopathic Parkinson's disease. *Neuropathology and Applied Neurobiology* 15(1) 27-44.

Giovannoni, G., et al. (1999). Bradykinesia akinesia inco-ordination test (BRAIN TEST): an objective computerised assessment of upper limb motor function. *Journal of Neurology Neurosurgery and Psychiatry* 67(5) 624-629.

Goetz, C. G. and Stebbins, G. T. (2004). Assuring interrater reliability for the UPDRS motor section: utility of the UPDRS teaching tape. *Movement Disorders* 19(12) 1453-1456.

Goetz, C. G., et al. (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(1) 41-47.

Goetz, C. G., et al. (2008). Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Movement Disorders* 23(15) 2129-2170.

Goetz, C. G., et al. (2010). Teaching program for the Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale: (MDS-UPDRS). *Movement Disorders* 25(9) 1190-1194.

Golbe, L. I. and Pae, J. (1988). Validity of a mailed epidemiological questionnaire and physical self-assessment in Parkinson's disease. *Movement Disorders* 3(3) 245-254.

Grafton, S. T., et al. (1995). Pallidotomy increases activity of motor association cortex in Parkinson's disease: a positron emission tomographic study. *Annals of Neurology* 37(6) 776-783.

Griffin, H. J., et al. (2011). The effect of real and virtual visual cues on walking in Parkinson's disease. *Journal of Neurology* 258(6) 991-1000.

Haaxma, C. A., et al. (2010). Timed motor tests can detect subtle motor dysfunction in early Parkinson's disease. *Movement Disorders* 25(9) 1150-1156.

Hallett, M., Shahani, B. T. and Young, R. R. (1975). EMG analysis of stereotyped voluntary movements in man. *Journal of Neurology Neurosurgery and Psychiatry* 38(12) 1154-1162.

Hallett, M. and Khoshbin, S. (1980). A Physiological Mechanism of Bradykinesia. *Brain* 103(2) 301-314.

Hallett, M. (2011). Bradykinesia: Why do Parkinson's patients have it and what trouble does it cause? *Movement Disorders* 26(9) 1579-1581.

Hardie, R. J. and Lees, A. J. (1988). Neuroleptic-induced Parkinsons syndrome – clinical features and results of treatment with levodopa. *Journal of Neurology Neurosurgery and Psychiatry* 51(6) 850-854.

Hassin-Baer, S., et al. (2001). Clinical characteristics of neuroleptic-induced parkinsonism. *Journal of Neural Transmission* 108(11) 1299-1308.

Heldman, D. A., et al. (2011). The modified bradykinesia rating scale for Parkinson's disease: Reliability and comparison with kinematic measures. *Movement Disorders* 26(10) 1859-1863.

Hellwig, S., et al. (2012). F-18 FDG-PET is superior to I-123 IBZM-SPECT for the differential diagnosis of parkinsonism. *Neurology* 79(13) 1314-1322.

Hely, M. A., et al. (1999). The Sydney multicentre study of Parkinson's disease: progression and mortality at 10 years. *Journal of Neurology Neurosurgery and Psychiatry* 67(3) 300-307.

Hely, M. A., et al. (2005). Sydney multicenter study of Parkinson's disease: non-L-dopa-responsive problems dominate at 15 years. *Movement Disorders* 20(2) 190-199.

Hely, M. A., et al. (2008). The Sydney multicenter study of Parkinson's disease: The inevitability of dementia at 20 years. *Movement Disorders* 23(6) 837-844.

Henderson, L., et al. (1991). Scales for rating motor impairment in Parkinson's disease- studies of reliability and convergent validity. *Journal of Neurology Neurosurgery and Psychiatry* 54(1) 18-24.

Hoehn, M. M. and Yahr, M. D. (1967). Parkinsonism: onset, progression and mortality. *Neurology* 17(5) 427-442.

Hoff, J. I., et al. (2001). Accelerometric assessment of levodopa-induced dyskinesias in Parkinson's disease. *Movement Disorders* 16(1) 58-61.

Hollingworth, L. (1914). Functional Periodicity. *Classics in the History of Psychology*.

Homann, C. N., et al. (2000). The Bradykinesia Akinesia Incoordination Test (BRAIN TEST), an objective and user-friendly means to evaluate patients with parkinsonism. *Movement Disorders* 15(4) 641-647.

Hughes, A. J., Lees, A. J. and Stern, G. M. (1990). Apomorphine test to predict dopaminergic responsiveness in parkinsonian syndromes. *Lancet* 336(8706) 32-34.

Hughes, A. J., Lees, A. J. and Stern, G. M. (1991). Challenge tests to predict the dopaminergic response in untreated Parkinson's disease. *Neurology* 41(11) 1723-1725.

Hughes, A. J., et al. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *Journal of Neurology, Neurosurgery, and Psychiatry* 55(3) 181-184.

Hughes, A. J., Daniel, S. E. and Lees, A. J. (2001). Improved accuracy of clinical diagnosis of Lewy body Parkinson's disease. *Neurology* 57(8) 1497-1499.

Hughes, A. J., et al. (2002). The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. *Brain* 125(4) 861-870.

Iansek, R., Huxham, F. and McGinley, J. (2006). The sequence effect and gait festination in Parkinson disease: Contributors to freezing of gait? *Movement Disorders* 21(9) 1419-1424.

Ikeda, A., et al. (1992). Movement related potentials recorded from supplementary motor areas and primary motor area: role of supplementary motor area in voluntary movements. *Brain* 115(4) 1017-1043.

Ikeda, A., et al. (1997). Dissociation between contingent negative variation (CNV) and Bereitschaftspotential (BP) in patients with parkinsonism. *Electroencephalography and Clinical Neurophysiology* 102(2) 142-151.

Iwasaki, Y., et al. (1999). Micrographia in Huntington's disease. *Journal of the Neurological Sciences* 162(1) 106-107.

Jackson, S. R., et al. (1995). The internal control of action and Parkinson's disease: a kinematic analysis of visually-guided and memory-guided prehension movements. *Experimental Brain Research.* 105(1) 147-162.

Jahanshahi, M., et al. (1995). Self-initiated versus externally triggered movements: I. An investigation using measurement of regional cerebral blood flow with PET and movement related potentials in normal and Parkinson's disease subjects. *Brain* 118(4) 913-933.

Jain, S., Lo, S. E. and Louis. E. D. (2006). Common misdiagnosis of a common neurological disorder: how are we misdiagnosing essential tremor? *Archives of Neurology* 63(8) 1100-1104.

Jankovic, J., et al. (1990). Variable expression of Parkinson's disease: a base-line analysis of the DATATOP cohort. *Neurology* 40(10) 1529-1534.

Jankovic, J., et al. (2000). The evolution of diagnosis in early Parkinson disease. *Archives of Neurology* 57(3) 369-372.

Jankovic, J., Schwartz, K. S. and Ondo, W. (1999). Re-emergent tremor of Parkinson's disease. *Journal of neurology, neurosurgery, and psychiatry* 67(5) 646-650.

Jenkins, I. H., et al. (1992). Impaired activation of the supplementary motor area in Parkinson's disease is reversed when akinesia is treated with Apomorphine. *Annals of Neurology* 32(6) 749-757.

Jobbagy, A., et al. (2005). Analysis of finger-tapping movement. *Journal of Neuroscience Methods* 141(1) 29-39.

Johnson, M. T. V., et al. (1991). Modulation of the stretch reflex during volitional sinusoidal tracking in Parkinson's disease. *Brain* 114 443-460.

Jones, Oliphant and Peterson. (2001). *SciPy:Open Source Scientific Tools for Python* [*online*]. Available at: http://www.scipy.org [Accessed 21 March 2013].

Kandori, A., et al. (2004). Quantitative magnetic detection of finger movements in patients with Parkinson's disease. *Neuroscience Research* 49(2) 253-260.

Kang, S. Y., et al. (2010). Characteristics of the Sequence Effect in Parkinson's Disease. *Movement Disorders* 25(13) 2148-2155.

Kaufman, M. J. and Madras, B. K. (1991). Severe depletion of cocaine recognition sites associated with the dopamine transporter in Parkinson's diseased striatum. *Synapse* 9(1) 43-49.

Keijsers, N. L., et al. (2000). Detection and assessment of the severity of levodopa-induced dyskinesia in patients with Parkinson's disease by neural networks. *Movement Disorders* 15(6) 1104-1111.

Kempster, P. A., et al. (2010). Relationships between age and late progression of Parkinson's disease: a clinico-pathological study. *Brain* 133(6) 1755-1762.

Khan, N. L., et al. (2003). Parkin disease: a phenotypic study of a large case series. *Brain* 126(6) 1279-1292.

Kim, E. J., et al. (2005). Micrographia on free writing versus copying tasks in idiopathic Parkinson's disease. *Parkinsonism & Related Disorders* 11(1) 57-63.

Kimber, T. E., et al. (1999). Voluntary movement after pallidotomy in severe Parkinson's disease. *Brain* 122(5) 895-906.

Kishore, A., et al. (2007). Unilateral versus bilateral tasks in early asymmetric Parkinson's disease: differential effects on bradykinesia. *Movement Disorders* 22(3) 328-333.

Krack, P., et al. (1998). Opposite motor effects of pallidal stimulation in Parkinson's disease. *Annals of Neurology* 43(2) 180-192.

Larson, W. (2009). *Genetic algorithms: Cool name and damn simple* [*online*]. Available at: http://lethain.com/genetic-algorithms-cool-name-damn-simple/ [Accessed 15 March 2013].

Leblois, A., et al. (2006). Temporal and spatial alterations in GPi neuronal encoding might contribute to slow down movement in Parkinsonian monkeys. *European Journal of Neuroscience* 24(4) 1201-1208.

Lekwuwa, G. U., et al. (1999). Progressive bradykinesia and hypokinesia of ocular pursuit in Parkinson's disease. *Journal of Neurology Neurosurgery and Psychiatry* 66(6) 746-753.

LePen, C., et al. (1999). Cost of illness and disease severity in a cohort of French patients with Parkinson's disease. *Pharmacoeconomics* 16(1) 59-69.

Levin, O. S., et al. (2003). Clinical and electromyographical analysis of peculiarities of parkinsonian syndrome in multiple system atrophy and Parkinson's disease. *Zhurnal Nevropatologii I Psikhiatrii Imeni S S Korsakova* 103(11) 4-9.

Ling, H., et al. (2012). Hypokinesia without decrement distinguishes progressive supranuclear palsy from Parkinson's disease. *Brain* 135(4) 1141-1153.

Litvan, I., et al. (1996). Accuracy of clinical criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome). *Neurology* 46(4) 922-930.

Litvan, I., et al. (1998). Accuracy of the clinical diagnoses of Lewy body disease, Parkinson disease, and dementia with Lewy bodies. *Archives of Neurology* 55(7) 969-978.

Liu, P., et al. (2011). Clinical heterogeneity in patients with early-stage Parkinson's disease: a cluster analysis. *Journal of Zhejiang University-Science B* 12(9) 694-703.

Logigian, E., et al. (1991). Does tremor pace repetitive voluntary motor behaviour in Parkinson's disease. *Annals of Neurology* 30(2) 172-179.

Lones, M.A., et al. (2013). Characterising neurological time series using biologically motivated networks of coupled discrete maps. *Biosystems* 112(2):94-101.

Longstaff, M. G., et al. (2003). Discrete and dynamic scaling of the size of continuous graphic movements of parkinsonian patients and elderly controls. *Journal of Neurology Neurosurgery and Psychiatry* 74(3) 299-304.

Louis, E. D., et al. (2001). Clinical correlates of action tremor in Parkinson disease. *Archives of Neurology* 58(10) 1630-1634.

Maetzler, W., Liepelt, I. and Berg, D. (2009). Progression of Parkinson's disease in the clinical phase: potential markers. *Lancet Neurology* 8(12) 1158-1171.

Majsak, M. J., et al. (1998). The reaching movements of patients with Parkinson's disease under self-determined maximal speed and visually cued conditions. *Brain* 121(4) 755-766.

Majsak, M. J., et al. (2008). Effects of a moving target versus a temporal constraint on reach and grasp in patients with Parkinson's disease. *Experimental Neurology* 210(2) 479-488.

Manson, A. J., et al. (2000). An ambulatory dyskinesia monitor. *Journal of Neurology, Neurosurgery, and Psychiatry* 68(2) 196-201.

Marras, C. and Lang, A. (2008). Invited Article: Changing concepts in Parkinson disease: moving beyond the Decade of the Brain. *Neurology* 70(21) 1996-2003.

Marsden, C. D. (1984). The pathophysiology of movement disorders. *Neurologic Clinics* 2(3) 435-459.

Marsden, C. D. and Obeso, J. A. (1994). The functions of the basal ganglia and the paradox of stereotaxic surgery in Parkinson's disease. *Brain* 117(4) 877-897.

Marshall, V. L., et al. (2009). Parkinson's Disease is Overdiagnosed Clinically at Baseline in Diagnostically Uncertain Cases: A 3-Year European Multicenter Study with repeat I-123 FP-CIT SPECT. *Movement Disorders* 24(4) 500-508.

Martinez-Martin, P., et al. (1994). Unified Parkinson's Disease Rating Scale - Characteristics and Structure. *Movement Disorders* 9(1) 76-83.

Martinez-Martin, P., et al. (2013). Expanded and independent validation of the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS). *Journal of Neurology* 260(1) 228-236.

Marttila, R. J. and Rinne, U. K. (1977). Disability and progression in Parkinson's disease. *Acta Neurologica Scandinavica* 56(2) 159-169.

Matsui, H., et al. (2005). Brain perfusion differences between Parkinson's disease and multiple system atrophy with predominant parkinsonian features. *Parkinsonism & Related Disorders* 11(4) 227-232.

Matsumoto, H., et al. (2011). Small saccades restrict visual scanning area in Parkinson's disease. *Movement Disorders* 26(9) 1619-1626.

McKinnon, J. H., et al. (2007). Sniffing out Parkinson disease - Can olfactory testing differentiate parkinsonian disorders? *Neurologist* 13(6) 382-385.

McLennan, J. E., et al. (1972). Micrographia in Parkinson's disease. *Journal of the Neurological Sciences* 15(2) 141-152.

Meara, J., Bhowmick, B. K. and Hobson, P. (1999). Accuracy of diagnosis in patients with presumed Parkinson's disease. *Age Ageing* 28(2) 99-102.

Merello, M., et al. (1999). GPi firing rate modification during beginning-of-dose motor deterioration following acute administration of apomorphine. *Movement Disorders* 14(3) 481-483.

Michell, A. W., et al. (2008). Hand tapping: A simple, reproducible, objective marker of motor dysfunction in Huntington's disease. *Journal of Neurology* 255(8) 1145-1152.

Morris, J. C., et al. (1989). Clinical and pathological aspects of parkinsonism in Alzheimers disease. A role for extranigral factors. *Archives of Neurology* 46(6) 651-657.

Morris, M. E., et al. (1996). Stride length regulation in Parkinson's disease normalization strategies and underlying mechanisms. *Brain* 119(2) 551-568.

Movement Disorders Society (2013) *MDS-UPDRS Training Program and Exercise.* Available at: http://www.movementdisorders.org/updrs/ [accessed on 26 March 2013].

Muir, S. R., et al. (1995). Measurement and analysis of single and multiple finger tapping in normal and Parkinsonian subjects. *Parkinsonism & Related Disorders* 1(2) 89-96.

NICE (June 2006). *Parkinson's disease: Diagnosis and management in primary and secondary care. Clinical Guideline 35*. National Institute for Health and Clinical Excellence.

Negrotti, A., Secchi, C. and Gentilucci, M. (2005). Effects of disease progression and L-dopa therapy on the control of reaching-grasping in Parkinson's disease. *Neuropsychologia* 43(3) 450-459.

Nicoletti, G., et al. (2006). Apparent diffusion coefficient measurements of the middle cerebellar peduncle differentiate the Parkinson variant of MSA from Parkinson's disease and progressive supranuclear palsy. *Brain* 129(1) 2679-2687.

Notermans, N. C., et al. (1994). Measuring ataxia: quantification based on the standard neurological examination. *Journal of Neurology Neurosurgery and Psychiatry* 57(1) 22-26.

Nussbaum, R. L. and Ellis, C. E. (2003). Genomic medicine: Alzheimer's disease and Parkinson's disease. *New England Journal of Medicine* 348(14) 1356-1364.

O'Sullivan, J. D., et al. (1998). Gait analysis in patients with Parkinson's disease and motor fluctuations: Influence of levodopa and comparison with other measures of motor function. *Movement Disorders* 13(6) 900-906.

Obeso, J. A., et al. (2008). Functional organization of the basal ganglia: Therapeutic implications for Parkinson's disease. *Movement Disorders* 23(S3) S548-S559.

Olanow, C. W., Obeso, J. A. and Stocchi, F. (2006). Drug insight: continuous dopaminergic stimulation in the treatment of Parkinson's disease. *Nature clinical practice. Neurology* 2(7) 382-392.

Oliveira, R. M., et al. (1997). Micrographia in Parkinson's disease: the effect of providing external cues. *Journal of Neurology Neurosurgery and Psychiatry* 63(4) 429-433.

Oliveira, R. M., et al. (1998). Hypometria in Parkinson's disease: Automatic versus controlled processing. *Movement Disorders* 13(3) 422-427.

Pal, P. K., et al. (2001). Alternating two finger tapping with contralateral activation is an objective measure of clinical severity in Parkinson's disease and correlates with PET F-18 -DOPA Ki. *Parkinsonism & Related Disorders* 7(4) 305-309.

Parkinson, J. (1817). An Essay on the Shaking Palsy.

Parkinson Study Group. (2002). Dopamine transporter brain imaging to assess the effects of pramipexole vs levodopa on Parkinson disease progression. *Journal of the American Medical Association* 287(13) 1653-1661.

Parkinson Study Group (2004). A controlled, randomized, delayed-start study of rasagiline in early Parkinson disease. *Archives of Neurology* 61(4) 561-566.

Parkinson's UK (2011). National Parkinson's Audit Report 2011.

***Polhemus:PATRIOT^{TM} Motion Tracking* [*online*].** Available at: http://www.polhemus.com/?page=motion_patriot [Accessed 18 March 2013].

Paviour, D. C., et al. (2007). Diffusion-weighted magnetic resonance imaging differentiates parkinsonian variant of multiple-system atrophy from progressive supranuclear palsy. *Movement Disorders* 22(1) 68-74.

Pennington, S., et al. (2010). The cause of death in idiopathic Parkinson's disease. *Parkinsonism & Related Disorders* 16(7) 434-437.

Piccini, P., et al. (1999). The role of inheritance in sporadic Parkinson's disease: Evidence from a longitudinal study of dopaminergic function in twins. *Annals of Neurology* 45(5) 577-582.

Plotnik, M., et al. (1998). Motor switching abilities in Parkinson's disease and old age: temporal aspects. *Journal of Neurology Neurosurgery and Psychiatry* 65(3) 328-337.

Poewe, W. and Mahlknecht, P. (2009). The clinical progression of Parkinson's disease. *Parkinsonism & related disorders* 15(S4) S28-S32.

Politis, M., et al. (2011). Graft-Induced Dyskinesias in Parkinson's Disease: High Striatal Serotonin/Dopamine Transporter Ratio. *Movement Disorders* 26(11) 1997-2003.

Ponsen, M. M., et al. (2004). Idiopathic hyposmia as a preclinical sign of Parkinson's disease. *Annals of Neurology* 56(2) 173-181.

Post, B., et al. (2005). Unified Parkinson's disease rating scale motor examination: Are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable? *Movement Disorders* 20(12) 1577-1584.

Postuma, R. B., et al. (2009). Quantifying the risk of neurodegenerative disease in idiopathic REM sleep behavior disorder. *Neurology* 72(15) 1296-1300.

Rajput, A. H., et al. (1984). Essential tremor in Rochester, Minnesota: a 45 year study. *Journal of Neurology Neurosurgery and Psychiatry* 47(5) 466-470.

Rajput, A. H., Rozdilsky, B. and Rajput, A. (1991). Accuracy of clinical diagnosis of parkinsonism – a prospective study. *Canadian Journal of Neurological Sciences* 18(3) 275-278.

Rajput, A. H., et al. (1993). Significance of parkinsonian manifestations in essential tremor. *Canadian Journal of Neurological Sciences* 20(2) 114-117

Ramaker, C., et al. (2002). Systematic evaluation of rating scales for impairment and disability in Parkinson's disease. *Movement Disorders* 17(5) 867-876.

R Development Core Team (2008). *R: A language and environment for statistical computing* [*online*]. Vienna, Austria. Available at: http://www.R-project.org [Accessed 21 March 2013].

Richards, M., et al. (1994). Interrater reliability of the Unified Parkinson's Disease Rating Scale motor examintion. *Movement Disorders* 9(1) 89-91.

Rodriguez-Oroz, M. C., et al. (2009). Initial clinical manifestations of Parkinson's disease: features and pathophysiological mechanisms. *Lancet Neurology* 8(12) 1128-1139.

Rothwell, J. C., et al. (1983). The behaviour of the long latency stretch reflex in patients with Parkinson's disease. *Journal of Neurology Neurosurgery and Psychiatry* 46(1) 35-44.

Roy, S. H., et al. (2013). High-resolution tracking of motor disorders in Parkinson's disease during unconstrained activity. *Movement Disorders* 28(8) 1080-1087.

Schettino, L. F., et al. (2004). Deficits in the evolution of hand preshaping in Parkinson's disease. *Neuropsychologia* 42(1) 82-94.

Schneider, S. A., et al. (2007). Patients with adult-onset dystonic tremor resembling parkinsonian tremor have scans without evidence of Dopaminergic deficit (SWEDDs). *Movement Disorders* 22(15) 2210-2215.

Schoenberg, B. S., Anderson, D. W. and Haerer, A. F. (1985). Prevalence of Parkinson's disease in the bircial population of Copiah County, Mississippi. *Neurology* 35(6) 841-845.

Schrag, A., Ben-Shlomo, Y. and Quinn, N. (2002). How valid is the clinical diagnosis of Parkinson's disease in the community? *Journal of Neurology, Neurosurgery, and Psychiatry* 73(5) 529-534.

Schrag, A., Quinn, N. P. and Ben-Shlomo, Y. (2006). Heterogeneity of Parkinson's disease. *Journal of Neurology Neurosurgery and Psychiatry* 77(2) 275-276.

Schrag, A., et al. (2007). Rate of clinical progression in Parkinson's disease. A prospective study. *Movement Disorders* 22(7) 938-945.

Schwingenschuh P., et al. (2010). Distinguishing SWEDDS patients with asymmetric resting tremor from Parkinson's disease: a clinical and electrophysiological study. *Movement Disorders* 25(5) 560-9.

Scolding, N. J. and Lees, A. J. (1994). Micrographia assocaited with a parietal lobe lesion in Multiple Sclerosis. *Journal of Neurology Neurosurgery and Psychiatry* 57(6) 739-741.

SIGN (January 2010). *Diagnosis and pharmacological management of Parkinson's disease. A national clinical guideline*. Scottish Intercollegiate Guidelines Network

Sheridan, M. R., Flowers, K. A. and Hurrell, J. (1987). Programming and execution of movement in Parkinson's disease. *Brain* 110(5) 1247-1271.

Shimoyama, I., Ninchoji, T. and Uemura, K. (1990). The finger tapping test. A quantitative analysis. *Archives of Neurology* 47(6) 681-684.

Siderowf, A., et al. (2002). Test-retest reliability of the Unified Parkinson's Disease Rating Scale in patients with early Parkinson's disease: Results from a multicenter clinical trial. *Movement Disorders* 17(4) 758-763.

Smith, S., et al. (2007). Diagnosis of Parkinson's disease using evolutionary algorithms. *Genetic Programming and Evolvable Machines* 8(4) 433-447.

Smith, S. L., Leggett, S and Tyrrell, A.M. (2005). An implicit context representation for evolving image processing filters. *Applications of Evolutionary Computing, Proceedings* 3449 407-416.

Smith, S. L. and Timmis, J. (2008). An immune network inspired evolutionary algorithm for the diagnosis of Parkinson's disease. *Biosystems* 94(1-2) 34-46.

Stiasny-Kolster, K., et al. (2005). Combination of 'idiopathic' REM sleep behaviour disorder and olfactory dysfunction as possible indicator for alpha-synucleinopathy demonstrated by dopamine transporter FP-CIT-SPECT. *Brain* 128(1) 126-137.

Taylor Tavares, A. L., et al. (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(10) 1286-1298.

Tomlinson, C. L., et al. (2010). Systematic Review of Levodopa Dose Equivalency Reporting in Parkinson's Disease. *Movement Disorders* 25(15) 2649-2653.

Tsanas, A., et al. (2012). Novel Speech Signal Processing Algorithms for High-Accuracy Classification of Parkinson's Disease. *IEEE Transactions on Biomedical Engineering* 59(5) 1264-1271.

UCL (2012) *University College London NHS Foundation Trust Provider Tarrifs 2012-2013*. Available at: www.uclh.nhs.uk/aboutus/Documents [accessed 6 April 2014].

Vaillancourt, D. E., et al. (2004). Effects of deep brain stimulation and medication on bradykinesia and muscle activation in Parkinson's disease. *Brain* 127(3) 491-504.

Vaillancourt, D. E., et al. (2006). Effects of deep brain stimulation and medication on strength, bradykinesia, and electromyographic patterns of the ankle joint in Parkinson's disease. *Movement Disorders* 21(1) 50-58.

Vaillancourt, D. E., et al. (2009). High-resolution diffusion tensor imaging in the substantia nigra of de novo Parkinson disease. *Neurology* 72(16).

Van Lieshout, P. H., Steele, C. M. and Lang, A. E. (2011). Tongue control for swallowing in Parkinson's disease: effects of age, rate, and stimulus consistency. *Movement Disorders* 26(9) 1725-1729.

Vitale, C., et al. (2001). Unawareness of dyskinesias in Parkinson's and Huntington's diseases. *Neurological Sciences* 22(1) 105-106.

Vossius, C., et al. (2011). The economic impact of cognitive impairment in Parkinson's Disease. *Movement Disorders* 26(8) 1541-1544.

Wagle Shukla, A., et al. (2012). Micrographia and related deficits in Parkinson's disease: a cross-sectional study. *BMJ Open* 2(3) doi: 10.1136/bmjopen-2011-000628.

Walter, U., et al. (2004). Sonographic discrimination of corticobasal degeneration vs progressive supranuclear palsy. *Neurology* 63(3) 504-509.

Wang, H. C., Lees, A. J. and Brown, P. (1999). Impairment of EEG desynchronisation before and during movement and its relation to bradykinesia in Parkinson's disease. *Journal of Neurology Neurosurgery and Psychiatry* 66(4) 442-446.

Wenning, G. K., et al. (2005). Prevalence of movement disorders in men and women aged 50-89 years (Bruneck Study cohort): a population-based study. *Lancet Neurology* 4(12) 815-820.

Whetten-Goldstein, K., et al. (1997). The burden of Parkinson's Disease on society, family, and the individual. *Journal of the American Geriatrics Society* 45(7) 844-849.

Whone, A. L., et al. (2003). Slower progression of Parkinson's disease with ropinirole versus levodopa: The REAL-PET study. *Annals of Neurology* 54(1) 93-101.

Wichmann, T. and Dostrovsky, J. O. (2011). Pathological basal ganglia activity in movement disorders. *Neuroscience* 198 232-244.

Williams, D. R., et al. (2005). Characteristics of two distinct clinical phenotypes in pathologically proven progressive supranuclear palsy: Richardson's syndrome and PSP-parkinsonism. *Brain* 128(6) 1247-1258.

Yokoe, M., et al. (2009). Opening velocity, a novel parameter, for finger tapping test in patients with Parkinson's disease. *Parkinsonism & Related Disorders* 15(6) 440-444.

Yoshita, M. (1998). Differentiation of idiopathic Parkinson's disease from striatonigral degeneration and progressive supranuclear palsy using iodine-123 meta-iodobenzylguanidine myocardial scintigraphy. *Journal of the Neurological Sciences* 155(1) 60-67.

Zou, K. H., O'Malley, A. J. and Mauri, L. (2007). Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation* 115(5) 654-657.

Abbreviations

3D	Three-dimensional
α-SN	Alpha-synuclein protein
ABN	Artificial biochemical network
AC	Alternating current
AD	Alzheimer's disease
ADL	Activity of daily living
AT	Akinesia time
AUC	Area under ROC curve
BG	Basal ganglia
BRAIN TEST©	Bradykinesia Akinesia Incoordination Test
BSP	Bereitschaftspotential
CBD	Corticobasal degeneration
CDS	Continuous dopaminergic stimulation
CGP	Cartesian genetic programming
CI	Confidence intervals
CNV	Contingent negative variation
COMT	Catechol-I-methyltransferase
COV	Coefficient of variance
CS	Closing speed
CT	Computed tomography
DA	Dopamine
DaT scan	Dopamine active transporter scan
DBS	Deep brain stimulation
DIP	Drug induced parkinsonism
DLB	Dementia with Lewy bodies
DoH	Department of Health
DS	Dysmetria score
DT	Dystonic tremor
DWI	Diffusion weighted imaging
EA	Evolutionary algorithm
ECG	Electrocardiogram

EEG	Electroencephalogram
EM	Electromagnetic
EMG	Electromyography
ET	Essential tremor
^{18}F fluorodopa	Fluorine-18-labelled-dopa
fMRI	Functional Magnetic Resonance Imaging
FN	False negative
FP	False positive
FP-CIT	Iodine^{123} labelled ioflupane
FT	Finger tapping
GABA	Gamma-aminobutyric acid
GP	General practitioner
GPe	Globus pallidus externus
GPi	Globus pallidus interna
HC	Healthy control
HIV	Human immunodeficiency virus
HY	Hoehn and Yahr
Hz	Hertz
IPD	Idiopathic Parkinson's disease
IS	Incoordination score
KS	Kinesia score
LA	Least affected hand
LB	Lewy body
LED	Light emitting diode
LEDD	levodopa equivalent daily dose
LID	Levodopa induced dyskinesia
LTHT	Leeds Teaching Hospitals NHS Trust
MA	Most affected hand
MAOI	Monoamine oxidase inhibitors
MBRS	Modified Bradykinesia Rating Scale
MDS	Movement Disorders Society
MDS-UPDRS	Movement Disorders Society Sponsored revision of

	the Unified Parkinson's Disease Rating Scale
MEMS	Micro-electromechanical systems
MIBG	Iodine -123 labelled meta-iodobenzylguanidine
MIDI	Musical instrument digital interface
MMSE	Mini Mental Screening Examination
MoCA	Montreal Cognitive Assessment
MPTP	1-methyl-4-phenyl-1,2,3,6-tetrahydropridine
MRI	Magnetic resonance imaging
MSA	Multiple System Atrophy
MSA -P	Multiple system atrophy - parkinsonism
MSA- C	Multiple system atrophy - cerebellar
MT	Movement time
N	Number
NHS	National Health System
NICE	National Institute for Health and Clinical Excellence
NPH	Normal pressure hydrocephalus
NPV	Negative predictive value
NRES	National Research Ethics Service
OKS	Optokinetic system
OS	Opening speed
PD	Parkinson's disease
PDD	Parkinson's disease with dementia
PDNS	Parkinson's disease nurse specialist
PET	Positron emission tomography
PIGD	Postural instability and gait disorder
PM	Post mortem
PMC	Primary motor cortex
PPB	Purdue Pegboard Test
PPV	Positive predictive value
PSP	Progressive Supranuclear Palsy
PwPD	People with Parkinson's disease
QOL	Quality of life
RBD	Rapid eye movement sleep behaviour disorder

ROC	Receiver operating characteristic (curve)
RT	Reaction time
SD	Standard deviation
SE	Sequence effect
SEU	Systems electronic unit
SIGN	Scottish Intercollegiate Guidelines Network
SMA	Supplementary motor area
SN	Substantia nigra
SNc	substantia nigra pars compacta
SNpr	substantia nigra pars reticulata
SOL	Space occupying lesion
SPECT	Single photon emission computed tomography
SPSS	IBM Statistical Package for the Social Sciences
STN	Subthalamic nucleus
SWEDD	Scans without evidence of dopaminergic deficit
TD	Tremor dominant
TGUGT	Timed Get up and Go Test
TMS	Transcranial magnetic stimulation
TN	True negative
TP	True positive
TUS	Transcranial ultrasound
UCSF	University of California-San Francisco
UDysRS	Unified Dyskinesia Rating scale
UKBBDC	UK Parkinsons Disease Society Brain Bank Clinical Diagnostic Criteria
UPDRS	Unified Parkinson's Disease Rating Scale
UPSIT	University of Pennsylvania Smell Identification Test
VL	Ventrolateral thalamic nucleus
VP	Vascular parkinsonism