A COMPARISON OF THREE DEPRESSION SCALES AS SCREENING TOOLS FOR DEPRESSION IN PARKINSON'S DISEASE

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Sarah Baillon BSc MPhil

Department of Health Sciences

University of Leicester

2013

Sarah Baillon

A comparison of three depression scales as screening tools for depression in Parkinson's disease

ABSTRACT

Background
This study aimed to evaluate the validity of the Edinburgh Depression Scale (EDS) as a screening measure for depression in patients with Parkinson's disease (PD) and to compare the performance with the two questions recommended for depression screening by the National Institute for Health and Clinical Excellence (NICE) and the 15-item Geriatric Depression Scale (GDS-15).

Methods
A sample of 120 patients attending PD out-patient clinics were interviewed in a standardised manner using the relevant sections of the Present State Examination-Schedules for Clinical Assessment in Neuropsychiatry to identify depression according to DSM-IV criteria. Participants then completed the three depression screening measures.

Results
The sensitivity, specificity, positive and negative predictive values (PPV and NPV) of each screening measure for the identification of major and minor depression were calculated for different cut-off scores, and receiver operating characteristics (ROC) analyses were conducted. The EDS gave sensitivity of 74%, specificity of 92% (PPV 64%, NPV 95%) at the cut-off score of 10/11. The questions recommended by NICE gave sensitivity of 100% and specificity of 84% (PPV 54%, NPV 100%) at a threshold of 0/1, and the GDS-15 showed sensitivity of 84% and specificity of 89% (PPV 59%, NPV 97%) at a cut-off of 5/6. The areas under the ROC curve were 0.90, 0.95 and 0.92 respectively. The relative benefits of a two-stage screening process were explored.

Conclusion
This study shows that the EDS, NICE questions and GDS-15 are valid screening measures for depression in PD provided the appropriate threshold scores are applied. The NICE questions showed the highest sensitivity but low PPV; a positive response to either of these two questions should highlight to the clinician that further diagnostic assessment of the patient may be warranted. A second screening measure, such as the EDS, could then be used to reduce the number of false-positives.

ACKNOWLEDGEMENTS

I would like to thank my supervisors Professor James Lindesay and Professor Mick Dennis for their invaluable advice and support throughout the whole process of this research and during the preparation of this thesis. I am also very grateful to Dr Nelson Lo for his input and assistance during the course of the study.

The completion of this study would not have been possible without the following people who between them carried out the diagnostic interviews with the participants: Dr Santanu Chakrabarti, Dr Jairaj Padmanabhan, Dr Aniruddha Rajkonwar, Dr Deepak Shukla, and in particular Dr Pravin Prabhakaran, Dr Oli Hands, Dr Jonathan Murray, Dr Suzanne Stacey and Dr Sandhya Gaur who completed the lion's share of the assessments.

Many thanks to Dr Lo, Dr Abbott, Dr Critchley and Jean Martey for taking the time in their busy clinics to approach their patients about participating in the study. I would also like to thank Maria Viskaduraki for her advice regarding statistical analysis.

I am very grateful to all the people who participated in this study – without their willingness to participate this thesis would not have been possible.

Finally, I would like to thank Bee Furber for her support and encouragement throughout the many years this project has taken.

Some of the data in this thesis are included in the following published paper:

Sarah Baillon; Michael Dennis; Nelson Lo; James Lindesay (2013) Screening for depression in Parkinson's disease: the performance of two screening questions, *Age and Ageing*; doi: 10.1093/ageing/aft152

Link to the published article:
http://ageing.oxfordjournals.org/cgi/reprint/aft152?ijkey=VpNcA1fcBbzVWOl&keytype=ref

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ADL	Activities of Daily Living
APA	American Psychiatric Association
AUC	Area Under the Curve
BDI	Beck Depression Inventory
BEDS	Brief Edinburgh Depression Scale
CBT	Cognitive Behaviour Therapy
CES-D	Centre for Epidemiologic Studies – Depression scale
CI	Confidence Interval
CSDD	Cornell Scale for Depression in Dementia
Dep	Depression
df	Degrees of freedom
DOR	Diagnostic Odds Ratio
DSM-III	Diagnostic and Statistical Manual of Mental Disorders, Third edition
DSM-IIIR	Diagnostic and Statistical Manual of Mental Disorders, Third edition revised
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders, Fourth edition
ECT	Electro-convulsive Therapy
FN	False Negative
FP	False Positive
GDS-15	Geriatric Depression Scale (15 item)
GDS-4	Geriatric Depression Inventory (4 item)
GP	General Practitioner

HADS	Hospital Anxiety and Depression Scale
HamD	Hamilton Depression Scale
HDI	Hamilton Depression Inventory
ICD-10	International Classification of Diseases, Tenth edition
IDA	Irritability, Depression and Anxiety scale
IDS-C	Inventory of Depressive Symptoms – Clinician rated
IDS-SR	Inventory of Depressive Symptoms – Subject Rated
IPD	Idiopathic Parkinson's Disease
MADRS	Montgomery-Asperg Depression Rating Scale
MDS	Movement Disorder Society
MMSE	Mini-Mental State Examination
MOA	Monoamine oxidase
N	Number
NCCMH	National Collaborating Centre for Mental Health
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
NIMH	National Institute of Mental Health
NINDS	National Institute for Neurological Diseases and Stroke
NLR	Negative Likelihood Ratio
NPV	Negative Predictive Value
nr	not reported
OMR	Overall Misclassification Rate
ONS	Office for National Statistics
PD	Parkinson's disease
PDS	Parkinson's Disease Society

PHQ-2	Patient Health Questionnaire (2 item)
PHQ-9	Patient Health Questionnaire (9 Item)
PLR	Positive Likelihood Ratio
PPV	Positive Predictive Value
PSE	Present State Examination
PSE-SCAN	Present State Examination – Schedules for Clinical Assessment in Neuropsychiatry
RCT	Randomised Controlled Trial
REC	Research Ethics Committee
REM	Rapid Eye Movement
ROC	Receiver Operating Characteristic
SCAN	Schedules for Clinical Assessment in Neuropsychiatry
SD	Standard Deviation
Se	Sensitivity
SNRI	Serotonin and Norepinephrine Reuptake Inhibitor
Sp	Specificity
SPECT	Single Photon Emission Computed Tomography
SSRI	Selective Serotonin Reuptake Inhibitor
TCA	Tricyclic antidepressant
TMS	Transcranial Magnetic Stimulation
TN	True negative
TP	True Positive
UK	United Kingdom
UKPDS	United Kingdom Parkinson's Disease Society
UK-PDS-BB	UK Parkinson's Disease Society Brain Bank

UPDRS	Unified Parkinson's Disease Rating Scale
USA	United States of America
USPSTF	United States Preventative Services Task Force
WHO	World Health Organisation

CHAPTER 1 – INTRODUCTION

1. PARKINSON'S DISEASE

Parkinson's disease (PD) is a chronic, progressive neurological disorder that primarily affects movement. It is characterised by resting tremor, muscular rigidity, slowness of movement and postural instability. These symptoms impact upon even the simplest physical tasks and thus affect nearly all aspects of a person's everyday life. In addition to these motor symptoms there are many 'non-motor' symptoms that can have just as much, if not more, impact upon an individual and their quality of life. For example, depression, anxiety, sleep disturbance, cognitive problems are all significant problems that are associated with the disease and may be experienced by someone who has PD.

PD is one of the most common neurodegenerative diseases in the elderly and it is estimated that it currently affects approximately 120,000 people in the UK (Parkinson's Disease Society, 2005). Whilst PD can occur in younger people, the risk of developing the disease increases with age, and in a population that is forecast to have an increasingly high proportion of older people, the numbers of people with PD will increase further and affect more and more people.

1.1 What is Parkinson's disease?

PD was first described in detail by Dr James Parkinson in 1817 in his 'Essay on the shaking palsy', and the condition later became known as Parkinson's disease. There is no diagnostic test to confirm a diagnosis of PD. Instead, diagnosis is based upon clinical history and examination. Specific clinical criteria have been developed to

diagnose the condition - the most widely accepted criteria being the United Kingdom Parkinson's Disease Society Brain Bank criteria (See Table 1.1).

Table 1.1 UKPDS Brain Bank criteria for the diagnosis of PD (Gibb & Lees, 1988)

Step 1: Diagnosis of a parkinsonian syndrome

- Bradykinesia and at least one of the following:
- Muscular rigidity
- Rest tremor (4-6 Hz)
- Postural instability unrelated to primary visual, vestibular, cerebellar or proprioceptive dysfunction

Step 2: Exclusion criteria for Parkinson's disease

- Repeated strokes with stepwise progression
- Repeated head injury
- Antipsychotic or dopamine-depleting drugs
- Definite encephalitis and/or oculogyric crises on no drug treatment
- More than one affected relative
- Sustained remission
- Negative response to large doses of levodopa (if malabsorption excluded)
- Strictly unilateral features after three years
- Other neurological features: supranuclear gaze palsy, cerebellar signs, early severe autonomic involvement, Babinski sign, early severe dementia with disturbances of language, memory and praxis
- Exposure to known neurotoxin
- Presence of cerebral tumour or communicating hydrocephalus on neuroimaging

Step 3: Supportive criteria for Parkinson's disease (three or more required for definite diagnosis)

- Unilateral onset
- Rest tremor present
- Progressive disorder
- Persistent asymmetry affecting the side of onset most
- Excellent response to levodopa
- Severe levodopa-induced chorea
- Levodopa response for over five years
- Clinical course over ten years

Adapted by permission from the BMJ Publishing Group Ltd . Journal of Neurology, Neurosurgery and Psychiatry, Gibb and Lees, 55, 181-184, 1988.

The diagnosis of PD can be difficult in the early stages of the disorder. The symptoms develop gradually and it may take some time before they develop to a point when the individual decides to consult their doctor about them. The symptoms may be subtle at first, and attributed to ageing or to other illnesses. Every person is different in the way that the disease affects them and so different people may present with different initial symptoms. The symptoms seen in PD can also have other causes and so it often takes some time before a diagnosis of PD is made, as the other possible causes for the symptoms have to be considered and ruled out.

In clinical practice, diagnosis is usually based upon the presence of physical signs of at least two of the following:

- Slowness of movement (bradykinesia) – people with PD often find that it is difficult to initiate movements, that movement is slower, and/or that coordinating repeated or sequential movements is difficult. As a result of bradykinesia people with PD often walk dragging one leg to some degree or with short shuffling steps.
- Muscular rigidity – people with PD often find it difficult to turn over in bed, or get out of a chair. They can find it difficult to make small finger movements, like those required to fasten a button or for handwriting, and may find that their facial muscles become stiff, making facial expressions more difficult.
- Resting tremor – this usually affects one hand or arm and most often occurs when the affected limb is at rest. The tremor will disappear when the limb is being used and is slower than other types of tremor, with a frequency of about 4-6 per second. Not all people with PD display tremor but around 70% do (Frucht, 2004).

Later in the development of the disease postural instability becomes a problem. Impaired postural reflexes, necessary for stability when standing and walking, and adjustment of stance and easy turning, lead to poor balance and may increase the risk of falls. People with PD can develop a stooped posture, typically with knees slightly flexed and elbows, wrists and hands flexed.

1.2 Parkinsonism

Parkinsonism is an umbrella term for conditions with symptoms of tremor, rigidity and slowness of movement. PD is the most common form of parkinsonism, with cerebrovascular disorders and drugs (antipsychotic) as the most common cause of secondary parkinsonism (Thanvi *et al.*, 2005; Thanvi and Treadwell, 2009). Illnesses such as multiple system atrophy and progressive supranuclear palsy can also cause parkinsonism. Careful diagnosis is needed to ensure the correct aetiology for the symptoms is identified.

1.3 The difficulty of diagnosing PD

The pattern and nature of these symptoms of PD vary from person to person, and the symptoms fluctuate from day to day. Differential diagnosis can be even more difficult in older people who are more likely to have confounding comorbid illness, but also are more likely to show mild extrapyramidal symptoms (abnormal involuntary movements, alterations in muscle tone, and postural disturbances) on neurological examination even without any neurological or psychiatric illness (Grosset, 2006)[1].

[1] extrapyramidal symptoms were found in 32 and 35% of over 65s who have no known neurological or psychiatric illness (Richards *et al.*, 1993; Bennett *et al.*, 1996 respectively).

In a study by Hughes *et al.* (1992a, 1992b) they found that 24% of patients diagnosed with PD were misdiagnosed. Since this study Hughes *et al.* (2002) have reported improved accuracy of diagnosis to 98.6% amongst movement disorder specialists. This improvement is most probably due to increased awareness of the difficulty of differential diagnosis in PD and the use of operational criteria for parkinsonian syndromes other than PD (Hughes *et al.*, 2002). However, research has shown that in 6-8% of cases the diagnosis made by a specialist is later changed (NICE, 2006).

There remains a problem of misdiagnosis in the community, however. Meara *et al.* (1999) showed a high error rate (47%) in the diagnosis of PD in the community where patients had been diagnosed by the GP. Schrag *et al.* (2002) found that in a sample taken from GP practices at least 15% of patients diagnosed with PD did not have PD, and that 19% of patients selected as having tremor or anti-PD drugs who had not been diagnosed with PD in fact were found to have PD. That is, false-positive diagnoses in 15% of patients and false-negative diagnosis in 19% - even though a significant proportion of each group (74% and 38%) had seen a specialist at some point in the past. This indicates the importance of early referral of patients suspected to have PD for specialist assessment, and reassessment of patients, even after a diagnosis is made, as well as the application of diagnostic criteria to rule out alternative diagnoses.

In some patients it may be unclear to the clinician whether symptoms of tremor are due to PD or to essential tremor (a common condition that is not related to PD). A specialist SPECT (single photon emission computed tomography) scan is available which can be used to assist clinicians in making the differential diagnosis. For this scan a radioactive isotope is administered to the patient by injection. The radioactive iodine compound,

called DaTSCAN, passes through the blood brain barrier and is taken up by the dopamine transporters in the striatum area of the brain. The uptake of DaTSCAN will indicate the extent to which there is a loss of dopamine neurones and dopamine transporters and indicate to the clinician whether the symptoms of tremor are due to a Parkinsonian syndrome or not. This technique has been shown to differentiate between Parkinsonian syndromes and essential tremor in 96.5% of cases (European Medicine Agency, 2007).

1.4 What is the cause of Parkinson's disease?

PD is caused by a reduction in the amount of dopamine, a chemical neurotransmitter, in the brain, which relays signals associated with muscular activity. The reduction in dopamine levels is caused by the depletion in the dopamine-producing cells in the substantia nigra. Without enough dopamine the brain is no longer able to coordinate the body's movements normally. It is thought that motor symptoms appear when at least 50% of the pars compacta neurons in the substantia nigra have been lost and dopamine levels have dropped by around 80% (Bernheimer *et al.*, 1973). Other areas of the brain also show cell degeneration, albeit less consistently and severely, and account for the other symptoms that can be associated with PD – for example, the dorsal raphe nuclei, locus coeruleus, nucleus basalis of Meynert and the dorsal motor nucleus of the vagus.

Acetylcholine is another neurotransmitter which works in conjunction with dopamine, and the balance of these two neurotransmitters is essential to produce smooth movement. In PD when the levels of dopamine drop there is no longer sufficient dopamine to maintain this dopamine-acetylcholine balance, resulting in inability to

control bodily movement and producing symptoms such as muscle tension, tremor, joint rigidity and slow movement.

A pathologic feature that is a 'marker' of PD is the Lewy body, which is an abnormal aggregate of protein that forms in the affected nerve cells. Lewy bodies are not specific to PD as they also occur in dementia with Lewy bodies, and in Alzheimer's disease, but their presence in the substantia nigra (on histopathology) would indicate idiopathic PD (IPD).

1.5 Why do people get PD?

The pathogenesis of IPD is unclear, but most researchers believe that multiple factors, genetic and environmental, as well as ageing, are involved.

It is thought that between 15 and 25 percent of people who have PD have a close relative who also has the condition. People who have one or more close relative with PD have an increased risk of developing the disease themselves, but the risk is thought to be between two to five percent (Marder *et al.*, 2003) – only slightly higher than for the general population, unless the family member has an atypical presentation of the disease such as early-onset PD. These features suggest that genetic factors are at least partially involved in causing PD. In recent years several genes have been identified as being involved in PD. These genetic mutations appear to be involved, in different ways, with mechanisms that contribute to cell death.

Environmental factors are also thought to be significant factors contributing to the risk of PD. Recent research has focussed on environmental neurotoxins in diet, soil or water

supplies, or on the use of industrial or agricultural chemicals, such as herbicides and pesticides, in industrialised countries. Whilst no association has been established that would independently account for the worldwide distribution of the disease, it is possible that environmental factors may contribute to PD in individuals who have a genetic or age-related susceptibility those toxins (Nutt *et al.*, 1992). For example, a genetic defect that results in inadequate detoxification of environmental toxins, which in turn causes damage to nerve cells.

1.6 Incidence and prevalence of PD

It is estimated that PD affects 100-180 people per 100,000 of the population, and that each year between 4 and 20 people per 100,000 people are newly diagnosed with PD (NICE, 2006). This means that in the UK approximately 120,000 people have PD and about 10,000 people are newly diagnosed with the condition each year (PDS, 2005). The illness is predominately found in older people (over 50 years old), but in 1 in 20 cases the patient is under 40 years old when they are diagnosed (PDS, 2005).

Estimates of prevalence vary, because of variations in study design - in particular, the diagnostic criteria and sample populations used (Table 1.2). However, the prevalence of PD clearly increases with age (Table 1.3). Recent research suggests that, since the introduction of levodopa therapy, life expectancy for the majority of people with IPD is close to normal. However, a small but significant number develop severe end-stage parkinsonism that leaves them bedridden with increased risk of fatal infections (Herlofsen *et al.*, 2004).

Table 1.2 - Comparison of UK studies. Taken from Porter *et al.* 2006

Location	Year	Sources	Diagnostic criteria	Pop. size	Crude prevalence estimate x10^5	Estimated number of Cases per 100,000[1]
Carlisle	1962	H, GP	Clinical judgement	71,101	112	97 (78–116)
Northants	1982	H, GP	Webster Scale	208,000	108	134 (111–157)
Aberdeen	1984	H, GP, PD soc., drug	Clinical judgement	151,616	164	196 (169–223)
Northants	1992	H, GP, drug	UK PDS Brain Bank criteria	302,000	121	142 (119–165)
London	1997	GP	UK PDS Brain Bank criteria	121,608	128	171 (145–197)
North Tyneside	2002	H, GP, drug	UK PDS Brain Bank criteria	108,597	148	139 (116–162)

Figures in parentheses indicate 95% CI. [1] Standardised to 2001 UK population.
H = Hospital records; GP = GP records; drug = pharmacy records; PD soc. = local PD society.

Table 1.3 - Estimated age- and sex-specific prevalence rates for PD per 100,000 of the population in the UK. From Porter *et al.*, 2006

Age, years	Prevalence in males	95% CI	Prevalence in females	95% CI	Overall prevalence	95% CI
0–39	0	0	0	0	0	0
40–49	7	2–12	0	0	7	2–12
50–54	51	37–65	51	37–65	51	37–65
55–59	99	79–119	32	21–43	65	49–81
60–64	220	191–249	235	205–265	228	198–258
65–69	387	348–426	353	316–390	369	331–407
70–74	641	391–691	789	734–844	724	671–777
75–79	1,415	1,341–1,489	906	847–965	1,115	1,050–1,180
80–84	956	895–1,017	732	679–785	814	758–870
85–90	1,393	1,320–1,466	598	550–646	837	780–894
90+	0	0	1,425	1,351–1,499	1,134	1,068–1,200
Crude prevalence	142	119–165	153	129–177	148	124–174
Age-adjusted prevalence[1]	130	108–152	145	121–169	139	116–162

[1] Adjusted to the 2001 UK population.

1.7 Burden and financial cost of the disease

As the population in the UK ages, the number of individuals with PD will increase, and so too will the personal burden and financial costs incurred. Findley *et al.* (2003) conducted a survey to estimate the costs of PD in 1998. This survey included information regarding financial expenditure by the individual and their family, use of health and social resources, as well as the impact upon employment of the individual and their family. The total annual cost per patient in 1998, including NHS, social services and private expenditure, was £5993 per year. Direct NHS costs accounted for 38% of that cost, and social services for 34%. Hidden costs such as lost wages and informal care were substantial but not included in the analysis.

It was evident from analysis of the data that costs increase with age and also with disease severity. All areas of cost increased with disease severity but NHS costs tended to fall with increasing age, whilst social services and personal costs increased. By extrapolating their data, based upon there being 100,000 people in the UK with PD, Finley *et al.* estimated that approximately £599,300,000 is spent per annum (in 1998).

Whetten-Goldstein *et al.* (1997) completed a study of the costs of PD in USA in terms of direct costs to society, family and the individual. The survey was of 109 people with PD and the mean total 'societal cost' per individual was approximately $6000 per year. Family burden was not primarily health care costs but burden of providing informal care and loss of earnings. The authors stated that if the informal care was converted into the cost of equivalent formal care it was the single most expensive element of burden attributable to PD after lost earnings.

A more recent study indicated that the total costs of PD in Europe are approximately €11,153 per patient (direct health care €5,626, direct non-medical €4,417 and indirect costs €1109) which amounts to a total of €13.9 billion over 30 European countries (Olesen *et al.,* 2012).

1.8 Non-motor symptoms of Parkinson's disease

Although PD is defined by its motor symptoms people who have the disease commonly also experience non-motor symptoms. These symptoms are poorly recognised in clinical practice and are frequently missed during specialist reviews which tend to focus on the motor symptoms experienced by the patient (Shulman *et al.*, 2002). Non-motor symptoms associated with PD are wide-ranging - from gastrointestinal and sexual dysfunction to cognitive and psychiatric problems (see Table 1.4). Such symptoms occur often (Shulman *et al.*, 2001 reported 88% of a sample of PD patients reported at least one non-motor symptom) and are a source of major impairment and distress to both the patient and their families (Weintraub *et al.*, 2004; Karlsen *et al.*, 1999).

In a study reported by the Global Parkinson's Disease Survey Steering Committee (2002) the non-motor symptoms associated with PD were shown to be linked to quality of life, carer distress, admission to care and economic costs. Patients have rated problems such as depression, sleep problems, pain, apathy, and memory problems as symptoms which negatively impact upon their lives more than the motor symptoms of PD (Chaudhuri *et al.*, 2006).

Table 1.4 - Non-motor symptoms associated with PD.

Autonomic dysfunction	Gastrointestinal dysfunction (dysphagia, constipation) Urinary dysfunction (incontinence, nocturia, frequency) Sexual dysfunction (impotence, diminished libido) Impaired thermoregulation (sweating dysfunction) Orthostatic hypotension
Sensory phenomena	Pain Akathisia (inner restlessness) Restless leg syndrome
Psychiatric symptoms	Cognitive impairment / dementia Anxiety Psychosis Sleep disturbance Behavioural disorders Depression

Autonomic dysfunction

Autonomic dysfunction affects 70-80% of PD patients (Zesiewicz *et al.*, 2003). Gastrointestinal disorders, urinary symptoms, sexual dysfunction, problems with thermoregulation and orthostatic hypotension can occur. Some more commonly occur in the more advanced stages of PD (e.g. dysphagia, urinary problems) whilst others occur at any stage of the illness (e.g. constipation). Whilst many of the symptoms themselves are often not considered serious they have a significant impact upon the patient's quality of life and can lead to life-threatening complications if left untreated. Treatment of these symptoms can be challenging and consideration of concomitant medications that exacerbate the symptoms is important.

Sensory symptoms

Research suggests that 38-50% of patients with PD experience pain which is often non-specific, poorly localized, cramp-like ache which usually affects the side that is more

affected by the motor symptoms of PD (Mirza and Biglan, 2006). Patients can also experience coldness, numbness, burning sensations, inner restlessness or restless legs syndrome. Olfactory dysfunction is common in PD with most patients experiencing a diminished sense of smell, often from the very early stages of the disease (Dewey, 2003).

Cognitive impairment and dementia

The majority of people with PD will experience some degree of cognitive impairment, but the impairment is typically less severe and more selective than in dementia. Cognitive impairment is often evident early in the course of PD and is often predominantly associated with problems with executive function and verbal memory (Bassett, 2006). However, in 20-40% of people with PD the pattern of cognitive impairment progresses to become more marked and to affect more cognitive domains (Aarsland *et al.*, 1996; Tröster and Kaufer, 2006), although estimates vary greatly due to differing operational definitions of dementia and study designs. The spectrum of cognitive impairments in PD is often described as a continuum (Marder and Jacobs, 2008) – some patients with PD never develop cognitive impairment, but most people with PD develop some impairment in selected cognitive domains, and in a proportion of these patients the deficits progress to dementia. Studies suggest that those who are older at onset of PD, with longer duration of PD and have more severe motor symptoms are more likely to develop dementia (Dewey, 2003).

Anxiety

Anxiety disorders are common in PD and are estimated to occur in 20-40% of patients (e.g. Stein *et al.*, 1990; Chaudhuri *et al.*, 2006), although prevalence rates range as high

as 75% depending on the methodology and diagnostic criteria used. The extent of anxiety disorder does not correlate with the severity of PD, and is significantly more prevalent than in similarly disabled non-PD patients suggesting that the anxiety is not merely a reaction to the impairment associated with PD but is related to the underlying neuropathology of the disease (Menza and Dobkin, 2006).

Psychotic symptoms

Up to a half of patients with PD experience psychotic symptoms (Marsh, 2006) such as hallucinations, predominantly visual or auditory, and delusions. Prevalence estimates vary according to the sample and methodology used but visual hallucinations are experienced by nearly 50% of PD patients at some point in their illness, with auditory hallucinations in far fewer (8-13%) (Fénelon *et al.*, 2000). Hallucinations may result in behavioural disturbances if the patient lacks insight into the 'reality' of what they have perceived. In a survey of 123 PD patients 20% reported experiencing hallucinations, and 12% delusions in the preceding month (Chaudhuri *et al.*, 2006). Whilst psychotic symptoms may be due to the disturbances in neurotransmitters in PD they may also be caused by the drugs that are used to treat the motor symptoms of PD.

Sleep disturbance

Sleep problems are very common in PD and include insomnia, broken sleep, vivid dreams, restless leg syndrome and REM behaviour disorder. In a survey of symptoms experienced by PD patients in the preceding month 28% reported daytime sleepiness, over 40% reported insomnia, 31% intense vivid dreams, 33% acting out dreams and 37% restless legs (Chaudhuri *et al.*, 2006). Whilst sleep problems are common in the elderly, and in patients with some medical disorders, research has shown consistently

that these symptoms are significantly more prevalent in PD (Menza *et al.*, 2006) and that, of the non-motor symptoms associated with PD, they are one of those more strongly associated with poor quality of life (Karlsen *et al.*, 1999).

Behavioural disturbance

Behaviours that are characterised by their repetitive and reward based nature are now recognised as being a feature of PD (Voon & Fox, 2008). Hypersexuality, pathological gambling, compulsive shopping or eating, hobbyism and punding (an intense fascination with repetitive, complex, non-goal oriented behaviours) are all examples of this type of behaviour disturbance seen in PD. These behaviours usually occur without any distress on the part of the person with PD, but due to their compulsive and excessive nature can lead to serious negative consequences for both the patient and their family. Prevalence has been estimated at 6% (Voon *et al.*, 2006).

Depression

Depression is the most prevalent neuropsychiatric symptom associated with PD. Estimates of the proportion of people with PD who suffer from depression vary massively, between 2-90%, depending on the research design used. Depression in PD is strongly associated with decreased quality of life (Kuopio *et al.*, 2000; Schrag *et al.*, 2000; Global Parkinson's Disease Survey Steering Committee, 2002), carer distress (Aarsland *et al.*, 1999) and impaired functioning (Weintraub *et al.*, 2004). Therefore, it is important that the issues surrounding the difficulties in identifying depression in PD are addressed, so that diagnosis and effective treatment are possible. It is this issue upon which this thesis focuses.

1.9 DEPRESSION

Depression is characterised by pervasive and persistent low mood, loss of interest and enjoyment in usually pleasurable activities, and a range of associated emotional, cognitive, physical and behavioural symptoms. The emotional symptoms include feeling of worthlessness, guilt, low self-esteem, helplessness and suicidal ideation or attempts at self-harm or suicide. Behavioural and physical symptoms include tearfulness, irritability, reduced sleep, decreased appetite, fatigue and agitation or feeling slowed down. People with depression may also experience poor concentration and reduced ability to think. Individuals will vary in the patterns of symptoms they experience and the impact it has upon them but the symptoms of depression are pervasive and have an increasingly negative effect on an individual's social and everyday functioning.

Depression is best described as occurring on a continuum of severity, ranging from what is considered to be 'normal' low mood to clinically significant depression (Lewinsohn *et al.*, 2000) and is classified according to the numbers of symptoms as well as duration, and the level of social and functional impairment that results. Depression tends to be categorised according to the commonly used diagnostic criteria (for example the Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV) and the International Classification of Diseases, 10th edition (ICD-10)) which classify the degree of impairment according to the severity (DSM-IV: Major and Minor Depressive Episode (referred to as Major and Minor Depression in this thesis); ICD-10: mild, moderate and severe depressive episode). However, the thresholds at which depression is regarded as 'clinically significant' differ between the two main diagnostic systems. Nonetheless depressive symptoms below these criteria thresholds can be distressing and

disabling if persistent (NICE, 2009b). As a result of this, a further category of 'sub-threshold depression' is also increasingly recognised as an important group to identify (NICE, 2009b) in which several depressive symptoms are present but the patient has maintained function.

Recently published National Institute for Health and Clinical Excellence (NICE) guidelines (NICE, 2009a, 2009b) describe four severity groupings which should be viewed as "pragmatic subdivisions" rather than fixed separate categories, which can be used to guide the diagnosis and treatment of depression. These guidelines emphasise that a count of symptoms should not be the sole basis for a diagnosis of depression, and that severity and duration of symptoms, as well as factors such as personal history and family history of depression, should be taken into account with the four severity groupings:

- Sub-threshold depression (2-4 symptoms with maintained function)
- Mild depression (few, if any, symptoms in excess of 5 and only minor functional impairment)
- Moderate depression (symptoms or functional impairment between 'mild' and 'severe')
- Severe depression (several symptoms in excess of 5 and the symptoms markedly interfere with functioning)

A further diagnostic category is included in both DSM-IV and ICD-10 criteria – that of Dysthymia or Dysthymic Disorder. Dysthymia is characterised by chronic, less severe depressive symptoms that have persisted for most days for at least two years. The

pattern of symptoms experienced by the individual does not meet criteria for a diagnosis of depression, and usually everyday functioning is maintained.

It is estimated that between 4 and 10% of people experience major depressive disorder at some point in their lifetime, with between 2.5 and 5% suffering from Dysthymia (Waraich *et al.*, 2004). Depression is more prevalent in women than men (1.5 to 2.5 times higher, Waraich *et al.*, 2004) and is associated with various socioeconomic factors (such as unemployment and social deprivation) (NICE, 2009a). The economic consequences of depression are huge (over £9 billion in the UK; Thomas & Morris, 2003).

The aetiology of depression has been explained as being due to biochemical, psychological, genetic and social factors. These factors interact to affect an individual's susceptibility to depression, and the role of each factor varies between individuals. There is clear evidence that family history of depression increases the likelihood of depression (Kendler *et al.*, 2001), and that factors such as adverse life experiences, personality and social circumstances all contribute to vulnerability to depression. It is known that some medical illnesses increase the risk of depression, which supports the hypothesis that there is a biochemical basis to the cause of depression also.

Research has shown that depression is associated with chronic physical disease, particularly in older people (Dent *et al.*, 1999). Chronic illness can cause and exacerbate depression, and studies have indicated that aspects of illness such as multiple sources of pain, functional disability, and physical changes associated with the disease (biochemical or degenerative changes) lead to increased depression (NICE, 2009b).

There is also evidence to suggest that depression can increase the risk of physical disease such as coronary heart disease, myocardial infarction, stroke, back pain and irritable bowel syndrome (NICE 2009b).

Depression can have a huge impact upon the sufferer and those around them in terms of their quality of life, and their social and occupational functioning. It also has a substantial effect upon the physical health of the sufferer, increases the risk of illness, adversely affects the pain, distress and disability associated with physical illness and has a negative impact upon the outcome of illness including increased mortality. There is also a broader impact in terms of considerable health care costs and lost working days (Moussavi *et al.*, 2007).

1.10 Prevalence of depression in PD

Studies into the prevalence of depression in PD have varied both in terms of the sample studied and the criteria used to identify depression. Consequently estimates of prevalence vary greatly. Previously, authors have reviewed and summarised the wealth of studies that have attempted to establish the rates of depression in PD. In his review of 26 studies Cummings (1992) reported the mean frequency of depression was 40%, with a range of 4-70%[2]. However, many of the earlier studies he included (the studies were published between 1922-1990) did not use standardised rating scales nor diagnostic criteria, and the later studies did not utilise "random or comprehensive sampling, and all were subject to patient selection biases" (p444).

[2] Prevalence estimates of Major Depression in adults in primary care settings range between 5 and 13%, with lower estimates in people over 55 yrs (O'Connor *et al.,* 2009).

In 2001 Slaughter *et al.* conducted a review of 45 studies published between 1922 and 1998 in order to assess prevalence of depression in PD. Many of the studies included here were the same as those reviewed by Cummings (1992). However, when Slaughter *et al.* included only those studies that used DSM criteria (DSM-III or DSM-IIIR; American Psychiatric Association, 1980, 1987) for depression, 24.8% were diagnosed with Major Depressive Episode (Major Depression). If numbers of patients with Minor Depression or Dysthymia were also included the proportion with depression increased to 42.4%.

Veazey *et al.* (2005) carried out a systematic review of articles reporting the prevalence of depression in PD. They included 16 studies. Of those that used clinical interviews to apply DSM criteria for Major Depression prevalence estimates ranged from 7.3% to 32%. The prevalence of Minor Depression (defined as Dysthymia) ranged from 9.7% to 31%. Studies based on self-report questionnaires resulted in higher prevalence of depression (27.3% to 69%).

More recently Reijnders *et al.* (2008) carried out a systematic review of prevalence studies and a meta-analysis of the data. Having excluded studies of insufficient quality or those based on the same datasets, 36 studies were included in their review. Prevalence rates across studies were analysed as weighted means (i.e. individual study prevalence rate multiplied by that sample size, divided by total sample size of all the studies in the analysis). In those studies that included a clinical interview to establish DSM criteria Major Depression was present in 17% of patients, Minor Depression in 22% and Dysthymia in 13%. Overall, including those studies which used a cut-off on a

rating scale to indicate depression, 35% of patients had clinically relevant depressive symptoms.

Reijnders *et al.* also investigated the difference that sampling techniques made to the prevalence of depression found. Population based studies reported the lowest prevalence both of major depressive disorder as well as clinically relevant depressive symptoms, and studies of outpatient and inpatient samples showed much higher prevalence (see Table 1.5). These significant differences, according to the sample studied, indicate the importance of taking into account the design of the study when considering the prevalence reported.

Table 1.5 – The prevalence of depression according to the type of sample. Taken from Reijnders *et al.*, 2008.

	Major Depression (%)	Clinically relevant depressive symptoms (%)
Community Sample	8.1	10.8
Outpatient Sample	24.0	40.4
Inpatient sample	21.7	54.3

The review by Reijnders *et al.* also highlighted the importance of the means by which patients were assessed for depression, and the impact that this may have on the prevalence of depression found. Those studies that used semi-structured clinical interview to establish DSM criteria found 2.3-55.6% (weighted mean 19%) of patients with Major Depressive disorder. Those that applied DSM criteria without a structured interview found 2.9-7.7% (weighted mean 7%) of patients with Major Depression and studies using rating scales resulted in the highest mean proportion with significant depressive symptoms (42%). See Table 1.6.

Table 1.6 – The prevalence of depression according to method of assessment, Reijnders *et al.*, 2008.

	Major Depression	Weighted mean	Clinically relevant depressive symptoms	Weighted mean
DSM criteria in semi-structured interview	2.3% -55.6%	19%	2.7%-57.8%	33%
DSM criteria	2.9% -7.7%	7%	7.3%-47%	27%
Cut-off on rating scale	--	--	13-89%	42%

Studies using cut-off scores on a depression rating scale vary in the cut off used to indicate Major Depression. This will obviously impact on the prevalence estimates between studies. Furthermore, such scales were not developed to be diagnostic instruments and may lead to inaccurate estimates of the frequency of diagnosable depression (Starkstein *et al.*, 1990a).

A study carried out by Braam *et al.* (2010) assessed the prevalence of depression and parkinsonism in a huge community sample of Europeans aged over 65 years old. Of those who had parkinsonism (1.9% of the sample) 22.2% had depression according to the depression scale cut-off used in the study, compared to 12.4% in those without parkinsonism. The depression rate when more stringent DSM-IV criteria were used was 12.3%. These results are in keeping with the ranges reported by Reijnders *et al.* (2008).

Some surveys of depression in PD have also attempted to identify the proportion of patients suffering from less severe forms of mood disturbance as well as those suffering from Major Depression. Liu *et al.* (1997) found the prevalence of Major Depression in a sample of PD outpatients was 16.5%, whereas 25% met criteria for Dysthymia or other depressive disorder (depressive disorder not otherwise specified). Another study

by Tandberg *et al.* (1996) found only 7.7% of their community sample of PD patients met DSM-IIIR criteria for Major Depression but that over 45% had a mild depressive symptoms according to the Montgomery Asberg Depression Rating scale. This suggests that a high proportion of patients with PD experience clinically significant depressive symptoms but do not necessarily meet diagnostic criteria for a depressive illness (sub-threshold depression) and that this group remains under-represented in surveys of depressive disorder in PD.

A recent study by Reiff *et al.* (2011) investigated a sample of 110 PD patients – 25.5% met criteria for sub-threshold depression and 26.4% met criteria for Major or Minor Depression or Dysthymia. They found that those with sub-threshold depression had significantly poorer quality of life compared to non-depressed patients, indicating that this is a group of patients who warrant identification despite not meeting the criteria for Minor or Major Depression.

Surveys focussing on quality of life of patients with PD have consistently reported high levels of depressive symptoms. Kuopio *et al.* (2000) found 53% of their community-based PD sample reported possible or probable depression according to the depression scale they used. They also found that depression was by far the most important impairing factor for quality of life. A population-based study of PD patients by Schrag *et al.* (2000) found 15% of their sample recorded depression scale scores that indicated moderate to severe depression, and that depression scores were the most important predictive factor of quality of life scores.

A large international survey of health related quality of life in PD found 50% of participants had significant depressive symptoms indicating at least mild depression (Global Parkinson's Disease Survey, 2002) although only 1% of patients considered themselves depressed. The survey found that depression had an "overwhelming" impact upon quality of life.

1.11 Aetiology of depression in PD

There are two theories regarding the aetiology of depression in PD. These theories are not mutually exclusive and it is likely that both are factors in the development of depression in PD. Firstly, depression can be "reactive" to the disability and psychological stress of having PD and the prognosis of a disease that is progressive and for which there is no cure. The patient and their families have to come to terms with the diagnosis, and adjust to managing the illness as well as the possible impact that it may have on their job, family, and social life. This theory is supported by research that has indicated a higher rate of depression in people at the time of first diagnosis (e.g. Starkstein *et al.*, 1990a).

However it seems that the pattern of depression in PD is not fully explained as a reaction to the psychosocial stress of the illness. Research has indicated that depressive symptoms are more common in PD than in other chronic medical illness with comparable impairment (Ehmann *et al.*, 1990, Nilsson *et al.*, 2002), suggesting that perhaps there is also a pathophysiological cause related to the neurodegenerative changes in the brain associated with PD (e.g. Weintraub *et al.*, 2006a). This is the second theory regarding the basis of depression in PD.

It has been suggested that neuropathological changes in the brain in PD (serotonergic, noradrenergic and dopaminergic deficits) contribute to depression in PD (Aasland and Cummings, 2002). Post-mortem and imaging studies have shown the degeneration of subcortical neurons in certain areas of the brain (for example the ventral tegmental area and substantia nigra) result in the depletion of dopamine, serotonin and noradrenaline (Brown and Gershon, 1993) – neurotransmitters that are known to be involved in depression.

In addition to such neurodegenerative changes, some treatments for PD (e.g. levodopa) lead to rapid changes in availability of dopamine. Some patients on levodopa experience depressive symptoms during 'off' periods (the period before the next dose when the therapeutic effects of the medication are waning) but not the rest of the time, and this fluctuating mood can sometimes involve marked changes – for example from catastrophic depression to euphoria within a very short time.

For these patients their depressive symptoms are associated with the levels of their medication and its affect upon their motor functioning, suggesting that changes in dopamine levels play a role (Brooks and Doder, 2001). However, research has shown that changes in mood are not consistently correlated with the motor fluctuations, which one would expect if dopamine deficiency were the only cause of the depressive symptoms. In addition, replacement of dopamine does not treat depression in the majority of depressed PD patients. This indicates that mood and motor symptoms have different aetiologies, and that dopamine levels are unlikely to be the sole explanation of depression in PD (Richard *et al.*, 2001; McDonald *et al.*, 2003).

Some research has indicated that depressive symptoms may precede the development of motor symptoms – suggesting some common biochemical aetiology (Nilsson *et al.*, 2001, Leentjens *et al.*, 2003a). Major Depression is more common in patients with specific clinical features of PD, i.e. akinetic-rigid type symptoms (Starkstein *et al.*, 1998) and in patients with right-sided motor symptoms (e.g. Starkstein *et al.*, 1990a), again supporting the hypothesis that there is a neurochemical dysfunction underlying the association with depression.

Research has failed to establish the pattern of the relationship between depression and stage of PD although some research has indicated higher rates of depression early in the illness and again in the later stage of the illness as symptoms have progressed (Starkstein *et al.*, 1990a; Brown and Jahanshahi, 1995; Weintraub, 2004). This pattern may reflect the psychological reaction to the diagnosis in the early stage of the illness, and in the later stages the depressive symptoms are as a result of the underlying neurodegenerative changes as well as a reaction the increased impairment of function and ADL (some research has also indicated an association between depression and cognitive impairment, van der Hoek *et al.,* 2011*).* Research has shown a correlation between depression and degree of impairment in ADL (Schrag *et al.*, 2001). However, it appears that patients' perceptions of handicap are even more important than actual disability (Ehmann *et al.*, 1990) indicating that whilst there may be a biological basis for the depression there is likely to be a significant influence of individual psychological factors too (Tandberg *et al.*, 1997).

1.12 Impact of depression in PD

Depression in PD has been shown to be associated with greater cognitive decline (Starkstein *et al.*, 1990a, 1990b, 1992), deterioration in functional ability (Liu *et al.*, 1997; Starkstein *et al.*, 1992, Weintraub *et al.*, 2004), worse motor symptoms of PD (Cubo *et al.*, 2000) and increased mortality (Hughes *et al.*, 2004), when compared to patients with no depression. It has also been shown to be associated with greater medical and psychiatric comorbidity (Chen *et al.*, 2007a). Depression has a negative impact on the quality of life of the patient and their family (e.g. Aarsland *et al.*, 1999; Kuopio *et al.*, 2000; Caap-Ahlgren & Dehlin, 2001; Schrag *et al.*, 2000) and was identified as the single most important factor affecting quality of life, over and above motor symptoms of the disease in a survey of people with PD (Global Parkinson's Disease Survey Steering Committee, 2002).

In light of these negative associations and consequences of depression in PD, and the fact that depressive symptoms respond to treatment, it is important that the diagnosis of depression is made and that the depression is treated specifically rather than focussing treatment exclusively on the motor symptoms of the disease (Schrag *et al.*, 2001).

1.13 Problems with diagnosing depression in chronic illnesses

The Diagnostic and Statistical Manual of Mental Disorders (DSM-IV, APA, 1994) criteria for depressive disorder are the most commonly accepted means by which a standardised diagnosis of depression is established (see Figure 1.1) although the International Classification of Diseases (ICD-10, WHO, 1992) is also widely used (see Appendix 1). The two classificatory systems are very similar but differ mainly in terms of their definitions of severity. This may result in the different classification of

individual patients, or identification of slightly different groups of patients, when using DSM-IV or ICD-10 diagnostic criteria (Andrews *et al.*, 2008). For example, in a study by Wittchen *et al.* (2001) significantly more patients were diagnosed depressed according the ICD-10 criteria than by DSM-IV criteria (11.3% cf 4.2%).

In order to meet criteria for DSM-IV Major Depression an individual needs to experience at least five of the nine specified symptoms, including a persistent and pervasive depressed mood, or loss of interest or pleasure in doing previously pleasurable activities, for a duration of at least two weeks. The nine specified symptoms of depression are emotional, behavioural, cognitive and somatic. A diagnosis of Minor Depression involves the presence of fewer symptoms (only two of the nine symptoms) and less impairment but the same duration as Major Depression. Dysthymia involves depressed mood accompanied by at least two of the specified symptoms for at least two years. (It is important that, particularly in clinical practice, a diagnosis of depression is not made purely based on a 'count' of criterion symptoms, but that it is made taking into account the duration, severity and level of impairment resulting from those symptoms).

Figure 1.1 DSM-IV Criteria for Major Depressive Episode

A. Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; a least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure.

   Note: Do not include symptoms that are clearly due to a general medical condition, or mood-incongruent delusions or hallucinations.

   1) Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g. feels sad or empty) or observation made by others (e.g. appears tearful). Note: in children and adolescents, can be irritable mood.
   2) Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation made by others).
   3) Significant weight loss when not dieting or weight gain (e.g. a change of more than 5% of body weight in a month), or decrease or increase of appetite nearly every day. Note: in children, consider failure to make expected weight gains.
   4) Insomnia or hypersomnia nearly every day.
   5) Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down).
   6) Fatigue or loss of energy nearly every day.
   7) Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick).
   8) Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others).
   9) Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.

B. The symptoms do not meet criteria for a Mixed Episode.
C. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
D. The symptoms are not due to the direct physiological effects of a substance (e.g. a drug of abuse, a medication) or a general medical condition (e.g. hypothyroidism).
E. The symptoms are not better accounted for by bereavement, i.e. after the loss of a loved one, the symptoms persist for longer than 2 months or are characterised by marked functional impairment, morbid preoccupation with worthlessness, suicidal ideation, psychotic symptoms, or psychomotor retardation.

Minor Depressive Episode

Requires only 2 of the 9 symptoms above, but one must be either depressed mood or loss of interest or pleasure.

The distinction between patients who have symptoms that meet criteria for a standardised diagnosis of depression and those who do not is increasingly recognised as being restrictive and lacking utility in routine clinical practice. This has resulted from the realisation that a substantial proportion of patients who have less severe forms of depression nevertheless have clinically significant symptoms that may benefit from treatment. In the recent NICE guidelines for the treatment of depression for example (NICE, 2009a), this group is defined as having between two and four DSM-IV criterion symptoms, with maintained function (sub-threshold depression).

Patients who have medical illnesses often have symptoms in common with those of depression, particularly the somatic symptoms, making it difficult to recognise depression when it occurs. For example, loss of appetite, lack of energy or fatigue, general slowing of movement or difficulty concentrating are all symptoms that are typical for people who are medically unwell, regardless of whether they are depressed or not (Koenig *et al.*, 1997). It is therefore dependent upon the clinician to identify whether such symptoms are indeed a direct physical consequence of the medical illness or instead, a symptom of depression. However, this is often not an easy distinction to make, particularly when a symptom, for example fatigue, may have both a physiological *and* a psychological component.

DSM-IV diagnostic criteria for Major Depression specify that, in patients with a general medical condition, symptoms should be 'counted' towards a diagnosis of Major Depressive Episode "*except when they are clearly and fully accounted for by a general medical condition*" (p323) (DSM-IV, American Psychiatric Association, 1994). DSM-

IV further specifies that Major Depressive Disorder should also be diagnosed if the depressive episode "*is considered to be psychological consequence of having the general medical condition or if there is no etiological relationship between the Major Depressive Episode and the general medical condition*" whereas 'Mood Disorder due to a General Medical Condition' should be diagnosed only "*if the mood disturbance is judged to be the direct physiological consequence of a specific general medical condition*" (p343).

This requires the clinician to make a distinction between whether a symptom is a *physiological* or a *psychological* consequence of the medical illness. In some symptoms it is possible that there is both a physiological and psychological component and it is down to the clinician to decide whether to attribute the symptom to the medical illness or a depressive episode. This requirement inevitably leads to inconsistency and unreliability in diagnoses of depression in people who are medically ill, as clinicians have varying thresholds of when to attribute symptoms to depression (Koenig *et al.*, 1997). This phenomenon is likely to be more pronounced in the diagnosis of depression in older people who are medically ill, in whom somatic symptoms of depression are more common (both in those who are depressed and not depressed) (Koenig *et al.*, 1993).

It is worth noting that diagnostic criteria in DSM-III and DSM-III-R offered less guidance on the attribution of somatic symptoms to a diagnosis of depression than in DSM-IV. In DSM-III-R clinicians were instructed to decide whether a symptom is attributable to psychiatric causes, i.e. a symptom counts towards the diagnosis of depression, only if the rater judges them clearly <u>not</u> to be due to a physical condition.

("diagnosis is made only if it cannot be established that an organic factor initiated and maintained the disturbance…" p219). This relies upon the clinician's experience and expertise in assessing the likely aetiology of a symptom and tends to result in a less inclusive approach (compared to DSM-IV) regarding counting somatic symptoms towards diagnosis of depression (Koenig *et al.*, 1995).

Much research has reported the under-recognition of depression in medically ill patients (e.g. Koenig *et al.*, 1992; Zimmerman *et al.*, 2006) and that the diagnosis (and indeed treatment) of depression in medically ill patients is dependent upon both clinician and patient characteristics. In particular, the clinician's experience with geriatric depression is a significant factor - suggesting that doctors who have less experience of depression in older patients are less confident or likely to attribute symptoms to depression (Koenig, 2007).

The issues and challenges regarding the diagnosis of depression in medically ill patients has led to the question of whether 'conventional' diagnostic criteria are appropriate in this patient group, and whether they should have the same 'benchmark status' as in other patient groups (Wilhelm *et al.*, 2004). Depending upon the interpretation of the diagnostic criteria, and attribution of symptoms, the rate of diagnosis of depressive disorder is greatly affected. For example in a study by Koenig *et al.* (1997) the rate of Major Depression varied between 21% (inclusive attribution of symptoms) to 16.5% (aetiologic attribution of symptoms) and for Minor Depression between 25% (inclusive) to 14% (aetiologic)[3]. This disparity between the two approaches casts uncertainty on

[3] inclusive attribution of symptoms means that all symptoms are counted towards the diagnosis of depression, regardless of whether the rater feels the symptom is due to medical (aetiological) or psychological causes. Under the aetiologic method (as per DSM-IV) criterion symptoms of Major

the intended 'gold standard' nature of the diagnostic criteria in people who are medically ill.

The use of an 'exclusive' method of diagnosis, where symptoms that are considered to be affected by physical illness are excluded from counting towards a diagnosis of depression, tends to lead to fewer patients meeting criteria for depression (Kathol *et al.*, 1990) largely because many patients no longer have the required number of depressive symptoms. Therefore it has been proposed that where an item is excluded it is replaced with a psychological symptom, in order that the number of symptoms that can contribute to a diagnosis of depression remains the same (Endicott, 1984). This system would allow for the adjustment of diagnostic criteria to take into account the relationship between somatic symptoms of depression and physical illness without restricting the identification of depression by simply excluding all somatic symptoms and it has been suggested that diagnostic criteria modified in this way should be the 'gold standard' for the diagnosis of depression in medically ill patients (Kathol *et al.*, 1990).

Cavanaugh (1995) proposed that additional clarifying guidelines for the DSM-IV criteria for depression would improve consistency of diagnosis in patients with medical illness, rather than substitution of somatic symptoms (which can "hamper communication if the diagnostic criteria are too different from established criteria"). Cavanaugh also highlighted the importance of careful consideration of the symptoms of depression within the context of what may be considered 'reasonable' given the severity of the patient's physical illness.

Depression were only counted if the clinician-rater judged that they were not explained by the patient's medical condition.

In light of these issues some studies have suggested that concentrating on the cognitive symptoms of depression in medically ill patients will discriminate those who are depressed from those who are not (Steer *et al.*, 1999; Parker *et al.*, 2001). Wilhelm *et al.* (2004) compared self-report depression scales, which were comprised of items relating to cognitive symptoms of depression, with clinicians' impressions and with standard DSM-IV diagnostic criteria. They found that the cognitive-symptom based scales were able to identify patients with Major Depressive Episode, as well as other DSM-IV affective diagnoses, as they did not rely on vegetative symptoms of depression that can be confounded by medical illness. However more research is needed before such a system could be regarded as a reliable substitute for a careful diagnostic interview.

1.14 Issues in diagnosing depression in PD

The problem of symptom overlap particularly applies in the diagnosis of depression in patients with PD. Symptoms such as tiredness, lack of energy, psychomotor retardation, mental slowing, impaired concentration, reduced appetite, and insomnia may occur in both depression and PD – making it difficult for clinicians to attribute symptoms to PD or a depressive illness. In addition, sadness and low mood may be an appropriate and transient reaction to having a chronic progressive illness like PD, and this must be differentiated from a persistent and limiting mood disorder. Literal application of diagnostic criteria for depression and interpretation of somatic symptoms in PD may theoretically result in either over-diagnosis (i.e. symptoms of PD are incorrectly attributed to depression) or under-diagnosis (i.e. depressive symptoms are incorrectly attributed to PD) of depression.

A study by Hoogendijk *et al.* (1998) assessed prevalence of Major Depression in PD using both inclusive and exclusive diagnostic methods (i.e. either including a depressive symptom regardless of its possible origin or excluding any depressive symptom that may be attributable to PD). The inclusive approach resulted in prevalence of 23% whereas the exclusive approach gave a prevalence of 13%. The inclusive approach has the advantage of being most likely to identify those cases of clinically significant depressive symptoms and does not require clinicians to make any judgements regarding the likely aetiology of symptoms. However, it may result in over-diagnosis of depression. This means that the diagnostic approach used will affect the sensitivity and specificity of the criteria, and the most appropriate method should be adopted depending on the aim of the research or patient care.

In the study by Hoogendijk *et al.* the criterion items that were most frequently 'excluded' and showed overlap with PD symptoms were those relating to loss of interest, psychomotor retardation and loss of energy, with loss of interest being the item that made the most difference to diagnosis as it is a mandatory item for Major Depression according to the DSM-III-R criteria used. The authors proposed that the loss of interest item should no longer be regarded as a valid sole mandatory item (i.e. in the absence of depressed mood) for the diagnosis of depression in PD as it is frequently seen as a symptom of PD in non-depressed patients.

In 2003 a group of experts in the field (from the National Institute for Neurological Diseases and Stroke (NINDS), and the National Institute of Mental Health (NIMH)) met to discuss diagnostic criteria for depression in PD (Marsh *et al.*, 2006). The Work

Group considered that when applying DSM criteria for Major Depression an inclusive approach should be taken to the attribution of depressive symptoms. This would "enhance the sensitivity and reliability of the diagnostic criteria" (p151), and would provide greater likelihood of identifying the most cases of clinically significant depression, and is more reliable and easier to implement as the clinician is not required to judge the likely aetiology of symptoms.

This advice is supported by the NICE guidelines on PD which recommends that clinicians should have a "low threshold" for diagnosing depression in PD (NICE, 2006).

Gómez-Esteban *et al.* (2009) applied the NINDS/NIMH working group recommendations regarding diagnostic criteria for depression in PD and found that it resulted in significantly lower frequency of Major Depression but higher Minor Depression (17.8% and 28.0% under DSM-IV criteria compared to 7.6% and 37.3% following the working group recommendations). The authors also reported difficulty in distinguishing between symptoms of anhedonia and apathy.

The Work Group also made recommendations relating to the diagnosis of those patients with clinically significant depressive symptoms that do not meet the criteria for Major or Minor Depression. The group recognised that such "non-major" forms of depression accounted for a high proportion of PD patients with depressive symptoms, but that it is commonly overlooked. As research has suggested untreated Minor Depression in PD commonly progresses to Major Depression (Starkstein *et al.*, 1992) identification and inclusion of this group of patients in research may be beneficial. In light of this, the Work Group further recommended that the non-major depressive disturbances be

included when considering depression in PD in order to reflect the full spectrum of depressive illness. They proposed that the core criterion of "markedly decreased interest" should be omitted when considering a diagnosis of Minor or subsyndromal depression due to its "diagnostic ambiguity" (because the symptom is common to depression, apathy or dementia).

Starkstein *et al.* (2008) carried out a study to validate the symptoms of depression that are used in DSM-IV diagnostic criteria for Major, Minor, and sub-threshold depression and Dysthymia in patients who have PD. They concluded that all the criteria for a diagnosis of DSM-IV Major Depression and Dysthymia were significantly associated with clinically significant sad mood, and that the full DSM-IV diagnostic criteria should be used to diagnose Major Depression and Dysthymia in patients with PD. They found that diagnostic criteria for Minor Depression were valid but only when sad mood was present, and that criteria for sub-threshold depression in PD needed further validation.

A recent paper published by Starkstein *et al.* (2011) described a study in which the pattern of depression symptoms (the 9 symptoms included in DSM-IV criteria for depression) in 259 PD patients were analysed using latent class analysis. The analysis demonstrated three patient classes – those with severe depression, those with moderate depression and those with no depression. The authors stated that the results of the analysis validated the use of DSM-IV criteria for Major Depression in PD without any modification.

There is some research which suggests that the pattern of depressive symptoms in PD is different to those in other neurological diseases or primary major depression (Zesiewicz

et al., 1999; Lemke *et al.*, 2004) with PD patients with depression showing less self-blame, guilt, delusions and sense of failure, and less frequent thoughts of self-harm but more anxiety (Henderson *et al.*, 1992) irritability, sadness, pessimism about the future, and suicidal ideation (Burn, 2002).

Anguenot *et al.* (2002) reported that although certain symptoms indicative of depression are frequently seen in non-depressed PD patients, there are symptoms that are never seen in PD patients who do not have depression. They concluded therefore, that feeling that life is not worth living, hopelessness, feeling worthless and incompetent, having low energy and morning sadness are characteristic of depression in PD.

Under-recognition of depression in PD by clinicians was reported by Shulman *et al.* (2002) where physicians only identified depression in 35% of those who scored above the threshold for depression on a standardised depression rating scale (the Beck Depression Inventory). However, the majority of those patients reporting depressive symptoms scored in the mild to moderate range for depression. The authors concluded that the under-identification of depression in PD is partly due to the fact that patients do not tend to volunteer information about their mood to clinicians spontaneously, and also that clinicians do not tend to specifically ask patients about such symptoms. This phenomenon was demonstrated in the Global PD Survey (2002) - although 50% of patients reported clinically significant depressive symptoms only 1% had revealed their symptoms to their clinician.

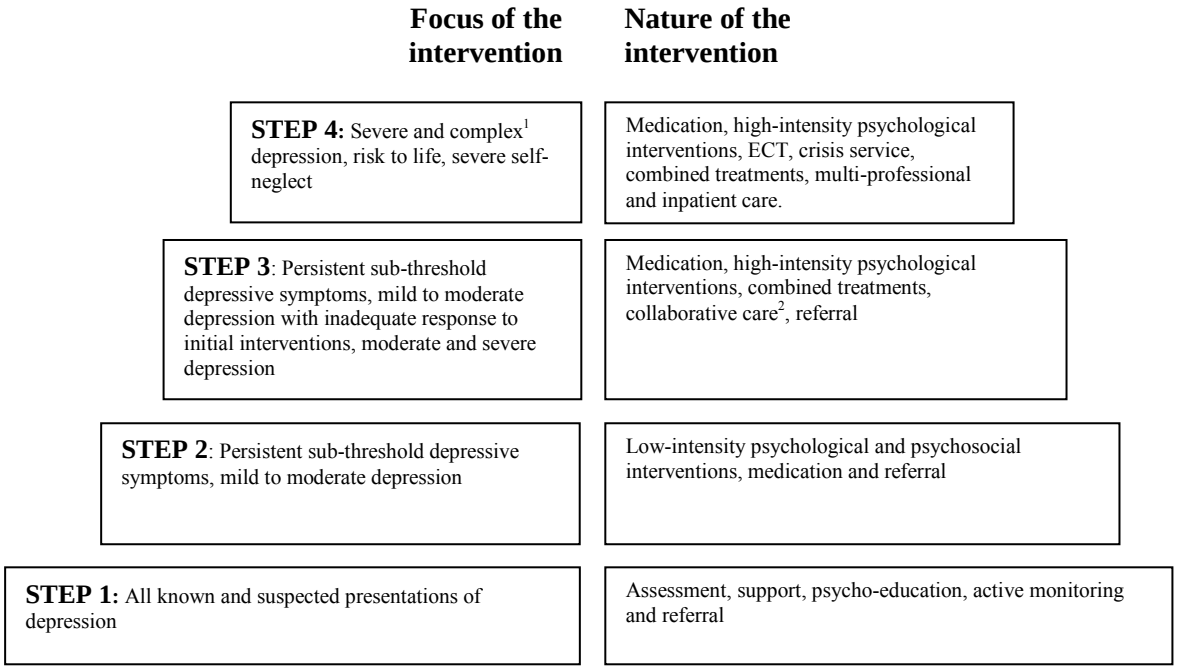
In a study by Weintraub *et al.* (2003) only 35% of PD patients who met diagnostic criteria for depression were receiving antidepressant treatment, and this included 56%

of those in the sample who had been diagnosed with Major Depression. This suggests that a high proportion of patients with PD who have depression are undiagnosed and untreated. A survey of antidepressant treatment in patients with PD found that the frequency of treatment was not significantly different from that of older patients without PD, suggesting that it is under-recognition of depression that accounts for the significant proportion of depressed PD patients who do not receive treatment for the depression (Chen *et al.*, 2007b).

1.15 Treatment for depression in PD

NICE guidance (NICE, 2009a) recommends a "stepped-care" model for the treatment for depression, in which the least intrusive (and least costly) intervention is provided first. If that intervention fails to benefit the patient, or if they decline that intervention, an intervention should then be offered from the next step (see Figure 1.2 below). Routine outcome monitoring is essential if such a model of care is followed, to ensure that those who have not benefitted from a low level intervention are then offered an intervention from the next step.

Figure 1.2 - The stepped care model. (NICE, 2009a)



[1] Complex includes depression with an inadequate response to multiple treatments, complicated by psychotic symptoms, and/or significant psychiatric comorbidity or psychosocial factors
[2] Only for depression associated with chronic physical illness and associated functional impairment

Antidepressant medication, psychological therapy and electro-convulsive therapy (ECT) are all treatments that are utilized for the treatment of depression. However, it is apparent that there is a lack of rigorous research evidence for all of these treatments in people with PD. The majority of trials tend to be small or poorly designed studies that fail to give clear evidence regarding the efficacy and tolerability of the treatment in PD. Clear guidance and evidence for antidepressant medication is important, particularly in light of the possibilities that some antidepressants may make PD symptoms worse and interactions between antidepressants and dopaminergic treatment for PD are possible.

There is some research that has investigated the antidepressant properties of some antiparkinsonian medications. Dopaminergic medications, such as seligiline, amantadine, bromocriptine and pramipexole can have an antidepressant effect but there

has been little formal evaluation of their effectiveness as a treatment for depression in PD (Burn, 2002). One large double-blind placebo-controlled study evaluated the efficacy of pramipexole on depressive symptoms in PD (Barone *et al.*, 2010). There were significantly greater improvements in depression scale, activities of daily living and quality of life scale scores for patients on pramipexole. The authors of that study suggested that optimal dopaminergic treatment should be the initial treatment strategy for depression in PD before prescription of antidepressant medication is considered (Barone, 2011).

There are three main groups of antidepressant drugs; tricyclic antidepressants (TCA), monoamine oxidase (MAO) inhibitors and selective serotonin reuptake inhibitors (SSRIs). Tricyclic antidepressants and monoamine oxidase inhibitors work by blocking the re-uptake of the neurotransmitters noradrenaline and serotonin, thus increasing the levels available in the brain. Tricyclic antidepressants can have a sedative effect and may produce other unwanted negative side-effects, such as worsening autonomic symptoms and cognitive impairment. They may also slow the absorption of levodopa, which has important implications for the motor symptoms of PD. SSRIs selectively block the re-uptake of serotonin, and some suggest that SSRIs cause worsening of PD symptoms. SSRI medications can interact with selegiline to cause serious side-effects ("serotonin syndrome") but this is now thought to be quite rare (0.24% of patients, Richard *et al.*, 1997). The characteristics of medications within each category vary in terms of their mode of action, anticholinergic effect and associated side-effects.

In 2002 the Movement Disorder Society commissioned a review of the available evidence regarding pharmacological treatment of depression in PD (Movement Disorder

Society, 2002). Although the task force identified a significant number of published studies, few met their quality criteria for inclusion in the review. They described one randomised-controlled trial (RCT) of nortriptyline (a TCA), which showed a significant reduction in depression scale scores. There was no worsening of PD symptoms during the treatment phase of the trial but orthostatic hypotension was a significant side-effect for two of the twenty-two patients (Andersen *et al.*, 1980). This study was small (22 participants) and did not use an established depression rating scale.

The task force review also described two trials of MOA inhibitors in PD (moclobemide and selegiline). One study showed significant improvements in depression scores, but did not have a placebo group, whilst the other did not show any significant improvement compared to placebo. Both trials were small (10 and 15 patients). Some of the patients in these trials experienced worsening of motor symptoms as well as nausea and hypotension. Two open-label studies of SSRI medication (paroxetine) were described in the review. Both showed clinically significant improvement in mood after treatment, but 20% of the patients discontinued treatment because of significant side-effects.

A Cochrane review of treatments for depression in PD was carried out by Ghazi-Noori *et al.* (2003). They identified only three randomised controlled trials (RCTs) of antidepressant medication and none for behavioural therapies that met their inclusion criteria. The review found that all three trials had flaws in the study design and so the authors felt unable to make any recommendations regarding the efficacy or safety of antidepressant medication in PD.

Although there are few RCTs of antidepressants in PD, there are several open-label trials that have indicated that antidepressant medications are beneficial in PD and are reasonably well tolerated (Weintraub *et al.*, 2006a). A more recent review of antidepressant treatment studies was described by Weintraub *et al.* in 2005, and included both open-label and RCT trials. Eleven studies were included in the meta-analysis, and the results indicated an equally large treatment effect for both the active medication and placebo. The authors felt that, in light of the fact that trials of antidepressant treatment in non-PD elderly patients show active treatment to be superior to placebo, the lack of superior performance of active treatment in PD patients in the meta-analysis was due to poorer response of PD patients to the active treatment. The majority of the studies included in the analysis were trials of SSRIs, limiting the conclusions that can be drawn about other classes of antidepressants. However, interestingly over 87% of all patients, including 86% of patients taking SSRIs, completed their treatment (compared to 79% of those on placebo) – suggesting that antidepressants, and in particular SSRIs, are well tolerated in PD patients.

In light of the paucity of rigorous research evidence for the treatment of depression in PD the NICE guidance for the management of PD (2006) stated that "there is insufficient evidence from RCTs of the efficacy or safety of any antidepressant therapy in PD" (p116). The NICE recommendations were based on the Cochrane review (Ghazi-Noori *et al.,* 2003), and two additional RCTs that compared antidepressant medication to placebo or another active medication, as the meta-analysis by Weintraub *et al.* (2005) was not available at the time the recommendations were prepared. There were no RCTs of electroconvulsive therapy or behavioural therapy in patients with PD. The NICE guidelines further recommend that whilst it is tempting to adopt the

guidelines for management of depression for people with physical disorders it is important to consider that depression in PD may require different strategies. The guidelines recommend that clinicians take into consideration the possibility that some antidepressants may make PD symptoms worse, and may interact with some PD medications.

In 2009 Menza *et al.* published the results of an RCT of nortriptyline (a TCA), paroxetine (a SSRI) and placebo for the treatment of depression in PD (Menza *et al.*, 2009a). Although a relatively small study (52 patients), it is the largest that has been carried out so far in this patient group. Nortriptyline showed significant benefit in terms of improvement in depression scale scores, whilst those treated with paroxetine did not. Both treatments were well tolerated although paroxetine caused more side effects than placebo, whereas nortriptyline did not. This trial was small in terms of the number of participants and of relatively short duration, assessing patients after 8 weeks of treatment. A larger sample size and longer follow-up would have perhaps better demonstrated the efficacy and tolerability of both drugs.

Another study (of 48 patients) of the TCA desipramine compared to the SSRI citalopram was carried out by Devos *et al.* (2008). Patients on the TCA showed benefit sooner (after 14 days treatment) than the SSRI but both medications showed significant improvement in depression scale scores after 30 days of treatment. However, there were twice as many mild adverse events for the TCA than for the SSRI.

A recent RCT (Richard *et al.*, 2012) was carried out to evaluate the efficacy of a SSRI and a serotonin and norepinephrine reuptake inhibitor (SNRI) for the treatment of

depression in 115 PD patients over a 12 week period. Both treatments showed greater reductions in depression scale scores compared to placebo, and did not show any significant side effects or worsening of motor symptoms.

A survey of treatment of depression in patients with PD (Chen *et al.*, 2007b) showed that clinicians tend to prescribe SSRI antidepressants far more than any other type of antidepressant medication. For example, out of 7868 patients 63% were prescribed SSRI medication, compared to only 7% of patients who were on a TCA, possibly reflecting the prescribing practice for depression in general. The RCT by Menza *et al.* however, suggests that the TCA nortriptyline may be a more effective and well tolerated treatment than the SSRI paroxetine. The study by Devos *et al*. indicated that both TCAs and SSRIs can be an effective treatment but suggested that the TCA was less well tolerated than the SSRI. On the whole, there remains a lack of substantial research evidence regarding the use of all types of antidepressant medications in patients who have PD (both in terms of efficacy and tolerability) and until such information is available clinicians will be unable to make informed, evidence-based decisions regarding their choice of treatment for depression in this patient group.

The Movement Disorder Society Task Force on evidence-based medicine recently carried out a review of treatments for the non-motor symptoms of PD (Seppi *et al.,* 2011). Up to date conclusions on the efficacy of treatments for depression in PD were made. The only treatment for which the Task Force members felt there was clear evidence of benefit was pramipexole. They felt there was evidence that tricyclic antidepressants nortriptyline and desipramine are *likely* to be effective for the treatment of depressive symptoms in PD but that there is insufficient evidence to make

conclusions regarding the efficacy of the TCA amitriptyline, all SSRIs and the newer types of antidepressants.

Another important factor for clinicians to consider is that the treatment given is effective and that patients are reviewed and offered alternative treatment if they fail to respond to the first course of treatment. Weintraub *et al.* (2003) surveyed the use of antidepressants in 100 PD patients attending an outpatient clinic. They found that 37% of patients taking antidepressants still met criteria for Major Depression – suggesting that a significant proportion of PD patients who are prescribed antidepressants either receive suboptimal treatment or do not respond to the treatment they are given. Considering that effective treatment will bring about improvements in quality of life, physical functioning, and emotional well-being (Menza *et al.*, 2009b) it is important that clinicians not only commence treatment with a suitable antidepressant but that they also review and optimise that treatment.

A high proportion of PD patients with depressive symptoms do not meet criteria for the diagnosis of Major Depression, and evidence for treatment of these patients remains to be demonstrated. Evidence of the efficacy of antidepressant medication and psychological treatments for patients with Minor Depression and Dysthymia is limited, and suggests only a small or moderate benefit (Ackerman & Williams, 2002; Williams *et al.*, 2000). Weintraub *et al.* (2005) suggest that studies assessing treatments of depression in PD that include patients with these diagnoses (as well as Major Depression) are likely to show less treatment effect, as patients with Minor Depression or Dysthymia are thought to be less likely to respond to treatment than those with more severe depressive symptoms.

1.16 Non-pharmacological and psychosocial treatments for depression

Electro-convulsive therapy (ECT) is an established treatment for depression, which is mainly used following failure of a patient to respond to treatment with antidepressant medication. Strong research evidence regarding the efficacy of ECT for the treatment of depression in PD is lacking. The MDS review of treatments described one study that compared depressed PD patients with depressed non-PD patients who received treatment with ECT. The authors reported significant improvement in depression for both study groups. The PD patients did not show a significant worsening of PD motor symptoms but did appear to be more susceptible to "ECT-related complications" (temporary delirium). This study was relatively small (25 patients in each group) and ratings were made by an unblinded research nurse. Several studies have demonstrated that transcranial magnetic stimulation (TMS), a treatment involving the delivery of electrical stimuli to highly localised region of the brain, can be beneficial in the treatment of depression in PD (Veazey *et al.*, 2005) but clear evidence of benefit remains to be established (Burn, 2002).

Given the possibility of side-effects, and interaction of antidepressants with PD medication psychosocial interventions for the treatment of depression may be preferable. Indeed many patients with PD may be reluctant to take another medication in addition to their medication for PD. However there has been little research into the efficacy of psychotherapeutic treatments in PD. Whilst some studies have described the use of such interventions (e.g. CBT) with patients with PD, in particular during the early 'adjustment' phase of the disease, until recently they have failed to assess the impact upon depressive symptoms. Recently Dobkin *et al.* (2011) conducted the first RCT of

CBT for depression in PD. The study included 80 patients who met DSM-IV criteria for depression who received either 10 weeks of CBT or clinical monitoring. The CBT group showed greater improvements in depression scale scores, and several secondary outcome measures, compared to those who received usual clinical monitoring.

Overall the evidence base for both pharmacological and non-pharmacological treatments for depression in PD is patchy and insufficient for definite guidelines for clinicians. There is significant support for the use of antidepressant medications from both open-label trials and clinical practice, but the benefit remains to be established in rigorous large-scale RCTs for each of the categories of antidepressant medication available. It is apparent that clinicians prescribe antidepressant medication for depression in patients with PD but clear evidence regarding effective doses and treatment options should initial trials fail to benefit is essential for the effective management of such patients.

1.17 Screening for depression

It is important to screen medically ill patients for depression because depression complicates their medical treatment, prolongs recovery, and increases disability and impacts upon their social functioning (Steer *et al.*, 1999). In addition, depression has a negative impact upon adverse health behaviour and adherence to health care regimens (Katon, 2003).

Screening involves the routine or systematic testing of patients with a measure designed to identify those individuals who are at risk of having or developing a particular disorder, and who may benefit from further investigation and appropriate management.

Screening tests do not diagnose a condition but indicate that an individual may have or be at risk of developing the condition in the future, and highlights that they should undergo further investigation, such as a diagnostic interview, to establish whether they do indeed have the disorder.

Earlier diagnosis, intervention and management will potentially reduce suffering, improve outcomes and reduce costs associated with the disorder. Usually, carrying out a full diagnostic assessment of every patient to detect an illness would be costly and difficult to do, so screening is a more viable alternative.

Screening measures are never 100% accurate and will result in some patients being incorrectly identified as being at risk, whilst others are incorrectly identified as being not at risk. Those identified as false-positives are subjected to further unnecessary testing, and those who are identified as false-negatives will not be given the further investigation that they need. The decision to carry out clinical screening for a condition involves assessment of the relative costs of errors, both false-positives and false-negatives. A false-negative result (i.e. missing a patient with the condition) will lead to continued morbidity and its associated outcomes. A false-positive (incorrectly identifying a patient with the condition) will lead to unnecessary distress to the patient, unwarranted time, effort and cost of diagnostic testing and perhaps even unnecessary treatment. The choice of screening instrument should be made with the relative cost of missing a diagnosis compared to making an incorrect diagnosis in mind.

The ability of a test to correctly identify patients with the condition being tested for is known as the sensitivity (expressed as the proportion of positive cases that are correctly

identified by the test). As sensitivity increases the number of false-negatives decreases. The ability of a test to correctly identify patients who do not have the condition is known as the specificity (expressed as the proportion of negative cases who are correctly identified by the test). As specificity increases the number of false-positives will decrease (Altman, 1999).

$$\text{Sensitivity} = \frac{\text{Number of True Positives}}{\text{Number of True Positives} + \text{Number of False Negatives}}$$

$$\text{Specificity} = \frac{\text{Number of True Negatives}}{\text{Number of True Negatives} + \text{Number of False Positives}}$$

For the purposes of screening it is important that the tool used has high sensitivity. Specificity is less important than sensitivity in a screening tool, as further investigation prompted by the positive screening result will then rule out any false-positives whilst having highlighted, ideally, all of the positive cases. However, low specificity will result in more well patients being subjected to unnecessary further investigation. Specificity is important in a diagnostic tool, as it is important that non-cases are identified as such, and not incorrectly classed as positive cases.

The positive predictive value (PPV) of a test is the proportion of positive screening test results that do actually have the illness. The negative predictive value (NPV) is the proportion of negative test results that correctly identify individuals who do not have the illness. Both the PPV and the NPV are dependent upon the prevalence of the illness in the population being studied. When the prevalence of an illness is high there will be a higher PPV and a lower NPV and even a measure that has good sensitivity and specificity will have low PPV if the population prevalence of the disorder is low

(<10%). This is because PPV increases as the prevalence of the disorder increases. Therefore although sensitivity and specificity are important and do not depend upon the prevalence of the illness, it is the PPV that is most relevant to clinicians as it is the proportion of patients with a positive result that actually have the illness that is important to them. Low PPV makes the test less acceptable to clinicians as it will result in a high proportion of patients being false-positives and therefore followed-up unnecessarily.

A high negative predictive value is important for screening tests as it reflects the proportion of patients with negative test results who do not have the condition being screened for – therefore a high NPV indicates that few patients with the condition will be missed.

$$\text{Positive Predictive Value} = \frac{\text{Number of True Positives}}{\text{Number of True Positives + Number of False Positives}}$$

$$\text{Negative Predictive Value} = \frac{\text{Number of True Negatives}}{\text{Number of True Negatives + Number of False Negatives}}$$

		Actual Diagnosis / Gold Standard		
		+ve	-ve	
Screening Test	+ve	True Positives (TP)	False Positives (FP)	→ TP/TP+FP = Positive Predictive Value (PPV)
	-ve	False Negatives (FN)	True Negatives (TN)	→ TN/TN+FN = Negative Predictive Value (NPV)
		→ TP/TP+FN = Sensitivity	→ TN/TN+FP = Specificity	

The prevalence of a condition in the population being tested is important when considering screening tests – in populations in which the prevalence is high the number of false-negative results will be higher, in populations with a lower prevalence the number of false-positives will be higher. Therefore the choice of screening test, the cut-off or threshold score for the test, and the likely prevalence of the condition being screening for are important factors in clinical screening.

Another statistic often used when considering the performance of a screening test are likelihood ratios. The likelihood ratio is the ratio of the probability of a specific test result in people who do have the disease to the probability in those who do not. In other words, the likelihood ratio summarises how many times more (or less) likely patients with the disease are to have a positive (or negative) result than patients without the disease (Deeks and Altman, 2004). A positive likelihood ratio (PLR) above 10 and a negative likelihood ratio (NLR) below 0.1 are considered to provide strong evidence of a diagnosis or no diagnosis respectively (Deeks and Altman, 2004). The Diagnostic Odds Ratio is the ratio of the odds of the test being positive if the patient has a disease relative to the odds of the test being positive if the patient does not have the disease (i.e. PLR/NLR). Likelihood ratios are not dependent upon the prevalence of the disease being screened for and indicate the value of a test for increasing certainty about a positive diagnosis. They are calculated as follows:

$$\text{Positive Likelihood Ratio} = \frac{\text{Sensitivity}}{1 - \text{Specificity}}$$

$$\text{Negative Likelihood Ratio} = \frac{1 - \text{Sensitivity}}{\text{Specificity}}$$

The evidence of the benefit of screening programmes, in primary care or in other care settings, is not as clear as one might expect. The use of a screening or case-identification tool must change clinician behaviour in a way that improves not only the recognition of an illness but also its management, and ultimately the clinical outcome, otherwise it is a waste of time (clinician and patient) and money (Gilbody *et al.*, 2001). The benefits of such a programme need to justify the cost both in terms of resources as well as the potential inconvenience, or indeed harm, to patients who are screened who do not have the condition. Arguments for screening for depression would seem strong - bearing in mind the prevalence of depression in some patient groups, the poor rates of recognition and treatment of depression, and the potential benefits such as improved quality of life, social and occupational functioning, and morbidity (Palmer and Coyne, 2003).

In 2002 Pignone *et al.* carried out a review of evidence regarding the efficacy of screening for depression in order to inform recommendations by the U.S. Preventative Services Task Force (USPSTF). They concluded that although use of a screening test for depression, followed by feedback of the results to the clinician, increased the recognition of depression it did not necessarily translate into an improvement in clinical outcomes unless the screening programme was associated with the facility for accurate diagnosis following a positive screening result, and appropriate treatment and follow-up of cases. There was evidence however, that screening for depression that is accompanied by enhanced systems and patient care, can have positive benefit (Wells *et al.*, 2000).

A later meta-analysis of studies of screening for depression by Gilbody *et al.* (2005, 2008) supported the conclusion that standalone screening programmes "have little or no impact on the detection and management of depression in clinicians" and suggested that "recommendations to adopt screening strategies using standardised questionnaires without organisational enhancements are not justified." (Gilbody *et al.*, 2008, p997). The authors did conclude that there is some evidence to suggest that the selective feedback of only positive screening results to the clinician may be more effective than the feedback of all screening results, and that the feedback of results of routine screening is most effective in the context of an educational programme for clinicians and specialised referral and management of cases (Gilbody *et al.*, 2005).

In 2002 the U.S. Preventative Services Task Force published recommendations regarding screening for depression (USPSTF, 2002). They recommended screening for depression in adults in primary care settings "that have systems in place for accurate diagnosis, effective treatment and follow-up". The group concluded that although small benefits of screening have been demonstrated in studies where the results of the screening test are fed-back to the clinician there can be significant benefits in clinical outcome where screening takes place with coordinated resources for diagnosis, treatment and follow-up. These recommendations were supported by O'Connor *et al.* when they conducted an updated review of the evidence for the USPSTF in 2009.

Since the review by Pignone *et al.* (2002) the Canadian Task Force on Preventative Health Care (MacMillan *et al.*, 2005) also made recommendations that include the use of screening or case-finding instruments for depression, where the screening is included as part of enhanced patient care. The UK National Institute of Health and Clinical

Excellence guidelines for depression in 2004, and the most recent guidance (2009a and 2009b), recommend screening for depression in patient groups that are known to be at high risk of depression.

Gilbody *et al.* (2008) question that screening is an essential component of enhanced care for depression, and propose that other aspects of improved patient care (for example, patient follow-up, staff trained in mental health care, and clinician support) also achieve positive results even in the absence of use of a screening tool (see also Gilbody *et al.*, 2006) and some authors have suggested it may be more cost-effective to invest in improving these aspects of care than in screening (Palmer and Coyne, 2003).

It has been suggested that a proportion of those patients identified as positive cases with a depression screening test, who do not meet diagnostic criteria for depression (i.e. false positive cases), will be patients whose depressive symptoms are sub-threshold or who are dysthymic, and whose depression is most likely to be missed during routine clinical care (O'Connor *et al.*, 2009). Whilst screening might result in identification of patients who have significant depressive symptoms who would otherwise have not been identified, they are also a patient group whose symptoms may resolve without treatment and for whom a period of 'watchful waiting' may be appropriate before active treatment is considered (NICE, 2009a) and for whom evidence of treatment benefit is limited (Palmer and Coyne, 2003). Furthermore, patients with mild depressive symptoms and less impairment who are identified by the screening tool may be less inclined to accept or benefit from treatment (Palmer & Coyne, 2003).

The effectiveness of a screening programme is dependent upon a number of factors including the tool used, feedback of screening results to the treating clinician, accurate diagnosis following screening, appropriate treatment and follow-up, clinician confidence in the benefits of screening (in order for them to invest the time required), and patient willingness to accept screening, and if diagnosed, willingness to accept treatment (Palmer & Coyne, 2003; Gilbody *et al.*, 2006). In light of these factors there is some question regarding whether screening for depression can be cost-effective – a screening programme must have low cost of screening (in terms of production of the test as well as clinician time), high prevalence of the disorder, and result in a high proportion of positive cases receiving treatment and high remission rates (Valenstein *et al.*, 2001). As the effectiveness of a screening strategy is dependent upon clear benefits in terms of successful treatment of a significant proportion of cases identified it is important that effective treatment and appropriate follow-up are in place prior to screening taking place.

Whilst it would be unethical to screen for an illness if clinicians fail to, or are unable to act appropriately upon the results, Hickie *et al.* (2002) suggest that failure of healthcare professionals to attempt to increase the identification of high prevalence mental disorders such as depression, which can have serious health consequences, may also be regarded as unethical.

1.18 Methods and scales used to screen for depression

The UK National Screening Committee specify that a screening tool should be simple, safe, precise and validated; that a suitable cut-off level should be defined and agreed for the target population; and that the test should be acceptable to the population (2003). If

routine screening is to be feasible and incorporated into clinical care the screening tool needs to be quick and short.

There are many relatively short instruments available that are commonly used for screening for depression ranging from 30-item scales to ultra-short scales consisting of only one or two items (e.g. Mahoney *et al.*, 1994; Chochinov *et al.*, 1997). Short screening tests have the advantage of being quick and therefore more acceptable than lengthier scales, and have lower cost in terms of clinician and patient time. Longer scales have the advantage of including items relating to different symptoms and so may have the advantage of fewer false positives (greater specificity). Mitchell and Coyne (2007) reviewed evidence of the efficacy of ultra-short depression screening tests and concluded that single-item tests performed poorly compared to two or three-item tests, although these produced rates of false-positives that the authors felt would be unacceptable in clinical practice (74% sensitivity and 75% specificity).

Pooled analysis of case-identification instruments carried out by Williams *et al.* (2002) concluded that use of screening measures in primary care would "advance current practice significantly" (in terms of recognition of depression) and that most of the instruments included in the study (ranging from ultra short to 30-item scales) showed acceptable performance and should be selected according to the demands of the situation in which they are to be used (e.g. reading ability, ease of response selection).

Previous NICE guidance (2004) has recommended use of two short questions to screen for depression in patient groups thought to be at higher risk of depression. These questions were taken from the Primary Care Evaluation of Mental Disorders Procedures

(Spitzer *et al.*, 1994) which were based upon the core criteria for depression in DSM, and are;

1) "*During the past month, have you often been bothered by feeling down, depressed or hopeless?"*
2) "*During the past month, have you often been bothered by having little interest or pleasure in doing things?"*

Whooley *et al.* (1997) compared the performance of these two screening questions with several other commonly-used case-identification instruments. The two-question method performed favourably (96% sensitivity and 57% specificity) compared to the other instruments, with similar results and had the advantage of its brevity and therefore high acceptability to clinicians and patients. Verbal presentation of these items in a primary care setting has also been validated (97% sensitivity and 67% specificity, Arroll *et al.*, 2003). Arroll *et al.* (2005) have suggested that the addition of a third item ("Is this something with which you would like help?"), following the two screening questions, increases the specificity of the screening test to 89%.

However, a study into screening patients with musculoskeletal pain questioned the effectiveness of screening using the two questions used by Whooley *et al.* in that patient group, and found significantly different results if the screening questions were presented verbally or for self-completion in written format (21% and 51% screened positive respectively, Mallen and Peat, 2008).

Recent updates of the NICE guidance (2009a and 2009b) continue to recommend that clinicians ask these two questions of patients who are at high risk of depression, in

particular "people with a past history of depression or a chronic physical health problem with associated functional impairment" (2009a, p108). These guidelines go on to recommend that if a patient responds positively to either question they should receive a further mental health assessment either by the clinician conducting the initial screening or by referral to a clinician with appropriate training and expertise to do so. This recommendation, to screen patients felt to be at high risk of depression, has been included in the incentive programme for GPs in the UK - the Quality and Outcomes Framework.

There are however many other scales that are commonly used for screening for depression. The recent revised NICE guidelines for depression (2009a, 2009b) included a meta-analysis of depression screening scales and compared their performance against either DSM-IV or ICD-10 diagnostic criteria (see p86, NICE 2009b). The analysis included the following scales:

- 9 and 2 item Patient Health Questionnaire (PHQ-9, Spitzer *et al.*, 1999; PHQ-2, Kroenke *et al.*, 2003)
- 2 Whooley questions (Whooley *et al.*, 1997),
- Beck Depression Inventory (21 items) (Beck *et al.*, 1961)
- Beck Depression Inventory fast-screen (13 Items) (Beck *et al.*, 1997)
- General Health Questionnaire (12 and 28 items) (Goldberg and Williams, 1991),
- Hospital Anxiety and Depression Scale (7 items) (Zigmond and Snaith, 1983),
- Centre for Epidemiological Studies Depression Scale (20 items) (Radloff, 1977),
- 15 and 30-item Geriatric Depression Scale (Yesavage *et al.*, 1983; Sheikh and Yesavage, 1986)
- some single-item measures.

Of the scales included in the analysis the Whooley two-question scale had the highest sensitivity (95%), but less good specificity (66%). The PHQ-9 and the GDS-15 had better specificity (83% and 75%) but inferior sensitivity (82% and 87%) compared to the Whooley items.

1.19 Screening for depression in PD

NICE guidance for depression (2004) recommended that "screening should be undertaken in primary care and general hospital settings for depression in high-risk groups" (p14) and the updated guidelines (NICE, 2009a) recommend that clinicians be "alert to possible depression (particularly in people with a past history of depression or a chronic physical health problem with associated functional impairment) and consider asking people who may have depression two questions…." (p108).

Because of the high rate of depression (either meeting diagnostic criteria or sub-threshold depressive symptoms) in PD, and the high proportion that are undiagnosed, it is important that clinicians endeavour to identify patients who are experiencing significant depressive symptoms. A logical and effective way of increasing the recognition of depression is through screening. It is important that if depression rating scales are used they are appropriate for use in this patient group and that the threshold/cut-off scores used are appropriate and set to maximise sensitivity of the instrument (McDonald *et al.*, 2003). Therefore it is essential that the threshold scores are appropriate for use with patients with PD as many depression scales such as BDI, and the Hamilton Depression Rating Scale include items relating to somatic symptoms and so may result in over-identification of depression due to symptom overlap. Screening for depression is further complicated by the fact that the majority of PD

patients with depressive symptoms do not meet criteria for Major Depression but have Minor Depression or Dysthymia, or sub-threshold depressive symptoms (McDonald *et al.,* 2003).

A number of studies have assessed the performance of various depression scales for the purposes of screening for depression in patients who have PD. The instruments that have been studied include the Hospital Anxiety and Depression Scale, the Hamilton Depression Scale, the Montgomery-Asberg Depression Rating Scale, the Beck Depression Inventory and the Geriatric Depression Scale (see Table 1.7 below).

The HADS (Hospital Anxiety and Depression Scale, Zigmond and Snaith, 1983) is a measure designed to assess anxiety and depression in people with physical health problems. It comprises two sub-scales of 7 items each. As it was designed specifically for use with hospital patients it has few items relating to somatic symptoms of depression, although the items "I feel as if I am slowed down" and "I can sit at ease and feel relaxed" are perhaps subject to symptom overlap with PD. Leentjens *et al.* (2001) felt the performance of this scale in patients with PD was "disappointing" as the cut-off that maximised sensitivity and NPV, suitable for screening for depression, had extremely low specificity (51%). The scale performed better however, when assessed by Mondolo *et al.* (2006) against the criterion of the HamD. Further validation of the scale has not been carried out, although the psychometric properties of the scale have been examined and deemed satisfactory (Rodriguez-Blazquez *et al.*, 2009).

The HamD (Hamilton Depression Scale, Hamilton, 1960) is a 21-item scale which requires completion by a trained interviewer in a semi-structured interview. The last

four items of the scale do not measure intensity of depression and so are commonly dropped to give a 17-item version the scale. There are several items pertaining to somatic symptoms. Leentjens *et al.* (2000b) found that the HamD-17 could be used as a valid screening instrument for depression in PD, and identified a disease-specific cut-off for screening for depression (sensitivity 94%, and specificity 75%), but felt that easier, self-report questionnaires would be more practical for routine screening (Naarding *et al.*, 2002).

The HDI (Hamilton Depression Inventory, Kobak and Reynolds, 1999) is a 17-item user-rated scale that is based upon the HamD-17. A study by Dissanayaka *et al.* (2007) showed satisfactory performance of the scale as a screening and diagnostic test, although it was slightly inferior to the HamD-17 and GDS-15.

The MADRS (Montgomery-Asberg Depression Rating Scale, Montgomery and Asberg, 1979) is a 10 item scale which is rated by a clinician, with three of the items relating to physical symptoms. Several studies have assessed the performance of the MADRS as a screening tool in PD. Leentjens *et al.* (2000b) found the concurrent validity of the scale with DSM-IV criteria was high and that if a low cut-off is selected it could be a valid screening tool (sensitivity 88%, specificity 89%). However, the scale was designed to be completed by a trained observer so is less practical for routine screening for depression than a self-rated scale.

Table 1.7 - Sensitivity, Specificity, PPV, NPV of depression scales in PD, and the prevalence of depressive illness and number of subjects in the study sample.

Scale	Study	criterion	Cut-off	Se %	Sp %	PPV %	NPV %	Prev. Dep.	N
HADS	Leentjens *et al.*, 2001	DSM-IV Δ	10/11	92	51	34	96	22%	55
	Mondolo *et al.*, 2006	HamD-17	10/11	100	95	71	100	11%	46
HamD-17	Leentjens *et al.*, 2000b	DSM-IV	11/12	94	75	56	97	25%	63
	Naarding *et al.*, 2002	DSM-IV Δ	9/10	95	72	51	98	24%	85
	Weintraub *et al.*, 2006b	DSM-IV	9/10	88	78	52	96	22%	148
	McDonald *et al.*, 2006	DSM-IV	12/13	81	82	58	93	23%	248 #
	Dissanayaka *et al.*, 2007	DSM-IV	9/10	94	74	52	98	23%	79
HDI	Dissanayaka *et al.*, 2007	DSM-IV	11.5/12	83	80	56	94	23%	79
MADRS	Leentjens *et al.*, 2000b	DSM-IV	14/15	88	89	74	96	25%	63
	Silberman *et al.*, 2006*	DSM-IV	8	72	82	72	82	39%	46
BDI	Leentjens *et al.*, 2000a	DSM-IV	8/9	92	59	39	96	23%	53
	Silberman *et al.*, 2006*	DSM-IV	10	78	61	56	81	39%	46
	Visser *et al.*, 2006	DSM-IV Δ	14/15	71	90	nr	nr	19%	92
	Tumas *et al.*, 2008	DSM-IV Δ	17/18	100	76	nr	nr	24%	50
	Schneider *et al.*, 2010	DSM-IV	14/15	78	85	49	96	15%	207
GDS-30	Ertan *et al.*, 2005	DSM-IV	8/9	91	60	71	86	51%	109
	Mondolo *et al.*, 2006	HamD-17	10/11	100	76	33	100	11%	46
	McDonald *et al.*, 2006	DSM-IV	9/10	81	84	58	94	23%	213 #
GDS-15	Weintraub *et al.*, 2006b	DSM-IV	4/5	88	85	61	96	22%	148
	Dissanayaka *et al.*, 2007	DSM-IV	4/5	94	75	53	98	23%	79
	Weintraub *et al.*, 2007. *<65yrs*	DSM-IV	4/5	85	84	nr	91	35%	57
	Weintraub *et al.*, 2007. *65-75yrs*	DSM-IV	4/5	89	82	nr	89	31%	88
	Weintraub *et al.*, 2007. *>75yrs*	DSM-IV	5/6	90	90	nr	96	25%	81

Table 1.7 continued - Sensitivity, Specificity, PPV, NPV of depression scales in PD, and the prevalence of depressive illness and number of subjects in the study sample.

Scale	Study	criterion	Cut-off	Se %	Sp %	PPV %	NPV %	Prev. Dep.	N
	Tumas *et al.*, 2008	DSM-IV Δ	8/9	91	92	nr	nr	24%	50
	Thompson *et al.*, 2011	DSM-IV	4/5	85	79	52	95	22%	214
CSDD	Williams & Marsh, 2009§	DSM-IV	6/7	83	73	54	92	27%	134
WHO-5	Schneider *et al.*, 2010	ICD-10	12/13	88	74	37	97	15%	210
PHQ-9	Thompson *et al.*, 2011	DSM-IV	--	54	85	49	87	22%	214

* this study focussed on optimal cut-off for diagnostic test, the optimal cut-off for screening is quoted here, taken from the reported results. nr not reported
# this study included multiple assessments (up to 4) of 52 subjects over 60 months.
§ this study, unlike all others listed above, included patients with cognitive impairment (37%)
Δ results reported for diagnosis of DSM-IV Major Depression only

The BDI (Beck Depression Inventory, Beck *et al.*, 1961) is a widely used depression rating scale that has 21 items. It comprises several (7) items that relate to somatic symptoms of depression. Levin *et al.* (1988) studied the performance of the BDI in PD and concluded that the inclusion of the many somatic items did not adversely affect the validity of the scale. Leentjens and colleagues (Leentjens *et al.*, 2000a; Visser *et al.*, 2006) examined the validity of the BDI as a screening tool in PD and concluded that "the psychometric properties of the BDI for PD populations are not ideal". They felt the standard cut-off score was not appropriate in a PD patient group and that no alternative cut-off score resulted in both high sensitivity and specificity.

The GDS (Geriatric Depression Scale, Yesavage *et al.*, 1983) is a scale developed specifically for use with older people. There are 30 and 15-item versions of the scale. Due to the higher frequency of somatic symptoms in older adults, even those who are not depressed, the GDS has only one item that relates to the physical symptoms of

depression ("Do you feel full of energy?"). Mondolo *et al.* (2006) evaluated the GDS-30 as a screening tool for depression in PD and concluded that its performance was adequate when judged against the criterion of the HamD-17. Ertan *et al.* (2005) and Weintraub *et al.* (2006b) have showed that the GDS-15 is an effective screening instrument in PD, which is brief and user-completed. Weintraub *et al.* (2007) assessed the discriminant validity of the scale in younger (<65yrs) compared to older (65-75 and 75+yrs) PD patients. They found the GDS performed similarly well in all age groups, although their results indicated a slightly different cut-off for optimal performance for the oldest age group.

The short WHO-5 well-being index was validated as a screening tool for depression in PD by Schneider *et al.* (2010). It is a 5-item scale that is positively worded – resulting in scale where a higher score indicates better well-being. The study indicated that the WHO-5 was a suitable screening tool for use in PD, with satisfactory sensitivity and specificity, although the PPV was quite low.

The Patient Health Questionnaire (PHQ-9) is a self-report scale that is specifically based upon DSM-IV criteria for depression. It was recently evaluated in a PD setting by Thompson *et al.* (2011) who found that it did not perform as well as the GDS-15 as a depression screening measure, which had significantly superior sensitivity.

The Cornell Scale for Depression in Dementia (CSDD, Alexopoulos *et al.*, 1988a) has been validated for use in patients with PD with varying degrees of cognitive impairment (Williams and Marsh, 2009). The CSDD is a scale designed for use with patients with dementia and has also been validated for use in non-demented patients (Alexopoulos *et*

al., 1988b). The scale ratings are completed based upon observation and interview with both the patient and an informant. The scale was shown to have satisfactory performance when used in a PD sample with varying cognitive impairment.

Some researchers have argued that a depression scale that incorporates somatic items will not necessarily lead to inaccurate results when used with patients with PD, provided that the cut-off score is adjusted appropriately for the patient group concerned. In fact, omission of somatic symptoms from a depression scale may reduce face validity, particularly as the standard diagnostic criteria (DSM and ICD) include several somatic symptoms (Marinus *et al.*, 2002), and limit its validity in severe depression (Schrag *et al.*, 2007).

Leentjens *et al.* (2003b) studied the degree to which the somatic items in the HamD and the MADRS contribute to the discriminant ability of the two scales. Their analyses indicated that most of the somatic items on both scales did not contribute substantially to the discriminant ability of the scales and that the core symptoms and the other non-somatic items were the most important. However, they did find that 'reduced appetite" and 'early morning wakening' were relatively sensitive indicators of depression and contributed to distinguishing depressed and non-depressed PD patients.

In light of the lack of clarity regarding the suitability of depression rating scales for use in PD the Movement Disorder Society commissioned a task force of experts in the field to review available scales and make recommendations regarding their use (Schrag *et al.*, 2007). The group considered the utility of the BDI, HamD, HADS, GDS, MADRS, CSDD, as well as the Unified Parkinson's Disease Rating Scale (UPDRS), Zung self-

rated depression scale and the Centre for Epidemiologic Studies Depression Scale (CES-D). At the time of their review, all except the last three had had some validation of their use in PD. The expert task force concluded that MADRS, BDI, and GDS are appropriate for screening for depression in patients with PD when a suitable cut-off is used, and that the HADS is moderately suitable. The HamD was described as being most suitable for assessing severity of depression but also as an adequate screening measure. These findings are summarised in Table 1.8 below. The authors felt that the observer-rated scales had better psychometric properties than the self-rated scales and so recommended their use if feasible.

Table 1.8 - Recommendations for depression rating scales in PD. Schrag *et al.*, 2007.

HamD	Most suitable for assessing severity of depression, but is an adequate screening measure. However, it requires training for the observer-rater.
MADRS	Is an appropriate screening tool if used with appropriate cut-off score. However, it is an observer-rated scale.
BDI	Is an appropriate screening tool if used with appropriate cut-off score. Also suitable for rating severity and monitoring change.
HADS	Moderately suitable for screening but its use as a severity rating scale is questionable.
GDS	Appropriate for clinical research and routine clinical care as a screening measure. Validity as a measure of severity is yet to be established.
Zung	Validity in PD not established.
Cornell	Validity in PD not yet established (at the time of the review). Requires some training of the observer-rater.
CES-D	Validity in PD not established.
UPDRS	Should only be used a crude screening tool.

Williams *et al*. (2012) recently published a study in which nine different depression rating scales were compared in community-based patients with PD. The measures that were studied included 6 self-report measures: the BDI, GDS-30, PHQ-9, Centre for Epidemiologic Studies Depression rating scale (CESD-R), Inventory of Depressive

Symptoms-Patient (IDS-SR), and the depression subscale of the Unified Parkinson's Disease Rating Scale (UPDRS) and 3 clinician-rated measures: the HamD-17, MADRS and the Inventory of Depressive Symptoms-Clinician (IDS-C). The study highlighted the differences between the measures in terms of time taken and ease of completion, and demonstrated better performance of some scales over others in terms of correct identification of patients with DSM-IV (Major or Minor) depression (see Table 1.9), and greater sensitivity of some scales to severity of depression. The authors supported the use of certain self-report scales to screen for depression in PD as alternatives to clinician-rated scales. They felt the GDS-30 and BDI performed as well as the clinician-rated scales and the GDS-30 in particular was quicker to administer and therefore best suited to use for screening in a clinical setting.

Table 1.9 – Performance of the depression scales in Williams *et al.* (2012)

Measure	Cut-off	Se %	Sp %	PPV %	NPV %
BDI	6/7	95	60	62	94
CESD-R	11/12	72	70	62	79
GDS-30	9/10	72	82	73	81
IDS-SR	13/14	90	60	61	90
PHQ-9	5/6	66	80	69	77
UPDRS-D	0/1	70	77	68	79
HamD-17	6/7	77	76	69	83
IDS-C	11/12	81	79	73	86
MADRS	7/8	74	88	81	83

One ultra-short depression screening measure has been validated for use in patients who have PD. Chagas *et al.* (2011) investigated the performance of the two–item self-report scale - the Patient Health Questionniare-2 (PHQ-2). The PHQ-2 is a two item measure taken from the PRIME-MD study (Spitzer *et al.*, 1994) in which patients are asked about having 'little interest or pleasure in doing things' and 'feeling down depressed or

hopeless' over the past two weeks, and items are rated on a scale from 0 (not at all) to 3 (nearly every day). The study found the two–item scale to be a valid screening measure in PD (sensitivity 75%, specificity 89%, PPV 70% and NPV 91% at a threshold score of 2/3 out of 6).

The purpose of screening patients with PD for depression is to increase the recognition of depression so that appropriate treatment can be given and clinical outcomes improved. Without effective treatments for depression a screening programme would be a waste of resources and patient time.

The UK National Screening Committee recommends that:

- There should be an effective treatment or intervention for patients identified through early detection, with evidence of early treatment leading to better outcomes than late treatment.
- There should be agreed evidence based policies covering which individuals should be offered treatment and the appropriate treatment to be offered.

Although evidence of the efficacy of antidepressant therapy in PD (including antidepressant medication, psychological therapy and ECT) is insufficient for definite guidance for the treatment of depression in people with PD (NICE, 2006, p116), there is some evidence that such treatments can be effective. The high proportion of PD patients who experience depression at some point in the illness and the marked impact that such symptoms have upon their quality of life and functioning could be argued to justify the need for screening of PD patients for depression, followed by careful treatment that is tailored to the individuals who are found to be suffering. However the

lack of rigorous (RCT-based) evidence of the effectiveness of treatments for depression in PD will have implications for the assessment of potential patient benefit from screening programmes for depression in PD. A lack of evidence of efficacy of an intervention does not necessarily mean that the intervention is not effective, only that there is a current lack of good-quality evidence and that the efficacy remains to be demonstrated.

1.20 AIMS OF THIS STUDY

Recognition of depressive disorders in patients who have PD is poor - so the availability of a depression screening instrument that has good operator characteristics, is acceptable to patients and is easy to use in clinical practice is important. Identification of a screening instrument for depression in PD, which has superior performance to those scales currently validated for use in this patient group, would be of benefit to clinicians working with this patient group. Of those scales that have currently been validated for depression screening in PD only the MADRS, BDI and GDS were judged to be suitable for use in PD (Schrag *et al.*, 2007). However the MADRS requires completion by a trained observer, whilst the BDI includes several items, and the GDS one item, that could be subject to symptom overlap between depression and PD.

This study has been designed to assess the relative efficacy of three screening tools for depression in patients with PD:

1. The Edinburgh Postnatal Depression Scale is a scale that has no items relating to the somatic symptoms of depression and so would not be subject to problems

due to symptom overlap with PD, and as such could have better face-validity with PD patients completing the scale.

2. Current NICE guidelines (NICE 2009a, 2009b) recommend screening for depression in high-risk patient groups using two questions, which were first evaluated by Whooley *et al.*, (1997). This screening tool has not yet been evaluated in a PD patient group, and due to its brevity may prove to be a more effective and efficient screening tool than the MADRS, BDI and GDS. Evaluation of the screening performance of the Whooley questions compared to the other two instruments included in this study will give valuable information for clinicians looking to implement the NICE recommendations.

3. The GDS-15 is the patient-completed screening measure that has performed the best in validation studies in PD patient groups. It will provide a useful comparison against which to judge the performance of the other screening instruments, as well as providing additional evidence of its performance against the criterion of a clinical diagnosis.

All three of the screening instruments studied here have potential for use in a variety of hospital settings due to the absence of items that relate to symptoms that are open to confusion with symptoms of medical illness or the side-effects of medications. There would be an obvious benefit in a single screening instrument for depression that is valid in a variety of settings in the general hospital - health professionals would be likely to become more familiar and comfortable with one tool, and therefore more likely to use it and increase their alertness to the possibility of depression in their patients.

1.20.1 The Edinburgh Depression Scale

The Edinburgh Postnatal Depression Scale, which is now also known as the Edinburgh Depression Scale (EDS), was originally developed and validated to identify depression in women during the post-natal period (Cox *et al.*, 1987). The scale has no items relating to somatic symptoms of depression and instead relies upon identifying cognitive-affective symptoms of depression. The EDS scale comprises 10 items; each item has four statements and the patient is requested to select one statement from each item that corresponds best to how they have felt over the preceding seven days. The scale is designed for self-completion by the patient and can be administered by staff members with no specialist training in mental health (see Appendix 2).

The EDS scale was found to have good sensitivity (86%) and specificity (78%) in postnatal women at a threshold score of 12/13, and was also found to be sensitive to change in severity of depression over time (Cox *et al.*, 1987). The authors indicated that lowering the cut-off to 9/10 gave the best performance in terms of case identification (i.e. for screening).

The EDS has been validated in further studies of postnatal women (e.g. Harris *et al.*, 1989; Murray and Carothers, 1990), in females outside the postnatal period (Cox *et al.*, 1996), antenatal women (Murray and Cox, 1990) and in the male partners of postnatal women (Matthey *et al.*, 2001) and optimal cut-off scores for depression have been identified for each group (see Table 1.10). The validity of the EDS in ante- and post-natal women has been reviewed (Gibson *et al.*, 2009; Eberhard-Gran *et al.*, 2001; Gaynes *et al.*, 2005) but due to the heterogeneity of the studies meta-analysis was not

possible. Gibson *et al.* felt that although the heterogeneity of the results suggest that the utility of the EDS varies depending upon the clinical setting, country or language of administration, it is still a useful screening tool for depression in these patient groups.

In light of the fact that the EDS has no items that relate to the somatic symptoms of depression it was thought to be a suitable depression screening tool for patients receiving palliative care. Often the symptoms of a serious medical illness and/or the side-effects of the medication prescribed for treatment of that illness may mask or complicate the identification of depressive symptoms, making the use of many other depression scales unsuitable for use in a palliative care setting. The EDS has therefore also been validated for use in palliative care patients with advanced metastatic cancer (Lloyd-Williams *et al.*, 2000, 2004). Lloyd-Williams and colleagues have validated the EDS in patients in palliative care in two studies. In the first, (Lloyd-Williams *et al.*, 2000) the performance of the EDS was assessed against the gold standard of an ICD-10 diagnosis of depression derived from the Present State Examination and it was established that a threshold EDS score of 13 or more gave optimal performance in this patient group (sensitivity of 81% and specificity of 79%). A threshold of 9/10 gave optimal performance (i.e. highest sensitivity and NPV) for screening (see Table 1.11).

A later study (Lloyd-Williams *et al.*, 2004) compared the performance of the EDS with two other screening tools (a visual analogue scale and a single question "are you depressed") in a similar palliative care sample. The EDS proved to be the most appropriate tool for screening for depression compared to the other two screening measures. However in this study the EDS gave slightly lower sensitivity and specificity than in 2000, perhaps because the gold standard in this study was diagnosis according to

the more stringent criteria of DSM-IV. It has also been established (Lloyd-Williams, private communication) that this study utilised a slightly modified version of the EDS where one item "*I have felt scared or panicky for no good reason*" was replaced with the item "*I get a sort of frightened feeling as if something awful is about to happen*". It is not known what effect this change has upon the performance of the scale but the operating characteristics were similar for the modified version when compared to the previous validation study which used the original scale [4].

Matthey *et al.* (2003, 2006) have identified that the EDS is frequently used in studies and in clinical practice with variations from the scale that was first published by Cox *et al.* (1987). This includes slight variations in the wording of some items, and variations in the way respondents are asked to indicate their responses to each item (e.g. tick box, underline response), or in the instructions at the beginning of the scale. Matthey *et al.* rightly point out that such variations may alter the acceptability, and therefore the performance, of the scale but the impact of such variations from the original is not known. They stressed that it is important that scales are validated in the population in which it is going to be used, and that the appropriate optimal cut-off for depression are established and applied for that patient group.

The performance of the EDS in different studies and in different patients groups is shown in Tables 1.10 and 1.11.

[4] The EDS was originally compiled from items taken from the HADS and the Irritability, Depression and Anxiety Scale (IDA, Snaith *et al.*, 1978). The original 'scared or panicky item' was from the IDA and the 'frightened feeling' item that Lloyd-Williams replaced it with was taken from the HADS.

Table 1.10 - The sensitivity (Se), specificity (Sp), PPV, NPV of the EDS and sample prevalence of depression for the standard recommended cut-off score (12/13) and recommended cut-off for screening (9/10) in women.

				Major and Minor Depression				Major Depression				
Author Study sample	**N**	**Prev**	**Cut -off**	**Se %**	**Sp %**	**PPV %**	**NPV %**	**Se %**	**Sp %**	**PPV %**	**NPV %**	**Criterion measure**
Cox *et al.*, 1987 Post-natal women	63	42%	12/13	86	78	73	88	--	--	--	--	RDC Major & Minor dep illness
Harris *et al.*, 1989 Post-natal women	147	15%	12/13	--	--	--	--	95	93	75	99	DSM-III Major Depression
			9/10	--	--	--	--	100	82	NR	NR	
Murray & Carothers, 1990 Post-natal women	646*	NR	12/13	68	96	67	NR	81	96	43	NR	RDC Major & Minor dep illness
			9/10	89	82	39	NR	96	82	23	NR	
Boyce *et al.*, 1993 Post-natal women	103	9%	12/13	--	--	--	--	100	96	69	100	DSM-III Major Depression
			9/10	--	--	--	--	100	89	47	100	
Murray & Cox, 1990 Pregnant women	100	14%	12/13	64	90	50	NR	100	87	33	NR	RDC Major & Minor dep illness
Cox *et al.*, 1996 Post-natal women	128	16%	12/13	62	89	52	NR	75	84	24	NR	RDC Major & Minor dep illness
			9/10	81	77	41	NR	88	71	17	NR	
Non post-natal women	136	14%	12/13	79	85	46	NR	88	80	21	NR	
			9/10	95	72	35	NR	88	66	14	NR	
Leverton and Elliott, 2000 Post-natal women	199	5%	12/13	70	93	33	98	--	--	--	--	PSE diagnosis of depression
			9/10	90	84	23	99	--	--	--	--	
Beck and Gable, 2001 Post-natal women	150	31%	12/13	--	--	--	--	78	99	93	96	DSM-IV
			9/10	59	86	64	82	--	--	--	--	Major & Minor depression
Becht *et al.*, 2001 Menopausal women	951	21%	12/13	66	89	62	NR	88	85	40	NR	RDC Major & Minor dep illness
			9/10	81	80	51	NR	92	74	28	NR	
Matthey *et al.*, 2001												
Post-natal women	230	10%	8/9	71	76	25	NR	--	--	--	--	DSM-IV Major & Minor
Male partners	208	3%	9/10	71	94	29	99	--	--	--	--	Depression

NR = not reported

*Not all subjects had standardised diagnostic interview. Se, Sp and PPV were estimated using logistic regression.

Lloyd-Williams and her colleagues have more recently validated a shorter version of the EDS, comprising only 6-items from the modified EDS (Lloyd-Williams *et al.*, 2007). This scale is known as the Brief Edinburgh Depression Scale (Brief EDS) and performed better than the 10-item modified EDS (EDS sensitivity 72%, specificity 74%, PPV 55% and NPV 85% at the optimal cut-off of 11/12 c.f. Brief EDS sensitivity 72%, specificity 83%, PPV 65% and NPV 87%) in a sample of patients with advanced cancer.

Table 1.11 - The sensitivity (Se), specificity (Sp), PPV, NPV and sample prevalence of depression for the standard recommended cut-off score (12/13) and recommended cut-off for screening (9/10) in palliative care.

Author	N	Prev. dep.	Cut-off	Se %	Sp %	PPV %	NPV %	Criterion measure
Lloyd-Williams *et al.*, 2000	100	22%	12/13	81	79	53	94	ICD-10 Depression
			9/10	100	52	36	100	
Lloyd-Williams *et al.*, 2004 #	74	27%	12/13	70	80	56	88	DSM-IV Major Depression
			9/10	70	54	36	83	
Lloyd-Williams *et al.*, 2007 #								
10-item EDS	246	30%	12/13	61	80	58	82	PSE diagnosis of depression
10-item EDS			9/10	89	54	46	92	
6-item BEDS			6/7	72	83	65	87	

used a modified version of EDS (one item changed).

The EDS does not include any items that relate to somatic symptoms of depression that may be confounded by somatic symptoms of PD. For this reason it was judged that it may perform well in a Parkinson's disease population. The scale has already been validated in palliative care settings, where the average ages of the study participants are likely to be more similar to a (predominantly older) PD population than postnatal

women. (Lloyd-Williams *et al.*, 2000 Average Age = 57.3 yrs, Lloyd-Williams *et al.*, 2004 Average Age = 67.9 yrs (median = 70.5), Lloyd-Williams *et al.*, 2007 Average Age = 61.9 yrs.) The modified version of the scale, used by Lloyd-Williams *et al*. in 2004 and 2007, was chosen for this study as it includes the 6 items taken to comprise the Brief EDS.

1.20.2 The 'Whooley Questions'

Another screening measure used in this study is the two depression screening questions that are recommended in the NICE guidelines for depression (1996) and validated by Whooley *et al.* (1997).

1) "*During the past month, have you often been bothered by feeling down, depressed or hopeless?"*
2) "*During the past month, have you often been bothered by having little interest or pleasure in doing things?"*

These questions are taken from the PRIME-MD questionnaire (Spitzer *et al.*, 1994) and require a Yes/No answer. Similar versions of the questions are included in the Patient Health Questionniare-2 (PHQ-2), a subscale of the PHQ-9, in which the questions are phrased: "Over the last 2 weeks, how often have you been bothered by little interest or pleasure in doing things" or "feeling down, depressed or hopeless". Respondents are given 4 response choices (not at all, several days, more than half the days, or nearly every day) (Kroenke *et al.*, 2003). The Whooley questions were developed as an even simpler alternative screening instrument for depression, with a positive response to either question indicating possible depressive illness. They have been shown to be an

effective screening tool in patients of a wide age-range in a primary care setting (Whooley *et al.*, 1997; sensitivity 96%, specificity 57%, PPV 33%, NPV 98%).

McManus, Pipkin and Whooley (2005) compared the performance of the PHQ-9, PHQ-2 and the Whooley questions in a sample of patients with coronary heart disease. They found that the two Whooley questions were equally effective as the longer, more time-consuming instruments for identifying Major Depression, with a sensitivity of 91%, specificity of 71% and NPV of 96%. A negative response to both items made "depression highly unlikely" but the lower specificity and PPV meant that less than half of patients who responded positively to either question would actually meet criteria for Major Depression - indicating the importance of a follow-up diagnostic interview.

These two screening questions, amended to "In the past 2 weeks" have also been validated in a long-term residential care setting against the criterion of DSM-IV diagnosis, and were found to have the best combination of brevity, sensitivity (80%) and ease of administration when compared to various measures, including the GDS-15 (Watson *et al.*, 2009).

The Whooley screening questions have been recommended for use in the NICE guidelines for depression in adults with chronic physical health problems (2009) but have not yet been validated in a sample of patients with PD. These questions have the advantage of brevity and are therefore potentially much more easily assimilated into a routine clinical assessment than a longer rating scale, are more likely to be acceptable to all patients and the limited response options are likely to be more acceptable and easy to use for older people (Brink *et al.*, 1982).

1.20.3 The Geriatric Depression Scale

The final screening tool chosen for this study is the 15-item Geriatric Depression Scale (GDS-15, see Appendix 3). The performance of this scale has already been assessed and validated in PD and was selected for inclusion in this study because of its established performance as a screening measure in PD (Weintraub *et al.*, 2006b; 2007; Schrag *et al.*, 2007; Williams *et al*., 2012; see Table 1.7) and because it is a brief, self-complete depression scale that contains only one item that may be subject to symptom overlap with PD.

The GDS-15 was considered to be a good screening measure in elderly patients in a recent review of its performance, with an average sensitivity of 80.5% and specificity of 75% (Wancata *et al.*, 2006). It appears to perform less well with patients who have marked cognitive impairment, and so the authors of the review recommended that the GDS is not used in such groups. The generally recommended cut-off for depressive disorder in a community primary care setting is 4/5, with a cut-off of 6/7 for hospital in-patients (Cullum *et al.*, 2006). In PD samples the optimal cut-off (for both screening and dichotomising patients) has been identified as 4/5 (Weintraub *et al.,* 2006b) although a slightly higher cut-off of 5/6 was indicated for patients over the age of 75yrs (Weintraub *et al.,* 2007).

Researchers have also explored the impact of oral versus written self-completion presentation of the items, as older or medically ill patients may often need assistance in this manner (Cannon *et al.*, 2002). Oral presentation of the scale appears to be more acceptable to elderly patients but may result in the endorsement of fewer items when

compared to written self-completion of the scale (O'Neill *et al.*, 1992; Cannon *et al.*, 2002).

Shorter single-item, 4-item and 10-item versions of the GDS have been validated (D'Ath *et al.*, 1994; Galaria *et al.*, 2000). The single-item version performed poorly but the 10 and 4-item versions have been found to be adequate screening measures (Marwijk *et al.*, 1995; Shah *et al.*, 1997) which have the advantage of being briefer, and therefore more practical for screening, than the longer versions.

1.20.4 The criterion measure – Diagnosis of depression

The performance of these three depression screening instruments was compared to the 'gold standard' criterion of a diagnosis of depression derived from the present State Examination - Schedules for Clinical Assessment in Neuropsychiatry (SCAN) interview (version 2.1; World Health Organisation, 1999; Wing *et al.*, 1990). This is a semi-structured interview designed to assess and classify psychiatric syndromes in adults. The SCAN interview comprises many sections covering different symptoms: somatoform, dissociative, anxiety, depressive and bipolar disorders, and problems relating to bodily functions and use of alcohol or other substances are included in Part 1 of the interview schedule. Part 2 includes sections which relate to psychotic and cognitive disorders.

The SCAN interview is conducted by clinicians specifically trained in its use. It is a semi-structured interview and an important aspect of this is that it retains the flexibility of a clinical assessment, and can be completed in relation to a designated period of time (present state, lifetime or a specified episode). Computer-run algorithms can then be

applied to the data obtained from the SCAN interview to give diagnoses according to DSM-IV and ICD-10 criteria.

SCAN is an instrument that has been developed by the World Health Organisation with the aim of improving the description and classification of psychiatric disorders (Wing, 1996). The core component of SCAN 2.1 is the Present State Examination version 10 (PSE10) which is a semi-structured interview developed from PSE9 (Wing *et al.*, 1974). The idea is that a semi-structured standardised clinical interview more closely represents a clinical assessment than a fully structured interview. As such PSE10 allows the interviewer to change the order and wording of the questions according to the specific situation, in order to establish which of a comprehensive list of symptoms and signs have been present during the specified period of time, and to what severity. The interviewer makes the decision regarding whether a symptom is present having explored the patient's experiences and matching them to the comprehensive clinical definitions provided in the associated glossary.

There is substantial evidence of the reliability of SCAN (which incorporates PSE10) and of PSE9 which have been used extensively both in population surveys and experimental research (Wing *et al.*, 1967, Kendell *et al.*, 1968, Cooper *et al.*, 1977, Wing *et al.*, 1977a, 1977b, 1978, 1990, Wing, 1996) and SCAN 2.1 has been shown to demonstrate the same levels of acceptability and reliability (Rijnders *et al.*, 2000).

This study therefore aims to assess the performance of these three screening instruments in a sample of PD patients, in terms of correctly indentifying those patients who are

rated as having a diagnosis of depression according to either DSM-IV or ICD-10 criteria, from the SCAN interview.

CHAPTER 2 – METHODS

2.1 Research Ethics Committee and Clinical Governance Approval

Before any study procedures could commence it was required to obtain approval from both the local Research Ethics Committee (REC) and the NHS Trust Research and Development department. This process entails completion of a detailed application form describing the study in terms of the purpose, design and methodology. The ethical issues relevant to carrying out the research (such as informed consent procedures) and any possible harm or inconvenience to potential participants need to be explained and justified. The planned timescale, as well as details of the personnel carrying out the study procedures (including the relevant qualifications of the staff to demonstrate their appropriate training to carry out the research) and responsibilities of those staff, are specified. Under the Research Governance Framework (DOH, 2001) it is required that a sponsor is identified for all research, so details regarding this were required for both the REC and NHS approval processes. All written materials presented to potential participants are also submitted for approval.

Both the local REC and NHS Clinical Governance approvals were obtained in May and July 2006 respectively (See Appendix 4). Any subsequent substantial changes to the research procedures or personnel required further submissions to be made to the local REC and NHS Trust Research and Development Office, and their approval obtained, before those changes could be put in place.

2.2 The setting

This research was conducted at Leicester General Hospital, in the UK between January 2007 and December 2011. There are two consultants at the hospital who run specialist outpatient clinics for patients with PD, covering the whole county[5]. Initially patients were recruited from those attending the PD outpatient clinic run by a consultant geriatrician.

Fifteen months after the start of the project arrangements were made for patients attending the PD outpatients clinic run by a consultant neurologist, at the Leicester General Hospital, to also be approached regarding taking part. This was to increase the number of patients referred to take part in the research.

2.3 The subjects

All patients attending the clinic who met initial inclusion and exclusion criteria were asked to take part in the research by the consultant or the doctor running the clinic on his behalf. Patients who were approached to take part in the research included both those who were newly diagnosed and those attending the clinic for follow-up or review. In order to avoid taking too much time from the clinic, explaining the research to every patient, the consultant neurologist opted to approach every fifth patient in the clinic about taking part in the study.

[5] The estimated population of Leicestershire in mid-2009 was 949,400 (ONS, 2009). Based on this and the prevalence of PD (NICE, 2006) there are estimated to be between 949-1709 people with PD in the county, with between 38 and 190 people being newly diagnosed with the condition each year.

The inclusion criteria were:

- The patient meets Brain Bank criteria of the UK Parkinson's Disease Society for the diagnosis of Parkinson's disease (UK-PDS-BB, Hughes *et al.*, 1992b; De Rijk *et al*., 1997).
- The treating doctor is of the opinion that the patient is able to give informed consent to participate in the research.

The exclusion criterion was:

- The patient is not able to speak English.

2.4 The procedure

If the consultant considered the patient to be suitable for the research according to the inclusion and exclusion criteria they asked them to consider taking part in the study. The doctor provided the patient with a brief explanation of the research and a copy of the Patient Information Sheet for them to read and take home (See Appendix 5). If the patient expressed an interest in taking part in the research their contact details were then passed on to the research team. A member of the research team would then contact the patient to make an appointment to visit them either at their own home or at the hospital, depending upon their preference.

Visit 1

At this appointment the researcher (a psychiatrist) explained the nature and purpose of the research again, and answered any questions the patient may have had. The patients were then required to give informed consent, and sign a consent form (see Appendix 6).

Once informed consent had been obtained the researcher administered the cognitive assessment section (Section 21) of the Present State Examination – Schedules for Clinical Assessment in Neuropsychiatry (PSE-SCAN; WHO, 1999) which includes the items of the Mini-Mental State Examination (MMSE; Folstein *et al*., 1975); if the patient scored less than 24 on the MMSE the interview was terminated and the patient excluded from the study. This further exclusion criterion was applied because a score of less than 24 is likely to indicate a degree of cognitive impairment, and this potentially would influence responses to the depression screening instruments rendering them unreliable.

Provided the patient attained a score of 24 or higher on the MMSE the researcher then interviewed the patient using the PSE-SCAN. Only those sections of the PSE-SCAN pertinent to the assessment of depressed mood were used. These were Sections 6 (depressed mood and ideation), 7 (thinking, concentration, energy, interests) and 8 (bodily functions).

Visit 2

Within a period of five working days of the PSE-SCAN interview the subject was then seen by another researcher (SB) who administered the EDS, the two Whooley screening questions and the GDS-15. This researcher was blind to the subject's performance on the PSE-SCAN. The three depression screening instruments were presented in the form of self-completion questionnaires in the order they are listed above (i.e. EDS, Whooley questions, GDS-15). However, in light of the difficulty many patients with PD experience with writing, all patients were given the options of completing the questionnaire themselves, reading the questions themselves but verbally indicating to

the researcher their chosen response, or having the questions and responses read to them by the researcher. If the last option was chosen by the patient the questions and responses were read verbatim by the researcher with no additional prompts or explanation.

Once the patient assessments had been completed the following information was retrieved from the patient's clinical notes: the Hoehn and Yahr stage of PD (Hoehn and Yahr, 1967; see Appendix 7), date of initial diagnosis of PD and the patient's current medication as recorded by the specialist doctor at the patient's most recent clinic appointment. The Hoehn and Yahr scale reflects the severity of PD symptoms with stage 1 being the mildest stage, in which there are only unilateral symptoms with little or no impairment of function, whereas stage 5 represents the most severe stage, in which patients are confined to bed or a wheelchair.

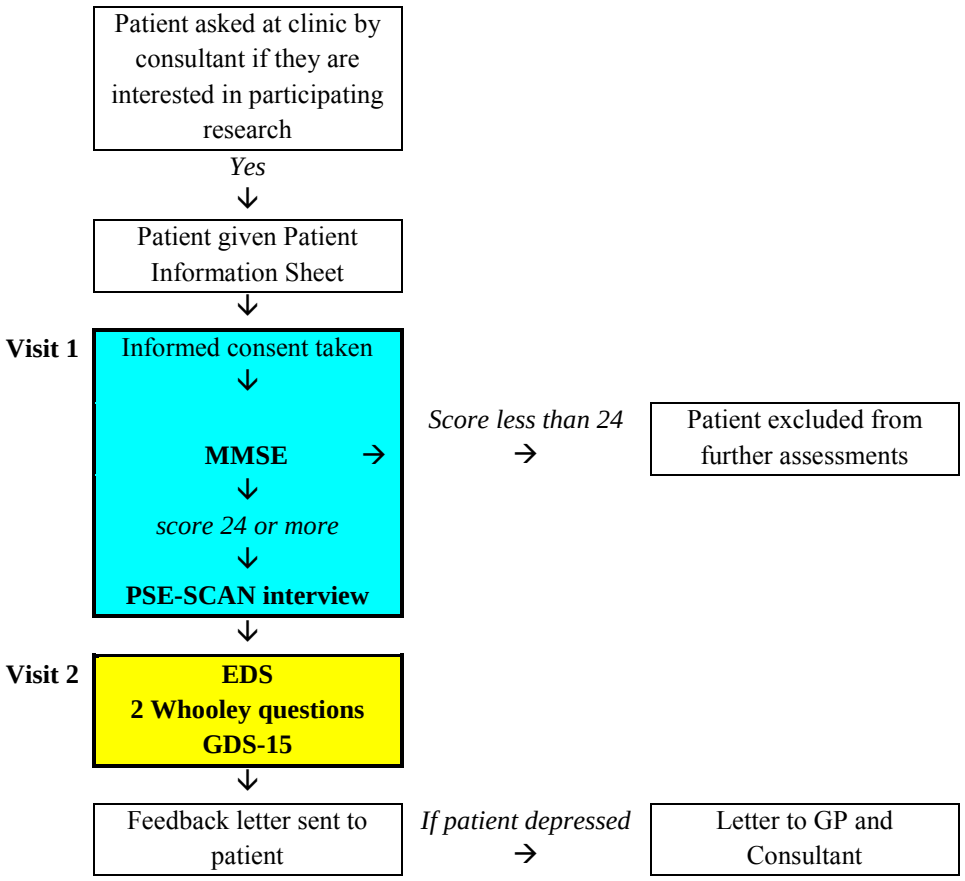
If the psychiatric assessment revealed the presence of depression then the patient's general practitioner was informed and advised on appropriate management. A letter was also sent to the patient following the second interview informing them of the results of their assessment - in the case of a positive diagnosis from the PSE-SCAN interview, the letter suggested the patient should make an appointment to see their general practitioner. A copy of this letter was also sent to the hospital consultant responsible for treating their PD.

If the psychiatric assessment indicated that the patient was not suffering from depression then the patient was sent a letter thanking them for their participation in the

trial and confirming that the assessments had not indicated they were suffering from depression.

A summary of the research procedure is represented in the flow chart below (See Figure 2.1)

Figure 2.1 – Flow chart of study procedures



2.5 The Research Team

There were several individuals involved in this research.

- Sarah Baillon, Research Fellow in the Department of Health Sciences at the University of Leicester, was responsible for planning, administration and coordination of the study and was also responsible for the completion of screening

questionnaires with all participants, feedback of the outcome of the assessment to participants (and their consultant/GP if necessary), as well as all data collation and analysis, and writing-up and dissemination of the study results.

The following were also involved in planning and review of the research:

- Professor Michael Dennis, Professor of Psychiatry for Older People at Swansea University. Professor Dennis was previously Senior Lecturer and Consultant Psychiatrist in the Department of Health Sciences, at the University of Leicester.
- Professor James Lindesay, Professor of Psychiatry for the Elderly and Consultant Psychiatrist in the Department of Health Sciences, at the University of Leicester.
- Dr Nelson Lo, Consultant Geriatrician at the Leicester General Hospital, University Hospitals of Leicester NHS Trust.

The following individuals carried out the first visit with participants (took informed consent and completed the MMSE and PSE-SCAN interview):

- Dr Santanu Chakrabarti, a Consultant Psychiatrist in Mental Health Services for Older People at Leicestershire Partnership NHS Trust and
- Dr Sandhya Gaur,
- Dr Oli Hands,
- Dr Jonathan Murray,
- Dr Jairaj Padmanabhan,
- Dr Pravin Prabhakaran,
- Dr Aniruddha Rajkonwar,
- Dr Deepak Shukla, and
- Dr Suzanne Stacey

all of whom were Specialist Registrars in Mental Health Services for Older People at Leicestershire Partnership NHS Trust.

All staff members carrying out the PSE-SCAN interview had received specialist training in the interview technique at a World Health Organisation designated SCAN Training and Reference Centre. Inter-rater reliability was assessed during the training course and raters had to achieve satisfactory levels of agreement to complete their training in the use of the interview technique.

2.6 Sample size

Determining adequate sample size is an important aspect of study design. The sample must be sufficient to enable identification of a clinically important difference with some certainty. A sample size that is too small will either fail to demonstrate a small but clinically significant effect or will produce an imprecise or inaccurate estimate, whereas an unnecessarily large sample will result in a waste of resources testing more individuals than would be needed to demonstrate the effect being studied.

Low numbers of cases (depressed patients) in a sample when validating a depression screening instrument will lead to a poor and inaccurate estimate of the sensitivity of the instrument. Therefore an adequate sample size is required to allow for a reasonable number of positive cases within the sample to demonstrate the test characteristics of the instruments being examined.

This validation study is a paired-design cohort study – meaning that the study sample is taken as a random sample of individuals from the population in which the screening

tools would be used (i.e. PD patients attending out-patient clinics) and all screening tools being investigated are applied to each participant (thus eliminating potential between-subject variance).

Calculation of the sample size required for such a study requires knowledge of the proportion of the cases (depressed patients) who test positive on the scales to be compared (Alonzo *et al*., 2002). However, this proportion is entirely unknown as neither the EDS nor the Whooley questions have ever been studied in a PD population before.

The mean prevalence of Major Depression in an outpatient sample of patients with Parkinson's disease is estimated to be 24% (Reijnders *et al.*, 2008). Therefore a sample of approximately 100 clinic attendees with a diagnosis of Parkinson's disease would yield approximately 24 depressed subjects. This was felt to be a sufficient number of depressed and non-depressed participants to demonstrate the operator characteristics of the screening tools being studied. Therefore the target sample size was 100 participants.

Increase in sample size

Once 84 participants had completed the study procedures the prevalence of depression in the sample was reviewed. 12/84 (14%) of participants were found to have a diagnosis of depression. This prevalence of depression was significantly lower than had been anticipated and it was decided to increase the target number of participants from 100 to 120 in order to increase the total number of participants who met the criteria for depression. Increasing the target recruitment figure for a study constitutes a substantial

amendment to the protocol – therefore an application was made to the local REC and NHS Trust Research and Development Office, and their approval obtained.

2.7 Statistical analysis

The PSE-SCAN interview was used as a criterion 'gold standard' measure to identify depression according to DSM-IV and ICD-10 diagnostic criteria, and provided the external validation for the screening tools. The sensitivity, specificity, positive and negative predictive values were calculated at various different cut-off points for each of the instruments to identify the optimal threshold scores for patients with PD. The likelihood ratios and diagnostic odds ratio for each of the screening tests were calculated using methods described by Altman (1999) and Altman *et al*., (2000). Receiver operating characteristic (ROC) curves were generated in order to compare the area under the curve (AUC) and the discriminating power of each scale in the sample population. The AUC for the scales were compared according to the method described by Hanley and McNeil (1983). These test characteristics were calculated for the identification of those participants with Major or Minor Depression, and for the identification of Major Depression.

These analyses were carried out using SPSS (version 18.0, IBM Corp, USA) and STATA (version 11.2, StataCorp, USA) statistical software. The 95% confidence intervals were generated using Confidence Interval Analysis software (version 2.1.2). These analyses were later verified using MedCalc for Windows (version 12.1.4.0, Mariakerke, Belgium).

2.8 Diagnosis according to DSM-IV and ICD-10 criteria

DSM-IV diagnosis

The diagnosis of depression according to DSM-IV and ICD-10 criteria were obtained by running the diagnostic algorithms incorporated into the SCAN software.

The DSM-IV algorithm produces a diagnosis of "Major Depressive Disorder, single episode", "Major Depressive Disorder, recurrent" or "Dysthymic Disorder" but does not produce a diagnosis of Minor Depression. In this study we wished to identify participants with Minor Depression as well as those with Major Depression and dysthymia. In order to do this the diagnostic algorithms for Major Depression used by the SCAN software were adapted and applied manually so that participants meeting the DSM-IV criteria for Minor Depression were identified using the same items responses upon which the diagnosis of Major Depression was based. See Appendix 8 for a summary of the SCAN diagnostic algorithms.

The algorithms for Major Depressive Disorder comprise a count of all items that indicate the presence of the nine symptoms included in the DSM-IV diagnostic criteria, for the required two week period (See Figure 1.0). Further algorithms then apply the criteria that there should be a total of at least five of the criterion symptoms present and that at least one of the symptoms that are present is either depressed mood or loss of interest or pleasure. If these conditions are met a diagnosis of Depressive Disorder is produced. (Further algorithms also included in this process apply the additional conditions specified in the DSM-IV diagnostic criteria for depression, i.e. do not meet criteria for a Mixed Episode, symptoms cause clinically significant distress or

impairment, symptoms are not due to a substance or general medical condition, or are better accounted for by bereavement.)

In order to generate a diagnosis of DSM-IV Minor Depression the algorithms for Major Depression were followed manually and the symptom count compiled in exactly the same way, but then required a total of two symptoms to be present, one of which was required to be either depressed mood or loss of interest or pleasure, in order for a diagnosis of Minor Depression to be made.

ICD-10 Diagnosis

The diagnosis of ICD-10 depression was produced using the diagnostic algorithms of the SCAN software. These algorithms produce a diagnosis of "Mild Depressive Episode", "Moderate Depressive Episode", "Severe Depressive Episode", or "Dysthymia".

Sub-threshold Depression

Diagnosis of sub-threshold depression was also produced manually following the criteria described in the NICE guidelines (NICE, 2009a) and by Judd *et al. (*1994). Participants who reported at least two of the criterion symptoms from DSM-IV criteria, daily over at least the previous two weeks, but with maintained function were diagnosed with sub-threshold depression.

CHAPTER 3 - RESULTS

3.1 THE PARTICIPANTS

176 patients were approached at their PD outpatient clinic and expressed an interest in participating in the research project. 136 of these patients were contacted by the research team, gave informed consent and participated in the study. 15/136 failed to score 24 or more on the MMSE, indicating possible cognitive impairment and so were excluded from completing the further study procedures. One patient scored over 24 on the MMSE and participated in the SCAN interview but later declined the second visit at which they would complete the depression screening scales. See Table 3.1.

Table 3.1 - Participants in the study

	N
Patients referred	176
Patients seen	136
- Both assessments	120
- MMSE less than 24	15
- Declined after SCAN	1
Patients not seen	40
- Unable to contact	16
- Declined	24

The demographic details of the study participants are shown in Table 3.2. The Hoehn and Yahr Stage rating (a standardised descriptive scale commonly used to describe the stage of PD) was available for 59 (49.2%) of the participants. For many of those participants who were not rated using the Hoehn and Yahr scale the consultant categorized them according to the stage of their care/treatment rather than the stage of their illness – this information was available for 46 (38.3%) participants. Therefore, in

order to describe the entire study sample in terms of the extent of their PD, the length of time since the initial diagnosis of PD was obtained from their medical notes. This is an unsatisfactory proxy measure for the stage of the illness but the only alternative available. The date of diagnosis was identified for all 120 participants – where only the year of diagnosis was reported in the PD notes an estimated date of 1st July that year was used.

Table 3.2 - Participant demographic details

Gender -	Male		67 (55.1%)
	Female		53 (44.9%)
Age -	Mean (SD)		73.1 yrs (9.40)
	Median (range)		74.0 yrs (41-91)
Hoehn & Yahr Stage (N, %) -		1	30 (25.0%)
		1.5	10 (8.3%)
		2	15 (12.5%)
		2.5	2 (1.7%)
		3	2 (1.7%)
		4	0
		5	0
		No rating available	61 (50.8%)
Alternative rating of PD – (N, %)		Untreated	3 (2.5%)
		Titration phase	19 (15.8%)
		Motor complications	23 (19.2%)
		Complex treatment	1 (0.8%)
		No rating available	74 (61.7%)
Time since diagnosis of PD -		Mean (SD)	48.6 months (52.1)
		Median (range)	25 (1-249)
PD medication (N, % Yes)			111 (92.5%)
Antidepressant medication (N, % Yes)			22 (18.3%)
MMSE Score -		Mean (SD)	27.9 (1.7)
		Median (range)	28.0 (24-30)

3.2 PREVALENCE OF DEPRESSION IN THE STUDY SAMPLE

The diagnosis of depression according to both DSM-IV and ICD-10 was obtained from the PSE-SCAN software. This software applies algorithms which provide diagnoses according to the established diagnostic criteria of DSM-IV and ICD-10. The software does not include an algorithm for the diagnosis of Minor Depression according to DSM-IV criteria - therefore diagnosis of Minor Depression was achieved by following and adapting the diagnostic algorithm for Major Depression manually (see Chapter 2.8).

A total of 19 participants (15.8%) were diagnosed with depression according to DSM-IV criteria – 5 with Minor depression and 14 with Major Depression. 14 participants were diagnosed with depression according to the ICD-10 criteria. All participants who were diagnosed with Major Depression according to the DSM-IV criteria also met criteria for depression on the ICD-10 criteria. No patients were diagnosed with dysthymia. See Table 3.3. The demographic details of those patients identified as depressed are shown in Table 3.4.

Table 3.3 - Diagnosis of depression derived from the SCAN interview

DSM-IV diagnosis	No diagnosis	101 (84.2%)
	Minor depression	5 (4.2%)
	Major Depression	14 (11.7%)
ICD-10 diagnosis	No diagnosis	106 (88.3%)
	Mild depression	8 (6.7%)
	Moderate depression	5 (4.2%)
	Severe depression	1 (0.8%)

Participants who were identified as depressed according to DSM-IV criteria were not significantly different in terms of gender (χ^2 =1.7, df=1, P=0.19), age (t= -0.80, df=118,

P=0.43), length of time since diagnosis of PD (t=0.44, df=118, P=0.66) nor MMSE score (t= 0.98, df=118, P=0.33) but had significantly higher EDS (t= -8.2, df=118, P<0.001), Brief EDS (t=-7.6, df=118, P<0.001), GDS-15 (t=-9.1, df=118, P<0.001) and GDS-4 score (t=-9.9, df=118, P<0.001). Those identified as depressed were also more likely to respond positively to (one or both questions of) the Whooley screening questions (χ^2 = 63.8, df=2, P<0.001).

Table 3.4 – Details of participants diagnosed with depression (DSM-IV Minor or Major Depression)

		Non-depressed (N=101)	Depressed (N=19)
Gender -	Male	59 (58.4%)	8 (42.1%)
	Female	42 (41.6%)	11 (57.9%)
Age (SD)		72.8 yrs (9.2)	74.6 yrs (10.5)
MMSE Score (Mean, SD)		27.9 (1.7)	27.5 (1.7)
Time since diagnosis of PD (Mean, SD)		49.5 (54.1)	43.7 (40.4)
Antidepressant medication (% Yes)		14 (13.9%)	8 (42.1%)
EDS Score (Mean, SD)		5.1 (3.7)	13.1 (3.7)
Brief EDS Score (Mean, SD)		3.3 (2.4)	8.16 (3.3)
Whooley questions	both 'No'	85 (84.2%)	0
	1 Question = Yes	11 (10.9%)	6 (31.6%)
	2 Question = Yes	5 (5.0%)	13 (68.4%)
GDS-15 Score (Mean, SD)		2.5 (2.5)	9.2 (4.7)
GDS-4 Score (Mean, SD)		0.3 (0.6)	2.2 (1.3)

3.2.1 Sub-threshold depression

Sub-threshold depression is described in the NICE guidelines for depression in chronic physical health problems (NICE, 2009b) as being those individuals who have "2-4 symptoms of depression with maintained function" (p22). Based on this definition (and using the nine criterion symptoms incorporated in DSM-IV diagnostic criteria), 5

participants had 2 or 3 symptoms of depression which did not interfere with their daily functioning (and neither of the two DSM-IV core symptoms) so therefore did not receive a diagnosis of Minor depression. For 4/5 participants the symptoms reported were disturbed sleep, retardation, or lack of energy.

3.2.2 Attribution of symptoms to depression or to PD

When completing the PSE-SCAN interview the presence or absence of each symptom is rated by the interviewer. It is also possible for the interviewer to indicate whether the symptom in question may be attributable to the presence of a physical illness. If a symptom has been attributed to a physical illness the diagnostic algorithms applied by the PSE-SCAN software exclude that symptom from contributing to a diagnosis of depression. It is possible that a symptom is discounted in this way when, if it were to be considered in the context of other symptoms, it would otherwise contribute to a diagnosis of depression. Therefore the data were examined to explore how many, if any, participants would have been diagnosed with depression if symptoms attributed to PD had been included in the diagnostic algorithm.

No additional participants received a diagnosis of Major Depression if all symptoms, regardless of attribution, were included in the diagnosis. However, two would have been diagnosed with Minor depression and two with sub-threshold depression.

3.3 PERFORMANCE OF THE DEPRESSION SCREENING MEASURES

All 120 participants completed the three depression screening measures. Table 3.5 summarises the results of the depression screening measures. Each participant was given the choice of either: completing the form themselves; reading the items

themselves and verbally indicating the response they would choose; or having the items read to them and then verbally indicating their choice of response. The majority of participants opted to read the items and indicate verbally which response they would choose – this was often because they experienced difficulty writing because of their PD. Those who opted to have the items and responses read to them tended to do so either because of poor eyesight or limited reading ability. Only two (1.7%) patients chose to complete the screening measures themselves, 104 (86.7%) chose to read the items and indicate their response verbally, and 14 (11.7%) opted to have the items read to them.

Table 3.5 - Screening Scale Scores

EDS Score -	Mean (SD)	6.4 (4.8)
	Median (range)	5.5 (0-22)
Brief EDS Score -	Mean (SD)	4.1 (3.1)
	Median (range)	3.5 (0-14)
Whooley questions –	'No' to both Qs	85 (70.8%)
	'Yes' to one question	17 (14.2%)
	'Yes' to both	18 (15.0%)
GDS-15 -	Mean (SD)	3.5 (3.8)
	Median (range)	2.0 (0-24)
GDS-4 -	Mean (SD)	0.6 (1.0)
	Median (range)	0.0 (0-4)

The screening performance of each measure was examined in terms of the ability of the scale to identify cases of depression at various cut-off thresholds. This was evaluated by calculating the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), the positive likelihood ratio (PLR), the negative likelihood ratio (NLR) and the diagnostic odds ratio (DOR). The overall misclassification rate (OMR) was also calculated. A receiver operating characteristic (ROC) curve analysis was also

performed for each screening measure to demonstrate the optimal cut-off for the identification of positive cases in this patient group. The ROC analysis also enabled calculation of the area under the curve (AUC) for each measure.

In light of the fact that when screening for depression in clinical practice one is likely to wish to identify cases of Minor/early depression this analysis was carried out focussing on the identification of those participants with Minor or Major Depression according to DSM-IV criteria.

3.3.1 The Edinburgh Depression Scale

Table 3.6 shows the distribution of scores on the EDS for both depressed and non-depressed participants and Table 3.7 summarises the performance characteristics of the EDS when different threshold scores are used to identify positive cases. Setting a low cut-off threshold clearly ensures the correct identification of positive cases (i.e. high sensitivity) but there is a corresponding trade-off in the specificity resulting in more false-positive cases. Selecting a higher cut-off threshold increases the specificity of the scale (giving fewer false-positives) but then leads to a reduction in the sensitivity (meaning that cases of depression are missed).

Table 3.6 - Distribution of participants' scores on the EDS

EDS Score	Non-depressed (N=101) (% scoring above)	Depressed (N=19) (% scoring above)
0	10 (90.1)	0 (100)
1	9 (81.2)	0 (100)
2	7 (74.3)	0 (100)
3	17 (57.4)	0 (100)
4	10 (47.5)	0 (100)
5	6 (41.6)	1 (94.7)
6	5 (36.6)	1 (89.5)
7	8 (28.7)	1 (84.2)
8	9 (19.8)	1 (78.9)
9	8 (11.9)	1 (73.7)
10	4 (7.9)	0 (73.7)
11	2 (5.9)	2 (63.2)
12	2 (4.0)	1 (57.9)
13	2 (2.0)	1 (52.6)
14	1 (1.0)	3 (36.8)
15	1 (0)	0 (36.8)
16	0	4 (15.8)
17	0	1 (10.5)
18	0	0 (10.5)
19	0	0 (10.5)
20	0	0 (10.5)
21	0	1 (5.3)
22	0	1 (0)

Table 3.7 - Test characteristics of the EDS for identification of Minor or Major Depression (DSM-IV)

Cut-off	Se (95% CI)	Sp (95% CI)	PPV (95% CI)	NPV (95% CI)	PLR	NLR	DOR	OMR
4/5	1.00 (0.83-1.00)	0.52 (0.43-0.62)	0.28 (0.19-0.40)	1.00 (0.93-1.00)	2.10	0.00	--	0.40
5/6	0.95 (0.75-0.99)	0.58 (0.49-0.68)	0.30 (0.20-0.43)	0.98 (0.91-1.0)	2.28	0.09	25.29	0.36
6/7	0.89 (0.69-0.97)	0.63 (0.54-0.72)	0.31 (0.21-0.45)	0.97 (0.90-0.99)	2.44	0.17	14.70	0.33
7/8	0.84 (0.62-0.95)	0.71 (0.62-0.79)	0.36 (0.23-0.50)	0.96 (0.89-0.99)	2.93	0.22	13.24	0.27
8/9	0.79 (0.57-0.92)	0.80 (0.71-0.87)	0.43 (0.28-0.59)	0.95 (0.89-0.98)	3.99	0.26	15.19	0.20
9/10	0.74 (0.51-0.88)	0.88 (0.80-0.93)	0.54 (0.36-0.71)	0.95 (0.88-0.98)	6.20	0.30	20.77	0.14
10/11*	0.74 (0.51-0.88)	0.92 (0.85-0.96)	0.64 (0.43-0.80)	0.95 (0.89-0.98)	9.30	0.29	32.55	0.11
11/12	0.63 (0.41-0.81)	0.94 (0.88-0.97)	0.67 (0.44-0.84)	0.93 (0.87-0.97)	10.63	0.39	27.14	0.11
12/13	0.58 (0.36-0.77)	0.96 (0.90-0.98)	0.73 (0.48-0.89)	0.92 (0.86-0.96)	14.62	0.44	33.34	0.10
13/14	0.53 (0.32-0.73)	0.98 (0.93-0.99)	0.83 (0.55-0.95)	0.92 (0.85-0.96)	26.58	0.48	55.00	0.09
14/15	0.37 (0.19-0.59)	0.99 (0.95-1.0)	0.88 (0.53-0.98)	0.89 (0.82-0.94)	37.21	0.64	58.33	0.11

*optimal cut-off

Figure 3.1 shows the ROC curve for the EDS identification of DSM-IV Minor or Major Depression. The optimal cut-off for the scale in this sample is indicated by the point that lies closest to the top left-hand corner of the plot, which represents the cut-off which maximises the sensitivity and specificity of the scale. In this instance the optimal cut-off is 10/11 which was found to correctly identify 14/19 (74%) participants who were diagnosed with depression, and incorrectly identified 8/101 (8%) non-depressed participants. At this threshold 2/5 participants with sub-threshold depression were

identified. The area under the curve, calculated from the ROC curve analysis, was 0.904 ($p<0.001$, 95% confidence interval 0.834 - 0.974).

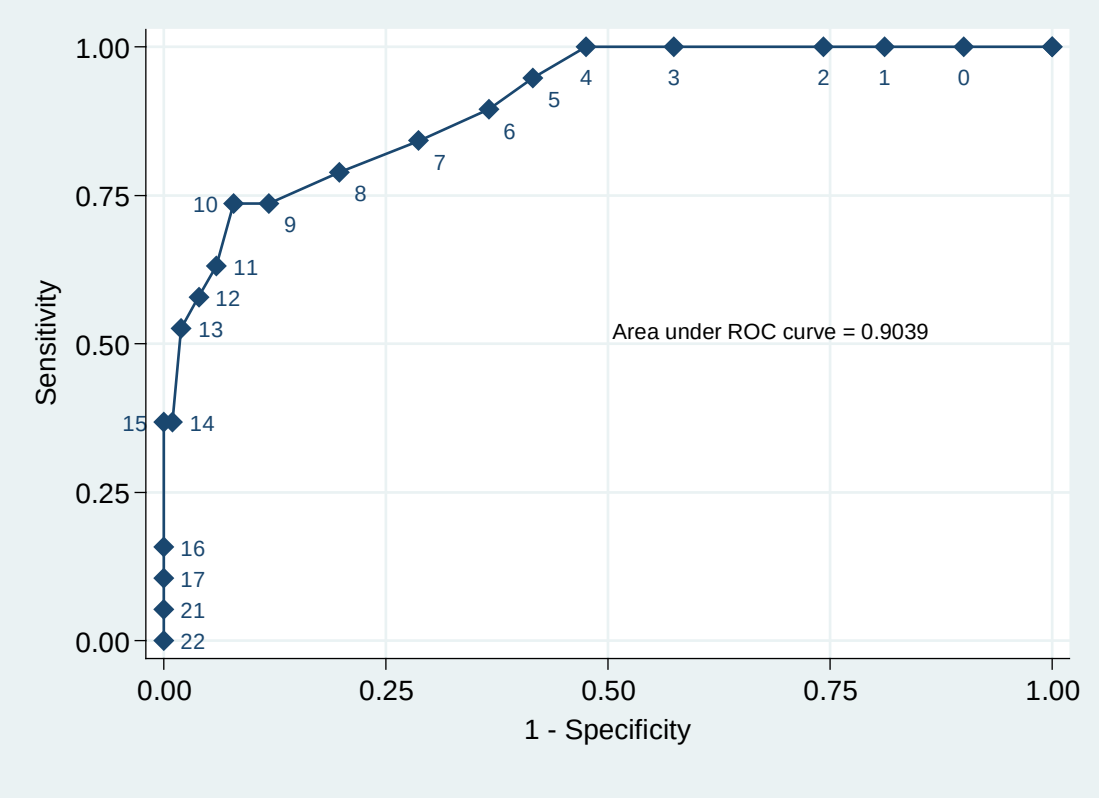
Table 3.8 shows the test characteristics of the EDS for the identification of those participants diagnosed with Major Depression (and ICD-10 depression). As for the identification of 'any depression' the optimal cut-off for the identification of Major Depression would probably be 10/11. Although a lower cut-off gives higher sensitivity it is at the cost of poor specificity – the 10/11 cut-off appears to give a better balance. At this cut-off 11/14 (79%) cases of Major Depression were correctly identified and 11/106 (10%) of non-depressed participants were incorrectly identified as cases.

Table 3.8 - Test characteristics of the EDS for identification of Major Depression (DSM-IV)

Cut-off	Se (95% CI)	Sp (95% CI)	PPV (95% CI)	NPV (95% CI)	PLR	NLR	DOR	OMR
4/5	1.00 (0.79-1.00)	0.50 (0.41-0.59)	0.21 (0.13-0.32)	1.00 0.93-1.00)	2.00	0.00	--	0.44
5/6	0.93 (0.69-0.99)	0.56 (0.46-0.68)	0.22 (0.13-0.34)	0.98 (0.91-1.0)	2.09	0.13	16.32	0.40
6/7	0.86 (0.60-0.96)	0.60 (0.51-0.69)	0.22 (0.13-0.35)	0.97 (0.90-0.99)	2.16	0.24	9.14	0.37
7/8	0.86 (0.61-0.96)	0.69 (0.60-0.77)	0.27 (0.16-0.41)	0.97 (0.91-0.99)	2.75	0.21	13.27	0.29
8/9	0.86 (0.60-0.96)	0.78 (0.70-0.85)	0.34 (0.21-0.51)	0.98 (0.92-0.99)	3.95	0.18	21.65	0.21
9/10	0.79 (0.52-0.92)	0.86 (0.78-0.91)	0.42 (0.26-0.61)	0.97 (0.91-0.99)	5.55	0.25	22.24	0.15
10/11*	0.79 (0.52-0.92)	0.90 (0.82-0.94)	0.50 (0.31-0.69)	0.97 (0.91-0.99)	7.57	0.24	31.67	0.12
11/12	0.64 (0.39-0.84)	0.92 (0.85-0.96)	0.50 (0.29-0.71)	0.95 (0.89-0.98)	7.57	0.39	19.40	0.12
12/13	0.57 (0.33-0.79)	0.93 (0.87-0.97)	0.53 (0.30-0.75)	0.94 (0.88-0.97)	8.65	0.46	18.86	0.11
13/14	0.50 (0.27-0.73)	0.95 (0.89-0.98)	0.58 (0.32-0.81)	0.94 (0.87-0.97)	10.60	0.52	20.20	0.10
14/15	0.71 (0.45-0.88)	0.96 (0.91-0.99)	0.71 (0.45-0.88)	0.96 (0.91-0.99)	18.93	0.30	63.75	0.07

*optimal cut-off

Figure 3.1 – ROC curve for the EDS identification of Minor and Major Depression



3.3.2 Brief Edinburgh Depression Scale (6-item)

Table 3.9 shows the distribution of scores for the Brief EDS. Table 3.10 summarises the performance of the Brief EDS across a range of thresholds. Figure 3.2 shows the ROC curve for the Brief EDS identification of DSM-IV Minor or Major Depression. The optimal cut-off for the scale (in terms of optimal combination of sensitivity and specificity) in this sample is 4/5 which was found to correctly identify 17/19 (90%) participants who were diagnosed with depression, but incorrectly identify 32/101 non-depressed participants. At this threshold 4/5 participants with sub-threshold depression were identified. The area under the curve for the Brief EDS, calculated from the ROC curve analysis, was 0.882 ($p<0.001$, 95% confidence interval 0.808 - 0.956). Examination of the ROC curve suggests that a cut-off of 5/6 may, depending on the

situation in which it is used, be preferable to the lower threshold score of 4/5. This would correctly identify 14/19 (74%) depressed participants (10/14 of those with Major Depression) and incorrectly identify 21/101 (21%) non-depressed.

Table 3.9 - Distribution of participants' scores on the Brief EDS

Brief EDS Score	Non-depressed (N=101) (% scoring above)	Depressed (N=19) (% scoring above)
0	13 (87.1)	0 (100)
1	10 (77.2)	0 (100)
2	21 (56.4)	0 (100)
3	15 (41.6)	1 (94.7)
4	10 (31.7)	1 (89.5)
5	11 (20.8)	3 (73.7)
6	11 (9.9)	3 (57.9)
7	5 (5.0)	1 (52.6)
8	3 (2.0)	0 (52.6)
9	1 (1.0)	3 (36.8)
10	1 (0)	3 (21.1)
11	0	1 (15.8)
12	0	1 (10.5)
13	0	0 (10.5)
14	0	2 (0)

Table 3.10 - Test characteristics of the Brief EDS for identification of Minor or Major Depression (DSM-IV)

Cut-off	Se (95% CI)	Sp (95% CI)	PPV (95% CI)	NPV (95% CI)	PLR	NLR	DOR	OMR
2/3	1.00 (0.83-1.0)	0.44 (0.34-0.53)	0.25 (0.17-0.36)	1.00 (0.92-1.0)	1.77	0.00	--	0.48
3/4	0.95 (0.75-0.99)	0.58 (0.49-0.68)	0.30 (0.20-0.43)	0.98 (0.91-1.0)	2.28	0.09	25.29	0.36
4/5*	0.89 (0.69-0.97)	0.68 (0.59-0.77)	0.35 (0.23-0.49)	0.97 (0.90-0.99)	2.82	0.15	18.33	0.28
5/6	0.74 (0.51-0.88)	0.79 (0.70-0.86)	0.40 (0.26-0.56)	0.94 (0.89-0.98)	3.54	0.33	10.67	0.22
6/7	0.58 (0.36-0.77)	0.90 (0.83-0.95)	0.52 (0.32-0.72)	0.92 (0.85-0.96)	5.85	0.47	12.51	0.15
7/8	0.53 (0.32-0.73)	0.95 (0.89-0.98)	0.67 (0.42-0.85)	0.91 (0.85-0.95)	10.63	0.50	21.33	0.12
8/9	0.53 (0.32-0.73)	0.98 (0.93-1.0)	0.83 (0.55-0.95)	0.92 (0.85-0.96)	26.58	0.48	55.00	0.09

*optimal cut-off

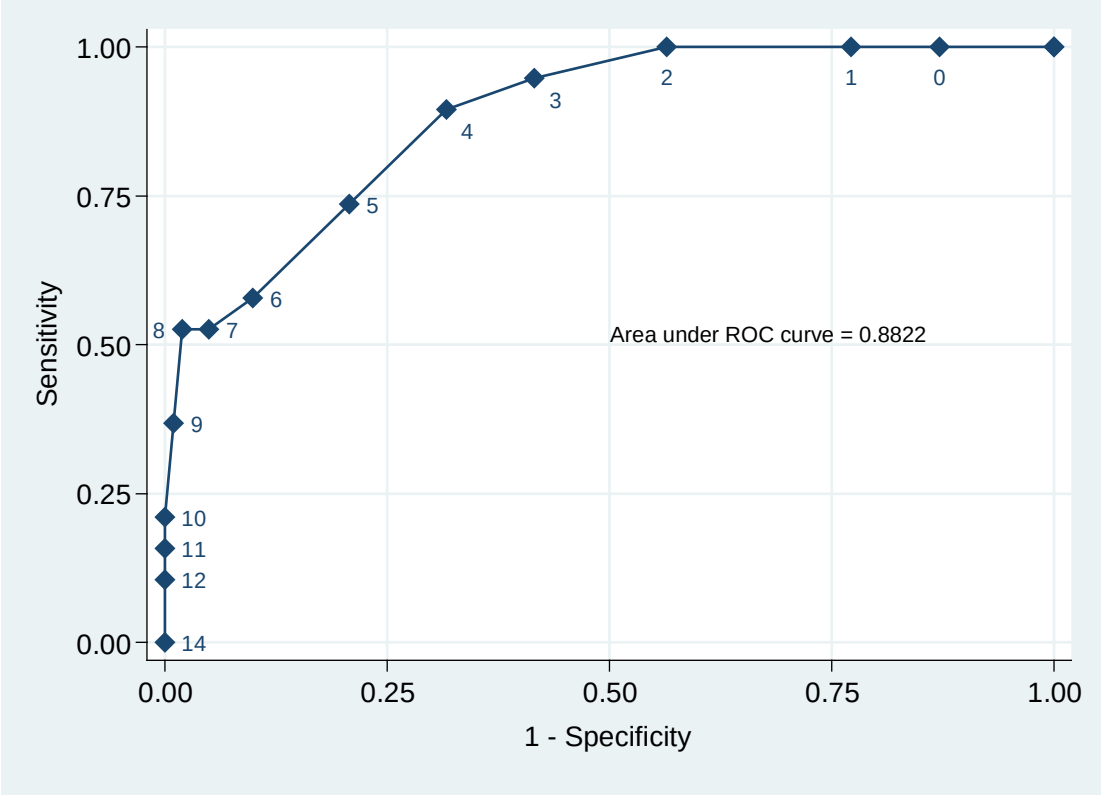
Table 3.11 shows the test characteristics of the Brief EDS for the identification of those participants diagnosed with Major Depression. The optimal cut-off for the identification of Major Depression was 4/5. At this cut-off 13/14 (93%) cases of Major Depression were correctly identified and 36/106 (34%) of non-depressed participants were incorrectly identified as cases.

Table 3.11 - Test characteristics of the Brief EDS for identification of Major Depression (DSM-IV)

Cut-off	Se (95% CI)	Sp (95% CI)	PPV (95% CI)	NPV (95% CI)	PLR	NLR	DOR	OMR
2/3	1.00 (0.79-1.0)	0.42 (0.33-0.51)	0.18 (0.11-0.29)	1.00 (0.92-1.0)	1.71	0	--	0.52
3/4	0.93 (0.69-0.99)	0.56 (0.46-0.65)	0.22 (0.13-0.34)	0.98 (0.91-1.0)	2.09	0.13	16.32	0.40
4/5*	0.93 (0.69-0.99)	0.66 (0.57-0.74)	0.27 (0.16-0.40)	0.99 (0.92-1.0)	2.73	0.11	25.28	0.31
5/6	0.71 (0.45-0.88)	0.76 (0.68-0.84)	0.29 (0.16-0.45)	0.95 (0.89-0.98)	3.03	0.37	8.10	0.24
6/7	0.57 (0.33-0.79)	0.88 (0.80-0.93)	0.38 (0.21-0.59)	0.94 (0.87-0.97)	4.66	0.49	9.54	0.16
7/8	0.50 (0.27-0.73)	0.92 (0.86-0.96)	0.47 (0.25-0.70)	0.93 (0.87-0.97)	6.63	0.54	12.25	0.13
8/9	0.50 (0.27-0.73)	0.95 (0.89-0.98)	0.58 (0.32-0.81)	0.94 (0.87-0.97)	10.60	0.52	20.20	0.10

*optimal cut-off

Figure 3.2 – ROC curve for the Brief EDS identification of Minor and Major Depression



3.3.3 Whooley screening questions

Table 3.12 shows the distribution of responses to the two Whooley screening questions. Table 3.13 summarises the performance of the two questions when a positive response to either or both are used to indicate a positive case. Figure 3.3 shows the ROC curve for the Whooley questions identification of DSM-IV Minor or Major Depression. The optimal cut-off for the scale in this sample is 0/1, (that is, a positive response to at least one of the two screening questions) which was found to correctly identify all 19 participants who were diagnosed with depression, and incorrectly identify 16/101 (16%) non-depressed participants. At this threshold 2/5 participants with sub-threshold depression were identified. The area under the curve for the Whooley questions, calculated from the ROC curve analysis, was 0.950 ($p<0.001$, 95% confidence interval 0.914 - 0.987).

Table 3.12 - Distribution of participants' responses on the two Whooley questions

Whooley questions	Non-depressed (N=101) (% scoring above)	Depressed (N=19) (% scoring above)
Both 'No' (score = 0)	85 (15.9)	0 (100)
'Yes' to one question (score = 1)	11 (5.0)	6 (68.4)
'Yes' to both questions (score = 2)	5 (0)	13 (0)

Table 3.13 - Test characteristics of the Whooley questions for identification of Minor or Major Depression (DSM-IV)

Cut-off	Se (95% CI)	Sp (95% CI)	PPV (95% CI)	NPV (95% CI)	PLR	NLR	DOR	OMR
0/1*	1.00 (0.83-1.0)	0.84 (0.76-0.90)	0.54 (0.38-0.70)	1.00 (0.96-1.0)	6.31	0	--	0.13
1/2	0.68 (0.46-0.85)	0.95 (0.89-0.98)	0.72 (0.49-0.88)	0.94 (0.88-0.97)	13.82	0.33	41.60	0.09

*optimal cut-off

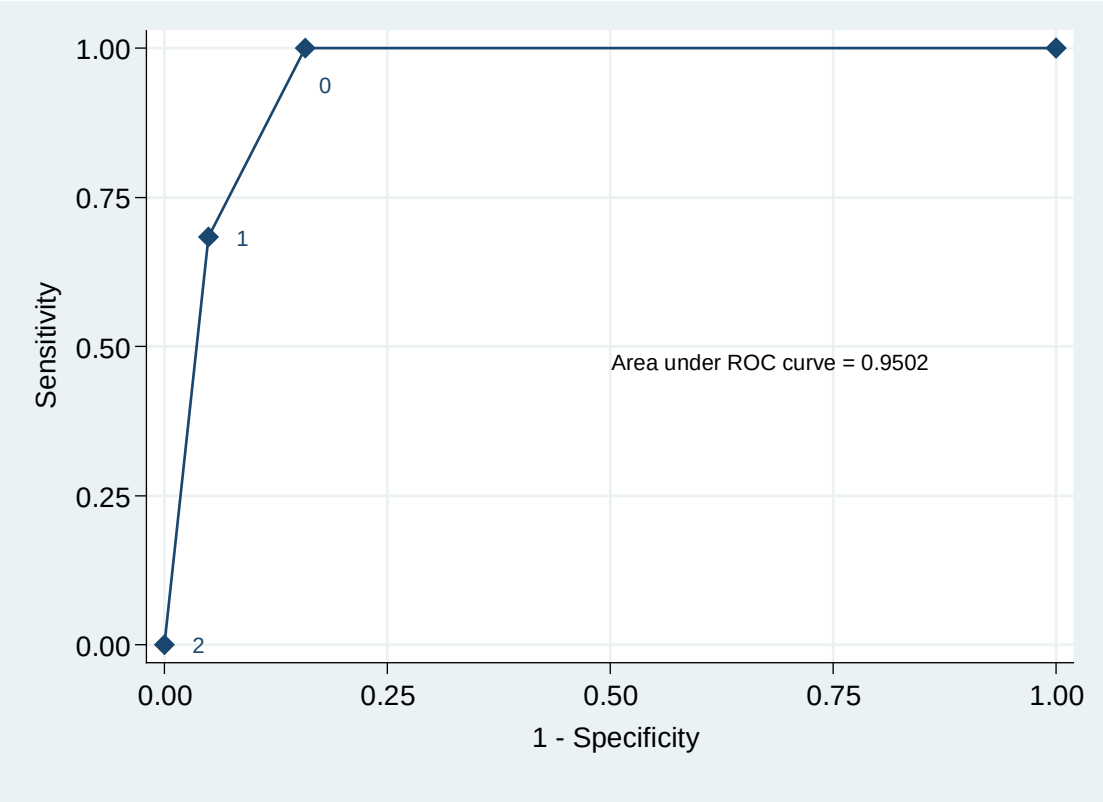
Table 3.14 shows the test characteristics of the Whooley questions for the identification of Major Depression. The optimal cut-off was 0/1 - at this cut-off 14/14 (100%) cases of Major Depression were correctly identified and 21/106 (20%) of non-depressed participants were incorrectly identified as cases.

Table 3.14 - Test characteristics of the Whooley questions for identification of Major Depression (DSM-IV)

Cut-off	Se (95% CI)	Sp (95% CI)	PPV (95% CI)	NPV (95% CI)	PLR	NLR	DOR	OMR
0/1*	1.00 (0.79-1.0)	0.80 (0.72-0.87)	0.40 (0.26-0.56)	1.00 (0.96-1.0)	5.05	0	--	0.18
1/2	0.64 (0.39-0.84)	0.92 (0.85-0.96)	0.50 (0.29-0.71)	0.95 (0.89-0.98)	7.57	0.39	19.40	0.12

*optimal cut-off

Figure 3.3 – ROC curve for the two Whooley questions identification of Minor and Major Depression



3.3.4 Geriatric Depression Scale (15-item)

Table 3.15 shows the distribution of scores on the GDS-15 for both depressed and non-depressed participants, and Table 3.16 summarises the performance characteristics of the GDS-15 when different threshold scores are used to identify positive cases.

Table 3.15 - Distribution of participants' scores on the GDS-15

GDS Score	Non-depressed (N=101) (% scoring above)	Depressed (N=19) (% scoring above)
0	23 (77.2)	0 (100.0)
1	23 (54.5)	1 (94.7)
2	21 (33.7)	0 (94.7)
3	5 (28.7)	0 (94.7)
4	5 (23.8)	1 (89.5)
5	13 (10.9)	1 (84.2)
6	5 (5.9)	1 (78.9)
7	2 (4.0)	2 (68.4)
8	1 (3.0)	2 (57.9)
9	2 (1.0)	3 (42.1)
10	0 (1.0)	3 (26.3)
11	0 (1.0)	1 (21.1)
12	0 (1.0)	2 (10.5)
13	0 (1.0)	1 (5.3)
14	1 (0)	0 (5.3)
15	0	0 (5.3)
16	0	0 (5.3)
17	0	0 (5.3)
18	0	0 (5.3)
19	0	0 (5.3)
20	0	0 (5.3)
21	0	0 (5.3)
22	0	0 (5.3)
23	0	0 (5.3)
24	0	1 (0.0)

Table 3.16 - Test characteristics of the GDS-15 for identification of Minor or Major Depression (DSM-IV)

Cut-off	Se (95% CI)	Sp (95% CI)	PPV (95% CI)	NPV (95% CI)	PLR	NLR	DOR	OMR
2/3	0.95 (0.75-0.99)	0.66 (0.57-0.75)	0.35 (0.23-0.48)	0.99 (0.92-1.0)	2.81	0.08	35.47	0.29
3/4	0.95 (0.75-0.99)	0.71 (0.62-0.79)	0.38 (0.26-0.53)	0.99 (0.93-1.0)	3.30	0.07	44.69	0.25
4/5	0.89 (0.69-0.97)	0.76 (0.67-0.84)	0.41 (0.28-0.57)	0.97 (0.91-0.99)	3.77	0.14	27.27	0.22
5/6*	0.84 (0.62-0.95)	0.89 (0.82-0.94)	0.59 (0.41-0.76)	0.97 (0.91-0.99)	7.73	0.18	43.64	0.12
6/7	0.79 (0.57-0.92)	0.94 (0.88-0.97)	0.71 (0.50-0.86)	0.96 (0.90-0.98)	13.29	0.22	59.38	0.08
7/8	0.68 (0.46-0.85)	0.96 (0.90-0.98)	0.76 (0.53-0.90)	0.94 (0.88-0.97)	17.28	0.33	52.54	0.08
8/9	0.58 (0.36-0.77)	0.97 (0.92-0.99)	0.79 (0.52-0.92)	0.92 (0.86-0.96)	19.49	0.43	44.92	0.09
9/10	0.42 (0.23-0.64)	0.99 (0.95-1.0)	0.89 (0.57-0.98)	0.90 (0.83-0.94)	42.53	0.58	72.73	0.10

*optimal cut-off

Figure 3.4 shows the ROC curve for the GDS-15 identification of DSM-IV Minor or Major Depression. The optimal cut-off for the scale in this sample is 5/6 which was found to correctly identify 16/19 (84%) participants who were diagnosed with depression, and incorrectly identify 11/101 (11%) non-depressed participants. At this threshold 2/5 participants with sub-threshold depression were identified. The area under the curve for the GDS-15, calculated from the ROC curve analysis was 0.922 ($p<0.001$, 95% confidence interval 0.849 - 0.996).

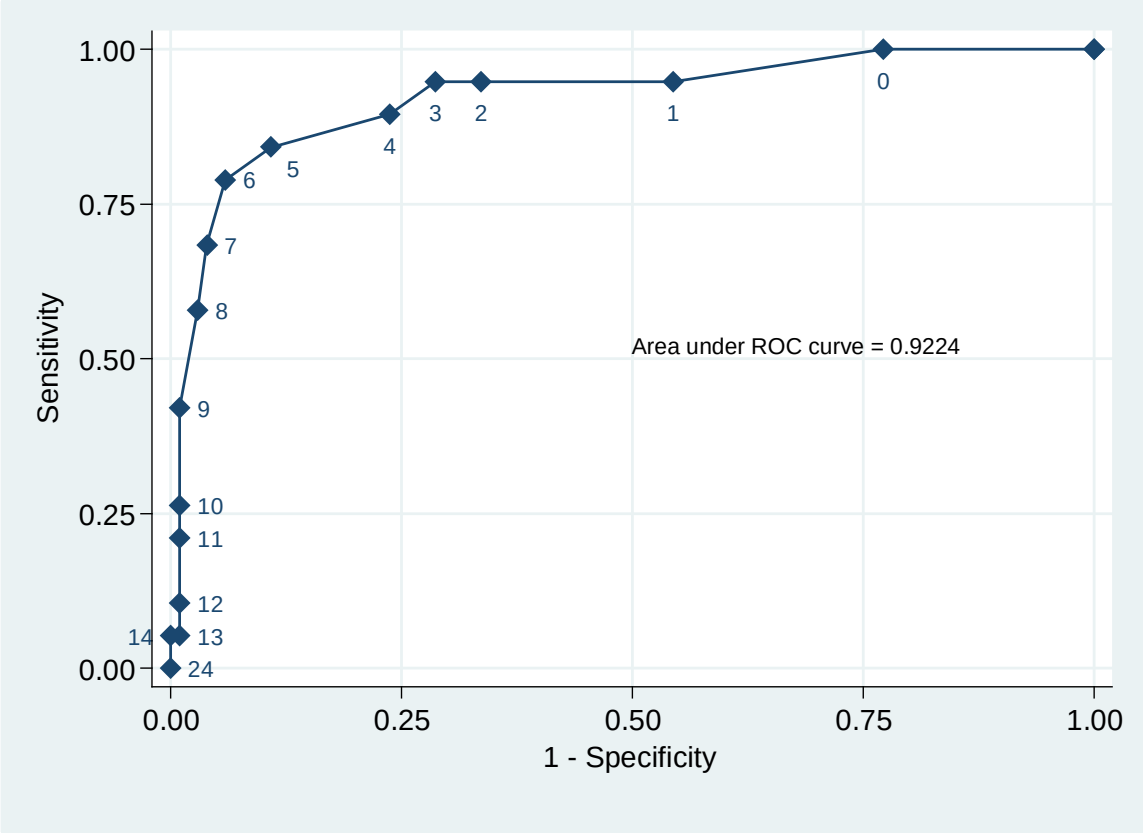
Table 3.17 - Test characteristics of the GDS-15 for identification of Major Depression (DSM-IV)

Cut-off	Se (95% CI)	Sp (95% CI)	PPV (95% CI)	NPV (95% CI)	PLR	NLR	DOR	OMR
2/3	1.00 (0.79-1.0)	0.64 (0.56-0.73)	0.27 (0.17-0.40)	1.00 (0.95-1.0)	2.79	0	--	0.32
3/4	1.00 (0.79-1.0)	0.69 (0.60-0.77)	0.30 (0.19-0.44)	1.00 (0.95-1.0)	3.21	0	--	0.28
4/5	0.93 (0.69-0.99)	0.74 (0.65-0.81)	0.32 (0.20-0.47)	0.99 (0.93-1.0)	3.52	0.10	36.21	0.24
5/6*	0.86 (0.60-0.96)	0.86 (0.78-0.91)	0.44 (0.28-0.63)	0.98 (0.93-0.99)	6.06	0.17	36.40	0.14
6/7	0.79 (0.52-0.92)	0.91 (0.84-0.95)	0.52 (0.32-0.72)	0.97 (0.92-0.99)	8.33	0.24	35.20	0.11
7/8	0.64 (0.39-0.84)	0.92 (0.86-0.96)	0.53 (0.31-0.74)	0.95 (0.89-0.98)	8.52	0.39	22.05	0.11
8/9	0.64 (0.39-0.84)	0.95 (0.89-0.98)	0.64 (0.39-0.84)	0.95 (0.89-0.98)	13.63	0.37	36.36	0.08
9/10	0.43 (0.21-0.67)	0.97 (0.92-0.99)	0.67 (0.35-0.88)	0.93 (0.86-0.96)	15.14	0.59	25.75	0.09

*optimal cut-off

Table 3.17 shows the test characteristics of the GDS-15 for the identification of those participants diagnosed with Major Depression. The optimal cut-off for the identification of Major Depression was also 5/6. At this cut-off 12/14 (86%) cases of Major Depression were correctly identified and 15/106 (14%) of non-depressed participants were incorrectly identified as cases.

Figure 3.4 – ROC curve for the GDS-15 identification of Minor and Major Depression



3.3.5 Geriatric Depression Scale (4-item)

Table 3.18 shows the distribution of scores on the GDS-4 for both depressed and non-depressed participants and Table 3.19 summarises the performance characteristics of the GDS-14 when different threshold scores are used to identify positive cases.

Table 3.18 - Distribution of participants' scores on the GDS-4

GDS-4 Score	Non-depressed (N=101) (% scoring above)	Depressed (N=19) (% scoring above)
0	77 (23.8)	3 (84.2)
1	19 (5.0)	2 (73.7)
2	3 (2.0)	4 (52.6)
3	2 (0)	8 (10.5)
4	0	2 (0)

Table 3.19 - Test characteristics of the GDS-4 for identification of Minor or Major Depression (DSM-IV)

Cut-off	Se (95% CI)	Sp (95% CI)	PPV (95% CI)	NPV (95% CI)	PLR	NLR	DOR	OMR
0/1	0.84 (0.62-0.95)	0.76 (0.67-0.84)	0.40 (0.26-0.55)	0.96 (0.90-0.99)	3.54	0.21	17.11	0.23
1/2*	0.74 (0.51-0.88)	0.95 (0.89-0.98)	0.74 (0.51-0.88)	0.95 (0.89-0.98)	14.88	0.28	53.76	0.08
2/3	0.53 (0.32-0.73)	0.98 (0.93-1.0)	0.83 (0.55-0.95)	0.92 (0.85-0.96)	26.58	0.48	55.00	0.09
3/4	0.11 (0.03-0.31)	1.00 (0.96-1.0)	1.00 (0.34-1.0)	0.86 (0.78-0.91)	--	0.89	--	0.14

*optimal cut-off

Figure 3.5 shows the ROC curve for the GDS-4 identification of DSM-IV Minor or Major Depression. The optimal cut-off for the scale in this sample is 1/2 which was found to correctly identify 14/19 (74%) participants who were diagnosed with depression, and incorrectly identify 5/101 (5%) non-depressed participants. At this threshold 0/5 participants with sub-threshold depression were identified. The area under the curve for the GDS-4, calculated from the ROC curve analysis was 0.876 ($p<0.001$, 95% confidence interval 0.767 - 0.984).

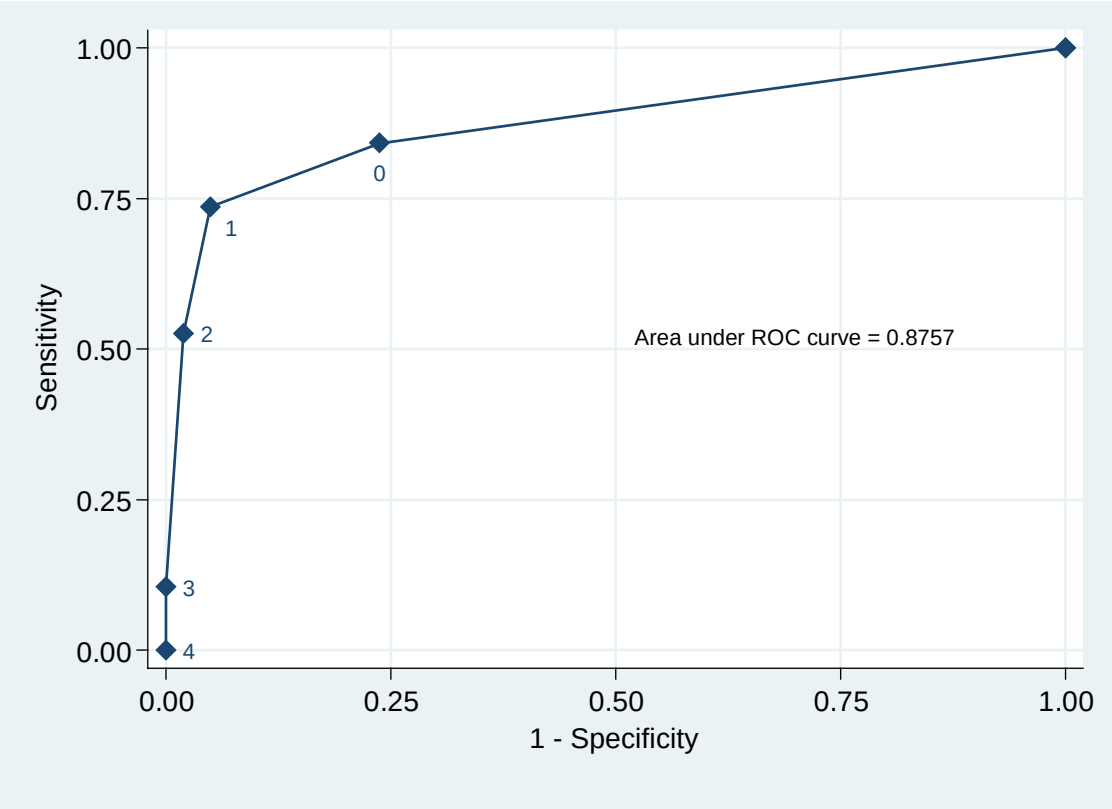
Table 3.20 shows the test characteristics of the GDS-4 for the identification of participants with Major Depression. The optimal cut-off for the identification of Major Depression was 1/2. At this cut-off 11/14 (79%) cases of Major Depression were correctly identified and 8/106 (8%) of non-depressed participants were incorrectly identified as cases.

Table 3.20 - Test characteristics of the GDS-4 for identification of Major Depression (DSM-IV)

Cut-off	Se (95% CI)	Sp (95% CI)	PPV (95% CI)	NPV (95% CI)	PLR	NLR	DOR	OMR
0/1	0.86 (0.60-0.96)	0.74 (0.65-0.81)	0.30 (0.18-0.45)	0.98 (0.91-0.99)	3.24	0.19	16.71	0.25
1/2*	0.79 (0.52-0.92)	0.92 (0.86-0.96)	0.58 (0.36-0.77)	0.97 (0.92-0.99)	10.41	0.23	44.92	0.09
2/3	0.50 (0.27-0.73)	0.95 (0.89-0.98)	0.58 (0.32-0.81)	0.94 (0.87-0.97)	10.60	0.52	20.20	0.10
3/4	0.07 (0.01-0.32)	0.99 (0.95-1.0)	0.50 (0.10-0.91)	0.89 (0.82-0.93)	7.57	0.94	8.08	0.12

*optimal cut-off

Figure 3.5 – ROC curve for the GDS-4 identification of Minor and Major Depression



3.4 COMPARISON OF THE PERFORMANCE OF THE SCREENING MEASURES

Each of the screening measures included in this study demonstrated good performance characteristics for the identification of depression according to DSM-IV criteria for Minor or Major Depression. Identification of the screening measure which performed the best in this sample group is best achieved by comparison of the area under the ROC curve.

3.4.1 Identification of Minor and Major Depression

Table 3.21 shows the area under the ROC curve for each of the screening measures for the identification of cases of Minor and Major Depression. The Whooley screening questions were shown to have the largest area under the curve (0.95), followed by the

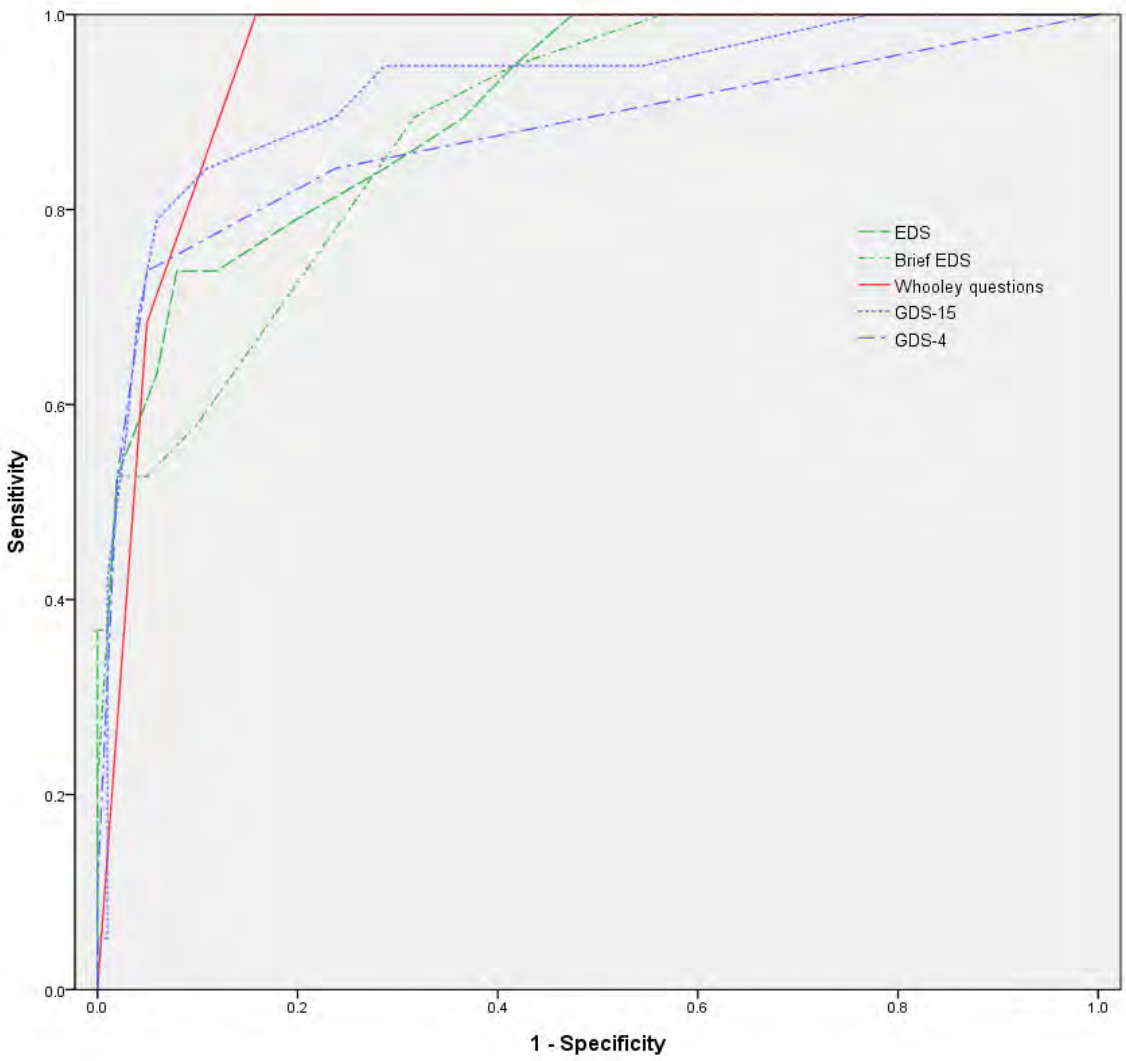
GDS-15 (0.92) and the EDS (0.90). The shortened versions of the EDS and the GDS performed less well than their longer counterparts. The ROC curves for the EDS, Brief EDS, Whooley questions, GDS-15 and the GDS-4 are displayed in Figure 3.6.

Table 3.21 – Area under the curve for the identification of Minor and Major Depression

	AUC	Std error	P	95% confidence interval
EDS	0.904	0.036	P<0.001	0.834-0.974
Brief EDS	0.882	0.038	P<0.001	0.808-0.956
Whooley Questions	0.950	0.019	P<0.001	0.914-0.987
GDS-15	0.922	0.037	P<0.001	0.849-0.996
GDS-4	0.876	0.055	P<0.001	0.767-0.984

The AUC for the EDS, Whooley questions and GDS were compared using the method for comparison of AUC derived from the same cases (Hanley & McNeil, 1983). Although the values of the AUC curves are similar the curves themselves may be significantly different. This calculation takes into account the fact that comparisons between measures completed on the same subjects will have an inherent correlation. When the AUC of the three main measures were compared none was significantly different from any of the others (EDS and Whooley questions z = 1.278, 95% CI -0.025 to 0.118, P=0.20; GDS-15 and Whooley questions z = 0.713, 95% CI -0.049 to 0.105, P=0.48; EDS and GDS-15 z = 0.465, 95% CI -0.060 to 0.097, P=0.64).

Figure 3.6 – Comparison of ROC curves for the screening measures' ability to identify Minor and Major Depression



3.4.2 Identification of DSM-IV Major Depression

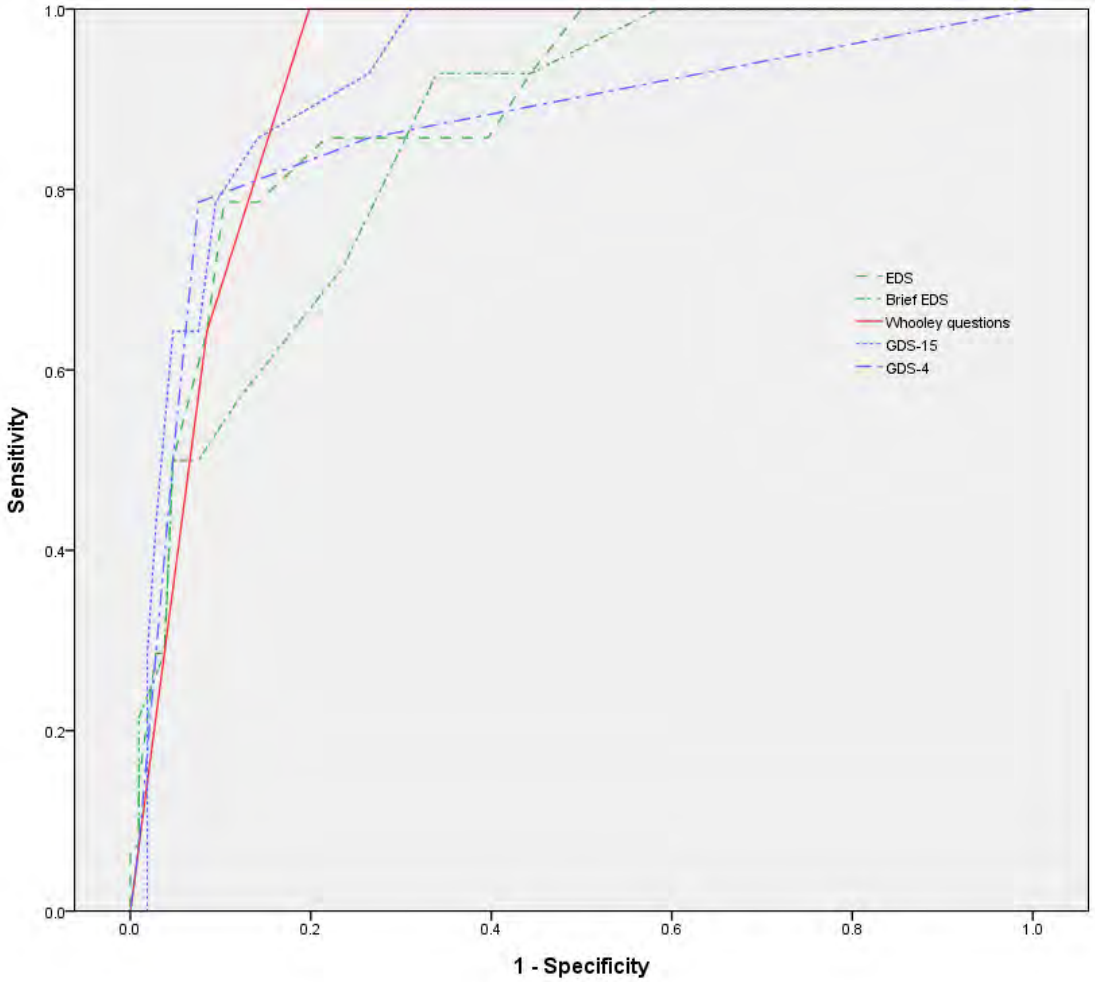
Table 3.22 shows the area under the ROC curve for each of the screening measures in terms of their ability to identify cases of Major Depression. The Whooley screening questions and the GDS-15 were shown to have the largest area under the curve (0.92 and 0.93), followed by the EDS (0.89). The shortened versions of the EDS and the GDS performed less well than their longer counterparts. The ROC curves for the EDS, Brief EDS, Whooley questions, GDS-15 and the GDS-4 are displayed in Figure 3.7.

Table 3.22 – Area under the curve for the identification of Major Depression

	AUC	Std error	P	95% Confidence interval
EDS	0.888	0.043	P<0.001	0.805-0.972
Brief EDS	0.856	0.045	P<0.001	0.768-0.945
Whooley Questions	0.922	0.025	P<0.001	0.873-0.971
GDS-15	0.928	0.027	P<0.001	0.874-0.981
GDS-4	0.868	0.062	P<0.001	0.746-0.989

When the AUC of the three main measures were compared none was significantly different from any of the others (EDS and Whooley questions z = 0.763, 95% CI -0.053 to 0.121, P=0.45; GDS-15 and Whooley questions z = 0.166, 95% CI -0.058 to 0.069, P=0.87; EDS and GDS-15 z = 1.01, 95% CI -0.038 to 0.116, P=0.31).

Figure 3.7 – Comparison of the ROC curves for the screening measures' ability to identify Major Depression



3.5 EXPLORATION OF A TWO-STAGE SCREENING PROCESS

NICE guidelines for depression in adults (NICE, 2009a) propose the use of a two-stage screening process. The guidelines recommend initial screening with the Whooley questions followed by a more detailed clinical assessment, which may comprise a second screening measure for those respondents who respond positively to either (or both) of the questions (page 108). Those who score above the threshold on the second measure can then be assessed further by their clinician or referred on for specialist assessment. A two-stage strategy has the advantage that a cut-off threshold for the

'first-stage' screening measure can be selected that would optimise the sensitivity of the scale (i.e. ensure identification of the optimum number of positive cases). If this choice of cut-off threshold leads to poor specificity (i.e. high false-positives) the cut-off threshold of the screening measure used as the second-stage of the process can be set to optimally exclude those respondents who have been incorrectly identified as depressed.

The data available from this study enables a simulation of such a strategy because all participants completed each of the screening measures. It is therefore possible to evaluate any potential additional benefit of using a two-stage process rather than using a single screening measure.

The AUC for the Whooley questions was the highest for all measures studied here. It is also evident from the data that a positive response to at least one of the two Whooley screening questions gave the best results – it correctly identified 19/19 (100%) of positive (depressed) cases, and incorrectly identified 16/101 (16%) non-cases. (It also has the advantage of being a very short measure which could easily be incorporated in to an assessment or routine review appointment in a clinic setting.) Therefore this was selected to be the preferred first-stage screening measure – in keeping with the recommendations of the NICE depression guidelines (2009). The performance of the remaining screening measures when used as a second-stage screening tool was then examined (for the identification of both Minor and Major Depression). This was achieved by examining the sensitivity and specificity, etc., of each measure for those respondents (N=35) who screened positive on the Whooley questions. The performance of the four measures in this scenario, when the optimal cut-off score is applied, is represented in Figure 3.9.

Table 3.23 shows the performance characteristics of the EDS as a second stage screening measure. Examination of the data indicates that the optimal cut-off score for the EDS is 10/11. This threshold represents the optimal specificity and PPV and, in the sample of 35 participants who were identified as positive cases by the Whooley questions, this cut-off correctly identified 14/19 depressed participants and 13/16 non-depressed. Using the cut-off 12/13, which has a slightly higher specificity and PPV, correctly identified 11/19 depressed and 14/16 non-depressed participants.

Table 3.23 - Performance of the EDS as second-stage screening measure

Cut-off	Se (95% CI)	Sp (95% CI)	PPV (95% CI)	NPV (95% CI)	DOR	Sp+ PPV	TP	TN
8/9	0.79 (0.57-0.92)	0.56 (0.33-0.77)	0.68 (0.47-0.84)	0.69 (0.42-0.87)	4.82	1.24	15/19	9/16
9/10	0.74 (0.51-0.88)	0.69 (0.44-0.86)	0.74 (0.51-0.88)	0.69 (0.44-0.86)	6.16	1.42	14/19	11/16
10/11*	0.74 (0.51-0.88)	0.81 (0.57-0.93)	0.82 (0.59-0.94)	0.72 (0.49-0.88)	12.13	1.64	14/19	13/16
11/12	0.63 (0.41-0.81)	0.81 (0.57-0.93)	0.80 (0.55-0.93)	0.65 (0.43-0.82)	7.43	1.61	12/19	13/16
12/13	0.58 (0.58-0.96)	0.88 (0.43-0.80)	0.85 (0.36-0.77)	0.64 (0.64-0.97)	9.63	1.72	11/19	14/16

TP= true positives i.e. proportion depressed correctly identified *optimal cut-off
TN= true negatives i.e. proportion non-depressed correctly identified

The optimal cut-off score for the Brief EDS as a second-stage screening measure was 7/8 (See Table 3.24). This threshold score correctly identified 10/19 depressed participants and gave only 2/16 false positives. Selection of a much lower threshold score, for example 5/6, yielded 14/19 correctly identified depressed cases but 7/16 false positives.

Table 3.24 - Performance of the Brief EDS as second-stage screening measure

Cut-off	Se (95% CI)	Sp (95% CI)	PPV (95% CI)	NPV (95% CI)	DOR	Sp+ PPV	TP	TN
4/5	0.89 (0.69-0.97)	0.31 (0.14-0.56)	0.61 (0.42-0.76)	0.71 (0.36-0.92)	3.86	0.92	17/19	5/16
5/6	0.74 (0.51-0.88)	0.56 (0.33-0.77)	0.67 (0.45-0.83)	0.64 (0.39-0.84)	3.60	1.23	14/19	9/16
6/7	0.58 (0.36-0.77)	0.75 (0.51-0.90)	0.73 (0.48-0.89)	0.60 (0.39-0.78)	4.13	1.48	11/19	12/16
7/8*	0.53 (0.32-0.73)	0.88 (0.64-0.97)	0.83 (0.55-0.95)	0.61 (0.41-0.78)	7.78	1.71	10/19	14/16
8/9	0.53 (0.32-0.73)	0.88 (0.64-0.97)	0.83 (0.55-0.95)	0.61 (0.41-0.78)	7.78	1.71	10/19	14/16
9/10	0.37 (0.19-0.59)	0.94 (0.72-0.99)	0.88 (0.53-0.98)	0.56 (0.37-0.72)	8.75	1.81	7/19	15/16

*optimal cut-off

When the GDS-15 is used as a second-stage screening measure (see Table 3.25) the optimal cut-off score would appear to be 6/7. At this threshold 15/19 depressed participants and 4/16 non-depressed are identified as positive cases.

Table 3.25 - Performance of the GDS-15 as second-stage screening measure

Cut-off	Se (95% CI)	Sp (95% CI)	PPV (95% CI)	NPV (95% CI)	DOR	Sp+ PPV	TP	TN
4/5	0.89 (0.69-0.97)	0.56 (0.33-0.77)	0.71 (0.51-0.85)	0.82 (0.52-0.95)	10.93	1.27	17/19	9/16
5/6	0.84 (0.62-0.95)	0.63 (0.39-0.82)	0.73 (0.52-0.89)	0.77 (0.50-0.92)	8.89	1.35	16/19	10/16
6/7*	0.79 (0.57-0.92)	0.75 (0.51-0.90)	0.79 (0.57-0.92)	0.75 (0.51-0.90)	11.25	1.54	15/19	12/16
7/8	0.68 (0.46-0.85)	0.81 (0.57-0.93)	0.81 (0.57-0.93)	0.68 (0.46-0.85)	9.39	1.63	13/19	13/16
8/9	0.58 (0.36-0.77)	0.81 (0.57-0.93)	0.79 (0.52-0.92)	0.62 (0.41-0.79)	5.96	1.60	11/19	13/16

*optimal cut-off

From examination of Table 3.26 the optimal cut-off for the GDS-4, in this scenario, would appear to be 1/2. At this cut-off 14/19 depressed participants and 2/16 non-depressed are identified as depressed. If a cut-off of 2/3 is applied only 10/19 depressed participants were identified as depressed, along with 1/16 non-depressed participant.

Table 3.26 - Performance of the GDS-4 as second-stage screening measure

Cut-off	**Se** (95% CI)	**Sp** (95% CI)	**PPV** (95% CI)	**NPV** (95% CI)	**DOR**	**Sp+ PPV**	**TP**	**TN**
0/1	0.84 (0.62-0.95)	0.31 (0.14-0.56)	0.59 (0.41-0.76)	0.63 (0.31-0.86)	2.42	0.91	16/19	5/16
1/2*	0.74 (0.51-0.88)	0.88 (0.64-0.97)	0.88 (0.64-0.97)	0.74 (0.51-0.88)	19.60	1.75	14/19	14/16
2/3	0.53 (0.32-0.73)	0.94 (0.72-0.99)	0.91 (0.62-0.98)	0.63 (0.43-0.79)	16.67	1.85	10/19	15/16

*optimal cut-off

ROC curve analysis of each screening measure in this smaller data sample (N=35) gave the AUC for each (see Table 3.27 and Figure 3.8). None of these AUC was significantly different from each other: EDS and Brief EDS z=0.874, 95% CI -0.043 to 0.112, P=0.382; EDS and GDS-15 z=0.112, 95% CI -0.190 to 0.213, P=0.911; EDS and GDS-4 z=0.050, 95% CI -0.188 to 0.198, P=0.960; Brief EDS and GDS-15 z=0.442, 95% CI -0.158 to 0.250, P=0.659; Brief EDS and GDS-4 z=0.389, 95% CI -0.159 to 0.238, P=0.697; GDS-15 and GDS-4 z=0.110, 95% CI -0.111 to 0.124, P=0.913.

Table 3.27 – Area under the curve for each measure as second-stage screening instrument (for the identification of Minor and Major Depression)

	AUC	Std error	P	95% confidence interval
EDS	0.786	0.078	P=0.004	0.633-0.939
Brief EDS	0.752	0.081	P=0.011	0.592-0.911
GDS-15	0.798	0.080	P=0.003	0.641-0.955
GDS-4	0.791	0.081	P=0.003	0.633-0.949

Figure 3.8 - Comparison of the ROC curves for each measure as a second-stage screening instrument following a positive result from initial screening with the Whooley questions

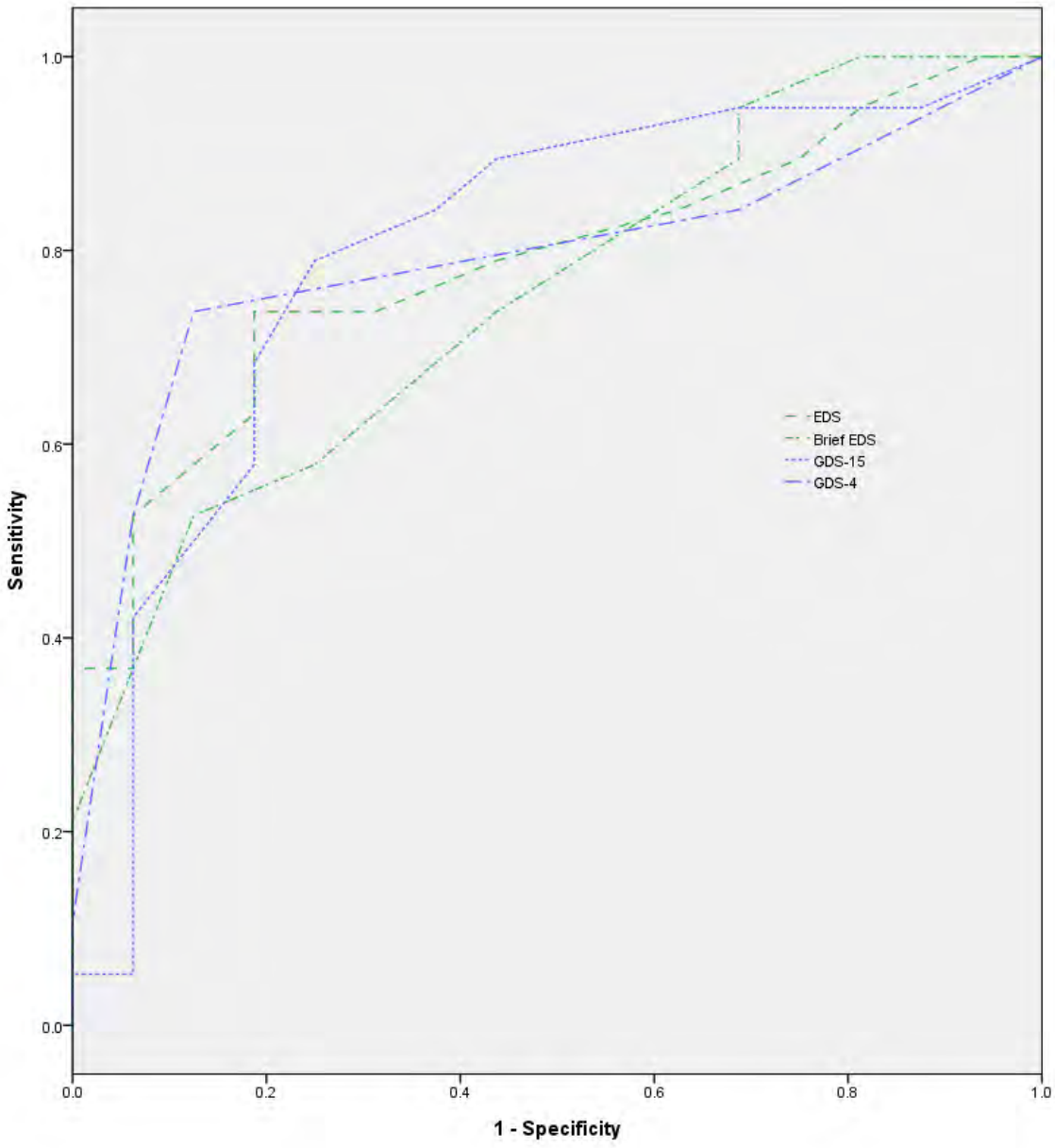
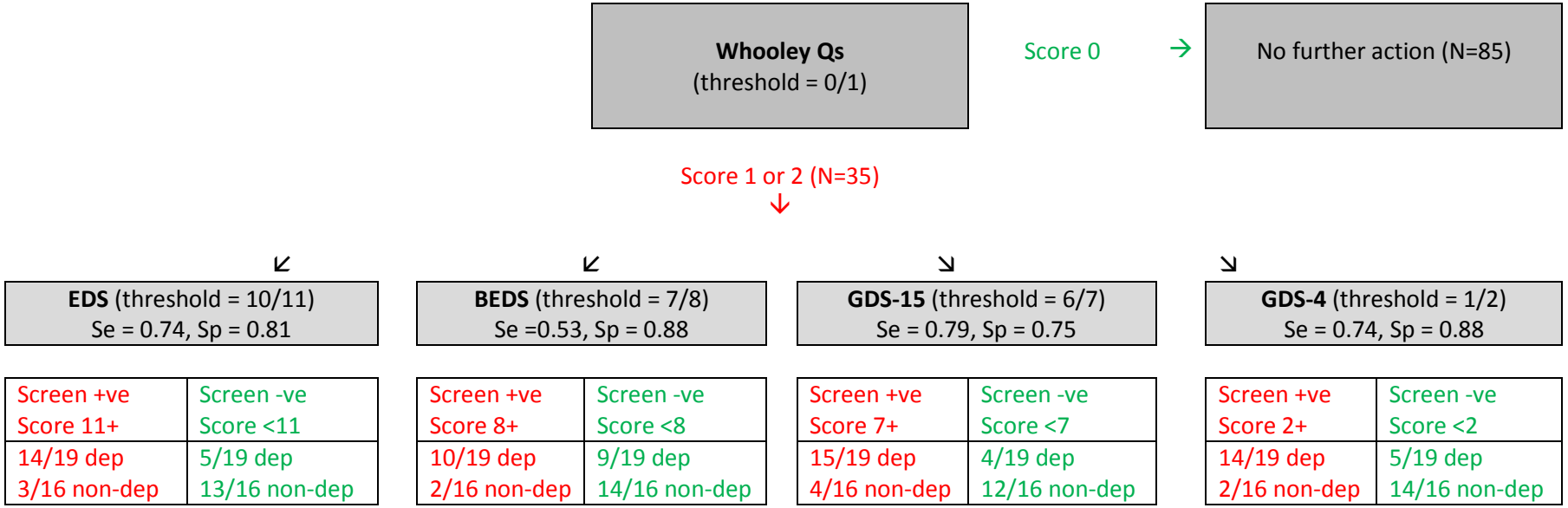


Figure 3.9 – Comparison of the screening measures in a two-stage screening scenario



CHAPTER 4 – DISCUSSION

This study was designed to assess the relative validity and utility of the EDS, the Whooley questions and the GDS-15 as screening measures for depression in patients with PD attending a hospital out-patient clinic. The participants all met standardised diagnostic criteria for PD and were not cognitively impaired.

4.1 The setting

Participants were recruited from the two outpatient clinics for PD in the local general hospital (with approximately equal numbers from each clinic). The majority of studies validating depression screening measures in PD have been conducted in out-patient neurology or movement disorder clinics, and there is no reason to believe that the clinics in this study would not be similar to PD outpatient clinics elsewhere. It is important that validation of a screening tool takes place in the setting and patient group in which it is intended to be used, both in terms of the type of patients included in the study and the prevalence of the condition being screened for.

4.2 The sample

If it is going to be possible to generalise the results of this study to other outpatient PD samples, it is important that the study sample is representative of PD out-patients (in terms of factors such as age, gender, MMSE score, stage of the illness and time since diagnosis), both in general and also of the local PD clinics where this research was carried out. On the whole, the study sample is comparable to those described in other published studies of a similar nature. The mean age of this study sample (73.1yrs) was a little higher than many other studies validating depression scales in PD (which range

between 58.4yrs to 73.4yrs)[6]. One of the clinics from which participants were recruited was a PD clinic run by a geriatrician. This clinic tended to include older patients than the other clinic and as it contributed over half of the study participants, this may be the reason why the average age was slightly higher than in other studies of this type. However, the study sample is similar to other studies in terms of the proportion of the sample that are male, and the mean MMSE score.

There is the possibility that selection bias occurred in this sample. Ideally, the study would have included consecutive patients attending the PD clinics, but there were breaks in the recruitment period due to lack of availability of researchers to conduct the PSE-SCAN interviews. The effect of these breaks in recruitment is unknown but is unlikely to have affected the prevalence of depression found and therefore should not have affected the scale performance measures (e.g. PPV/NPV).

Information regarding the severity of PD symptoms (Hoehn and Yahr stage) was not available for the whole sample. This is because the Hoehn & Yahr scale was not used by all consultants at the PD clinics – partly for the reason that it is not as reliable in older patients due to comorbid illness/frailty impairing mobility. Therefore, only one of the PD consultants used this rating on a regular basis. When planning the protocol we had been led to believe this information would be available for all participants, but it was too late to add it to the study procedure once it was realised it was not available for all. Lack of this information for all participants prevents the 'easy' description of the sample. Most other studies evaluating screening measures in PD use the Hoehn and Yahr scale to describe their sample and so direct comparison, in terms of PD severity,

[6] The mean age of symptom onset in PD is 62 to 70 years (Muangpaisan *et al.,* 2011)

with other studies is not possible. However, for those patients for whom Hoehn and Yahr stage was available, there were similar proportions of each stage compared to other studies in PD – that is, the sample was predominantly stage I-III. Several studies have also reported duration of PD symptoms (which range between 5.8 and 8.0 years) so this study sample has a somewhat lower duration (4.1 years) of PD compared to those studies.

All patients attending the PD clinics during the recruitment period who met the inclusion criteria were asked by their consultant if they would be willing to take part in the study. Many patients declined at this point and no information is available regarding these patients. It is possible that patients who are experiencing depressive symptoms are more likely to decline to participate in research than patients who are not feeling depressed. Whilst there is little evidence of the extent of this problem when conducting any research with patients who have depression, as in itself it is difficult to assess, many depressive symptoms, such as loss of self-confidence and self-esteem, social withdrawal, apathy and indifference could affect the likelihood of patients agreeing to take part in the research. Some patients who initially indicated to their consultant that they would be willing, later declined to participate when telephoned to arrange the first study visit. Many of these patients gave reasons for not wishing to take part (these included reasons such as physical illness of themselves or family member or the inconvenience of taking part because they were still working), but it is not possible to draw conclusions regarding whether they were experiencing depressive symptoms or not.

4.3 Prevalence of depression in the sample

The prevalence of depression in the study sample was much lower than expected. The meta-analysis by Reijnders *et al.* (2008) reported the prevalence of depression in out-patient samples of PD to be 24% (and 40% clinically significant depression) and in all studies using DSM-IV criteria 17% Major Depression, 22% Minor Depression and 13% Dysthymia[7]. The rate of depression found in this study, 12% DSM-IV Major Depression, and 4% Minor Depression, was considerably lower than this. This study also found lower prevalence of depression than reported in many studies validating depression scales in PD (e.g. Leentjens *et al*., 2000a, 2000b, 2001, 2003b; Naarding *et al.*, 2002; Tumas *et al.*, 2008; Williams and Marsh, 2009), which reported between 22 to 27% prevalence in their outpatient clinic samples (which all used DSM-IV depression criteria). However, some studies have reported similar prevalence in their validation studies (e.g. Visser *et al.*, 2006; Dissanayaka *et al.*, 2007).

Those participants who were found to be depressed according to DSM-IV criteria for Major Depression also all met diagnostic criteria for ICD10 depression (i.e. application of ICD-10 and DSM-IV diagnostic criteria identified the same participants). In addition to the 14 participants identified as having Major Depression there were five who met DSM-IV diagnostic criteria for Minor Depression and five who met criteria (as described in NICE, 2009b) for sub-threshold depression. This would give a prevalence of 20% of the study sample having 'clinically significant' depressive symptoms – still substantially lower than reported in prevalence studies of depression in PD (although the validity of criteria for sub-threshold depression remains to be established in PD and

[7] Other studies of prevalence of depression in PD - Slaughter *et al.* (2001) reported a prevalence of 24.8% DSM-IV Major Depression and 42% all depression, and Veazey *et al.* (2005) reported 3-32% Major Depression and 9-31% Minor Depression.

may be of limited clinical significance). As this study was not designed to assess the prevalence of depression in this setting this, whilst of interest, is not a major concern and does not pose a major issue in the interpretation of the results.

The reasons why the prevalence of depression in the study sample was lower than expected are unknown. It could be that the clinicians working in the local PD clinic are better at identifying and treating depression in their patients than those where other studies have taken place. However, over half of the patients who met diagnostic criteria for depression were not on antidepressant medication, so this may not necessarily be the case. It may be that the manner in which patients were approached and recruited to the study made it more likely that many who were suffering from depressive symptoms chose to decline from taking part. This may have resulted in patients with more severe depressive symptoms being less likely to participate in the study, and under-represented in the sample.

A significant proportion (14%) of non-depressed patients were prescribed antidepressant medication – indicating that depressive symptoms had been identified and effectively treated in at least some patients in this clinic. However, 58% of patients who had depression were not prescribed antidepressant medication. This also means that 42% of patients who met diagnostic criteria for depression were on antidepressant medication but evidently still had significant depressive symptoms to the extent that they still met diagnostic criteria for depression. This suggests that often depressive symptoms are not identified in PD patients, and even if they are, they do not always

receive effective treatment[8]. This is in keeping with the findings of Weintraub *et al.* (2003) who found a significant proportion of patients who were on antidepressant medication but still met criteria for depression – suggesting that such PD patients receive suboptimal treatment or do not respond to treatment.

This study was not designed to be a prevalence study, but a test of the validity and efficacy of the screening measures in this sample population. Whatever the reasons for the low prevalence of depression in the study sample, this study provided a test of the performance characteristics of the screening measures in a sample with a lower prevalence of depression. The prevalence of a disorder does not affect the sensitivity or specificity of a measure but it does impact on the PPV (the proportion of patients with a positive test result who do have the disease) of a screening measure. At the optimal cut-off identified in this sample (with prevalence of 16%) the PPV for the EDS was 64%, for the Whooley questions it was 55% and for the GDS-15 was 60%.

Using the following formula:

PPV = sensitivity x prevalence / sensitivity x prevalence + (1-specificity) x (1-prevalence)

it is possible to calculate the effect that disease prevalence has on the PPV of a measure (Altman, 1999). At a lower prevalence (e.g. 10% prevalence) the PPV for the EDS would be 51%, the Whooley questions 41% and the GDS-15 46%. The use of any of these screening measures in a population with a lower prevalence will lead to a higher

[8] Information was not available regarding for how long patients had been prescribed antidepressant medication – it is possible that some patients had not had sufficient time to respond to treatment (which may take 4-6 weeks on a therapeutic dose) or that patients had only partially responded to treatment and were yet to have their treatment reviewed.

proportion of false-positives and the benefit derived from use of the measure will depend upon the costs involved with false-positive results, and the impact of lower PPV on clinician confidence in, and inclination to use, the scale.

Bearing in mind that the prevalence of depression found in this sample appeared lower than indicated by other studies in similar study settings it would be more likely that screening would take place in a population with higher disease prevalence. This would mean that, in a population with 25% prevalence of depression for example, the PPV of the rating scales would be higher (EDS 76%, Whooley questions 68% and GDS-15 72%) and the proportion of false-positives lower.

4.4 The performance of the screening measures in the study sample

This study has demonstrated that all three screening measures were acceptable and feasible to use in this patient population. A useful way to quantify the relative global diagnostic accuracy of measures is to consider the area under the ROC curve (Zweig and Campbell, 1993). A measure with no discriminatory power will have an AUC of 0.5 (no better than chance or flipping a coin). A measure with perfect discriminatory power would have an AUC of 1.00. Therefore, in reality, a screening test will have an AUC somewhere between 0.5 and 1.0, and the closer to 1.0 the AUC is the better the diagnostic accuracy of the test. The AUC for the screening measures in this study were all between 0.88 and 0.95 for the identification of participants with Minor or Major Depression. (This means that for the EDS, with an AUC of 0.9, a randomly selected participant with depression would have an EDS score higher than a randomly selected participant without depression 90% of the time.)

With AUCs between 0.88 and 0.95 the EDS, Brief EDS, GDS-15, GDS-4 and Whooley questions all showed very good levels of diagnostic accuracy. The EDS performed well but the GDS-15 performed slightly better, and the Whooley questions slightly better than both of these – although the differences between the AUC for the measures were small. Comparison of the AUCs (according to the method of Hanley & McNeill, 1983) showed no significant difference between the measures – meaning they are not significantly different in terms of their diagnostic accuracy. However, one must bear in mind that the AUC is based upon the whole area of the ROC plot and does not take into account the shape of the curve - so judging the performance of the measure also requires consideration of characteristics such as the sensitivity and specificity. That is, if one needed to select one of the scales for screening in a PD outpatient setting, one would choose between them on other criteria, according to the demands of the situation, for example sensitivity, specificity, and perhaps aspects such as the length of the measure or the simplicity of the response options.

Mallett *et al.* (2012) stress that whilst the ROC AUC can be a useful representation of diagnostic accuracy across a range of thresholds, it can also be misleading as clinically relevant thresholds are combined with thresholds that are not clinically significant. Clinically it is more useful to know how a test performs at a particular threshold than a summary of its performance over all possible thresholds – therefore it is more pertinent to consider the performance characteristics at clinically relevant thresholds and the associated true positive and false positive rates.

When considering the sensitivity and specificity (and other performance characteristics such as the PPV and NPV) of the measures, it is essential to identify the optimal

threshold score at which a respondent would be identified as a positive case. The choice of this threshold/cut-off score will depend upon the demands and priorities of the circumstances in which it is to be used. Often the ROC curve is used to indicate the optimal value, where the cut-off score that features closest to the top left corner of the ROC plot represents the maximum combination of sensitivity and specificity. However, using this cut-off value is based on the assumption that the 'costs' of false-positives and false-negatives are equal – but this is not always the case. When screening for a condition it is a priority to maximise the identification of patients who have the condition (true-positives); patients who are wrongly identified by the process and don't have the condition (false-positives) can be ruled out by further investigations. A balance needs to be achieved between the benefits of correctly identifying patients with the condition and the costs of further investigation of those wrongly identified by the screening process, and also the cost of failing to identity patients who have the condition.

In the results section of this thesis the optimal cut-off has been indicated based upon the point that gives the highest combination of sensitivity and specificity. For the purposes of screening it is generally better to select a threshold score that offers the best combination of sensitivity and NPV (Altman, 1999). This optimises the correct identification of patients who have depression and the proportion of patients who screen negative who do not have depression. In clinical practice however, the optimal balance between sensitivity and specificity depends greatly upon the circumstances of the situation in which the screening is taking place and the relative perceived 'costs' of failing to identify a patient with the illness versus incorrectly identifying a patient as being at risk of having the illness when they do not. The PPV is also an important

consideration as it reflects the proportion of patients who screen positive who are correctly identified – a poor PPV may result in the unnecessary further assessment of many patients who are not depressed, which is costly.

4.4.1 The EDS and Brief EDS

The optimal cut-off score for the EDS in terms of maximum sensitivity and specificity, when identifying Minor and Major Depression, was 10/11. At this threshold the scale had a PLR of 9.3, a NLR of 0.29 and DOR of 32.6 indicating that at this threshold the test has good accuracy and is likely to be useful clinically[9]. However a lower cut-off score would give a better combination of sensitivity and NPV. It is not clear from the data from this study which lower value would be best for screening. A threshold of 4/5 would give a sensitivity of 100% (identifies all positive cases) and an NPV of 100% (all those who screen negative are not cases) – but this cut-off score gives poor specificity (52%) and very poor PPV (28%), with the overall misclassification rate of 40%. This means that 40% of patients would be incorrectly classified and there would be a high number of false-positives. A cut-off somewhere between 4/5 and 10/11 would be best, perhaps 8/9 which gives sensitivity of 79% and NPV of 95%, but with a more acceptable specificity of 80% and overall misclassification rate of 20%. (This threshold would give a PLR of 3.99, NLR of 0.26 and DOR of 15.19). The choice of best cut-off would depend upon the situation in which the measure is being used and the costs of further investigation for patients identified upon screening as possibly having depression. If the further investigation would simply involve application of a further quick screening measure or further diagnostic questions from the clinician then a higher

[9] In the NICE guidelines (2009b, p44) a PLR greater than 5 and a NLR less than 0.3 are taken to indicate that a test is relatively accurate, and a DOR that is greater than 20 is taken to indicate that a measure is likely to be useful for clinical practice.

number of false-positives may be acceptable in order to maximise the likelihood of identifying all cases of depression, whilst having confidence that all who screen negative are likely not to be cases. For the Brief EDS the optimal cut-off was identified as 4/5 (sensitivity 89%, specificity 68%, PPV 35% and NPV 97%) whilst for the purposes of screening the optimal cut-off would be 3/4 which would give slightly higher sensitivity (sensitivity 95%, specificity 58%, PPV 30% and NPV 98%). At both these thresholds the PLR is lower than 5.0, and the PPV is likely to be too low to be acceptable in clinical practice.

Both the cut-off scores identified in this study for use of the EDS in a PD population (10/11, and 8/9 for screening) are lower than that reported as optimal by studies of other patient groups (postnatal women and palliative care), which is usually 12/13 and 9/10 for screening. The Brief EDS has not been widely used but Lloyd-Williams *et al.* (2007) used a much higher cut-off of 6/7 in their study in palliative care.

The results of this study show that the optimal cut-off scores for the EDS and the Brief EDS in a PD out-patient sample are lower than those indicated to be optimal in other patient groups. This demonstrates the importance of validation of a measure before it is used in a new clinical population rather than applying the cut-off that has been used different patient groups.

4.4.2 The Whooley questions

The results for the Whooley questions are more straightforward to interpret because of the limited score range. The optimal cut-off of 0/1 (i.e. a positive response to either, or both, questions) delivered the best results and would also be the optimal cut-off for

screening, and is in keeping with other studies of the measure in other patient samples. This threshold gave good sensitivity, NPV and specificity (100, 100 and 84% respectively) and an overall misclassification rate of 13%. The PLR and NLR were also satisfactory. These performance characteristics are stronger than those demonstrated by Whooley *et al.* (1997) in their study carried out in a primary care setting (sensitivity 96%, specificity 57%, AUC 0.82) and by McManus *et al.* (2005) in patients with coronary heart disease (sensitivity 90%, specificity 69%, AUC 0.84).

It is important to note that the sensitivity of this two-item measure is likely to be inflated due to the fact that the two questions are part of (and in fact are essential items in) the criterion standard. This leads to 'incorporation bias' in the results (Williams *et al.*, 2002). However, when the number of positive cases identified and the relatively modest number of non-cases incorrectly identified is considered, this measure still has superior screening performance when applied using the 0/1 threshold.

4.4.3 The GDS-15 and GDS-4

In this study the GDS-15 demonstrated an optimal cut-off at 5/6 for the purpose of dichotomising patients (sensitivity 84%, specificity 89%, PPV 59%, NPV 97%), and a lower score of 3/4 for screening (sensitivity 95%, specificity 71%, PPV 38%, NPV 99%). These cut-off scores gave good sensitivity, specificity and overall misclassification rate (12% and 25% respectively). The PLR, NLR and DOR also indicated that the scale performed well when these thresholds were applied. The optimal cut-off that has been identified in previous studies of the measure in PD patients was 4/5 (with the same score also being optimal for screening) although Weintraub *et al.* (2007) found a higher score of 5/6 best for patients who are over the age of 75 years.

Given the mean age of the study sample here, these results are therefore not too dissimilar (Weintraub *et al.*, 2006b Se 88%, Sp 85%, PPV 61%, NPV 96%, OMR 13%, and Weintraub *et al.*, 2007 Se 90%, Sp 90%, NPV 96%, OMR 10%).

The optimal threshold score for the GDS-4 was 1/2, and a lower score of 0/1 would perhaps be better for the purposes of screening, although this is at the expense of a marked increase in the overall misclassification rate (and poorer PLR and NLR) for only a moderate improvement in sensitivity. Other studies of the GDS-4 (in different patient groups) have indicated an optimal cut-off value of either 1/2 (Marwijk *et al.*, 1995) or 0/1 (Shah *et al.*, 1997) but no studies have validated the performance of the GDS-4 in a PD population before.

4.4.4 Identification of DSM-IV Major and Minor Depression

The identification of patients who have symptoms that meet the criteria for a diagnosis of Major Depression is particularly important both because of the distress and impact their symptoms have upon them, and because these patients tend to respond better to treatment than those with less severe symptoms (Weintraub *et al.*, 2005). When the performance of the screening measures is assessed for the identification of DSM-IV Major Depression the EDS, GDS and Whooley questions performed well with AUC between 0.85-0.93. The optimal cut-offs for identification of Major Depression were the same as those indicated for Minor and Major Depression. In the context of screening for depression one of the most important considerations is the false-negative rate – this is the number of people with depression who are missed and would therefore not receive the treatment that they need. The EDS and GDS-4 missed 3/14 (21%) cases of Major Depression, the GDS-15 2/14 (14%), the Brief EDS 1/14 (7%) and the

Whooley questions none. If the lower screening cut-off scores were used, the results were slightly better in terms of false-negative rate (EDS and GDS-4 2/14, Brief EDS 1/14 and the Whooley questions and the GDS-15 no false-negatives), but this was at the expense of a much higher number of false-positives. The psychometric properties of the scales were not appreciably different when the identification of participants with Minor Depression was excluded from the analyses – that is, the performance characteristics of the scales were very similar for the identification of both Minor and Major Depression and of Major Depression only, and the optimal cut-off scores for the scales remained the same.

4.4.5 Identification of sub-threshold depression

It can be argued that it is also important, when screening for depression, to try to identify individuals with sub-threshold depression. These are the patients whose depressive symptoms are most likely to be missed during routine clinical care (O'Connor *et al.*, 2009) but whom, research indicates, often progress to developing depression that meets diagnostic criteria. The NINDS/NIMH Work group on depression and PD stressed that the identification of non-major depression is beneficial as it accounts for a high proportion of PD patients who have depressive symptoms which often progresses to become Major Depression (Marsh *et al.*, 2006). Schrag *et al.* (2007) point out that it is important to evaluate screening scales that can be used to identify the more diverse depressive disorders (Minor, sub-threshold depression and Dysthymia) in PD, rather than just Major Depression. However, because somatic symptoms of depression are frequently reported by patients with no depression it is very difficult to classify sub-threshold depression in PD in a consistent and reliable manner (Nation *et al.*, 2009).

There were five participants who met the criteria for sub-threshold depression in this study sample. If the cut-off scores identified in this study were used the EDS, Whooley questions, and the GDS-15 identified 2/5, the Brief EDS identified 3/5, and the GDS-4 did not identify any. If the lower 'screening' cut-off scores were used the EDS, Brief EDS, GDS-15 and GDS-4 identified 3/5 cases of sub-threshold depression (although not the same three cases) and the Whooley questions identified 2/5.

Overall, the number of participants in the study sample who met criteria for sub-threshold depression is small, and because of this, and the difficulty in applying criteria for sub-threshold depression in PD, it is hard to draw any meaningful conclusions from these data. A larger sample size would be required to investigate the performance of the screening measures in terms of identifying such patients.

4.4.6 Comparison with other depression screening measures in PD

There have been many studies that have validated various screening measures for depression for use with patients who have PD (see Table 1.7), and Schrag *et al.* (2007) have made recommendations regarding the suitability of scales for the purpose of screening or for rating severity of depression in patients with PD. They judged that observer-rated scales (such as the MADRS) have better psychometric properties but that such scales are less practical for use as routine screening measures because of the demands on clinician time. The BDI, HADS, and the GDS-30 and GDS-15 were felt to be suitable for screening, with the GDS receiving the strongest endorsement as being suitable for screening in routine clinical care. Comparison of the performance characteristics of the screening tools in this study indicates that the EDS, Whooley

questions and the GDS-15 have performed at least as well as scales validated in other studies, in terms of sensitivity, specificity, PPV and NPV.

The recent paper by Williams *et al.* (2012), which evaluated the performance of nine depression scales in the identification of both Minor and Major Depression in a community-based sample of PD patients, has demonstrated similar levels of sensitivity and specificity for all of the scales. The authors concluded that the GDS-30 was the most efficient scale to use for screening in clinic, because it was a self-report scale, had favourable psychometric properties (sensitivity of 72%, specificity of 82%, PPV 73% and NPV 81%) and was quick to complete and administer. The EDS, Whooley questions and GDS-15 have shown slightly better performance in this study sample (except perhaps in terms of PPV), and all three have the advantage of being shorter scales than the GDS-30.

It appears that, until recently, no studies had investigated the utility of an ultra-short scale, such as the 2-item Whooley questions, in patients with PD. Chagas *et al.* (2011) describe a study in which they evaluated the performance of the PHQ-2. This two-item measure, like the Whooley questions, has its origins in the PRIME-MD study (Spitzer *et al.*, 1994) and as such the two measures are very similar. In the PHQ-2 patients are asked about having "little interest or pleasure in doing things" and "feeling down depressed or hopeless" but the time period they are asked to consider is the past two weeks. Patients indicate their response to each item on a scale from 0 (not at all) to 3 (nearly every day) resulting in a score between 0 and 6. Chagas and colleagues found the PHQ-2 to be a valid screening measure in PD (sensitivity 75%, specificity 89%, PPV 70% and NPV 91% at a threshold score of 2/3 out of 6). The data in this study

indicates that the Whooley questions offer superior sensitivity when compared to other measures, as well as being very short and quick to complete, and with a threshold score that is very easy to apply.

4.4.7 Presentation of the screening measures

All participants were given the option of either completing the screening measures themselves, or responding verbally. Nearly all the participants (98%) preferred to complete the scales verbally – usually opting to read the items themselves and indicating their response verbally. The high proportion of participants who chose to do this was most often due to the physical symptoms of PD – in particular, tremors and problems with hand-writing. It is possible that this will have had an impact on the results. Research has shown that subjects are less likely to endorse depression items when asked to respond verbally than when they complete a self-completion questionnaire. This has been observed in both the GDS (O'Neill *et al.*, 1992; Cannon *et al.*, 2002) and the Whooley questions (Mallen & Peat, 2008). This will not have impacted upon the number of participants diagnosed with depression but may have reduced the scores on the screening measures. If the screening measures were to be used in a clinical screening programme then the mode of presentation may be important. In a clinical setting it is possible that the Whooley questions would be presented verbally, as part of a routine review or assessment, but longer measures such as the EDS or the GDS-15 are more likely to be given to the patient to complete themselves. Consideration should also be given to the acceptability of the presentation method of a screening measure in a clinical setting – evidently the participants in this study preferred to complete the scale verbally.

4.5 Implications for clinical practice

The scales in this study have been evaluated regarding their performance in terms of the identification of individuals who are at risk of minor and major depressive disorder so that that they can receive further assessment for depression. Use of a screening scale should not be a substitute for proper diagnostic assessment, but should serve to highlight those whom the clinician should assess further in light of their above-the-threshold-score on the screening measure.

All three measures (as well as the shorter versions of the EDS and GDS) have been shown to have utility for screening for depression in patients who have PD. The AUC and other performance characteristics for the different measures were all good and demonstrate that each could be used as valid and effective screening tools in such a patient group if the identified optimal cut-off scores are applied. The choice of which screening tool to use would depend upon the particular circumstances and demands of the situation in which it is to be used. In most clinical settings, given similar psychometric properties, the measure that is shortest and easiest to use would usually be preferred - because of the length of time it takes to complete and its acceptability to both patient and clinician.

The Whooley questions, as recommended by NICE, had marginally stronger performance characteristics, and have the obvious advantage of being quick and easy to present to the patient, to the extent that they could be easily included in a routine review appointment at an outpatient clinic. In this sample, the Whooley questions identified all depressed cases and wrongly identified only 16/101 (16%) non-depressed cases. This level of false-positives may be acceptable depending upon what action would be taken

as a result of a positive screening result. If it would prompt further brief assessment by the clinician it may be an acceptable cost, but if it would result in referral for specialist assessment then that may not be an acceptable level of false-positives and secondary screening may be beneficial.

It is also possible that a very short screening measure could be easily incorporated into routine review of a patient and be administered at each patient contact. This would increase the chance of detecting emerging or worsening depressive symptoms or depression that was missed at the previous review. For example, most patients attending the out-patient PD clinics from which the study sample were recruited attend the clinic on a six or 12-month basis – screening for depression at each appointment would increase the chances of the identification of depressive symptoms not only at initial assessment but during the patients' on-going clinical management and care. If the screening measure fails to identify a patient with depressive symptoms at one visit, it may at the next. Depression has peaks and troughs in its severity and so repeat screening in this way would hopefully identify patients if their symptoms persist or worsen.

Broader usage of such an easy to use screening measure could also be considered. The Whooley questions are relevant to and, subject to evaluation, could be used with many other patient groups; there would be benefits associated with the use of one depression measure for patients in different settings. Clinicians would become increasingly familiar with the measure and be increasingly willing to use it.

It is important to consider that clinicians are most influenced by the PPV of a screening test. The PPV varies with the prevalence of the disorder being screened for. If the PPV is low clinicians will be reluctant to act on a positive screening result as a high proportion of screen-positive cases will be false-positives. Clinicians need the positive reinforcement of true-positive cases to actually influence their behaviour – if that only happens infrequently the screening is perceived to be less effective (Gilbody *et al.*, 2001). Clinicians need to have confidence in the need to identify and treat depression in order to invest time during their routine care to carry out screening so their perceptions of the utility of a screening measure are very important. In this study the Whooley questions had a PPV of 54% which means that 46% of positive screening results are false-positives.

4.5.1 A two-stage screening process

It is important that a screening tool has optimal sensitivity and NPV, so that (ideally) all possible positive cases of the condition are identified as such and that (ideally) all patients who are identified as negative are indeed non-cases. The 'price' of choosing a cut-off score that optimises these characteristics is that the specificity and PPV may be lower than they might be if another cut-off score is chosen. This will mean that more patients are initially identified as 'positive cases' when in fact they are not (false-positives). If the level of false-positives is not acceptable (and that depends on the action taken with a patient who is identified by the screening measure, and the 'cost' of that action) a two-stage screening process may be worth considering, as recommended in NICE guidelines for depression (2009).

If a two-stage screening process is implemented, a second screening tool is presented to all cases who were identified as positive by the first screening measure. Ideally, the second screening tool will then eliminate the false-positives but again identify the true positives. In such a scenario it may be necessary to apply a different threshold score for the second measure to do this (i.e. one that optimises specificity and PPV). Patients who then score as positive on the second screening measure could perhaps then have further, more detailed assessment by the clinician or referral for specialist assessment for depression. The NICE guidelines for depression (2009) recommend that the two Whooley questions are used as the primary screening measure, and the strong performance of the two questions in this study sample suggest that would indeed be a suitable choice in a PD out-patient population.

The data available from this study enable simulation of this process. In this sample, when a positive answer to either or both of the Whooley screening questions is taken as an indicator of depression all 19 cases of Minor and Major Depression were identified, along with 16 non-depressed participants. The responses of these 35 participants on each of the other screening measures were then examined in order to establish which would be the best secondary screening measure, as well as whether they would then have been correctly identified as positive cases or not.

If the EDS were to be the secondary screening measure, the optimal cut-off score (in terms of maximum specificity and PPV) would be 10/11. This would result in 14/19 depressed participants being correctly identified (sensitivity of 74%) and 13/16 non-depressed being correctly identified (specificity of 81%). From the 120 patients screened for depression the number of false-positives would be reduced from 16/35

(46%) to 3/35 (9%). However, this reduction in false-positives would be at the cost of 'missing' 5/19 (26%) depressed participants who would be wrongly identified as not having depression on the secondary screening measure. Closer examination of the data shows that in this sample the five 'false-negatives' comprised three patients diagnosed with Major Depression and two with Minor Depression, so it was not the case that those who were less severely depressed were the ones who were missed by the second screening measure.

If the GDS-15 was used as the second screening tool, and a threshold of 6/7 was used, this resulted in the correct identification of 15/19 depressed (sensitivity of 79%) and 12/16 non-depressed participants (specificity of 75%). As a result of the application of the second screening measure the number of false-positives was reduced from 46% (16/35) to 11% (4/35) but this was at the cost of 'missing' 4/19 (21%) depressed participants who were incorrectly identified as non-depressed by the GDS-15 (3 patients with Major Depression and 1 with Minor). If a lower threshold of 5/6 was applied then 16/19 depressed participants and 10/16 non-depressed participants were correctly identified (sensitivity 84%, specificity 63%). This would reduce the number of false-positives to 17% (6/35) at the cost of 'missing' 3/19 (16%) of depressed participants (two patients with Major Depression and one with Minor Depression).

Neither the Brief EDS nor the GDS-4 performed as well as second-stage screening measures as their longer counterparts (Brief EDS: Se 53%, Sp 88% and GDS-4: Se 74%, Sp 88%).

4.5.2 The relative benefit of two-stage screening process over using the Whooley questions as a single screening measure.

Sixteen non-depressed participants screened positive on the Whooley questions. However, eight of those 16 participants (who had not been classified as depressed according to the DSM-IV diagnostic criteria) were noted to have significant mental health issues that may have accounted for their responses on the two screening questions. Two were recently bereaved, two had only very recently been diagnosed with PD and were very distressed by it, two had significant anxiety symptoms and two had some depressive symptoms but not to a degree that met diagnostic criteria (i.e. sub-threshold depression). This would suggest that although 16/101 non-depressed participants were wrongly identified by the screening measure, a significant proportion of them may actually benefit from further attention or support regarding their mood – either because of 'low-level' problems with depression or anxiety (which may in time and without intervention develop into more significant problems) or because of issues such as coming to terms with their diagnosis or a recent bereavement.

It may be that the cost of further investigation of all patients identified by the Whooley questions, despite the likelihood that a significant proportion (46%) would not meet criteria for depression, would be worthwhile, in light of the fact that a proportion of those 'non-depressed' false-positives may benefit from further investigation even though they may not be depressed at that time. This would mean that using a second screening measure, which importantly carries the cost of some false-negatives, may not give significant benefits in terms of reducing false-positives because many of those false-positives may benefit from further assessment of their mood anyway. All of the screening measures in this study, when used as a secondary screening measure after the

Whooley questions, would have excluded the majority of those participants who were not depressed but may have benefitted from further assessment.

In this sample, using the Whooley questions followed by either the EDS or GDS-15 as a second-stage screening measure gave better results than when either the EDS or GDS-15 were used as single screening measures (the EDS resulted in 8 false-positives and 5 false-negatives, GDS-15 gave 11 false-positives and 3 false-negatives). When used as a second-stage screening measure both measures reduced the number of false-positives (from 16 to 3 with the EDS or to 4 with the GDS-15) but resulted in several false-negatives (5 with the EDS, 4 with the GDS-15). This demonstrated that, in this sample at least, the two-stage screening process comprising the Whooley questions and then either the EDS or GDS-15 gives better results than using the EDS or the GDS-15 on their own.

The added benefit of a two-stage screening process over using the Whooley questions as a single screening measure would depend upon the action taken following a positive screening result. If the clinician is experienced and confident in the further evaluation of patients for depression then a second longer screening measure may not add significant gains following initial screening – in such a situation the clinician can further assess the mood of any patient who 'screens positive' and rule out any who are not depressed. However, if all patients who are identified as depressed by the Whooley questions are, for example, referred for specialist assessment of their mood the use of a second screening measure to reduce the number of false-positives would save unnecessary referrals and use of resources.

It may therefore be more effective to screen for depression using the Whooley questions and ensure that clinicians have sufficient skills, experience and confidence to assess and treat any patient who screens positive than to utilise a two-stage screening process. Many physicians specialising in geriatric medicine or neurology, working in PD outpatient clinics, would have the skills to treat uncomplicated depression themselves rather than relying of referral to specialist mental health services and investment in ensuring such skills and experience may be a better use of resources than implementation of a two-stage screening process.

4.6 The benefits of a screening programme

A screening programme is pointless if the consequence of a positive screening does not result in benefit for the patient. A screening programme must not only change clinician behaviour in terms of improved recognition of the condition but also result in improved clinical outcome, and it must justify the cost (time and money) of the screening (Gilbody *et al.*, 2001). Evidence suggests that screening and feedback of the results to the clinician alone do not improve depression outcomes even though it improves recognition of depression (O'Connor *et al.*, 2009). Clinicians who receive the screening results need to be familiar with the disorder they are screening for, and either be confident in recognising it or be able to refer patients for specialist assessment.

It is important that any depression screening programme takes place in the context of clear systems for accurate diagnosis, effective treatment and follow-up. In order to achieve positive clinical benefits a screening programme for depression in PD would need:

- An appropriate tool with established validity for use in PD, which has high PPV, that is acceptable to patients,
- To be accompanied by education of the clinicians involved to recognise depression, and refer to mental health services if necessary,
- Availability of effective treatment and resources to ensure adequate follow-up of cases.

It is wrong to screen for an illness if the required systems do not exist to enable clinicians to take appropriate action as a result of a positive screening result (National Screening Committee, 2003). However, it is also unethical for clinicians not to attempt to improve their correct identification of high prevalence disorders (Hickie *et al.*, 2002).

The development of a screening programme for depression would need to take place with due consideration of the "effectiveness, the ethical and clinical implications and the impact upon finite healthcare resources" (Gilbody *et al.*, 2006, p1027). To this end the National Screening Committee (2003) has developed criteria regarding different aspects involved in a screening programme that should be met before embarking upon such a programme. They specify that the screening test should be safe, simple, precise and validated, and a suitable cut-off should be used. The test should also be acceptable to the population with which it is to be used. All the screening measures used in this study meet these criteria and suitable cut-off scores have been indicated in this study. Our impression of the acceptability of the scales to PD patients is limited to those who opted to take part in the research; their acceptability to the wider PD outpatient population is not known.

The National Screening Committee criteria regarding the treatment available for the condition being screened for specify that effective treatment should be available and there should be evidence that treatment leads to better outcomes. NICE guidelines for the treatment of depression (NICE, 2009a, 2009b) recommend psychological intervention and support as the initial 'steps' of care for patients with sub-threshold or mild depression because the evidence of benefit from antidepressant medication for this group is poor. Antidepressant medication is recommended for those whose symptoms persist or who have more severe depressive symptoms.

There is a need for good evidence that pharmacological treatment of depression is effective in PD, and clear evidence of benefit of pharmacological treatment for patients with Minor or sub-threshold depression remains to be established. So far, the limited studies of treatment efficacy in PD have focussed on Major Depression. Weintraub *et al.*, (2005) suggest that patients with sub-threshold depression are less likely to respond to treatment (i.e. they show only small or moderate benefit) and therefore studies designed to assess the efficacy of treatment which include patients with these diagnoses are likely to show less treatment effect. However, the recent Movement Disorder Society Evidence Based Medicine review of treatments for depression in PD (Seppi *et al.*, 2011) incorporated studies which included patients with Minor Depression, Dysthymia and clinically-significant depressive symptoms – and this review concluded that some treatments are efficacious (pramipexole) or likely to be efficacious (some TCA). This, along with the evidence from the more recent RCTs of SSRI and SNRI (e.g. Richard *et al.,* 2012) and of CBT (Dobkin *et al.*, 2011), suggests that good quality evidence of treatment benefit for PD patients with Major and non-major depression is mounting.

Without good evidence that treatment or interventions are effective for Major and non-major depression in PD, the investment of time and resources involved in a screening programme may not be justified. However, it could be argued that it is not ethical not to screen for an illness just because sufficient studies have not been carried out to establish the efficacy of treatment for some of the patient group. In the meantime there is evidence that anti-depressant treatments are effective for physically well patients with non-major depression (NICE, 2009a) and so, until clearer evidence for PD patients is available it may be best for clinicians to extrapolate from non-PD patients or from what evidence there is so far for the treatment of PD patients who have Major Depression.

Gilbody *et al.*, (2006) suggest that many cases of depression identified by screening would have been identified later during routine consultations so the benefit derived from screening patients would be reduced. The National Steering Committee stress that steps should be taken to optimise treatment of patients before a screening programme is introduced – research suggests that many patients are prescribed sub-therapeutic doses of anti-depressants and receive poor follow-up and so optimising existing systems for the treatment and management of those patients whose depression is identified during routine care would be an important step before investment in a screening programme.

It is likely that a proportion of cases identified by a screening test will be patients with sub-threshold depression and whose depression is most likely to be missed during routine clinical care. These are a patient group whose symptoms may resolve without treatment but who would benefit from 'watchful waiting' so that treatment can be considered if their depressive symptoms remain unresolved. However, this patient

group are less likely to accept treatment (because their symptoms are less severe and they are less impaired) and less likely to benefit from treatment (Palmer & Coyne, 2003). It is important to consider the identification of these patients in a screening programme and to bear in mind that their inclusion will affect the results in terms of clinical outcome, but nonetheless identification of such cases is important.

The National Screening Committee identifies several important criteria for a screening programme. These include: there should be RCT-based evidence that the screening programme effectively reduces morbidity; it should be clinically, ethically and socially acceptable to both patients and clinicians; the benefits of screening should outweigh the (physical and psychological) harm, the costs in terms of medical resources should represent value for money.

Whilst this study has established that the screening measures considered are suitable for screening for depression in PD (and represent little inconvenience or 'harm' to the patients), at the current time, it is clear that many of the core criteria of the National Screening Committee cannot be met when considering a screening programme for depression in PD - in particular, the lack of strong evidence of the effectiveness of depression treatments in PD. Also, further high quality research is needed to establish whether screening for depression in PD patients results in more successful identification and treatment of depression, and ultimately in improvement in depressive symptoms and morbidity in general, than if there is no screening programme.

4.7 The strengths of this study

4.7.1 The setting

It is essential to validate a screening measure in the population in which it is to be used. This is in order to establish that the measure is an effective indicator of the condition in that specific patient group. In this context it was important to investigate whether the somatic symptoms of PD affected the efficacy of the depression screening tools, and to establish the appropriate cut-off scores to apply to indicate a participant is at risk of depression. Therefore, this study was carried out in the setting in which a screening programme for patients with PD is most likely to take place – the out-patient clinic. This ensures that the sample is likely to be representative of PD patients who attend out-patient clinics and enables generalisation of the results to other out-patient PD settings. There were few inclusion criteria, thus ensuring that the sample is as representative as possible of patients attending a PD outpatient clinic.

4.7.2 The gold standard measure

An important consideration in a validation study is the choice of the criterion measure. The study was designed with the PSE-SCAN, a standardised semi-structured clinical interview, as the gold standard. This is more consistent than a clinical psychiatric interview, was applied by psychiatrists who have been specially trained the use of the interview and resulted in the diagnosis of depression according to both DSM-IV and ICD10 criteria. Although there is some debate regarding the suitability of DSM-IV and ICD10 diagnostic criteria for depression in patients who have PD, due to the overlap of somatic symptoms of the two illnesses, recent studies have concluded that the criteria

are indeed valid in PD without any modification being necessary (Starkstein *et al.*, 2008, 2011).

It is essential in a study designed to establish the validity of a screening scale to utilise the best possible criterion or gold standard. In the absence of a better alternative DSM-IV diagnostic criteria were applied in this study in keeping with the recommendations made by the NINDS/NIMH working group - that is, an inclusive approach was taken regarding somatic symptoms unless the interviewer was clear that the symptoms were solely attributable to PD. The diagnostic interviews were conducted by psychiatrists who would be familiar with the concepts involved. The diagnosis of depression was reached through application of the algorithms incorporated in the PSE-SCAN software. This means that the diagnostic criteria were more reliably and consistently applied than if made by clinician.

The analysis of the data from this study has focussed on DSM-IV criteria for depression rather than ICD10 criteria. The participants who were identified as meeting DSM-IV criteria for Major Depression were the same individuals who met the ICD10 criteria – despite the criteria being slightly different and therefore potentially identifying slightly different patients (although in this study this was not the case). This thesis has focussed on the DSM-IV criteria because it was also possible to apply diagnostic criteria for Minor Depression (enabling the identification of that less severely affected group of patients), and also because most other diagnostic studies of depression use these criteria (see Table 1.7).

4.7.3 The study design

One strength of the study design was that the gold-standard assessment was always completed first and was therefore always blind to the participant's performance on the screening measures. This ensures there was no chance of 'verification bias' (where the diagnosis of depression is influenced by knowledge of a positive result on the screening test). Also, the completion of the screening scales was carried out later by one researcher who was always blind to the outcome of the gold-standard (PSE-SCAN) assessment. This reduces the possibility of 'interviewer bias' (where the researcher potentially influences the responses of the participant due to their knowledge of the participant's PSE-SCAN result).

A further advantage of the study was that the screening measures being studied were all completed by the same subjects. This eliminates between-subject differences and enables comparisons between the performances of the measures.

4.8 The limitations of this study

4.8.1 Sample size

The sample size in a study of this nature is an important consideration. The sample must be large enough to demonstrate a clinically important difference with an acceptable level of certainty. However, a sample that is too large results in testing more individuals than is necessary to demonstrate the difference being studied and may require too much time and resources to be achievable. In a study designed to validate a screening instrument a sample size that is too small will give an inaccurate and misleading estimate of the sensitivity of the instrument. The sample needs to be large

enough to allow for a sufficient number of positive cases within the sample to demonstrate the test characteristics of the screening measures being examined.

Alonzo *et al.* (2002), in their paper on sample size calculations for studies that compare diagnostic tests, indicate that in order to calculate the necessary sample size the proportion of cases who will test positive on the scales needs to be known. However, for this study, this proportion was unknown as neither the EDS nor the Whooley questions had ever been studied in PD patient population before.

Metz (1978) stated that a larger sample size is needed to demonstrate subtle differences in diagnostic performance than to demonstrate gross differences. However he goes on to state that "common sense and experience suggest that meaningful qualitative conclusions can be drawn from ROC experiments performed with as few as about 100 clinical cases" (p293). Based upon the prevalence of depression in PD reported by Reijnders *et al.* (2008) it was estimated that a sample of 100 patients would include approximately 24 with depression. It was felt that this would be a sufficient number of depressed and non-depressed participants to enable accurate estimates of the performance characteristics to be calculated. However, during the course of the study it became apparent that the prevalence of depression in the study sample was somewhat lower than the 24% reported by Reijnders *et al.* (2008). Therefore the sample size was increased to 120 with the expectation that it would result in an increase in the number of patients with depression in the study sample. With 120 participants this study had a relatively small sample but this is comparable with all other studies in this patient population (see Table 1.7) and it was essential that the sample size was feasible.

Following completion of data collection, consultation with a biostatistician enabled access to up-to-date software that could carry out sample-size calculations based upon methods described in Machin *et al.*, (2009). The sensitivity and specificity (at the optimal cut-off score) of each of the screening measures being studied, and the prevalence of depression found, were now available and were input into the sample-size software. The results indicated that the ideal sample size would have been much larger. According to the software, a study designed to compare the EDS with the Whooley questions would require a sample size of 121, and for comparison of the Whooley questions with the GDS-15 a sample of 203 was indicated, but for comparison of the EDS with the GDS-15 a much larger sample of 1766 would be required (due to the lower sensitivity of each measure). Although the ideal sample size indicated by these calculations show a much larger sample would be needed, this would not have been feasible or achievable. The main purpose of the study was to investigate the validity and efficacy of the screening measures in a PD patient population rather than compare the screening measures against each other, which is what these sample size calculations assume to be the case.

The sample size in this study, although small, was sufficient to enable meaningful estimates of the psychometric properties of the different screening scales in patients who have PD. Comparison of the performance between the scales (in terms of the AUC) may however be at risk of type II error (failure to detect small but clinically significant differences) as the sample size was far from the ideal size indicated for such comparisons.

4.8.2 Delays in recruitment

Recruitment of the target number of participants took five years. This was substantially longer than was planned or anticipated at the start of the project. The long recruitment period was largely due to the fact that recruitment was not continuous - due to lack of availability of doctors who were trained to complete the PSE-SCAN interview. On several occasions doctors left the project and there was a delay before another could be trained and available. Another limiting factor was that the doctors conducting the PSE-SCAN interviews had to make time for the study assessments during their very busy clinical workload. The length of time taken to recruit the required number of participants supports the stipulation that a larger sample size, although desirable, would have not have been feasible.

4.8.3 Was the sample truly representative?

Not all of the patients who were asked about participation in the study by their PD consultant expressed an interest in doing so and declined to have their details passed on to the research team. Some patients who agreed for their details to be passed to the research team later declined to participate in the study. Information about these non-participants is not available for analysis and it is possible that patients who are experiencing depressive symptoms are more likely to decline to participate in research. This is an unavoidable issue in research into depression and not one to which there is an easy solution. If a patient was contacted with a view to participating in the study but declined and gave a reason then this was noted – the reasons given usually related to medical illness of themselves or a family member, or to the inconvenience of taking part in the study.

4.8.4 Generalisability of the results to all patients with PD

This study excluded patients who scored less than 24 on the MMSE, so it is not possible to generalise the results to patients with cognitive impairment in PD - which may be significant proportion (30-40%) of patients with PD (Emre, 2003). Separate validation of the screening measures in a patient group comprising PD patients with memory impairments or dementia would be necessary in order to assess whether the screening measures are useful in such patients. However, if screening is likely to include patients with significant cognitive impairment, use of a dementia-specific measure would be more suitable. Williams and Marsh (2009) validated the Cornell Scale for Depression in Dementia (CSDD) for use in PD patients with and without dementia. They found that the scale was a valid tool for identifying depression in patients with PD across a spectrum of cognitive impairment (sensitivity 75% and specificity 82%, PPV 60%, NPV 90%). The CSDD is a rater-administered measure and so would not be as practical to use as a routine screening measure compared to a self-rated scale, but informant and observation-based rating scales may be the only valid methods of assessing depressive symptoms in patients who have significant cognitive impairment (Williams and Marsh, 2009). (However, some attempt has been made to validate the GDS in patients with dementia, and results indicated that it may be a valid measure of mild to moderate depressive symptoms in patients with mild to moderate dementia (Feher *et al.*, 1992)).

The majority of patients in this study sample had mild to moderate PD. Therefore it is not possible to generalise the results of this study to patients with more advanced PD. However, patients with more advanced PD do not tend to attend outpatient clinics (certainly here in Leicester) and the object of the study was to investigate the utility and

validity of these scales in an outpatient clinic sample. If we wish to screen patients with more advanced PD for depression we would need to validate the scales in such a sample first.

4.8.5 The validity of the criterion standard in PD

There is no doubt that depression is a problem for many patients who have PD but the attribution of some symptoms, which are common to both conditions, is difficult and there is much debate regarding the most effective way to apply diagnostic criteria in such cases. Symptoms of lethargy, sleep disorder and retardation were often reported by participants in this study and in clinical practice it is perhaps only in the context of other non-somatic symptoms of depression that a confident diagnosis of depression can be made.

Having to distinguish between the physiological or psychological cause of a symptom may lead to inconsistency in diagnoses because clinicians have varying thresholds at which they attribute symptoms to depression. This problem is more pronounced in the elderly because somatic symptoms are more common. A clinician's experience of geriatric depression is important (Koenig, 2007) and in this study the raters completing the diagnostic interview were all trained psychiatrists with experience of working with older people.

DSM-IV criteria for depression specify that symptoms should be counted "except when they are clearly and fully accounted for by a general medical condition" (p323) (i.e. if in doubt, count the symptom). The NINDS/NIMH working group also recommended that an inclusive approach should be used when diagnosing depression in people with PD.

The PSE-SCAN interview enables the interviewer to record the attribution of a symptom to a physical condition and as a result the diagnostic algorithm will 'discount' that symptom from contributing towards the diagnosis of depression in that individual. The doctors conducting the diagnostic interview were instructed to record all symptoms of depression and to only attribute the symptom to PD if it was clear that there was no association with depressed mood. This ensured that the diagnosis of depression in this study was made according to current DSM-IV guidelines for the diagnosis of depression. Although one could argue there may be a degree of subjectivity in the attribution of symptoms to PD rather than depression, this is the most suitable measure of depression that could be used as the gold standard in this study.

In order to investigate the potential difference that using this approach to the diagnostic criteria made to the numbers of positive cases the diagnosis was also made using a completely inclusive approach to those symptoms that are subject to overlap with PD. It was found that no participants would have been diagnosed with Major Depression using this approach that had not been diagnosed already, but two more would have been diagnosed with Minor Depression and two with sub-threshold depression.

4.8.6 PSE-SCAN interviewers

The PSE-SCAN interviews were not all carried out by the same rater. Over the course of the study there were nine different raters involved in the completion of the PSE-SCAN interviews. This was not ideal but was unavoidable due to the competing demands on the psychiatrists' time and the length of time it took to complete the study. All of the PSE-SCAN raters were qualified psychiatrists who had completed PSE-SCAN interview training at a World Health Organisation registered training site. All of

the raters had completed an inter-rater reliability check as part of their training, although this was not necessarily based on interviews conducted with patients similar to the study sample. The PSE-SCAN is a standardised instrument that has established reliability and the use of clinicians with an equivalent experience in psychiatry should make results more reliable.

4.8.7 Presentation of the screening measures

Most of the participants chose to respond to the screening instruments verbally and there is limited evidence of how this might affect their responses. A study comparing written versus oral administration of the GDS-30 (Cannon *et al.*, 2002) suggested that patients score significantly higher when it is used as a written rather than orally completed questionnaire, possibly due to a tendency for patients to suppress depressive responses when answering orally. This would suggest that written administration of the screening scales would be preferable and would elicit a more honest and accurate score. Mallen and Peat (2008) also found significantly different results if the Whooley questions were presented verbally rather than in a written format. However, requiring written completion of the scales is perhaps not as acceptable or practical in older, medically ill groups or a PD population (O'Neill *et al.*, 1992) and this is supported by the proportion of participants who chose to respond to the screening questions verbally rather than in written format. Arroll *et al.*, (2003) validated verbal presentation of the Whooley questions and reported sensitivity of 97% and specificity of 67%. They also found that general practitioner diagnosis following the two questions had sensitivity of 77% and specificity of 86%. In a later study Arroll *et al.* found that the addition of a third question "is this something you would like help with" (with the response options No, Yes or Yes, but not today) improved the sensitivity and specificity of doctors'

subsequent diagnosis of depression to 79% and 94% respectively (Arroll *et al.,* 2005) suggesting that the addition of this extra question ruled out many of the false-positives identified by responding positively to one of the two initial questions[10].

The fact that most respondents chose to respond to the screening measures verbally does not necessarily represent a weakness of the study; it means that the performance characteristics of the scales represent their performance when presented in that format and that this must be a consideration when making a choice of screening measure for use in clinical practice or when comparing the performance with other measures.

It would have been possible to vary the order of presentation of the items or the scales to patients. However, it was decided that if a patient tired or refused to complete all of the questions we would not have a complete set of data for any of the three measures. Also, the scales refer to different reference periods (last week or last month) so random presentation order of the items would have been confusing for the patient.

4.8.8 Timing of the screening measures

As participants were visited at their homes it was not possible for the PSE-SCAN interview and the screening measures to be completed on the same day. However, there was a maximum of a five-day interval between completion of the diagnostic interview and the screening questionnaires. It was anticipated that the diagnostic status of the participants would remain stable during that time. All the screening measures ask the

[10] The authors suggest that any patient who answers yes to one or both of the screening questions or answers yes or yes, but not today to the help question should be further assessed for depression by the clinician.

respondent to consider how they have felt over either 'the last week' or 'the last month' so it was judged that this would be sufficient.

4.8.9 Which version of the EDS?

The version of the EDS that was used in this study was same as that used by Lloyd-Williams (2004). This version of the scale varied slightly from the original version by Cox *et al.* (1987) in that one item *" I have felt scared or panicky for no good reason*" was replaced with the item "*I get a sort of frightened feeling as if something awful is about to happen*". It is not known what effect this change has upon the performance of the scale but the modified scale was an effective screening measure for depression in palliative care (Lloyd-Williams *et al.*, 2004) and so this version was used in the present study of PD patients. This also made it possible to examine the performance of the shorter version of the scale, the Brief EDS, which includes the new item from the modified EDS. It is important to remember that it is the modified version of the EDS that has been validated in this PD sample and not the original version.

4.9 Conclusion

Whilst depression rating scales are generally not suitable for diagnostic purposes, and are no alternative to proper clinical assessment, they can be essential tools when screening for depression. In clinical practice the choice of a depression screening scale is very important as it will inevitably have an impact on the identification (or not) of patients who have a depressive disorder (Schrag, 2011).

This study has demonstrated that the EDS, the Whooley questions and the GDS-15 can be used as valid tools for identifying clinically significant (DSM-IV Major and Minor)

depression, and they performed well as screening measures for depression in patients who have PD in an out-patient setting. Optimal cut-off scores for the scales were identified that maximised sensitivity and specificity, and data is provided to enable selection of more suitable cut-off scores for screening, depending upon the particular situation in which the scale is to be used.

Selection of an appropriate depression screening scale for use with patients who have PD is complicated by the overlap of symptoms of depression with those of PD. The EDS is a scale that was designed to have no somatic items and as such has been validated for use in post-natal women and in palliative care. The hypothesis of this study was that due to the lack of items relating to the somatic symptoms of depression, the scale would show good performance as a screening measure in patients who have PD. This has been demonstrated to be the case, although the psychometric properties of the scale were not significantly superior to the other measures in this study, or to others described in previous studies in PD, which also include items relating to somatic symptoms. What appears to be most important is the use of an appropriate cut-off score, identified through validation of the scale in patients with PD.

It is unlikely that any one depression screening scale can have psychometric properties that are suitable for all purposes. Instead, the choice of scale will depend upon the particular demands of the situation in which it is to be used, taking into account factors such as its acceptability, ease of use, length of time to administer as well as performance characteristics like sensitivity, specificity, PPV and NPV.

This study has indicated that all the screening measures were acceptable to the participants (particularly if they are given the option to respond verbally to the items). The EDS has more complicated response options than both the Whooley questions and the GDS-15 (which have Yes/No responses), and as research has suggested that simpler response options are more suitable for older people (Brink *et al.*, 1982) these other scales may be a preferred choice for screening for depression in patients with PD for that reason.

Whilst the EDS and GDS-15 do not take long to administer, scales such as the Whooley questions (two items), and even the Brief EDS (six items) and the GDS-4 (four items) are significantly shorter than most validated self-report scales. Very short scales may be more suitable for routine (and repeated) clinical use because they are quicker to complete, whilst still showing effective screening characteristics. The Whooley questions demonstrated good psychometric properties in this study, whilst the Brief EDS and GDS-4 did not perform quite as strongly as their longer counterparts. The PHQ-2 is an ultra-short screening measure that has already been evaluated in patients with PD (Chagas *et al.*, 2011) with acceptable performance characteristics (sensitivity 75%, specificity 89%, PPV 70%, NPV 91%) but the simpler response options of the Whooley questions and a threshold that is easier to apply are likely to make that measure, recommended in NICE guidelines, an attractive choice for screening PD patients for depression in routine clinical practice.

It is likely that, due to its brevity, the Whooley questions would be the easiest of the scales studied here to incorporate into routine clinical assessments – perhaps with a secondary screening measure to reduce the number of false positives. This was

something that we were able to explore with the data from this study. In this sample the Whooley questions identified all 19 patients with depression plus 16 non-depressed patients - the addition of the EDS or the GDS as a second-stage screening tool reduced the number of false-positives significantly but also resulted in some false-negatives. In a clinical screening programme the correct identification of patients who have the illness is usually the paramount concern so this may not represent a worthwhile benefit – it would depend upon the particular priorities and circumstances of the situation.

The two Whooley questions can be presented verbally by the clinician during the routine assessment or review appointment, adding only a couple of minutes to the consultation and, if a patient responds positively to either of the questions, it should prompt the clinician to question them further regarding their mood. In this study half of the 'non-depressed' participants who screened positive on the Whooley questions had either sub-threshold depression or other mental health issues. These patients would benefit from further investigation or support despite not meeting diagnostic criteria for depression and would clearly be worthwhile identifying in a clinical context. The PPV of the Whooley questions indicates that its use in clinical practice would result in a high number of false-positives and this may affect clinician confidence in the scale. The clinical challenge would be ensuring appropriate further assessment and accurate diagnosis of patients who screen positive, either by the PD specialist themselves or by a mental health professional. Further research is needed however, to establish the utility of the two-item measure in screening for depression in PD patients who are in a more advanced stage of the illness or who have cognitive impairment.

NICE guidelines recommend screening of patients who are at high risk of depression, and although research gives wide-ranging estimates of the prevalence of depressive disorders in PD, it is clear that depression is common in this illness. Research has indicated that depression in PD patients is under-recognised and, if identified, is often sub-optimally treated and reviewed. A screening programme which incorporates a short screening measure into routine clinical assessments should increase recognition of depression. However, before implementation of a screening programme for depression in PD patients attending out-patient clinics using one of the measures studied here it is imperative that further research takes place in order to establish the true benefit (in terms of outcome for the patient) of a depression screening programme.

4.10 Future research

There are now several depression scales that have been validated for use in patients with PD, which have shown good psychometric characteristics. Whilst some further research may be necessary to establish the efficacy of these measures in particular groups (for example, in PD patients in the advanced stage of the illness and those with dementia) the focus of research now should be on the pragmatic evaluation of patient benefit from a depression screening programme. This would need to be coupled with on-going high quality research into the efficacy of both pharmacological and non-pharmacological/psychosocial treatments for depression in PD.

APPENDICES

APPENDIX 1

ICD-10 CRITERIA FOR DIAGNOSIS OF DEPRESSIVE EPISODE.
World Health Organisation (1992)

The individual usually suffers from:

- depressed mood
- loss of interest and enjoyment
- reduced energy leading to increased fatiguability and diminished activity.

Other common symptoms are:

- reduced concentration and attention;
- reduced self-esteem and self-confidence;
- ideas of guilt and unworthiness (even in a mild type of episode);
- bleak and pessimistic views of the future;
- ideas or acts of self-harm or suicide;
- disturbed sleep;
- diminished appetite.

Mild depressive episode:
Involves at least 2 of the 3 typical symptoms plus at least 2 other symptoms. None should be present to an intense degree. An individual with a mild depressive episode is usually distressed by the symptoms and has some difficulty in continuing with ordinary work and social activities, but will probably not cease to function completely. The diagnosis is further categorised to specify the presence of a somatic syndrome (4 or more somatic symptoms. (or 2 or 3 symptoms at a severe level)).

Moderate depressive episode:
Involves at least 2 of the 3 typical symptoms plus at least 3 (preferably 4) other symptoms. Several of the symptoms are likely to be present to a marked degree. An individual with a moderately severe depressive episode will usually have considerable difficulty in continuing with social, work or domestic activities. The diagnosis is further categorised to specify the presence of a somatic syndrome (4 or more somatic symptoms. (or 2 or 3 symptoms at a severe level)).

Severe depressive episode:
Involves all 3 of the 3 typical symptoms plus at least 4 other symptoms, some of which should be of severe intensity. During a severe depressive episode it is very unlikely that the sufferer will be able to continue with social, work, or domestic activities, except to a very limited extent. The somatic syndrome will almost always be present in a severe depressive episode. The diagnosis is further categorised to specify the presence or absence of psychotic symptoms (delusions, hallucinations or depressive stupor).

For all categories: the minimum duration of episode is 2 weeks but if the symptoms are particularly severe and of very rapid onset, it may be justified to make this diagnosis after less than 2 weeks.

APPENDIX 2

THE EDINBURGH (POSTNATAL) DEPRESSION SCALE (Cox *et al.,* 1987)

The Edinburgh Postnatal Depression Scale is not included here due to copyright restrictions. Please refer to the following link for the original validation paper published in the British Journal of Psychiatry:

http://bjp.rcpsych.org/content/150/6/782.full.pdf+html

In this study the item '**I have felt scared or panicky for no very good reason'** was replaced by the item below, as in the research by Lloyd-Williams *et al.* (2004, 2007)

> **I get a sort of frightened feeling as if something awful is about to happen**
>
> Very definitely and quite badly
> Yes but not too badly
> A little, but it doesn't worry me
> Not at all

The following items were selected to comprise the Brief EDS (Lloyd-Williams *et al.,* 2007*).*

- Things have been getting on top of me
- I have been so unhappy that I have had difficulty sleeping
- I have blamed myself unnecessarily when things go wrong
- I have felt sad or miserable
- The thought of harming myself has occurred to me
- I get a sort of frightened feeling as if something awful is about to happen

APPENDIX 3

THE GERIATRIC DEPRESSION SCALE (GDS-15) (Sheikh and Yesavage, 1986)

Please choose the answer that best describes how you have felt over the last week.

Please answer all the following questions by ringing either "Yes" or "No"

1.*	Are you basically satisfied with your life?	Yes/No
2.	Have you dropped many of your activities and interests?	Yes/No
3.*	Do you feel that your life is empty?	Yes/No
4.	Do you often get bored?	Yes/No
5.	Are you in good spirits most of the time?	Yes/No
6.*	Are you afraid that something bad is going to happen to you?	Yes/No
7.*	Do you feel happy most of the time?	Yes/No
8.	Do you often feel helpless?	Yes/No
9.	Do you prefer to stay at home, rather than going out and doing new things?	Yes/No
10.	Do you feel you have more problems with memory than most?	Yes/No
11.	Do you think it is wonderful to be alive now?	Yes/No
12.	Do you feel pretty worthless the way you are now?	Yes/No
13.	Do you feel full of energy?	Yes/No
14.	Do you feel that your situation is hopeless?	Yes/No
15.	Do you think that most people are better off than you are?	Yes/No

* These items contribute to the GDS-4

APPENDIX 4

Copies of the Local Research Ethics Committee and NHS Trust Research & Development Office approval letters

Leicestershire, Northamptonshire & Rutland Research Ethics Committee 1
1 Standard Court
Park Row
Nottingham
NG1 6GN

Telephone: 0115 912 3344 ext 49435
Facsimile: 0115 9123300

05 May 2006

Dr M Dennis
Senior Lecturer and Honorary Consultant Psychiatrist
University of Leicester,
Department of Health Sciences
Psychiatry for the Elderly
Leicester General Hospital
Gwendolin Rd, Leicester
LE5 4PW

Dear Dr Dennis

Full title of study: **The performance of the Edinburgh Depression Rating Scale (EDRS) as a screening tool for depression in people with Parkinson's disease**

REC reference number: **06/Q2501/33**

Thank you for your letter of 25 April 2006, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised.

Ethical review of research sites

The favourable opinion applies to the research sites listed on the attached form.

Conditions of approval

The favourable opinion is given provided that you comply with the conditions set out in the attached document. You are advised to study the conditions carefully.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Application	1	07 February 2006
Investigator CV		01 November 2005
Protocol	7	25 April 2006
Covering Letter		07 February 2006

Participant Information Sheet	2	21 April 2006
Participant Consent Form	2	25 April 2006
Response to Request for Further Information		25 April 2006
Letter informing patients - negative		
Letter informing patients - positive		
Edinburgh Depression Scale	1	02 November 2005
Geriatric Depression Scale	1	02 November 2005
NICE Depression Screening Questions	1	18 November 2005
MMSE	1	20 January 2006
Schedules for Clinical Assessment in Neuropsychiatry	2.1	

Research governance approval

The study should not commence at any NHS site until the local Principal Investigator has obtained final research governance approval from the R&D Department for the relevant NHS care organisation.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

06/Q2501/33	**Please quote this number on all correspondence**

With the Committee's best wishes for the success of this project

Yours sincerely

Dr C Edwards / Ms L Ellis
Chair / Committee Co-ordinator

Email: linda.ellis@rushcliffe-pct.nhs.uk

Enclosures: *Standard approval conditions*
Site approval form

Copy to: R&D Department for NHS care organisation at lead site – Dr D Clarke

Leicestershire, Northamptonshire & Rutland Research Ethics Committee 1

LIST OF SITES WITH A FAVOURABLE ETHICAL OPINION

For all studies requiring site-specific assessment, this form is issued by the main REC to the Chief Investigator and sponsor with the favourable opinion letter and following subsequent notifications from site assessors. For issue 2 onwards, all sites with a favourable opinion are listed, adding the new sites approved.

REC reference number:	06/Q2501/33	**Issue number:**	1	**Date of issue:**	05 May 2006
Chief Investigator:	Dr Michael Dennis				
Full title of study:	The performance of the Edinburgh Depression Rating Scale (EDRS) as a screening tool for depression in people with Parkinson's disease				

This study was given a favourable ethical opinion by Leicestershire, Northamptonshire & Rutland Research Ethics Committee 1 on 05 May 2006. The favourable opinion is extended to each of the sites listed below. The research may commence at each NHS site when management approval from the relevant NHS care organisation has been confirmed.

Principal Investigator	*Post*	*Research site*	*Site assessor*	*Date of favourable opinion for this site*	*Notes* (1)
Dr Michael Dennis	Senior Lecturer and Honorary Consultant Psychiatrist	University Hospitals of Leicester NHS Trust (Leicester General Hospital)	Leicestershire, Northamptonshire & Rutland Research Ethics Committee 1	05/05/2006	

Approved by the Chair on behalf of the REC:

P Wheeler on behalf of ..... (Signature of ~~Chair~~/Administrator)

(delete as applicable)

TRISH WHEATLEY (Name)

(1) *The notes column may be used by the main REC to record the early closure or withdrawal of a site (where notified by the Chief Investigator or sponsor), the suspension of termination of the favourable opinion for an individual site, or any other relevant development. The date should be recorded.*

FINAL APPROVAL (NHS INDEMNITY)

Leicestershire Partnership NHS
NHS Trust

Research & Development Office
Daisy Peake Building
Towers Hospital
Gipsy Lane
Leicester

Tel: 0116-225-3743
David.Clarke@leicspart.nhs.uk

Dr. Mick Dennis
Psychiatry for the Elderly
University of Leicester
Leicester General Hospital
Gwendolen Road
Leicester LE5 4PW

DC/MD/ADMH0409/001

23rd March 2006

Dear Dr. Dennis

Re: The performance of the Edinburgh depression Rating Scale (EDRS) as a screening tool for depression in people with Parkinson's disease? {REC Ref: 06/Q2501/33}

Thank you for applying for Research Governance Approval for the above study. I am pleased to inform you that the formal review of the project is now complete, and has been formally approved to be undertaken at Leicestershire Partnership NHS Trust. Your activity in relation to this project is covered by NHS Indemnity as set out in HSG(96)48 and your study has been entered onto the Trust database.

This study has now been validated and reviewed according to the Standard Operating Procedure for research appraisal. The study therefore has been granted the following level of approval:

Full Approval	☒	Approval in Principle	☐	Approval refused	☐

This approval applies to all documents (as related to research protocol V6) as supplied to the Reviewing ethical committee, and subject to their favourable opinion..

Your responsibilities are set out in the attached agreement, which must be signed and returned to the Research Office. You should keep a copy for your records. All research must be managed in accordance with the requirements of the Dept. of Health Research Governance Framework (RGF), and to ICH-GCP standards. In order to ensure compliance with these standards, the trust may randomly select your study for audit against these standards at any time, and may employ an external agency for this purpose

The duration of this approval extends only to the date specified in your NRES submission, and you should inform the Trust if this is to be extended. Action may be taken to suspend Trust approval if not conducted to these standards, and the study must commence within two years of the REC approval date, and within six months of R&D Approval.

I hope the project goes well, and if you need any help or assistance during its course, please do not hesitate to contact me.

Regards,

Dr. Dave Clarke
[Associate Director of Research & Development]

Copy to: Sarah Baillon (Research Associate)

University Hospitals of Leicester NHS
NHS Trust

DIRECTORATE OF RESEARCH AND DEVELOPMENT

Director: Professor D Rowbotham
Assistant Director: John Hampton
Co-ordinator: J Horsley
Direct Dial: 0116 2588239
Fax No: 0116 2584226
E-mail: jill.horsley@uhl-tr.nhs.uk

Leicester General Hospital
Gwendolen Road
Leicester
LE5 4PW

Tel: 0116 249 0490
Fax: 0116 258 4666
Minicom: 0116 258 8188

19 July 2006

Dr Michael Dennis
Senior Lecturer/Honorary Consultant Psychiatrist
Psychiatry for the Elderly
Department of Health Sciences
The General Hospital
Gwendolen Road, Leicester, LE5 4PW

Dear Dr Dennis

ID: 10057 **The performance of the Edinburgh Depression Rating Scale (EDRS) as a screening tool for depression in people with Parkinson's disease**

LREC Ref: 06/Q2501/39

We have now been notified by the Ethics Committee that this project has been given a favourable opinion by the Ethics Committee (please see the attached letter, dated 5.5.06, from LNR REC 1).

Since all other aspects of your UHL R+D notification are complete, I now have pleasure in confirming full approval of the project on behalf of University Hospitals of Leicester NHS Trust.

This approval means that you are fully authorised to proceed with the project, using all the resources, which you have declared in your notification form.

Please note: for ICH-GCP compliance, please could you forward the signed proforma (if viewed the on-line presentation) or a copy of the certificate (if attended the course) for Dr Nelson Lo, Professor Carol Jagger, Dr Santanu Chakrabarti, Miss Sarah Baillon & Dr Pravin Prabhakavan

The project is also now covered by Trust Indemnity, except for those aspects already covered by external indemnity (e.g. ABPI in the case of most drug studies).

We will be requesting annual and final reports on the progress of this project, both on behalf of the Trust and on behalf of the Ethical Committee.

Please make sure if you or other researchers have an honorary contract with the Trust that this stays within date whilst working on the research study.

If you want to extend the study's end date, 31.5.08, you will have to submit an annual report available through the R&D website which will be forwarded and noted by the Trust and the relevant Ethics Committee. This allows you to continue working on the study under the previous arrangements covered by Trust Indemnity. Please note ethics approval is only granted until the proposed end date as reflected in A3 of the COREC form. You are no longer indemnified pass this date unless you have submitted the annual report form detailing this extension.

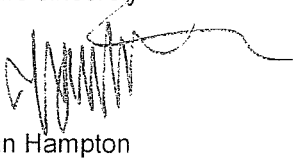
In the meantime, in order to keep our records up to date, could you please notify the Research Office if there are any significant changes to the start or end dates, protocol, funding or costs of the project.

I look forward to the opportunity of reading the published results of your study in due course.

Below is a list of the Researchers Approved to work on this Application within UHL

Miss Sarah Baillon
Dr Santanu Chakrabarti
Dr Micheal Dennis
Dr Carol Jagger
Dr N Lo
Dr Pravin Prabhakaran

Yours sincerely



John Hampton
Assistant Director for Research and Development

APPENDIX 5

Leicestershire Partnership NHS Trust University Hospitals of Leicester NHS Trust

Screening for Depression in Parkinson's disease using the Edinburgh Depression Rating Scale

INFORMATION SHEET FOR PARTICIPANTS

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

What is the purpose of this study?

The study has been set up to identify depression in people who have Parkinson's disease. It is hoped that this will lead to the development of services that will be more sensitive to the needs of this population group.

Why have I been chosen?

We will be asking all patients who attend the out-patient clinic at the Leicester General Hospital for Parkinson's disease to consider taking part in the research. We will be including approximately 120 patients in the study.

Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. You will also be given a copy of this to keep. If you decide to take part you are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect the standard of care you receive.

What will happen to me if I take part?
A member of the research team will telephone you to arrange to visit you at a convenient time at your home. At this visit the researcher will complete an interview with you, this will take approximately 60 minutes and will involve you answering questions about how you are feeling, your mood, and testing your memory. The purpose of the interview is to find out if you are suffering from any mental health difficulties, in particular depression. Within the following few days another researcher will arrange to visit you at home, and at that visit you will be asked to complete three brief questionnaires (also about your mood and how you are feeling), this should only take 10-15 minutes maximum.

If, following the interview and the questionnaires, we feel that you may be suffering from depression we will contact your GP and advise them regarding appropriate treatment.

Once both these visits have taken place the researcher will contact your consultant (Dr Lo or Dr Critchley) for clinical information about your Parkinson's disease and your current medication.

What are the possible benefits of taking part?
We do not anticipate any specific benefits for those who take part in the research. However, if we feel you are suffering from depression we will be able arrange suitable treatment for you. The information we get from this study may help us to be better at detecting when future patients with Parkinson's disease are suffering from depression.

What if I have a complaint about the research?
If you wish to complain, or have any concerns about any aspect of the way you have been approached or treated during the course of this study, the normal National Health Service complaints mechanisms should be available to you.

If something goes wrong
In the event that something does go wrong and you are harmed during the research study there are no special compensation arrangements. If you are harmed and this is due to someone's negligence then you may have grounds for a legal action for compensation against the Leicestershire Partnership NHS Trust but you may have to pay your legal costs. The normal National Health Service complaints mechanism will still be available to you.

Will my taking part in this study be kept confidential?
All information which is collected about you during the course of the research will be kept strictly confidential. Your GP will be notified of your participation in the study.

What will happen to the results of the study?
The results of this study will be reported to appropriate colleagues locally, and will be published in the relevant scientific journals. Findings from the study will be presented in a way that will not identify individual participants in any way.

Who is organising and funding the research?
This research has been organised by the University of Leicester, Leicestershire Partnership NHS Trust and University Hospitals of Leicester NHS Trust.

Thank you for taking the time to read this information sheet.

Contact for further information:
Please keep this sheet for your information. If you have any questions, please do not hesitate to contact the following:

Sarah Baillon, Research Fellow
Psychiatry for the Elderly, Department of Health Sciences
Leicester General Hospital.
Telephone: 0116 258 8161

For independent advice regarding participating in research you can contact the local Patient Advice and Liaison Service (PALS) at Leicester General Hospital – 0116 2588295.

APPENDIX 6

Leicestershire Partnership NHS Trust University Hospitals of Leicester NHS Trust

Screening for Depression in Parkinson's disease using the Edinburgh Depression Rating Scale

CONSENT FORM

Please initial <u>each</u> box

1.	I confirm that I have read and understand the information sheet dated _____ (version __) for the above study and have had the opportunity to ask questions.	☐
2.	I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and without my medical care or legal rights being affected.	☐
3.	I agree to my GP being informed of the findings of the study, if it is felt appropriate by the research team.	☐
4.	I agree to information regarding my Parkinson's disease and current medication, which is recorded in my medical notes, being made available to the researchers.	☐
5.	I agree to take part in the above study.	☐

_____	_____	_____
Name of patient	Date	Signature
_____	_____	_____
Name of researcher	Date	Signature

APPENDIX 7

MODIFIED HOEHN AND YAHR STAGING SCALE FOR PARKINSON'S DISEASE (Hoehn and Yahr, 1967; Jankovic *et al.*, 1990)

Stage 1.0: Unilateral involvement only.

Stage 1.5: Unilateral and axial involvement.

Stage 2.0: Bilateral involvement, without impairment of balance.

Stage 2.5: Mild bilateral disease with recovery on pull test.

Stage 3.0: Mild to moderate bilateral disease; some postural instability; physically independent.

Stage 4.0: Severe disability; still able to walk and stand unassisted.

Stage 5.0: Wheelchair bound or bedridden unless aided.

Taken from Goetz *et al.* (2004).

APPENDIX 8

SUMMARY OF THE SCAN DIAGNOSTIC ALGORITHMS FOR DSM-IV MAJOR DEPRESSION

Numbers in the format N_NNN relate to specific items of the SCAN interview, e.g. 6_033 = Section 6, item 33. (WHO, 1999)
If data is missing the algorithm assumes a score of 0
This information is taken from a detailed algorithm printout - for further information contact the author.

Variable	**algorithm**	***explanation***
MDEPE_A1	[6_001 = 2 or 3 *or* 6_003 = 2 or 3] *and*	*depressed mood or tearfulness are present*
	[6_028 = 0 *and*	*symptoms not due to an organic cause*
	13_038 = 0 or 1]	*symptoms not attributed to somatic cause*
MDEPE_A2	[6_004 = 2 or 3 *and*	*anhedonia is present*
	6_028 = 0 *and*	*symptoms not due to an organic cause*
	13_038 = 0 or 1] *or*	*symptoms not attributed to somatic cause*
	[7_004 = 2 or 3 *and*	*loss of interest is present*
	7_010 = 0 *and*	*symptoms not due to an organic cause*
	13_039 = 0 or 1]	*symptoms not attributed to somatic cause*
MDEPE_A3	8_006 = 2 *or*	*loss of weight present*
	8_007 = 2 or 3 or 4 *or*	*gain of weight present*
	8_005 = 2 or 3 or 5	*change in appetite present*
MDEPE_A4	[8_009 = 1 *or*	*sleep problem is associated with depressed mood*
	8_011 = 1 or 2 *or*	*delayed sleep present*
	8_013 = 1 or 2 *or*	*middle insomnia present*
	8_014 = 1, 2 or 3 *or*	*early waking present*
	8_016 = 1 or 2] *and*	*hypersomnia present*
	8_022 = 0 *and*	*symptoms not due to an organic cause*
	13_040 = 0 or 1	*symptoms not attributed to somatic cause*
MDEPE_A5	[3_006 = 2 or 3 *or*	*restlessness present*
	7_005 = 2 or 3] *or*	*feeling slowed down is present*
	[(7_005 = 3 and 7_010 = 0) *or*	*feeling slowed down not attributed to an organic cause*
	(3_006 = 3 and 3_016 = 0)]	*restlessness not attributed to an organic cause*
MDEPE_A6	[7_006 = 2 or 3 *or*	*loss of energy is present*
	3_007 = 2 or 3] *and*	*fatigue is present*
	7_010 = 0 *and*	*symptoms not due to an organic cause*
	3_016 = 0	*symptoms not due to an organic cause*

MDEPE_A7	[6_013 = 2 or 3 *or* 6_014 = 2 or 3 *or* 6_015 = 2 or 3 *or* 6_017 = 2 or 3] *and* 6_028 = 0 *and* 13_038 = 0 or 1	*self blame is present* *guilt is present* *loss of self-confidence is present* *loss of self-esteem is present* *symptoms not due to an organic cause* *symptoms not attributed to somatic cause*
MDEPE_A8	[7_002 = 2 or 3 *or* 7_003 = 2 or 3] *and* 7_010 = 0 *and* 13_039 = 0 or 1	*loss of concentration is present* *inefficient thinking is present* *symptoms not due to an organic cause* *symptoms not attributed to somatic cause*
MDEPE_A9	[6_010 = 2 or 3 *or* 6_011 = 1,2,3 or 4 *or* 6_012 = 2 or 3] *and* 6_028 = 0 *and* 13_038 = 0 or 1	*preoccupation with death is present* *thoughts of suicide or self-harm are present* *feeling life is not worth living is present* *symptoms not due to an organic cause* *symptoms not attributed to somatic cause*
MDEPE_A	count (MDEPE_A >= 5) *and* (MDEPE_A1 or MDEPE_A2 = 1) *and* (6_005>= 2 or 6_026>= 14)	*there are 5 or more criterion symptoms present* *either dep mood or loss of interest or pleasure is present* *duration of symptoms is over 2 weeks*
MDEPE_B	6_036 != 4 or 5	*symptoms are not due to a Mixed Episode*
MDEPE_C	already coded in MDEPE_A	
MDEPE_D	count (MDEPE_D >= 5)	*the count of criterion symptoms is >=5*
MDEPE_E	13_052!= 3	*symptoms are not due to bereavement*
MDEPE	MDEPE_A *and* MDEPE_B *and* MDEPE_C *and* MDEPE D *and* MDEPE E	*MDEPE is positive if these are all true/positive*
MDEPDSEA	MDEPE *and* 6_033 = 0 or 1	*MDEPE and there have not been two or more depressive episodes*
MDEPDSEB	SCHADIS = 0 *and* SCHIZ = 0 *and* SCHPDIS = 0 *and* DELDIS = 0	*none of these diagnostic criteria are positive (schizoaffective disorder, schizophrenia, schizophreniform disorder, delusional disorder)*
MDEPDSEC	6_035 = 0 or 1 *and* 6_036 = 0 or 1	*no manic episodes and no mixed episodes*
MDEPDSE (Major Depressive Disorder, single episode)	MDEPDSEA *and* MDEPDSEB *and* MDEPDSEC	*MDEPDESE is positive if these are all true/positive*

SUMMARY OF THE SCAN DIAGNOSTIC ALGORITHMS FOR ICD-10 DEPRESSIVE EPISODE

Variable	**algorithm**	***explanation***
F32_0B1	6_001 = 2 or 3 *and*	*low in spirits present*
	6_008 != 1	*loss of reactivity is present*
F32_0B2	[7_004 = 2 or 3 *and*	*loss of interest is present*
	7_010 = 0] *or*	*symptoms not due to an organic cause*
	6_004 = 2 or 3	*anhedonia present*
F32_0B3	[7_006 = 2 or 3 *or*	*loss of energy present*
	7_007 = 2 or 3] *and*	*feeling overwhelmed*
	7_010 = 0	*symptoms not due to an organic cause*
F32_0C1	6_015 = 2 or 3 *or*	*lack of self-confidence present*
	6_017 = 2 or 3	*lack of self-esteem present*
F32_0C2	6_013 = 2 or 3 *or*	*feelings of guilt present*
	6_014 = 2 or 3	*feelings of self blame present*
F32_0C3	6_010 = 2 or 3 *or*	*preoccupation with death present*
	6_011 = 1 or 2 or 3 or 4 *or*	*thoughts of suicide or self-harm present*
	6_012 = 2 or 3	*tiredness with life present*
F32_0C4	[7_002 = 2 or 3 *or*	*loss of concentration present*
	7_003 = 2 or 3] *and*	*inefficient thinking present*
	[7_008B >=14 *or*	*duration of over 14 days*
	7_008D >=2] *and*	*duration of over 2 weeks*
	7_010 = 0	*symptoms not due to an organic cause*
F32_0C5	7_005 = 2 or 3 *and*	*retardation present*
	[7_008B >= 14 *or*	*duration of over 14 days*
	7_008D >= 2] *and*	*duration of over 2 weeks*
	7_010 = 0	*symptoms not due to an organic cause*
F32_0C6	[8_009 = 1 *or*	*symptoms are associated with depressed mood*
	8_011 = 1 or 2 *or*	*delayed sleep present*
	8_012 = 1 or 2 *or*	*pooor quality sleep is present*
	8_013 = 1 or 2 *or*	*middle insomnia is present*
	8_014 = 1 or 2 or 3 *or*	*early morning waking is present*
	8_016 = 1 or 2] *and*	*hypersomnia is present*
	[8_020B >= 14 *or*	*duration of over 14 days*
	8_020D >= 2] *and*	*duration of over 2 weeks*
	8_022 = 0	*symptoms not due to an organic cause*

F32_0C7	[8_005 = 2 or 3 or 5 *and* 8_006 = 1 or 2 or 3] *or* [8_005 = 4 or 5 *and* 8_007 = 1 or 2]	*loss of appetite present* *loss of weight* *increased appetite is present* *weight gain is present*
F32G (Depressive episode)	F32_G1 *and* F32_G2 *and* F32_G3	*F32G is positive if these are all true/positive*
F32_G1	6_005>= 2 *or* 6_026B>= 14 *or* 6_026D>= 2	*duration for over 2 weeks*
F32_G2	6_035 = 0 *or* 1	*no history of hypomania*
F32_G3	6_028 = 0 *and* 13_038 = 0 or 1 *and* 23_004 = 0	*no organic brain or substance use etiology*
F32_0 (Mild depressive episode)	F32_0A *and* F32_0B *and* F32_0C *and* F32_0E	*F32_0 is positive if these are all true/positive*
F32_0A	F32G	*F32_G must be positive*
F32_0B	count of (F32_0B>= 2)	*At least two of the three typical symptoms are present*
F32 _0C	count of (F32_1C >= 4) *and* count (F32_1C <= 5)	*Total number of depressive symptoms >= 4 and <=5*
F32_0E	6_030 = 0 or 1 or 8 or 9	*no previous affective episode*
F32_1 (Moderate depressive illness)	F32_1A *and* F32_1B *and* F32_1C *and* F32 1D	*F32_1 is positive if these are all true/positive*
F32_1A	F32G	*F32G must be positive*
F32_1B	F32_0B	*At least two of the three typical symptoms are present*
F32_1C	[count of (F32_0B=2) *and* count (F32_1C>=6)] *or* [count of (F32_0B=3) *and* count (F32_1C>=6) *and* count (F32_1C<8)	*2 or 3 of the typical symptoms are present and the total number of depressive symptoms is >=6 or is <8*
F32_1D	F32_0E	*no previous affective episode*

F32_2 (severe depressive episode)	F32_2A *and* F32_2B *and* F32_2C *and* F32 2D *and* F32 2E	*F32_2 is positive if these are all true/positive*
F32_2A	F32G	*F32G must be positive*
F32_2B	count (F32_0B>=3)	*all three typical symptoms are present*
F32_2C	count (F32_1C>=8)	*total number of depressive symptoms >= 8*
F32_2D	F32_3D = 0	*no hallucinations or delusions*
F32_2E	F32_0E	*no previous affective episode*

BIBLIOGRAPHY

Aarsland, D., Larsen, J. and Tandberg, E. (1996) Frequency of dementia in Parkinson's disease, *Archives of Neurology,* 53, 538-542.

Aarsland, D., Larsen, J., Karlsen, K., Lim, N.G., and Tandberg, E. (1999) Mental symptoms in Parkinson's disease are important contributors to caregiver distress, *International Journal of Geriatric Psychiatry,* 14, 866-874.

Aarsland, D. and Cummings, J.L. (2002) Depression in Parkinson's disease. *Acta Psychiatrica Scandinavica*, 106, 161-162.

Ackerman, R., and Williams, J. (2002) Rational treatment choices for non-major depressions in primary care. An evidence-based review. *Journal of General Internal Medicine*, 17, 293-301.

Alexopoulos, G., Abrams, R., Young, R. and Shamoian, C.A. (1988a) Cornell scale for depression in dementia. *Biological Psychiatry*, 23, 271-284.

Alexopoulos, G., Abrams, R., Young, R. and Shamoian, C.A. (1988b) Use of the Cornell scale in non-demented patients. *Journal of the American Geriatrics Society*, 36, 230-236.

Alonso, T., Pepe, M. and Moskowitz, C. (2002) Sample size calculations for comparative studies of medical tests for detecting presence of disease. *Statistics in Medicine,* 21, 835-852.

Altman, D. (1999) *Practical Statistics for Medical Research*. Chapman & Hall/CRC, London.

Altman, D., Machin, D., Bryant, T. and Gardner, M. (2000) *Statistics with confidence. Confidence intervals and statistical guidelines. Second Edition.* London: BMJ Books.

American Psychiatric Association (1980) Diagnostic and Statistical Manual of Mental Disorders. Third Edition. American Psychiatric Association, Washington, DC.

American Psychiatric Association (1987) Diagnostic and Statistical Manual of Mental Disorders. Third Edition, Revised. American Psychiatric Association, Washington, DC.

American Psychiatric Association (1994) Diagnostic and Statistical Manual of Mental Disorders. Fourth Edition. American Psychiatric Association, Washington, DC.

Andersen. J., Aabro, E., Gulmann, N., Hjelmsted, A. and Pedersen, H. (1980) Antidepressive treatment in Parkinson's disease: a controlled trial of the effect of nortriptyline in patients with Parkinson's disease treated with L-dopa. *Acta Neurologica Scandinavica*, 62, 210-219.

Andrews, G., Anderson, T., Slade, T. and Sunderland, M. (2008) Classification of anxiety and depressive disorders: problems and solutions. *Depression and Anxiety*, 25(4), 274-281.

Anguenot, A., Loll, P., Neau, J., Ingrand, P., and Gil, R. (2002) Depression et Maladie de Parkinson: étude d'une série de 135 parkinsoniens, *Canadian Journal of Neurological Sciences*, 29(2), 139-146.

Arroll, B., Khin, N. and Kerse, N. (2003) Screening for depression in primary care with two verbally asked questions: cross-sectional study. *British Medical Journal*, 327, 1144-1146.

Arroll, B., Goodyear-Smith, F., Kerse, N., Fishman, T. and Gunn, J. (2005) Effect of the addition of a "help" question to two screening questions on specificity for diagnosis of depression in general practice: diagnostic validity study. *British Medical Journal*, 331, 884-887.

Barone, P. (2011) Treatment of depressive symptoms in Parkinson's disease. *European Journal of Neurology*, 18(Suppl. 1), 11-15.

Barone, P., Poewe, W., Albrecht, S., Debieuvre, C., Massey, D., Rascol, O., Tolosa, E. and Weintraub, D. (2010) Pramipexole for the treatment of depressive symptoms n patients with Parkinson's disease: a randomised, double-blind, placebo-controlled trial. *Lancet Neurology*, 9, 573-580.

Bassett, S.S. (2006) Non-motor somatic symptoms. In *Psychiatric Issues in Parkinson's disease. A Practical Guide*. Menza. M. and Marsh, L. (Eds). Taylor & Francis, London.

Becht, M.C., van Erp, C.F., Teeuwisse, T.M., van Heck, G.L., van Son, M.J. and Pop, V.J. (2001) Measuring depression in women around menopausal age. Towards a validation of the Edinburgh Depression Scale. *Journal of Affective Disorders*, 63, 209-213.

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J. and Erbaugh, J. (1961) An inventory for measuring depression. *Archives of General Psychiatry,* 4(6)*,* 561–571.

Beck, A.T., Guth, D., Steer, R.A. and Ball, R. (1997) Screening for major depression disorders in medical inpatients with the Beck Depression Inventory for primary care. *Behaviour Research and Therapy,* 35, 785-791.

Beck, C.T. and Gable, R.K. (2001) Comparative analysis of the performance of the Postpartum Depression Screening Scale with two other depression instruments. *Nursing Research*, 50(4), 242-250.

Bennett, D.A., Beckett, L.A., Murray, A.M., Shannon, K.M., Goetz, C.G., Pilgrim, D.M. and Evans, D.A. (1996) Prevalence of parkinsonian signs and associated mortality in a community population of older people. *New England Journal of Medicine*, 334, 71-76.

Bernheimer, H., Birkmayer, W., Hornykiewicz, O., Jellinger, K., and Seitelberger, F. (1973) Brain dopamine and the syndromes of Parkinson and Huntington – clinical, morphological and neurochemical correlations. *Journal of Neurological Science*, 20, 415-425.

Boyce, P., Stubbs, J. and Todd, A. (1993) The Edinburgh Postnatal Depression Scale: validation for an Australian sample. *Australian and New Zealand Journal of Psychiatry*, 27, 472-476.

Braam, A.W., Beekman, A.T.F., Dewey, M.E., Delespaul, P.A.E.G., Fichter, M., Lobo, A. Magnusson, H., Peres, K., Reischies, F.M., Roelands, M., Saz, P., Schoevers, R.A., Skoog, I. and Copeland, J.R.M. (2010) Depression and parkinsonism in older Europeans: results from

the EURODEP concerted action. *International Journal of Geriatric Psychiatry*, 25(7), 679-687.

Brink, T., Yesavage, J., Lum, O., Heersema, P., Adey, M. and Rose, T. (1982) Screening test for geriatric depression, *Clinical Gerontologist*, 1(1), 37–43.

Brooks, D.J. and Doder, M. (2001) Depression in Parkinson's disease. *Current Opinions in Neurology*, 14, 465-470.

Brown A.S. and Gershon S. (1993) Dopamine and depression. *Journal of Neural Transmission, General Section*, 91, 75-109.

Brown, R. and Jahanshahi, M., (1995) Depression in Parkinson's disease: a psychosocial viewpoint. *Advances in Neurology*, 65, 61-84.

Burn, D.J. (2002) Beyond the Iron Mask: Towards better recognition and treatment of depression associated with Parkinson's disease. *Movement Disorders,* 17(3), 445-454.

Caap-Ahlgren, M., and Dehlin, O. (2001) Insomnia and depressive symptoms in patients with Parkinson's disease. Relationship to health-related quality of life. An interview study of patients living at home. *Archives of Gerontology and Geriatrics*, 32, 23-33.

Cannon, B., Thaler, T. and Roos, S. (2002) Oral versus written administration of the Geriatric Depression Scale. *Aging and Mental Health*, 6(4), 418-422.

Cavanaugh, S. (1995) Depression in the medically ill. Critical issues in diagnostic assessment. *Psychosomatics*, 36, 48-59.

Chagas, M., Crippa, J., Loureiro, S., Hallak, J., Meneses-Gaya, C., Machado-de-Sousa, J., Rodrigues, G.R., Filho, A.S., Sanches, R.F. and Tumas, V. (2011) Validity of the PHQ-2 for the screening of major depression in Parkinson's disease: Two questions and one important answer. *Aging and Mental Health*. 15(7), 838-43.

Chaudhuri, K.R., Martinez-Martin, P., Schapira, A.H.V., Stocchi. F., Sethi, K., Odin, P., Brown, R.G., Koller, W., Barone, P., MacPhee, G., Kelly, L., Rabey, M., MacMahon, D., Thomas, S., Ondo, W., Rye, D., Forbes, A., Tluk, S., Dhawan, V., Bowron, A., Williams, A.J. and Olanow, C.W. (2006) International multicentre pilot study of the first comprehensive self-completed nonmotor symptoms questionnaire for Parkinson's disease: the NMSQuest Study. *Movement Disorders,* 21(7), 916-923.

Chen, P., Kales, H.C., Weintraub, D., Blow, F.C., Jiang, L., Ignacio, R.V. and Mellow, A.M. (2007a) Depressions in veterans with Parkinson's disease: frequency, co-morbidity, and healthcare utilization. *International Journal of Geriatric Psychiatry,* 22, 543-548.

Chen, P., Kales, H.C., Weintraub, D., Blow, F.C., Jiang, L., Ignacio, R.V. and Mellow, A.M. (2007b) Antidepressant treatment of veterans with Parkinson's disease and depression: analysis of a national sample. *Journal of Geriatric Psychiatry and Neurology,* 20, 161-165.

Chochinov, H., Wilson, K., Enns, M. and Lander, S. (1997) "Are you depressed?" Screening for depression in the terminally ill. *American Journal of Psychiatry*, 154(5), 674-676.

Cooper, J., Copeland, J., Brown, G., Harris, T. and Gourlay, A. (1977) Further studies on interviewer training and inter-rater reliability of the Present State Examination (PSE). *Psychological Medicine*, 7, 517-523.

Cox, J., Holden, J.M., and Sagovsky, R. (1987) Detection of postnatal depression: Development of the 10-item Edinburgh Postnatal Depression Scale. *British Journal of Psychiatry,* 150,782-786

Cox, J., Chapman, G., Murray, D. and Jones, P. (1996) Validation of the Edinburgh postnatal depression scale (EPDS) in non-postnatal women. *Journal of Affective Disorders*. 39, 185-189.

Cubo, E., Bernard, B., Leurgans, S. and Raman, R. (2000) Cognitive and motor function in patients with Parkinson's disease with and without depression. *Clinical Neuropharmacology*, 23(6), 331-334.

Cullum, S., Tucker, S., Todd, C. and Brayne, C. (2006) Screening for depression in older medical inpatients. *International Journal of Geriatric Psychiatry*, 21, 469-476.

Cummings, J.L. (1992) Depression and Parkinson's disease: A review. *American Journal of Psychiatry*, 149(4), 443-454.

D'Ath, P., Katone, P., Mullan, E., Evans, S. and Katona, C. (1994) Screening, detection and management of depression in elderly primary care attenders. I: The acceptability and

performance of the 15-item Geriatric Depression Scale (GDS15) and the development of short versions. *Family Practice*, 11(3), 260-266.

Deeks, J.J. and Altman, D.G. (2004) Diagnostic tests 4: likelihood ratios. *British Medical Journal*, 329, 168-169.

Dent, O., Waite, L., Bennett, H., Casey, B., Grayson, D., Cullen, J., Creasey, H. and Broe, G. (1999) A longitudinal study of chronic disease and depressive symptoms in a community sample of older people. *Aging and Mental Health*, 3(4), 351-357.

Department of Health (2001) *Research Governance Framework for Health and Social Care*. Department of Health, London.

Devos, D., Dujardin, K., Poirot, I., Moreau, C., Cottencin, O., Thomas, P., Destee, A., Bordet, R. and Defebvre, L. (2008) Comparison of desipramine and citalopram treatments for depression in Parkinson's disease: a double-blind, randomised, placebo-controlled study. *Movement Disorders*, 23, 850-857.

Dewey, R.B. (2003) Nonmotor symptoms of Parkinson's disease. In: Pahwa, R, Lyons, K.E, and Koller, W.C. (eds) *Handbook of Parkinson's disease,* 3rd Edition. Marcel Dekker Inc. New York; 109-126.

Dissanayaka, N., Sellbach, A., Matheson, S., Marsh, R., Silburn, P., O'Sullivan, J., Byrne, G. and Mellick, G. (2007) Validity of Hamilton Depression inventory in Parkinson's disease. *Movement Disorders*, 22(3), 399-403.

Dobkin, R., Menza, M., Allen, L., Gara, M., Mark, M., Tiu, J., Bienfait, K. and Friedman, J. (2011) Cognitive-behavioural therapy for depression in Parkinson's disease: A randomised, controlled trial. *American Journal of Psychiatry*, 168, 1066-1074.

Eberhard-Gran, M., Eskild, A., Tambs, K., Opjordsmoen, S. and Samuelsen, S.O. (2001) Review of validation studies of the Edinburgh Postnatal Depression Scale. *Acta Psychiatrica Scandinavica*, 104, 234-349.

Ehmann, T.S., Beninger, R.J., Gawel, M.J. and Riopelle, R.J. (1990) Depressive symptoms in Parkinson's disease: A comparison with disabled control subjects. *Journal of Geriatric Psychiatry and Neurology*, 3, 3-9.

Emre, M. (2003) Dementia associated with Parkinson's disease, *Lancet Neurology*. 2, 229-237.

Endicott, J. (1984) Measurement of depression in patients with cancer. *Cancer*, 53(Suppl), 2243-2248.

Ertan, F., Ertan, T., Kiziltan, G and Uygucgil, H. (2005) Reliability and validity of the Geriatric Depression Scale in depression in Parkinson's disease. *Journal of Neurology, Neurosurgery and Psychiatry*, 76, 1445-1447.

European Medicines Agency (2007) European Public Assessment Report (EPAR) – DATSCAN. European Medicines Agency, London.

Feher, E., Larrabee, G. and Crook, T. (1992) Factors attenuating the validity of the Geriatric Depression Scale in a dementia population, *Journal of the American Geriatric Society*, 40(9), 906-909.

Fénelon, G., Mahieux, F., Huon, R. and Ziégler, M. (2000) Hallucinations in Parkinson's disease. Prevalence, phenomenology, and risk factors. *Brain*. 123, 733-745.

Findley, L., Aujla, M., Baker, M., Beech, C., Bowman, C., Kingdom, W.K., Macmahon, D.G. and Playfer, J.R. (2003) Direct economic impact of Parkinson's disease: A research survey in the United Kingdom. *Movement Disorders*, 18(10), 1139-1189.

Folstein, M., Folstein, S., and McHugh, P. (1975) Mini-mental State – a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research,* 12, 189-198.

Frucht, S.J., (2004) Clinical assessments. In: Pahwa R, Lyons KE, Koller (eds) *Therapy of Parkinson's disease,* 3rd Edition. Marcel Dekker Inc. New York; 37-52.

Galaria, I., Casten, R. and Rovner, B. (2000) Development of a shorter version of the geriatric depression scale for visually impaired older patients. *International Psychogeriatrics*, 12(4), 435-443.

Gaynes, B., Gavin, N., Meltzer-Brody, S., Lohr, K., Swinson, T., Gartlehner, G., Brody, S. and Miller, W. (2005) *Perinatal depression: Prevalence, screening accuracy, and screening*

outcomes. Evidence Report/ Technology Assessment No. 119. AHRQ Publication No. 055-E006-2. Rockville, MD: Agency for Healthcare Research and Quality.

Ghazi-Noori, S., Chung, T., Deane, K., Rickards, H. and Clarke, C. (2003) Therapies for depression in Parkinson's disease. *Cochrane Database of Systematic Reviews.* 2, Art. No.: CD003465. DOI: 10.1002/14651858.CD003465.

Gibb, W. and Lees, A. (1988) The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease, *Journal of Neurology, Neurosurgery and Psychiatry*, 51, 745-752.

Gibson, J., McKenzie-McHarg, K., Shakespeare, J., Price, J. and Gray, R. (2009) A systematic review of studies validating the Edinburgh Postnatal Depression Scale in antepartum and postpartum women. *Acta Psychiatrica Scandinavica*, 119, 350-364.

Gilbody, S., House, A. and Sheldon, T. (2001) Routinely administered questionnaires for depression and anxiety: systematic review. *British Medical Journal*, 322, 406-409.

Gilbody, S., House, A. and Sheldon, T. (2005) Screening and case finding instruments for depression. *Cochrane Database of Systematic Reviews*. Issue 4, Art. No. CD002792. DOI: 10.1002/14651858.CD002792.pub2.

Gilbody, S., Sheldon, T. and Wessely, S. (2006) Should we screen for depression? *British Medical Journal,* 332, 1027-1030.

Gilbody, S., Sheldon, T. and House, A. (2008) Screening and case-finding instruments for depression: a meta-analysis. *Canadian Medical Association Journal,* 178(8), 997-1003.

Global Parkinson's Disease Survey Steering Committee (2002) Factors impacting on quality of life in Parkinson's disease. *Movement Disorders*, 17(1), 60-67.

Goetz, C.G., Poewe, W., Rascol, O., Sampaio, C., Stebbins, G.T., Counsell, C., Giladi, N., Holloway, R.G., Moore, C.G., Wenning, G.K., Yahr, M.D. and Seidl, L. (2004) Movement Disorder Society Task Force Report on the Hoehn and Yahr Staging Scale: Status and recommendations. *Movement Disorders*, 19(11), 1020-1028..

Goldberg, D.P. and Williams, P. (1991) *A User's Guide to the General Health Questionnaire.* Windsor: NFER-Nelson.

Gómez-Esteban, J., Tijero, B., Somme, J., Bilbao, I., Fernández, J., Boyero, S., Velasco, F., Lezcano, E. and Zarranz, J. (2009) Application of depression criteria (DSM-IV) in patients with Parkinson's disease. *Clinical Neurology and Neurosurgery*, 111, 665-669.

Grosset. D. (2006) Diagnosing Parkinson's disease. *Geriatric Medicine.* 36(11), 3-6.

Hamilton, M. (1960) A rating scale for depression. *Journal of Neurology Neurosurgery and Psychiatry*, 23, 56-62.

Hanley, J.A. and McNeil, B.J. (1983) A method of comparing the Areas under the Receiver Operating Characteristic curves derived from the same cases. *Radiology,* 148, 839-843.

Harris, B., Huckle, P., Thomas, R., Johns, S. and Fung, H. (1989) The use of rating scales to identify post-natal depression. *British Journal of Psychiatry*, 154, 813-817.

Henderson, R., Kurlan, R., Kersun, J.M. and Como, P. (1992) Preliminary examination of the comorbidity of anxiety and depression in Parkinson's disease. *Journal of Neuropsychiatry and Clinical Neurosciences*, 4, 257-264.

Herlofsen, K., Lie, S.A., Årsland, D. and Larsen, J.P. (2004) Mortality and Parkinson disease. A community based study. Neurology, 62, 937-942.

Hickie, I.B., Davenport, T.A. and Ricci, C.S. (2002) Screening for depression in general practice and related medical settings. *Medical Journal of Australia*, 177 (7 Suppl), S111-S116.

Hoehn, M.M. and Yahr, M.D. (1967) Parkinsonism: onset, progression and mortality. *Neurology*, 17, 427-442.

Hoogendijk W.J.G., Sommer, I.E.C., Tissingh, G., Deeg, D.J.H. and Wolters, E.C. (1998) Depression in Parkinson's disease. The impact of symptom overlap on prevalence. *Psychosomatics*, 39, 416-421.

Hughes, A.J., Daniel, S.E., Kilford, L., Lees, A.J. (1992a) Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *Journal of Neurology Neurosurgery and Psychiatry*, 55, 181-184.

Hughes, A.J., Ben-Shlomo, Y., Daniel, S.E., Lees, A.J. (1992b) What features improve the accuracy of clinical diagnosis in Parkinson's disease: a clinicopathologic study. *Neurology*, 42(6), 1142-1146.

Hughes, A.J., Daniel, S.E., Ben-Shlomo, Y. and Lees, A.J. (2002) The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. *Brain*, 125, 861-870.

Hughes, T.A., Ross, H.F., Mindham, R.H. and Spokes, E.G. (2004) Mortality in Parkinson's disease and its association with dementia and depression. *Acta Neurologica Scandinavica*, 110, 118-123.

Jankovic, J., McDermott, M., Carter, J., Gauthier, S., Goetz, C., Golbe, L., Huber, S., Koller, W., Olanow, C., Shoulson, I., Stern, M., Tanner, C., Weiner, W. and the Parkinson Study Group (1990) Variable expression of Parkinson's disease: a baseline analysis of the DATATOP cohort. *Neurology*, 40, 1529-1534.

Judd, L.L., Rapaport, M.H., Paulus, M.P. and Brown, J.L. (1994) Subsyndromal symptomatic depression: a new mood disorder? *Journal of Clinical Psychiatry*, 55(S), 18-28.

Karlsen, K.H., Larsen, J.P., Tandberg, E. and Maeland, J. (1999) Influence of clinical and demographic variables on quality of life in patients with Parkinson's disease. *Journal of Neurology Neurosurgery and Psychiatry*, 66, 431-435.

Kathol, R.G., Noyes, R., Williams, J., Mutgi, A., Carroll, B. and Perry, P. (1990) Diagnosing depression in patients with medical illness. *Psychosomatics*, 31, 434-440.

Katon, W.J. (2003) Clinical and health services relationships between major depression, depressive symptoms and general medical illness. *Biological Psychiatry*, 54, 216-226.

Kendell, R., Everett, B., Cooper, J., Sartorius, N. and David, M. (1968) The reliability of the "Present State Examination". *Social Psychiatry*, 3(3), 123-129.

Kendler, K., Gardner, C., Neale, M. and Prescott, C. (2001) Genetic risk factors for major depression in men and women: similar or different heritabilities and same or partly distinct genes? *Psychological Medicine*, 31, 605-616.

Kobak, K. and Reynolds, W. (1999) Hamilton Depression Inventory. In: Maruish, M. (Ed) *The use of psychological testing for treatment planning and outcomes assessment*. 2nd Edition. Lawrence Erlbaum Associates, London.

Koenig, H. (2007) Recognition of depression in medical patients with heart failure. *Psychosomatics*, 48(4), 338-347.

Koenig, H., Goli, V., Shelp, F., Kudler, H., Cohen, H. and Blazer, D. (1992) Major depression in hospitalized medically ill older men: documentation, management and outcome. *International Journal of Geriatric Psychiatry*, 7(1), 25-34.

Koenig, H., Cohen, H., Blazer, D., Krishnan, K. and Sibert, T. (1993) Profile of depressive symptoms in younger and older medical inpatients with major depression. *Journal of the American Geriatrics Society*, 41(11), 1169-1176.

Koenig, H., Pappas, P., Holsinger, T and Bachar, J. (1995) Assessing diagnostic approaches to depression in medically ill older adults: How reliably can mental health professionals make judgements about the cause of symptoms? *Journal of the American Geriatrics Society*, 43(5), 472-478.

Koenig, H., George, L., Peterson, B., and Pieper, C. (1997) Depression in medically ill hospitalised older adults: Prevalence, characteristics and course of symptoms according to six diagnostic schemes. *American Journal of Psychiatry*, 154(10), 1376-1383.

Kroenke, K., Spitzer, R.L. and Williams, J.B. (2003) The Patient Health Questionnaire-2: validity of a two-item depression screener. *Medical Care,* 41(11), 1284-1292.

Kuopio, A.M., Marttila, R.J., Helenius, H., Toivonen, M. and Rinne, U.K. (2000) The quality of life in Parkinson's disease. *Movement Disorders*, 15(2), 216-223.

Leentjens, A., Verhey, F., Luijckx, G-J. and Troost, J. (2000a) The validity of the Beck Depression Inventory as a screening and diagnostic instrument for depression in patients with Parkinson's disease. *Movement Disorders*, 15(6), 1221-1224.

Leentjens, A., Verhey, F., Lousberg, R., Spitsbergen, H. and Wilmink, F. (2000b) The validity of the Hamilton and Montgomery-Asberg depression rating scales as screening and

diagnostic tools for depression in Parkinson's disease. *International Journal of Geriatric Psychiatry*, 15, 644-649.

Leentjens, A., Lousberg, R. and Verhey, F. (2001) The psychometric properties of the Hospital Anxiety and Depression Scale in patients with Parkinson's disease. *Acta Neuropsychiatrica*, 13(4), 83-85.

Leentjens, A., Van den Akker, M., Metsemakers, J., Lousberg, M. and Verhey, F. (2003a) Higher incidence of depression preceding the onset of Parkinson's disease: a register study. *Movement Disorders*, 18(4), 414-418.

Leentjens, A., Marinus, J., van Hilten, J., Lousberg, R. and Verhey, F. (2003b) The contribution of somatic symptoms to the diagnosis of depressive disorder in Parkinson's disease: A discriminant analytic approach, *Journal of Neuropsychiatry and Clinical Neurosciences,* 15, 74-77.

Lemke, M., Fuchs, G., Gemende, I., Herting, B., Oehlwein, C., Reichman, H., Rieke, J. and Volkmann, J. (2004) Depression and Parkinson's disease. *Journal of Neurology,* 251, S6, S24-S27.

Leverton, T.J. and Elliott, S.A. (2000) Is the EPDS a magic wand? a comparison of the Edinburgh Postnatal Depression Scale and health visitor as predictors of diagnosis on the Present State Examination. *Journal of Reproductive and Infant Psychology,* 18(4), 279-296.

Levin, B., Llabre, M. and Weiner, W. (1988) Parkinson's disease and depression: psychometric properties of the Beck Depression Inventory. *Journal of Neurology, Neurosurgery and Psychiatry*, 51, 1401-1404.

Lewinsohn, P.M., Solomon A., Seeley J.R. and Zeiss, A. (2000) Clinical implications of "subthreshold" depressive symptoms. *Journal of Abnormal Psychology,* 109, 345-351.

Liu, C.H., Wang, S.J., Fuh, J.L., Lin, C.H., Yang, Y.Y. and Liu, H.C. (1997) The correlation of depression with functional activity in Parkinson's disease. *Journal of Neurology*, 244,493-498.

Lloyd-Williams, M., Friedman, T. and Rudd, N. (2000) Criterion validation of the Edinburgh postnatal depression scale as a screening tool for depression in patients with advanced metastatic cancer. *Journal of Pain and Symptom Management,* 20, 259-265.

Lloyd-Williams, M., Dennis, M. and Taylor, F. (2004) A prospective study to compare three depression screening tools in patients who are terminally ill. *General Hospital Psychiatry.* 26, 384-389.

Lloyd-Williams, M., Shiels, C. and Dowrick, C. (2007) The development of the Brief Edinburgh Depression Scale (BEDS) to screen for depression in patients with advanced cancer. *Journal of Affective Disorders,* 99, 259-264.

McDonald, W.M., Richard, I.H. and DeLong, M.R. (2003) Prevalence, etiology, and treatment of depression in Parkinson's disease. *Biological Psychiatry*, 54, 363-375.

McDonald, W., Holtzheimer, P., Haber, M., Vitek, J., McWhorter, K. and DeLong, M. (2006) Validity of the 30-item Geriatric Depression rating Scale in patients with Parkinson's disease. *Movement Disorders,* 21(10), 1618-1622.

McManus, D., Pipkin, S. and Whooley, M. (2005) Screening for depression in patients with coronary heart disease. *American Journal of Cardiology*, 96, 1076-1081.

MacMillan, H., Patterson, C. and Wathern, C. (2005) The Canadian Task Force on Preventative Health Care. Screening for depression in primary care: recommendation statement from the Canadian Task Force on Preventative Health Care. *Canadian Medical Association Journal*, 172, 33-35.

Machin, D., Tan, S-B. and Campbell, M.J. (2009) *Sample size tables for Clinical Studies* (3rd Edition), BMJ Books, Hoboken, NJ, USA.

Mahoney, J., Drinka, T., Abler, R., Gunter-Hunt, G., Matthews, C., Gravenstein, S. and Carnes, M. (1994) Screening for depression: Single question versus GDS. *Journal of the American Geriatrics Society*, 42, 1006-1008.

Mallen, C. and Peat, G. (2008) Screening older people with musculoskeletal pain for depressive symptoms in primary care. *British Journal of General Practice*, 58, 668-693.

Mallett, S., Halligan, S., Thompson, M., Collins, G.S. and Altman, D.G. (2012) Interpreting diagnostic accuracy studies for patient care. *British Medical Journal*, 345: e3999.

Marder, K., Levy, G., Louis, E., Mejia-Santant, H., Cote, L., Andrews, H., Harris, J., Waters, C., Ford, B., Frucht, S., Fahn, S. and Ottman, R. (2003) Familial aggregation of early- and late- onset Parkinson's disease. *Annals of Neurology,* 54, 507-513.

Marder, K.S. and Jacobs, D.M. (2008) Dementia. In *Parkinson's disease. Diagnosis and Clinical Management*. Factor, S.A and Weiner, W.J. (eds) Demos Medical Publishing, New York.

Marinus, J., Leentjens, A., Visser, M., Stiggelbout and van Hilten, J. (2002) Evaluation of the Hospital Anxiety and Depression Scale in patients with Parkinson's disease. *Clinical Neuropharmacology,* 25(6), 318-324.

Marsh, L. (2006) Psychosis. In *Psychiatric Issues in Parkinson's disease. A Practical Guide*. Menza. M. and Marsh, L. (Eds). Taylor & Francis, London.

Marsh, L., McDonald, W.M., Cummings, J., Ravina, B. and the NINDS/NIMH Work Group on Depression and Parkinson's disease (2006) Provisional diagnostic criteria for depression in Parkinson's disease: Report of an NINDS/NIMH Work group. *Movement Disorders*, 21(2), 148-158.

Marwijk, H., Wallace, P., Bock, G., Hermans, J., Kaptein, A. and Mulder, J. (1995) Evaluation of the feasibility, reliability and diagnostic value of shortened versions of the geriatric depression scale. *British Journal of General Practice*, 45, 195-199.

Matthey, S., Barnett, B., Kavanagh, D. and Howie, P. (2001) Validation of the Edinburgh Postnatal Depression scale for men, and comparison of item endorsement with their partners. *Journal of Affective Disorders*, 64, 175-184.

Matthey, S., Barnett, B. and White, T. (2003) The Edinburgh Postnatal Depression Scale. Correspondence. *British Journal of Psychiatry*, 182, 368.

Matthey, S., Henshaw, C., Elliott, S. and Barnett, B. (2006) Variability in use of cut-off scores and formats on the Edinburgh Postnatal Depression Scale – implications for clinical and research practice. *Archives of Women's Mental Health*, 9, 309-315.

Meara, J., Bhowmick, B.K. and Hobson, P. (1999) Accuracy of diagnosis in patients with presumed Parkinson's disease. *Age and Ageing*, 28, 99-102.

Menza, M. and Dobkin, R.D. (2006) Non-motor somatic symptoms. In *Psychiatric Issues in Parkinson's disease. A Practical Guide*. Menza. M. and Marsh, L. (Eds). Taylor & Francis, London.

Menza, M., Marin, H. and Dobkin, R.D. (2006) Sleep. In *Psychiatric Issues in Parkinson's disease. A Practical Guide*. Menza. M. and Marsh, L. (Eds). Taylor & Francis, London.

Menza, M., Dobkin, R., Marin, H., Mark, M., Gara, M., Buyske, S., Bienfait, K. and Dicke, A. (2009a) A controlled trial of antidepressants in patients with Parkinson's disease and depression. *Neurology*, 72, 886-892.

Menza, M., Dobkin, R., Marin, H., Mark, M., Gara, M., Buyske, S., Bienfait, K. and Dicke, A. (2009b) The impact of treatment of depression in quality of life, disability and relapse in patients with Parkinson's disease. *Movement Disorders*, 24(9), 1325-1332.

Metz, C.E. (1978) Basic principles of ROC analysis, *Seminars in Nuclear Medicine,*8(4), 283-298.

Mirza, N. and Biglan, K.M. (2006) Non-motor somatic symptoms. In *Psychiatric Issues in Parkinson's disease. A Practical Guide*. Menza. M. and Marsh, L. (Eds). Taylor & Francis, London.

Mitchell, A. J. and Coyne, J.C. (2007) Do ultra-sort screening instruments accurately detect depression in primary care? A pooled analysis and meta-analysis of 22 studies. *British Journal of General Practice*, 57(535), 144-151

Mondolo, F., Jahanshahi, M, Granà, A., Biasutti, E., Cacciatori, E. and Di Benedetto, P. (2006) the validity of the hospital anxiety and depression scale and the geriatric depression scale in Parkinson's disease. *Behavioural Neurology,* 17, 109-115.

Montgomery, S.A. and Asberg, M. (1979) A new depression scale – designed to be sensitive to change. *British Journal of Psychiatry*, 134, 382-389.

Moussavi, S., Chatterji, S., Verdes, E., Tandon, A., Patel, V. and Ustun, B. (2007) Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *Lancet*, 370, 851-858.

Movement Disorder Society (2002) Treatment of depression in idiopathic Parkinson's disease. *Movement Disorders*, 17(Suppl 4), S112-S119.

Muangpaisan, W., Mathews, A., Hori, H. and Seidel, D. (2011) A systematic review of the worldwide prevalence and incidence of Parkinson's disease. *Journal of the Medical Association of Thailand*, 94(6), 749-755.

Murray, L. and Carothers, A. (1990) The validation of the Edinburgh post-natal depression scale on a community sample. *British Journal of Psychiatry*. 157, 288-290.

Murray, L. and Cox, J. (1990) Screening for depression during pregnancy with the Edinburgh Depression Scale (EPDS). *Journal of Reproductive and Infant Psychology*, 8, 99-107.

Naarding P., Leentjens, A., van Kooten, F. and Verhey, F. (2002) Disease-specific properties of the Hamilton Rating Scale for depression in patients with stroke, Alzheimer's dementia and Parkinson's disease. *Journal of Neuropsychiatry and Clinical Neurosciences*, 14(3), 329-334.

Nation, D., Katzen, H., Papapetropoulos, S., Scanlon, B. and Levin, B. (2009) Subthreshold depression in Parkinson's disease. *International Journal of Geriatric Psychiatry*, 24, 937-943.

National Institute for Health and Clinical Excellence (2004) *Depression: Management of depression in Primary and Secondary Care – NICE guidance. Clinical Guideline 23*. National Institute for Health and Clinical Excellence, London.

National Institute for Health and Clinical Excellence (2006) *NICE Clinical Guideline 35. Parkinson's disease: Diagnosis and management in primary and secondary care*. National Institute for Health and Clinical Excellence, London.

National Institute for Health and Clinical Excellence (2009a) *Depression in Adults (update). Depression: the treatment and management of depression in adults. National Clinical Practice Guideline 90*. National Institute for Health and Clinical Excellence, London.

National Institute for Health and Clinical Excellence (2009b) *The Treatment and Management of Depression in Adults with Chronic Physical Health Problems. National Clinical Practice Guideline 91.* National Institute for Health and Clinical Excellence, London.

National Screening Committee (2003) *The UK National Screening Committee criteria for appraising the viability, effectiveness and appropriateness of a screening programme.* HMSO, London. www.nsc.nhs.uk/pdfs/criteria.pdf

Nilsson, F.M., Kessing, L.V. and Bolwig, T.G. (2001) Increased risk of developing Parkinson's disease for patients with major affective disorder – a register study. *Acta Psychiatrica Scandinavica*, 104, 380-386.

Nilsson, F.M., Kessing, L.V., Sorensen, T.M., Andersen, P.K. and Bolwig, T.G. (2002) Major depressive disorder in Parkinson's disease: a register-based study. *Acta Psychiatrica Scandinavica*, 106, 202-211.

Nutt, J.G., Hammerstad, J.P. and Gancher, S.T. (1992) *Parkinson's Disease. 100 Maxims.* Edward Arnold, London.

O'Connor, E., Whitlock, E., Beil, T. and Gaynes, B. (2009) Screening for depression in adult patients in primary care settings: a systematic evidence review. *Annals of Internal Medicine*, 151(11), 793-803.

Office for National Statistics (2009) Population estimates: Selected age groups for local authorities in the United Kingdom; estimated resident population. http://www.statistics.gov.uk/statbase/product.asp?vlnk=151406

Olesen, J., Gustavsson, A., Svensson, M., Wittchen, H.U., and Jönsson, B. on behalf of the CDBE2010 study group and the European Brain Council (2012) The economic cost of brain disorders in Europe, *European Journal of Neurology*, 19, 155-162.

O'Neill, D., Rice, I., Blake, P., Walsh, J. and Coakley, D. (1992) The Geriatric Depression Scale: rater-administered or self-administered? *International Journal of Geriatric Psychiatry*, 7, 511-515.

Palmer, S. and Coyne, J. (2003) Screening for depression in medical care. Pitfalls, alternatives and revised priorities. *Journal of Psychosomatic Research*, 54, 279-287.

Parker, G., Hilton, T., Hadzi-Pavlovic, D. and Bains, J. (2001) Screening for depression in the medically ill: the suggested utility of a cognitive-based approach. *Australian and New Zealand Journal of Psychiatry*, 35, 474-480.

Parkinson's Disease Society (PDS) (2005) Parkinson's and you. An introduction to Parkinson's disease.
http://www.parkinsons.org.uk/advice/publications/booklets/living-with-parkinsons.aspx

Pignone, M., Gaynes, B., Rushston, J., Burchell, C., Orleans, C., Mulrow, C. and Lohr, K. (2002) Screening for depression in Adults: A summary of the evidence for the U.S. Preventative Services Task Force. *Annals of Internal Medicine*, 136(10), 765-776.

Porter, B., Macfarlane, R., Unwin, N. and Walker, R. (2006) The Prevalence of Parkinson's Disease in an Area of North Tyneside in the North-East of England, *Neuroepidemiology*, 26, 156–161.

Radloff, L.S. (1977) The CES-D scale: a self-report depression scale for research in the general population. Applied Psychological Measurement, 1, 385-401.

Reiff, J., Schmidt, N., Riebe, B., Breternitz, R., Aldenhoff, J., Deuschl,G. and Witt, K. (2011) Subthreshold depression in Parkinson's disease. *Movement Disorders*, 26(9), 1741-1744.

Reijnders, J., Ehrt, U., Weber, W., Aasland, D. and Leentjens, A. (2008) A systematic review of prevalence studies of depression in Parkinson's disease. *Movement Disorders*, 23(2), 183-189.

Richard, I.H., Kurlan, R., Tanner, C., Factor, S., Hubble, J., Suchowershy, O. and Waters, C. (1997) Serotonin syndrome and the combined use of deprenyl and an anti-depressant in Parkinson's disease. *Neurology*, 48, 1070-1077.

Richard, I.H., Justus, A.W. and Kurlan, R. (2001) Relationship between mood and motor fluctuations in Parkinson's disease. *Journal of Neuropsychiatry and Clinical Neuroscience*, 13, 35-41.

Richard, I.H., McDermott, M.P., Kurlan, R., Lyness, J.M., Como, P.G., Pearson, N., Factor, S.A., Juncos, J., Serrano Ramos, C., Brodsky, M., Manning, C., Marsh, L., Shulman, L., Fernandez, H.H., Black, K.J., Panisset, M., Christine, C.W., Jiang, W., Singer, C., Horn, S., Pfeiffer, R., Rottenberg, D., Slevin, J., Elmer, L., Press, D., Hyson, H.C., McDonald, W. (2012) A randomised, double-blind, placebo-controlled trial of antidepressants in Parkinson's disease. *Neurology*, 78(16), 1229-1236.

Richards, M., Stern, Y. and Mayeux, R. (1993) Subtle extrapyramidal signs can predict the development of dementia in elderly individuals. *Neurology*, 43, 2184-2188.

de Rijk, M., Rocca, W., Anderson, D., Melcon, M., Breteler, M. and Maraganore, D. (1997) A population perspective on diagnostic criteria for Parkinson's disease. *Neurology*, 48(5), 1277-1281.

Rijnders, C., van den Berg, J., Hodiamont, P., Nienhuis, F., Furer, J., Mulder, J. and Giel, R. (2000) Psychometric properties of the schedules for clinical assessment in neuropsychiatry (SCAN-2.1). *Social Psychiatry and Psychiatric Epidemiology*, 35, 348-352.

Rodriguez-Blazquez, C., Frades-Payo, B., Forjaz, M., Pedro-Cuesta, J., and Martinez-Martin, P. (2009) Psychometric attributes of the Hospital Anxiety and Depression Scale in Parkinson's disease, *Movement Disorders,* 24(4), 519-525.

Schneider,C.B., Pilhatsch, M., Rifati, M., Wodarz, F., Djundja, D., Odin, P., Wolz, M., Bauer, M., Reichman, H. and Storch, A. (2010) Utility of the WHO-Five Well-being Index as a screening tool for depression in Parkinson's disease. *Movement Disorders,* 25(6), 777-783.

Schrag, A. (2011) Apathy and depression scales in Parkinson's disease: Are they good enough? *Journal of the Neurological Sciences*, 310, 216-219.

Schrag, A., Jahanshahi, M. and Quinn, N. (2000) What contributes to quality of life in Parkinson's disease? *Journal of Neurology, Neurosurgery and Psychiatry*, 69, 308-312.

Schrag, A., Jahanshahi, M. and Quinn, N. (2001) What contributes to depression in Parkinson's disease? *Psychological Medicine,* 31, 65-73.

Schrag, A., Ben-Shlomo, B. and Quinn, N. (2002) How valid is the clinical diagnosis of Parkinson's disease in the community? *Journal of Neurology, Neurosurgery and Psychiatry*, 73, 529-534.

Schrag, A., Barone, P., Brown, R., Leentjens, A., McDonald, W., Starkstein, S., Weintraub, D., Poewe, W., Rascol, O., Sampaio, C., Stebbins, G. and Goetz, C. (2007) Depression rating scales in Parkinson's disease: critique and recommendations. *Movement Disorders*, 22(8), 1077-1092.

Seppi, K., Weintraub, D., Coelho, M., Perez-Lloret, S., Fox, S.H., Katzenschlarger, R., Hametner, E-M., Poewe, W., Rascol, O., Goetz, C.G. and Sampaio, C. (2011) The Movement Disorder Society Evidence-Based Medicine Review Update: Treatments for the non-motor symptoms of Parkinson's disease. Movement Disorders, 26(S3), S42-S80.

Shah, A., Herbert, R., Lewis, S., Mahendran, R., Platt, J. and Bhattacharyya, B. (1997) Screening for depression among acutely ill geriatric inpatients with a short geriatric depression scale. *Age and Ageing*, 26, 217-221.

Sheikh, J. and Yesavage, J. (1986) Geriatric Depression Scale: recent evidence and development of a shorter version, *Clinical Gerontologist*, 5, 165-173.

Shulman, L.M., Taback, R.L., Bean, J. and Weiner, W.J. (2001) Comorbidity of the non-motor symptoms of Parkinson's disease. *Movement Disorders*, 16(3), 507-510.

Shulman, L.M., Taback, R.L., Rabinstein, A.A. and Weiner, W.J. (2002) Non-recognition of depression and other non-motor symptoms in Parkinson's disease. *Parkinsonism and Related Disorders*, 8, 193-197.

Silberman, C., Laks, J., Capitão, C., Rodrigues, C., Moreira, I. and Engelhardt, E. (2006) Recognising depression in Parkinson's disease. Accuracy and specificity of two depression rating scales. *Arq Neuropsiquiatr*, 64(2-B), 407-411.

Slaughter, J.R., Slaughter, K.A., Nichols, D., Holmes, S.E. and Martens, M.P. (2001) Prevalence, clinical manifestations, etiology and treatment of depression in Parkinson's disease. *Journal of Neuropsychiatry and Clinical Neuroscience,* 13(2), 187-196.

Snaith, R., Constantopoulos, A., Jardine, M. and McGuffin, P. (1978) A clinical scale for the self-assessment of irritability. *British Journal of Psychiatry*, 132, 164-171.

Spitzer, R.L., Williams, J.B., Kroenke, K., Linzer, M., deGruy, F., Hahn, S., Brody, D. and Johnson, J.G. (1994) Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study. *Journal of the American Medical Association*. 272(22), 1749-1756.

Spitzer, R.L., Kroenke, K., and Williams, J.B., (1999) Validation and utility of a self-report version of the PRIME-MD: the PHQ primary care study. *Journal of the American Medical Association*, 282, 1737–1744.

Starkstein, S.E., Preziosi, T.J., Bolduc, P.L. and Robinson, R.G. (1990a) Depression in Parkinson's Disease. *Journal of Nervous and Mental Disease.* 178(1), 27-31.

Starkstein, S.E., Bolduc, P.L., Mayberg, H.S., Preziosi, T.J. and Robinson, R.G. (1990b) Cognitive impairments and depression in Parkinson's disease. *Journal of Neurology, Neurosurgery and Psychiatry.* 53, 597-602.

Starkstein, S.E., Mayberg, H.S, Leiguarda, R., Preziosi, T.J. and Robinson, R.G. (1992) A prospective longitudinal study of depression, cognitive decline, and physical impairments in patients with Parkinson's disease. *Journal of Neurology, Neurosurgery and Psychiatry*, 55, 377-382.

Starkstein, S.E., Petracca, G., Chemerinski, E., Teson, A., Sabe, L., Merello, M and Leiguarda, R. (1998) Depression in classic versus akinetic-rigid Parkinson's disease. *Movement Disorders*, 13(1), 29-33.

Starkstein, S.E., Merello, M., Jorge, R., Brockman, S., Bruce, D., Petracca, G. and Robinson, R.G. (2008) A validation study of depressive syndromes in Parkinson's disease. *Movement Disorders*, 23(4), 538-546.

Starkstein, S.E., Dragovic, M., Jorge, R., Brockman, S., Merello, M., Robinson, R.G., Bruce, D. and Wilson, M. (2011) Diagnostic criteria for depression in Parkinson's disease: A study of symptom patterns using latent class analysis. *Movement Disorders*, 26(12), 2239-2245.

Steer, R., Cavalieri, T., Leonard, D. and Beck, A. (1999) Use of the Beck Depression Inventory for primary care to screen for Major Depression Disorders. *General Hospital Psychiatry*, 21, 106-111.

Stein, M.B., Heuser, I.J., Juncos, J.L. and Uhde, T.W. (1990) Anxiety disorders in patients with Parkinson's disease. *American Journal of Psychiatry*, 147, 217-220.

Tandberg, E., Larsen, J.P., Aarsland, D. and Cummings, J.L. (1996) Risk factors for depression in Parkinson Disease. *Archives of Neurology,* 54(4), 625-630.

Tandberg, E., Larsen, J.P., Aarsland, D., Laake, K. and Cummings, J.L. (1997) The occurrence of depression in Parkinson Disease: A community-based study. *Archives of Neurology,* 53(2), 175-179.

Thanvi, B., Lo, N. and Robinson, T. (2005) Vascular parkinsonism – an important cause of parkinsonism in older people. *Age and Ageing*, 34, 114-119.

Thanvi, B. and Treadwell, S. (2009) Drug induced parkinsonism: a common cause of parkinsonism in older people. *Postgraduate Medical Journal*, 85, 322-326.

Thomas, C. and Morris, S. (2003) Cost of depression among adults in England in 2000. *British Journal of Psychiatry*, 183, 514-519.

Thompson, A.W., Liu, H., Hays, R.D., Katon, W.J., Rausch, R., Diaz, N., Jacob, E.L., Vassar, S.D. and Vickery, B.G. (2011) Diagnostic accuracy and agreement across three depression assessment measures for Parkinson's disease. *Parkinsonism and Related Disorders*, 17, 40-45.

Tröster, A. and Kaufer, D. (2006) Non-motor somatic symptoms. In *Psychiatric Issues in Parkinson's disease. A Practical Guide*. Menza. M. and Marsh, L. (Eds). Taylor & Francis, London.

Tumas, V., Rodrigues, G., Farias, T. and Crippa, J. (2008) The accuracy of diagnosis of major depression in patients with Parkinson's disease. A comparative study among the UPDRS, the Geriatric Depression Scale and the Beck Depression Inventory. *Arq Neuropsiquiatr*, 66(2-A), 152-156.

U.S. Preventative Services Task Force (2002) Screening for depression: Recommendations and Rationale. *Annals of Internal Medicine*, 136 (10), 760-764.

Valenstein, M., Vijan, S., Zeber, J., Boehm, K. and Buttar, A. (2001) The cost-utility of screening for depression in primary care. *Annals of Internal Medicine,* 134(5), 345-360.

Van der Hoek, T.C., Boudewijn, A.A., Matui, P., van der Marck, M.A., Esselink, R.A. and Tendolkar, I. (2011) Prevalence of depression in Parkinson's disease: Effects of disease stage, motor subtype and gender. *Journal of the Neurological Sciences*, 310(1-2), 220-224.

Veazey, C., Aki, S.O.E., Cook, K.F., Lai, E.C. and Kunik, M.E. (2005) Prevalence and treatment of depression in Parkinson's disease. *Journal of Neuropsychiatry and Clinical Neuroscience*, 17(3), 310-323.

Visser, M., Leentjens, A., Marinus, J., Stiggelbout, A. and van Hilten, J. (2006) Reliability and validity of the Beck Depression Inventory in Patients with Parkinson's disease. *Movement Disorders*, 21(5), 668-672.

Voon, V. and Fox, S.H. (2008) Impulse-control disorders. In *Parkinson's disease. Diagnosis and Clinical Management*. Factor, S.A and Weiner, W.J. (eds) Demos Medical Publishing, New York.

Voon, V., Hassan, K., Zurowski, M., de Souza, M., Thomsen, T., Fox, S., Lang, A.E. and Miyasaki, J. (2006) Prevalence of repetitive and reward –seeking behaviours in Parkinson's disease. *Neurology,* 67, 1254-1257.

Wancata, J., Alexandrowicz, R., Marquart, B., Weiss, M. and Friedrich, F. (2006) The criterion validity of the Geriatric Depression Scale: a systematic review. *Acta Psychiatrica Scandinavica, 114, 398-410.*

Waraich, P., Goldner, E.M., Somers, J.M. and Hsu, L. (2004) Prevalence and incidence studies of mood disorders: A systematic review of the literature. *Canadian Journal of Psychiatry*, 49(2), 124-138.

Watson, L., Zimmerman, S., Cohen, L. and Dominik, R. (2009) Practical depression screening in residential care/assisted living: Five methods compared with gold standard diagnoses. *American Journal of Geriatric Psychiatry*, 17(7), 556-564.

Weintraub, D. (2004) Diagnosing and treating depression in patients with Parkinson's disease. *Psychiatry Annals*, 34(4), 299-304.

Weintraub, D., Moberg, P.J., Duda, J.E., Katz, I.R. and Stern, M.B. (2003) Recognition and treatment of depression in Parkinson's disease. *Journal of Geriatric Psychiatry and Neurology,* 16, 178-183.

Weintraub, D., Moberg, P.J., Duda, J.E., Katz, I.R. and Stern, M.B. (2004) Effect of psychiatric and other nonmotor symptoms on disability in Parkinson's disease. *Journal of the American Geriatrics Society*, 52(5), 784-788.

Weintraub, D., Morales, K., Moberg, P., Bilker, W., Balderston, C., Duda, J., Katz, I. and Stern, M. (2005) Antidepressant studies in Parkinson's disease: a review and meta-analysis. *Movement Disorders*, 20(9), 1161-1169.

Weintraub, D., Holtzheimer, P.E. and McDonald, W.M. (2006a) Depression. In *Psychiatric Issues in Parkinson's disease. A Practical Guide*. Menza. M. and Marsh, L. (Eds). Taylor & Francis, London.

Weintraub, D., Oehlberg, K., Katz, I. and Stern, M. (2006b) Test characteristics of the 15-item Geriatric Depression Scale and the Hamilton Depression Rating scale in Parkinson's disease. *American Journal of Geriatric Psychiatry*, 14(2), 169-175.

Weintraub, D., Saboe, K. and Stern, M. (2007) Effect of age on Geriatric depression Scale performance in Parkinson's disease. *Movement Disorders*, 22(9), 1331-1335.

Wells, K., Sherbourne, C., Schoenbaum, M., Duan, N., Meredith, L., Unutzer, J., Miranda, J., Carney, M.F. and Rubenstein, L.V. (2000) Impact of disseminating quality improvement programmes for depression in managed primary care: a randomised controlled trial. *Journal of the American Medical Association*, 283, 212-220.

Whetten-Goldstein, K., Sloan, F., Kulas, E., Cutson, T. and Schenkman, M. (1997) The burden of Parkinson's disease on society, family, and the individual. *Journal of the American Geriatrics Society*, 45(7), 844-849.

Whooley, M., Avins, A. Miranda, J. and Browner, W. (1997) Case-finding instruments for depression. Two questions as good as many. *Journal of General Internal Medicine,* 12, 439-445.

Wilhelm, K., Kotze, B., Waterhouse, M., Hadzi-Pavlovic, D. and Parker, G. (2004) Screening for depression in the medically ill: A comparison of self-report measures, clinician judgement, and DSM-IV diagnoses. *Psychosomatics*, 46(6), 461-469.

Williams, J., Barrett, J, Oxman, T., Ellen, F., Katon, W., Sullivan, M., Cornell, J. and Sengupta, A. (2000) Treatment of dysthymia and minor depression in primary care: a randomised controlled trial in older adults. *Journal of the American Medical Association,* 284(12), 1519-1526.

Williams, J., Pignone, M., Ramirez, G. and Perez Stellato, C. (2002) Identifying depression in primary care: a literature synthesis of case-finding instruments, *General Hospital Psychiatry*, 24(4), 225-237.

Williams, J. and Marsh, L. (2009) Validity of the Cornell scale for depression in dementia in Parkinson's disease with and without cognitive impairment. *Movement Disorders*, 24(3), 433-437.

Williams, J.R., Hirsch, E.S., Anderson, K., Bush, A.L., Goldstein, S.R., Grill, S., Lehmann, S., Little, J.T., Margolis, R.L., Palanci, J., Pontone, G., Weiss, H., Rabins, P. and Marsh, L. (2012) A comparison of nine scales to detect depression in Parkinson's disease. *Neurology,* 78(13), 998-1006.

Wing, J. (1996) SCAN and the PSE tradition. *Social Psychiatry and Psychiatric Epidemiology*, 31, 50-54.

Wing, J., Birley, J., Cooper, J., Graham, P. and Isaacs, A. (1967) Reliability of a procedure for measuring and classifying "present psychiatric state". *British Journal of Psychiatry*, 113, 499-515.

Wing, J., Cooper, J. and Sartorius, N. (1974) *The description and manual for the PSE and CATEGO system.* Cambridge University Press: London.

Wing, J., Henderson, A. and Winckle, M. (1977a) The rating of symptoms by a psychiatrist and a non-psychiatrist: a study of patients referred from general practice. *Psychological Medicine*, 7, 713-715.

Wing, J., Nixon, J., Mann, S. and Leff, J. (1977b) Reliability of the PSE (ninth edition) used in a population study. *Psychological Medicine*, 7, 505-516.

Wing, J., Mann, S., Leff, J. and Nixon, J. (1978) The concept of a 'case' in psychiatric population surveys. *Psychological Medicine*, 8(2), 203-17.

Wing, J., Babor, T., Brugha, T., Burke, J., Cooper, J., Giel, R., Jablenski, A., Regier, D. and Sartorius, N. (1990) SCAN. Schedules for Clinical Assessment in Neuropsychiatry. *Archives of General Psychiatry*, 47, 589-593.

Wittchen, H., Höfler, M. and Meister, W. (2001) Prevalence and recognition of depressive syndromes in German primary care settings: poorly recognised and treated? *International Clinical Psychopharmacology*, 16, 121-135.

World Health Organisation (1992) *International Classification of Diseases*. 10th Edition. WHO, Geneva.

World Health Organisation (1999) Schedules for Clinical Assessment in Neuropsychiatry Interview, Version 2.1. WHO, Geneva.

Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M. and Leirer, V.O. (1983) Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research,* 17, 37-49.

Zesiewicz, T., Gold, M., Chari, G. and Hauser, R. (1999) Current Issues in depression in Parkinson's disease. *American Journal of Geriatric Psychiatry*, 7(2), 110-118.

Zesiewicz, T., Baker, M., Wahba, N. and Hauser, R. (2003) Autonomic nervous system dysfunction in Parkinson's disease. *Current Treatment Options in Neurology*, 5(2), 149-160.

Zigmond, A.S. and Snaith, R.D. (1983) The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica,* 67*,* 361-370.

Zimmerman, M., Chelminski, I., McGlinchey, J. and Young, D. (2006) Diagnosing major depressive disorder X: can the utility of the DSM-IV symptom criteria be improved? *Journal of Nervous and Mental Diseases,* 194(12), 893-897.

Zwieg, M. and Campbell, G. (1993) Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39(4), 561-577.