

BINDING SERVICES
Tel +44 (0)29 2087 4949
Fax +44 (0)29 20371921
e-mail bindery@cardiff.ac.uk

Validation of Patient-Reported Outcomes in Parkinson's Disease: Comparisons of Generic and Disease-Specific Health-Related Quality of Life Instruments.

A thesis submitted in accordance with the conditions governing candidates for the degree of

DOCTOR OF PHILOSOPHY

In the

UNIVERSITY OF WALES

Presented by

TAUHID ALI

Division of Clinical Pharmacy

Welsh School of Pharmacy

Cardiff University

April 2005

UMI Number: U584151

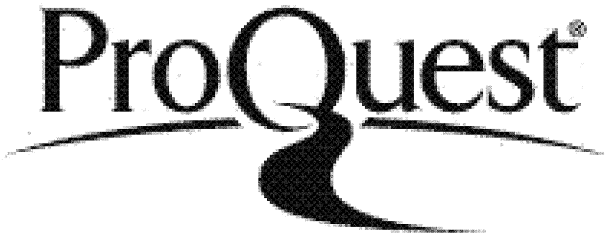
All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.



UMI U584151
Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ACKNOWLEDGEMENTS

بسم الله الرحمن الرحيم

I wish to express my gratitude to Dr. Sam Salek, my supervisor, for his guidance, and encouragement throughout the course of this research as well as his endless support during the period of completing this thesis. My special thanks go to Professor David Luscombe, Head of School, for allowing me to undertake this PhD programme and for his valuable support.

I would like to thank Dr. Russell Lane for his continued interest and support of the research project and for allowing me access to his patients at Ashford Hospital, Middlesex. Also, I would like to thank Professor Leslie Findlay who was particularly interested in the development of the Parkinson's Impact Scale and allowed me access to his patients at the Harold Wood Hospital, Romford, Essex and his PD specialist nurse, Sarah Morgan who worked diligently with patients to complete the questionnaires. My particular thanks and appreciation to all patients who participated in this research, including those from the out patient clinics, and the members of the UK Parkinson's disease Society. I am also grateful to Glaxo SmithKline Pharmaceuticals for agreeing to my involvement in their health outcome trial in Parkinson's disease and permitting me to use the data towards my thesis. I wish to thank James Lewis for his invaluable statistical advice.

I wish to thank Professor Jonathan Brostoff of the Middlesex Hospital, London, for his initial encouragement and recommendation to undertake this research and I would like to thank Dr. Wilson Totten for his support in providing me with such an opportunity to pursue postgraduate research whilst at Shire Pharmaceuticals Ltd.

Lastly, my wholehearted gratitude goes to my family for their support and encouragement, in particular to my wife for her endless belief in my ability and my children who persevered with my endless hours of absence whilst writing up this research.

ABSTRACT

Health-related quality of life (HRQOL) is central to the management of chronic diseases such as Parkinson's disease (PD). Several disease specific and generic instruments have been used to measure quality of life in PD, but nobody has to date identified which disease-specific or generic measure has the most relevant domains and optimal psychometric properties (reliability, validity and sensitivity) for use in the PD population. The aims of the study were therefore to compare the measurement properties of available instruments in PD population and recommend the most appropriate ones for use in future clinical research and routine practice.

The UKSIP a generic HRQoL instrument which had not been previously assessed in the PD population or compared against the PDQ-39 demonstrated good test-retest reliability (Spearman's r >0.70) and convergent validity ($p<0.001$) of the UKSIP overall score with all domains of the PDQ-39 in a postal survey of PD patients (n=91; mean age = 70; 64male). The UKSIP overall score correlated most strongly with the PDQ-39 mobility and activities of daily living scales. The strong Cronbach's a values (>0.70) for the UKSIP physical and psychosocial summary scores gave an indication that their items were internally consistent. Interestingly, much of the dysfunction reported by PD patients was related to non-physical difficulties according to the UKSIP and vice versa for the PDQ-39. This study highlighted the robustness and reliability of the UKSIP in this PD population. When comparing the UKSIP against a benchmark generic instrument, the SF-36, in a sample of patients (n=36; mean age = 69; 24male) from a neurology out-patients clinic, it had comparable psychometric properties with a PD clinical measure of disease severity, the Hoehn and Yahr (H&Y) scale. The two questionnaires were also able to discriminate between stage I, II and III/IV of PD as measured by the H&Y scale ($p<0.05$) for most domains. However, the UKSIP exhibited evidence of responsiveness to change over time in some domains of interest ($p<0.05$) amongst patients reporting deterioration as assessed by the H&Y scale, a pattern that was not repeated by the SF-36. Furthermore, items in the SF-36 were identified as redundant as they had little bearing in the elderly population whereas the UKSIP detected impairment in areas of interest to PD patients; such as sexual activity, impairment in cognition, and sleep disturbance which are not addressed in the SF-36. Having established that the UKSIP was the more suitable

generic instrument to use in PD the established disease-specific instrument, the PDQ-39 was assessed against the relatively new disease-specific Parkinson's Impact Scale (PIMS). The PIMS exhibited varying levels of internal consistency reliability which were found to be erratic between baseline and follow-up visits (n=47; mean age = 65; 25male). Also, the internal consistency of the PIMS column which was completed by fluctuating patients were wide-ranging (a 0.284 – 0.731) for each of the PIMS dimensions at baseline. On the other hand, all PDQ-39 domains exhibited good internal consistency reliability (Cronbach's a >0.70) which was stable across the baseline and follow-up visits. Though there was evidence of convergent validity for the PIMS with strong correlations with the PDQ-39 domain scores this was only for the non-fluctuating patients. Also, none of the correlation coefficients were found to be statistically significant at $p<0.05$ between PIMS domains and the clinical measure, the H&Y, indicating a lack of association between this clinical outcome measure and the domains of HRQoL measured by the PIMS. The PDQ-39 exhibited strong correlation ($p<0.005$) against the UPDRS score and against the H&Y. The PIMS did highlight that there was a difference in dimension scores between stable and fluctuating patients. An important observation from this study was that the PIMS was not well comprehended by the PD patients as was reflected by the high rate of missing data in what is a relatively short instrument. The study which was conducted across four neurology out-patient centres ($n=145$; mean age = 65; 80male) revealed that the PDQ-39 was able to identify that patients with fluctuating disease and dyskinesia experience greater impairment in the relevant domains of HRQoL compared to stable patients, and in particular for the domain of emotional well-being. The responsiveness of the PDQ-39 was partially confirmed as levels of significance ($p<0.05$) were reached for the domains of mobility in the motor fluctuation group with no dyskinesia; and communication for the stable group, who were deemed to have deteriorated at the follow-up visit as indicated by an increase in the total UPDRS score. Finally, the study observed that greater caregiver distress was evident for the motor fluctuating and dyskinetic groups compared to the carers of the stable patients. The PDQ-39 was shown to have good utility in a trial setting differentiating between distinct PD populations and showing suggestions of responsiveness to change.

In conclusion, the UKSIP and PDQ-39 instruments were found to be the most applicable generic and disease-specific instruments, respectively for use in PD studies.

ABBREVIATIONS

ADL	Activities of Daily Living
BDI	Beck Depression Inventory
CBI	Care-giver Burden Interview
DSRS	Dyskinesia Subjective Rating Scale
ES	Effect size
HRQoL	Health Related Quality of Life
H&Y	Hoehn and Yahr
LOCF	Last observation carried forward
MCID	Minimum clinically important difference
MFD	Motor fluctuations and dyskinesia
MFND	Motor fluctuations and no dyskinesia
MMSE	Mini Mental State Examination
NDMF	No dyskinesia or motor fluctuations
NHP	Nottingham Health Profile Scale
PD	Parkinson's Disease
PDQ-39	Parkinson's Disease Questionnaire
PDQUALIF	Parkinson's Disease Quality of Life Instrument

PIMS	Parkinson's Impact Scale
QoL	Quality of Life
SF-36	Short Form-36 Health Status Questionnaire
SRM	Standardised Response Mean
UKPDS	United Kingdom Parkinson's Disease Society
UKSIP	United Kingdom Sickness Impact Profile
UPDRS	Unified Parkinson's Disease Rating Scale

GLOSSARY OF TERMS

Activities of Daily Living (ADL)	Functions that are typically performed as part of a persons daily routine, such as dressing, bathing, eating, toileting, leisure activities, socializing, and other functions of daily life.
Applicability	Describes the content and emphasis of an instrument in terms of appropriateness of wording, clarity and simplicity of language.
Bradykinesia	The gradual loss of spontaneous movement
Bonferroni correction	A *post hoc* adjustment to the *P*-value to take account of the number of tests performed in multiple hypothesis testing situations
Concurrent Validity	Describes criterion validity that is demonstrated by administrating the new measure at the same time as the 'criterion' measure.
Construct Validity	Describes the extent to which a measure behaves in a manner consistent with it's relevant theoretical or logical properties by comparison with other measures of population characteristics.
Content Validity	Describes how adequately the items sampled represent the range of each domain assessed by an instrument.
Convergent Validity	Describes the extent of agreement between the instrument and measures of related and similar constructs.
Comprehensiveness	Describes how thoroughly the domain(s) of interest in a measure are sampled by the items or questions included in that measure.

Criterion Validity	Describes the ability of a HRQoL instrument to distinguish between different populations with different health related characteristics.
Cross sectional study	Those that are carried out at a single point in time
Divergent Validity	Describes the extent of agreement between the instrument and measures of related yet different constructs.
Dopamine	A neurotransmitter that controls movement and balance and is essential to the proper functioning of the central nervous system (CNS). Dopamine assists in the effective transmission of electrochemical signals from one nerve cell (neurone) to another.
Dopamine agonist	A drug that acts like dopamine. DAs combine with dopamine receptors to mimic dopamine actions. Such medications stimulate dopamine receptors and produce dopamine-like effects.
Dyskinesia	An abnormal involuntary movement that results in jerky or repetitive motions. Dyskinesias are common in many neurologic movement disorders including late Parkinson's disease (PD), restless legs syndrome, etc.
Face Validity	Describes the extent to which the items, on the face of it, appear to be measuring the variables they claim to measure.
Idiopathic	Refers to a disease of unknown cause that appears to be spontaneous in origin.

Internal Consistency	Describes the average degree of association or homogeneity between the items in the test.
Interpretability	Describes the meaningfulness of HRQOL scores produced by an instrument.
Kruskal-Wallis test	A non-parametric alternative to the one-way ANOVA; used to compare the distributions of more than two independent groups of observations
Longitudinal study	Following individuals over a period of time
Non-parametric tests	Hypothesis tests that do not make assumptions about the distribution of the data.
Null hypothesis	The statement that assumes no effect in the population
***P*-value**	The probability of obtaining our results, or something more extreme, if the null hypothesis is true.
Parkinsonism	A constellation of the following symptoms: tremor, rigidity, bradykinesia (slow movements), and loss of postural reflexes. Although classically seen in Parkinson's disease, parkinsonism may have other causes. In the elderly, parkinsonism may be caused by dopamine-blocking drugs, multiple system atrophy, striatonigral degeneration, Shy-Drager syndrome, cortico basal degeneration, diffuse Lewy body disease, and Alzheimer's disease with parkinsonism. In younger people, parkinsonism may be caused by juvenile-onset dystonia/parkinsonism, Westphal variant of Huntington's disease, Wilson's disease, L-dopa-responsive dystonia, Hallervorden-Spatz disease, and progressive pallidal degeneration.

Practicality	Describes the feasibility of using an instrument in its intended population and clinical setting.
Responsiveness/ Sensitivity	Describes instruments sensitivity to small clinically important changes in an individual's health-related quality of life.
Rigidity	Stiffness and resistance to movement. May be a symptom of a neurologic movement disorder such as Parkinson's disease.
Side effect	An effect of a drug that is not the main or intended effect. Side effects may be of no concern, or they may be bothersome or even dangerous, in which case they may limit the upper dose a patient can tolerate. Side effects are also called adverse effects.
Specificity	Describes the ability of an instrument to discriminate between a population with a particular disease and a healthy population.
Substantia Nigra	A dark band of grey matter deep within the brain where cells manufacture the neurotransmitter dopamine for movement control. Degeneration of cells in this region may lead to a neurologic movement disorder such as Parkinson's disease.
Test – Re-test Reliability	Describes the stability of an instrument over time by assessing the extent to which a score can be replicated under constant conditions or equivalent testing.
Spearman's rank Correlation coefficient	The non-parametric alternative to the Pearson correlation coefficient; it provides a measure of association between two variables

Tremor	Involuntary, small repetitive movements. May be a symptom of a neurologic movement disorder such as Parkinson's disease, essential tremor, etc.
Type I error	Rejection of the null hypothesis when it is true.
Type II error	Non-rejection of the null hypothesis when it is false.
Unpaired (two sample) *t*-test	Tests the null hypothesis that two means from an independent groups are equal
Wilcoxon rank sum test	A non-parametric test comparing the distribution of two independent groups of observations. It is equivalent to the Mann-Whitney *U* test.
Wilcoxon signed rank test	A non-parametric test comparing paired observations

CONTENTS

DECLARATION	ii
ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
ABBREVIATIONS	vi
GLOSSARY OF TERMS	viii
Chapter 1 – General Introduction	1
AIMS OF THE STUDY	24
Chapter 2 – Rationale and methodological framework of the study	26
Chapter 3 – Reliability and validity of the Parkinson's Disease Questionnaire (PDQ-39) and the United Kingdom Sickness Impact Profile (UKSIP) in patients with Parkinson's disease (PD)	56
Chapter 4 – Comparison of psychometric properties of the UK-SIP and SF-36 in idiopathic Parkinson's disease	99
Chapter 5 – An evaluation of the reliability and validity of the Parkinson's Impact Scale (PIMS) in patients with idiopathic Parkinson's disease (PD)	185
Chapter 6 – The burden of dyskinesia and motor fluctuations in patients with Parkinson's disease: a longitudinal study	239
Chapter 7 – General discussion	273
References	295
Publications	319
Appendices	321

Chapter 1

General Introduction

BACKGROUND

Since 1948, when the World Health Organisation (WHO) defined health as being not only the absence of disease and infirmity but also the presence of physical, mental, and social well being (World Health Organisation, 1991), quality of life issues have become steadily more important in health care practice and research. Based on it's definition of health, it is not surprising that the WHO defines quality of life (QoL) as an "individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns" (WHOQOL group, 1993). There has been a rapid increase in the use of quality of life evaluation as a technique of clinical research since 1973, when only 5 articles listed "quality of life" as a reference key word in the Medline database; during the subsequent five year periods there were 195, 273, 490 and 1252 such articles (Testa and Simonson, 1996). The growing field of outcomes research evaluate the efficacy, cost effectiveness, and net benefit of new therapeutic strategies to determine whether the associated increases in expenditures for health care are justified (Thier, 1992). Also, interest in the field of quality of life measures has grown partly in response to problems in managing chronic diseases and conditions. If we are unable to cure a patient, the aim should therefore be to provide good care, and part of such care would involve an assessment of how the disease condition impacts on broad aspects of the patients' lives. Quality of life assessment measures changes in physical, functional, mental and social health in order to evaluate the human and financial costs and benefits of new programs and interventions. The terms "quality of life" and more specifically "health related quality of life" refer to the physical, psychological, and social domains of health seen as distinct areas that are influenced by a person's experiences, beliefs, expectations and perceptions (Brook et al., 1983, Patrick et al., 1973).

WHAT IS HEALTH-RELATED QUALITY OF LIFE (HRQOL)

While there is general agreement on the potential value of QoL measures as key evaluation variables, there is an absence of clear agreement on a definition of QoL though such definitions do exist and share a common thread: definitions of QoL in the health context are mostly vague or absent and this ambiguity of definition has probably hindered the uptake of HRQoL measurement in the clinical setting. As noted

by Deyo and Patrick (Deyo and Patrick, 1989) conceptions relevant to health and QoL are diverse, scattered through many disciplines, and use many different labels (e.g. health status, functional status, disability scale, quality of life). Bergner (Bergner, 1989) notes that the notion of Quality of Life (QoL) has been a category in Index Medicus (a bibliographic listing of references to articles from biomedical journals worldwide) since 1966, yet QoL is usually not defined in the reports of clinical trials, and "definitions must be deduced from the dimensions assessed", and that "each investigator that purports to address quality of life actually examines a very narrow and specific set of factors". Generally notions of quality of life are not specified, but are considered to be implicit in the measure used, i.e. they are more inferred than explained.

Nevertheless, there seems to be acceptance that HRQoL is a 'multi-dimensional concept that encompasses the physical, emotional, and social components associated with an illness or treatment' (Revicki, 1989) (the common thread). Which precise dimensions to include are less agreed. For example, Torrance (Torrance, 1987) states that physiological and emotional functioning contribute directly to quality of life, and "taken together these two constitute health-related quality of life", with social functioning (e.g. social role and social contacts) outside the scope of health-related quality of life. On the other hand Kaplan et al (Kaplan et al., 1989) use the term health-related quality of life to refer to the impact of health conditions on function, but include social role, although suggesting that health-related quality of life may be independent of quality of life relevant to work setting, housing, or similar factors. Though the WHO definition of health is often cited, Calman gave one of the most cited definitions of HRQoL: "Quality of life measures the difference, or gap, at a particular period of time, between the hopes and expectations of the individual and that individual's experiences" (Calman, 1984). For Schipper et al., "QoL in clinical medicine represents the functional effect of an illness and its consequent therapy upon a patient as perceived by the patient (Schipper et al., 1996). From a pragmatic point of view, "QoL means the perception and evaluation, by the patient themselves, of the impact of disease and its consequences have on their life (Martinez-Martin, 1998). HRQoL is multidimensional, that is, it has numerous components. Illness compromises the biological integrity of patients and the prime objectives of medical intervention is to restore health and/or ameliorate the symptoms. The quality of life dimension is not separate from the clinical outcome but a complementary aspect of

the same condition (Malek, 1997). Therefore, the dichotomy between the clinical and quality of life endpoints is very much an artificial one. To include a dimension as a HRQoL component it should be susceptible of modification by therapeutic or health interventions (Martinez-Martin, 2001).

Authors do generally concur about the components of health which should be included in any measure of HRQoL, and these include as mentioned psychological, social and physical health; but also, duration of life; impairments; functional status; health perceptions and opportunities. These are health related, in that they are influenced by disease, injury, treatment or health policy (Patrick and Bergner, 1990). Such items reflect states that are felt to be universally desirable (Faden and Leplege, 1992). Other widely valued aspects of human existence that might be included in some measure of 'quality of life' are not generally domains of HRQoL. These include safe environment; adequate housing; guaranteed income and freedom. Such global concerns may adversely affect or be affected by disease, injury, treatment or policy, but are often unrelated to or distant from health or medical concern. HRQoL generally distinguishes the social, familial and behavioural factors and processes that influence it, particularly when health is viewed as an outcome. It is from the outcome perspective that most health status measures are developed and applied (Patrick and Bergner, 1990). Table 1.1 produces a comprehensive summary of some of the wide variety of components that have been included in operationalised measures of HRQoL.

WHY MEASURE HRQOL IN PARKINSON'S DISEASE

Pharmaceutical and technologic developments have contributed to longer life spans. As the population ages, the incidence and prevalence of chronic and disabling illnesses will likely rise, demanding increasing attention to the topic of quality of life (QoL). The number of people affected with Parkinson's disease (PD) continues to grow as the longevity of the patient population increases. We have an increase in the number of clinical trials assessing new treatments for PD, which is suggestive of advances in developing newer pharmacological treatments and advances in our understanding of therapeutic benefits.

An important part of understanding the therapeutic benefit of anti-parkinsonian drugs is understanding their impact on QoL.

Table 1.1 Components of Health Status and HRQoL Adapted from Ware(Ware, 1987)

Physical health Ability to perform physical activity and self care activities (e.g. eating, bathing and dressing), and the limitations caused by illness and pain.
Mental health Intensity of symptoms of psychological distress and behavioural dysfunction consequent upon this. Includes not just psychological distress, but psychological well-being.
Social functioning Social contacts and other activities (e.g. visits with friends and relatives), and social ties or resources (e.g. close friends and relatives that can be relied upon for tangible and intangible support).
Role functioning The performance (or ability to perform) usual role activities, including formal employment, school work, etc.
General perceptions of health and well-being How people evaluate overall health and well-being. Representing an individual appraisal or overall evaluation of the above factors.
Cognitive capacity and function Orientation, memory, comprehension, abstract reasoning and problem solving.

This is to say that therapeutic benefit is not solely the improvement of health but also an improvement to all aspects of a patient's existence which is affected by their health or in other words, their health related quality of life (HRQoL). Salek and Luscombe (Salek and Luscombe, 1992) state that health is viewed by some professionals in terms of simple, ordinary, normal or good psychological function. The more common definition of health, as previously stated, has been defined by the World Health Organisation (WHO) as "complete physical, social and psychological well being and not merely the absence of disease or infirmity (World Health Organisation, 1991). This however, does not provide a framework to measure HRQoL in a research setting and also the definition proposed focuses on the absence of good health or rather, outlines the negative aspects of health. The importance of positive health has been emphasised, and is considered to be more than just the absence of disease or disability but is instead concerned with completeness, full functioning or efficiency of mind, body and social adjustment (Bowling, 1997). Few attempts have been made within HRQoL instruments to achieve this, although some do contain items which emphasise positive aspects of HRQoL such as the SF-36 (Ware et al., 1994, Ware and Sherbourne, 1992). Martin (Martinez-Martin, 2001) has stated that HRQoL is a concept which has arisen as the

counterpoint to that of indiscriminate "quantity" of life for those illnesses that are chronic, progressive and without cure. Many health care workers now realise that when elimination or cure of a disease is not possible, a major aim of therapy must be to limit the complications of the illness so that the patient is able to carry on a comfortable, functional and satisfying lifestyle (Wenger et al., 1984). The alleviation of symptoms and prolongation of a patients life is not sufficient and must consider whether the patient can lead a satisfying life and to assess this HRQoL must be introduced and utilised in the medical care of patients.

A patient with chronic illness is likely to face many challenges in the course of or remainder of their life. No two experiences will be the same but what these patients encounter will impact on their physical, social, psychological and perhaps also their economic well being. Each aspect of well-being will be affected to different degrees by the severity of the disease or by external factors such as the availability of healthcare, access to affordable medication, support and care or perhaps the patients own expectations. The patient with Parkinson's disease (PD) is unlikely to be fit enough to hold down a full time job, and in an era where the age of retirement is increasing this can have an adverse financial impact on the sufferer and their immediate family. Such a patient is also more likely to need help in the house and become more dependent on the spouse or close family member as symptoms of the disease or medication impact on the patients' activities of daily living. As the disease advances the patient is unlikely to venture out of their home or away from their inner circle which could lead to social isolation and feeling of helplessness as they become more dependant on someone else. All these factors could contribute to a depressive state or lead to anxiety. In other words such a patient is very likely to experience deterioration in their Health Related Quality of Life (HRQoL). The impact of disease can also affect the carer who is burdened with the consequences of such an illness which can in effect impede the quality of life of the carer too. As the patient's ability to perform self-care and remain independent is diminished, caregiving burdens fall on the family and other caregivers, frequently straining interpersonal relationships. It is because of these entangled causes and effects that PD has on patient well-being that research into HRQoL is necessary and has gained considerable interest over the last ten years.

In fact, in certain diseases, such as cancer, quality of life may be the most important health outcome to consider in assessing treatment efficacy. Quality of life is a construct commonly viewed as having several dimensions or domains. The impact of PD reaches far beyond motor function and physical disability, ultimately affecting quality of life for many years. When assessing a chronic condition such as PD, the importance of recognising the subjective impact of disease, as well as the side effects of long term medical therapy for chronic illness has become the criterion.

BURDEN OF PARKINSON'S DISEASE

With an ageing population one should expect to see a rise in diseases associated with the elderly such as PD, Alzheimer's and cancer to name a few. Coupled with this increase in disease incidence is an increasing burden in manpower and finances to the healthcare system and society. The reality of the situation is who is willing to finance the new more expensive medications and what is the advantage to society in providing such a level of healthcare to such a population. Not all treatments are affordable to the state and there is a philosophy that extending the life of a patient does not necessarily result in good clinical care and ensure functionality (physical and psychosocial) for patients. Additionally, if the benefits of interventions are short lived, associated with toxicities, ineffective or result in the diversion of funds away from diseases or populations that are deemed to be more justifiable, economical or curable. The need to find the most cost-effective use of resources has therefore become a challenge to society, due to the inevitable finite level of funding that can be devoted to the health care of the nation. PD poses a major financial impact on society which is expected to increase in future years as the age distribution shifts to older age groups. On the basis of a prevalence of PD of 183 per 100,000, Dodel et al (Dodel et al., 1998) calculated an annual expenditure of €1.53 billion for the direct medical costs of PD in Germany. These costs were attributed primarily to drug therapy, care and nursing, and in-patient hospital stay.

CHOICE OF TREATMENT AND HRQOL

One of the potential ways of choosing between two treatments which produce similar outcomes in terms of cost, physiological outcomes and mortality is by assessing their impact upon patients HRQoL. For example, a more expensive therapy may be more

cost-effective because it improves quality of life, allows patients to return to employment, reduces hospitalisation and causes fewer drug related adverse events. Therapeutic success should be based on the criteria of clinical success, improvement in quality of life and economic effectiveness, rather than unidimensional cheaper alternatives (Salek, 1993). It should also be acknowledged that drug therapy is likely to have a negative impact on the HRQoL of individuals due to side effects as well as a positive effect. Measurement of HRQoL allows the net impact of a treatment both positive and negative, on the patients day to day life to be assessed (Salek, 1993).

The World Health Organisation (WHO) all encompassing view of health highlights the limitations of traditional approaches to quantifying not only health and the severity of disease states, but also the measures used to determine therapeutic efficacy. The expanded view of health proposed by the WHO has highlighted the need for the assessment of pertinent, although more subjective, health related concepts, such as disability, handicap and quality of life. Disease severity, physical function, and therapeutic efficacy have traditionally been measured in both clinical practice and research by disease outcomes or by specific symptoms. The effectiveness of new pharmacological agents and interventions have routinely been determined using these conventional clinical end points. However, there is now increasing interest in measuring the expanded view of health as outlined by the WHO that goes beyond clinical symptoms, including the psychosocial and economic cost of disease (Hobart et al., 1996). This expanded view of health and its relationship to disease impact is often referred to as HRQoL.

TREATMENT COMPLIANCE AND HRQOL

HRQoL is not merely about cost containment and the identification of the optimal treatment intervention but perhaps more importantly it is about what the patients own desire is with respect to the treatment process. This is particularly important in view of the problems with compliance which hamper effective treatment of chronic diseases. Col et al (Col et al., 1990) interviewed 315 consecutive elderly patients admitted to an acute care hospital to determine the percentage of elderly hospital admissions due to non-compliance with medication regimens or adverse drug reactions, their causes, consequences, and predictors. Eighty-nine of the elderly admissions (28.2%) were drug related 36 due to non-compliance (11.4%), and 53

due to adverse drug reactions (16.8%). One hundred three patients had a history of non-compliance (32.7%). Factors statistically associated with a higher risk of hospitalisation due to non-compliance were poor recall of medication regimen, seeing numerous physicians, use of numerous medications, and having the opinion that medications are expensive or unnecessary. Factors associated with an increased risk of an admission due to an adverse drug reaction were use of numerous different medications, higher medication costs, and not receiving any home services. Hence there is an economic and humanistic requirement to address such problems. The net impact of benefits versus side effects can be measured through the assessment of HRQoL. Research has also demonstrated that patient expectation and preference are important components of quality-of-life assessment when the purpose of this assessment is to predict or explain human behaviour (Testa, 2000). Such assessment is relevant, only if based on the patients own perception and value judgement.

PATIENTS PERCEPTION – PILLAR OF HRQOL ASSESSMENT

There is also a divergence of perception on disease status or impact or the effectiveness of an intervention between the healthcare practitioner and patient. The clinical evaluation of PD is usually carried out using scales and simple objective tests. The information obtained from clinical evaluation includes a variety of functional and mental aspects, in addition to data gathered in the patient's examination. However, many factors that are important to the patient may remain un-assessed. Emotional and psychosocial consequences of the disease are usually omitted from the physicians evaluation (Martinez-Martin, 1998). The frustrations and restrictions caused by the disease are heavy burdens for patients and their families. Clinical rating may provide data on severity of the disease or on the degree of disability, but in no way are the medical evaluations equivalent to the QoL status perceived by the patient. Quality of life data allows the physician to gather pertinent information on the patient that he would not otherwise be aware of from clinical measurements alone; such as the expectation of treatment from the patient's perspective. For example the physician may observe an improvement in clinical motor rating scores based on the cardinal symptoms of PD but the patient may be experiencing drug related side effects such as somnolence and involuntary movements which is not equating to satisfactory therapy from the patients perspective. Such important facets of the

health burden as HRQoL use may be seriously misjudged if not carefully measured but inferred from clinical observations alone (Chrischilles et al., 1998).

Hence the main reasons for the growing interest in HRQoL and the reason to measure HRQoL can be summarized to the following:

- Ability to measure the subjective perception of health in a reliable way
- Unique information supplied by HRQoL measures
- Relationship between perceived health and use of resources
- Usefulness for evaluation of outcomes and decision making

HEALTH RELATED QUALITY OF LIFE MEASUREMENT

The methodology supporting the development and validation of HRQoL instruments makes possible that subjective data may be produced and analysed with a reliability similar to that accepted for hard data (e.g. laboratory tests) (Fitzpatrick and Alonso, 1999). Methods for the measurement of QoL are classified as generic or specific depending on the population and health conditions which are to be investigated. Generic instruments include health profiles which are designed as questionnaires that attempt to assess the most important aspects of QoL. Generic instruments are multidimensional, cover a variety of areas and can be applied to different health states and diseases (Martinez-Martin, 1998). These tools are considered more appropriate for studies on health care programmes, surveys of general health status, and comparisons between diseases. Their potential drawbacks are the lack of sensitivity in some areas of interest and their low responsiveness to specific changes (Guyatt et al., 1996). Some of the most frequently used generic health assessment profiles are the Sickness Impact Profile (Bergner et al., 1981), the Nottingham Health Profile (Hunt et al., 1981) and the Short Form Health Survey (SF-36) (Ware and Sherbourne, 1992).

However, specific instruments focus on the aspects considered most relevant to the respective area of interest. These may be specific for a given disease, population, function or condition. Their main advantages derive from the prevalence and responsiveness of areas under assessment and are thus more appropriate than generic instruments for clinical trials (Martinez-Martin, 1998). The evaluation of

restricted domains and the limitation to specific populations and therefore their limited scope are potential disadvantages of these instruments (Guyatt et al., 1996). Examples of this type of instrument used in Parkinson's disease are the Parkinson's Disease Questionnaire (PDQ-39) (Jenkinson, 1997), Parkinson's Impact Scale (PIMS) (Calne et al., 1996) and the Parkinson's Disease Quality of Life Scale (PDQUALIF) (Welsh et al., 2003).

DISEASE-SPECIFIC VERSUS GENERIC INSTRUMENTS

Measures vary in the degree to which they are developed to measure a specific disease or to be capable of application to many or all illness states. As noted by Deyo and Patrick (Deyo and Patrick, 1989), disease specific measures have greater salience for physicians, better focus on functional areas of particular concern, and may possess greater responsiveness to disease-specific interventions. On the other hand generic measures permit comparisons across interventions and diagnostic conditions, which is particularly important for policy makers [resource allocation]. They also allow dysfunction to be quantified for an individual experiencing several disease conditions (Bombardier et al., 1986, Temkin et al., 1989).

There is some evidence that generic measures can be as responsive in some settings as disease-specific measures. Kaplan et al. (Kaplan et al., 1989) criticise diseasespecific methods on the grounds that all diseases and disabilities affect overall quality of life, and the purpose of QoL measure is not to identify clinical information relevant to the disease but to determine the impact of the disease on general function. General QoL measures are proposed as having a wider scope in capturing a wide variety of dysfunction that might be in different systems, i.e. not specific to the disease condition (e.g. confusion, tiredness, sexual impotence, depression). Furthermore, general measures have the ability to capture side effects and benefits that might not have been anticipated (Kaplan and Anderson, 1988) although once identified a disease-specific measure could be prepared to more exactly assess the dimension of interest.

An important question for the design of clinical trials is whether generic or specific patient completed measures should be used, or possibly some combination of the

two. It is thus necessary to appreciate the advantages and disadvantages of the two types of instruments.

Advantages and disadvantages of Generic measures

A major advantage of generic measures is that they can be used across a wide range of patient groups and also in general populations. This is because the content covers issues of general applicability such as physical mobility, sleep, emotional problems, social relationships and feelings of vitality. Some widely used generic measures also have reference values available (McKenna, 1993). These are derived from studies in which random populations have been surveyed and asked to complete the measure. Results from particular groups of patients can be compared with the reference values grouped for age, sex and socio-economic status. Where the measure is in the form of an index, specific areas of health status can be compared. The results of any one particular study can also be compared with those of another where the patients are suffering from similar disorders. This can be of great importance in economic decision making. Where choices have to be made between, for example, funding the treatment of two different illnesses, the benefit of each treatment can be determined by the change score on the generic measure. The treatment providing the greatest improvement on the measure can then be selected for funding.

Generic measures provide information on a wide spectrum of the problems that can be experienced by patients. Consequently, there is a limit to the detail in which each of these aspects can be described. Because of this limited coverage, the measure will lack sensitivity in aspects of health status which are important in a particular condition. Moreover, generic measures may also contain many items which are of no relevance to the condition under study and this could have the effect of alienating respondents. A further important consideration is the age of the most common generic measures. Most were developed around 20 years ago and the content may become less pertinent over time. The way in which patients conceptualise their problems, and the terms in which they express themselves can change in a generation. This is influenced by changes in values and standards in society and the resultant change in individuals' expectations.

Advantages and disadvantages of Specific measures

Specific measures can focus directly on relevant aspects of health problems under study. If they have been carefully developed, they should not omit important aspects of the respondent's experience. Because they do not need to enquire into a wide range of possible problems, specific measures can investigate the relevant areas in sufficient depth. As their content is all relevant, specific measures will be more sensitive to change in condition. This is particularly important where two similar treatments are being compared, as is the case in most clinical trials. An added advantage is that these measures should not include any items which are not relevant to the respondent, avoiding the associated ambiguous responses and missing data. Again, because of the relevance of the items, these measures are more acceptable to respondents. The questions should allow patients to express their experiences and feelings and feel that they are providing helpful responses.

By definition, specific measures have limited applicability. Some can be used in a group of related conditions but others will only be appropriate to one health problem. Such measures cannot be used to compare different types of medical conditions, although this is not a consideration with clinical trials. Disease-specific instruments take a narrow view of HRQoL as they miss important effects of the disease that are not intuitively obvious from a clinical point of view and yet are important aspects of HRQoL which could be captured by generic instruments (Coons and Kaplan, 1992). There is an argument that suggests HRQoL measurement should evaluate overall well being and satisfaction and not clinical information hence HRQoL measured by a generic instrument is relevant as it does not concern itself with a disease-specific symptom or characteristic. General reference values will not be available as there would be little point in using such measures with general populations. Disease-specific measures often suffer from a lack of published study data relating to their psychometric measurement properties, and tend to have undergone far less rigorous testing than their established generic counterparts (Salek and Luscombe, 1992). Some established disease-specific scales tend to be less well validated than benchmark generics such as the SIP and NHP.

Pros and Cons of the use of generic versus specific measures in clinical trial

Researchers have attempted to adapt generic measures to suit their particular research need. For example, sections have been omitted from measures where they

are thought not to be relevant to the population being studied, as was the case in a study of stroke patients where sections were removed from the NHP (Ebrahim et al., 1986). Others have added questions to fill in missing areas of importance. For example, Vickrey and colleagues (1992) added a few items to the SF-36 measure and then referred to it as a disease-specific measure (Vickrey et al., 1992). There are of course major problems with such approaches. When generic measures have been carefully developed they constitute a complete package and have been tested for reliability and validity in this form (McKenna, 1993). Adding or subtracting items changes the psychometric properties of the instrument as well as its nature.

A common recommendation is to include both disease-specific and generic measures in a study (Fletcher et al., 1992) and ideally a few instruments should be established as standards for use in most studies, although there is as yet no consensus on which should be selected. However, McKenna (1993) states "it is surprising that it has been suggested that disease-specific questions or measures should be used in addition to a generic measure when the latter is not appropriate for the therapeutic area or regimen being investigated" (McKenna, 1993). Indeed, this is an isolated view.

In the event where one wishes to utilise a generic measure for specific health problems or when generic measures are required to permit comparisons across interventions there is then a tendency to use more than one measure with the intention of covering (between them) all relevant aspects of health status. This approach has its problems. By increasing the number of measures used the burden on respondents is increased, as is the complexity of the clinical trial. The analysis from multiple instruments poses interesting challenges too. How are the results from different types of measures or instruments added together, particularly when they are covering similar areas? How do the researchers compare different patterns in changes in score, especially when two treatments produce improvements in different measures?

The first and most important issue when selecting an instrument is how well it will perform in the required situation (Table 1.2). This can be assessed from the instruments psychometric properties such as validity and reliability which are necessary.

PSYCHOMETRIC ATTRIBUTES OF HRQOL

There are a number of measurement properties which are required by HRQoL instruments. The development and validation of a QoL measure require methods to check the acceptability, reliability, validity, responsiveness and interpretability of the instrument (Juniper et al., 1996, Kirshner and Guyatt, 1985). The aim of testing scales for such attributes is essentially to ensure that they measure what they intend in a consistent and accurate manner. A summary of the attributes required by HRQoL scales is presented in Table 1.3. One must not assume a scale exhibits the required psychometric properties as psychometric properties may be affected by the population under evaluation, the method of administration, or other external variables.

Table 1.2: Factors influencing selection of instruments(Fletcher et al., 1992)

• Good measurement properties
  ○ Validity
  ○ Reliability
  ○ Responsiveness

• Type of instrument
  ○ Generic
  ○ Disease specific
  ○ Dimension specific
  ○ Items specific to study

• Method of administration
  ○ Self administered
  ○ Interviewer administered

• Cultural setting

Table 1.3: Psychometric Attributes Required by HRQoL Scales

Attribute	Description
Reliability	For any test or measure of health status to be useful it must be reliable, that is, repeat measurements made under constant conditions need to give the same result i.e. the instrument is measuring the given attribute in a consistent and reproducible manner. As put by Anastasi(Anastasi, 1990), measures of test reliability allow an estimate of the proportion of test variability that is due to error variance, where error variance is change in scores due to anything other than the characteristic of interest.
Validity	The instrument must be measuring what it is supposed to measure(Kaplan et al., 1976, Kirshner and Guyatt, 1985). There are a number of different forms of validity (see below).
Criterion validity	Criterion-related validation procedures indicate the effectiveness of a test in predicting an individual's performance in specified activities. It entails performance on the test being checked against a *criterion,* that is a direct and independent measure of that which the test is designed to predict(Anastasi, 1990) and validity can be estimated based on the correlation coefficient between predictor and criterion score. Correlation coefficients are used to measure the agreement between the instrument being tested and the "gold standard"(Streiner and Norman, 1995). In HRQoL there are currently no gold standard measures and so criterion validation is not pursued.
Construct validity	An ongoing process, where the association between the instrument and outcomes of related interest are tested against hypothetical constructs. The aim is to build a body of evidence which supports the claim that the instrument is measuring what it purports to measure.(Streiner and Norman, 1995). This is the cornerstone of the validation of HRQoL instruments. *Convergent* construct validity is expected with related measures, and *divergent* construct validity with unrelated ones.
Content validity	Ensures the right content and emphasis. A major priority in constructing any test is to ensure that all the elements necessary to measure the construct have been included. With a multidimensional construct there is a need to develop an appropriate weighting of these dimensions. This is the area of content validation. Content validation involves the systematic examination of the test content to determine whether it covers a representative sample of the behaviour domain to be measured(Anastasi, 1990).
Face validity	An indication of whether or not, on the face of it, the instrument appears to be assessing the desired areas of functioning(Streiner and Norman, 1995). This assesses whether the scale makes sense.
Responsiveness	Refers to the instruments ability to detect clinically important changes in quality of life and is determined by two properties(Guyatt et al., 1987): reproducibility and changeability (i.e. it must register changes in score when a subject's QoL improves or deteriorates(Guyatt et al., 1989).

In a clinical trial setting an instrument must demonstrate construct validity but also and equally importantly, it must exhibit responsiveness and reliability. If an instrument is not is not responsive to change then it has no utility in the clinical trial setting; alternatively, in a cross sectional survey this attribute is less important.

PARKINSON'S DISEASE AND ITS EPIDEMIOLOGY

In 1817, Parkinson, in the classic "Essay on the Shaking Palsy", described a novel condition that now bears his name (Parkinson, 1817). Parkinsonism is a clinical syndrome defined by the presence of tremor, rigidity and bradykinesia. The presence of two out of three cardinal motor signs and a robust response to an adequate dose of L-dopa are considered by most experts in the field to be essential for the diagnosis of PD and to distinguish it from other causes of parkinsonism (e.g. multiple system atrophy) (Colcher and Simuni, 1999). A set of exclusion criteria has been developed to make the diagnosis more accurate (Table 1.4).

PD is characterised by a chronic, slowly progressive course, and is initiated by an as yet unknown event which results in a localised degeneration of pigmented dopaminergic neurons in the substantia nigra pars compacta. The aim of therapy is to replace the loss of dopamine and is currently symptomatic.

PD is one of the most common neurodegenerative diseases. It has an estimated incidence of 5 to 21 per 100,000 per year (Rajput et al., 1984) and a prevalence of 66 to 258 per 100,000 (Sutcliffe et al., 1985) with a steady increase in older age groups (prevalence 60 to 64 years: 239.8 per 100,000; 70 to 74 years: 707.4 per 100,000; >85 years: 2205 per 100,000) (Mutch et al., 1986). PD affects manly older patients, but nevertheless, up to 10% of the patients are younger than 50 years old (Oertel and Quinn, 1996). Initially, symptoms occur unilaterally characterised by tremor, rigidity, slowness of movement (bradykinesia) and postural disturbances. As the disease progresses, it slowly but unremittingly involves other body parts (Maier Hoehn, 1983). This bilateral condition is reached after mean disease duration of 9 years. After a mean duration of 12 years patients require assistance with activities of daily living, and after approximately 18 years they may become confined to bed or a wheelchair. PD has no cure and despite the variety of pharmacological and surgical treatment options (Oertel and Quinn, 1996), the disease results in severe disability.

Another crucial occurrence in the progression of PD marks the onset of motor complications, which may present as fluctuations in motor ability or dyskinesias. These were rarely documented in the pre-levodopa era, but have now become a major concern in the long term management of PD.

Table 1.4: Clinical Diagnosis of Parkinson's Disease(Hughes et al., 1992).

Step 1	**Diagnosis of Parkinsonian Syndrome**
Bradykinesia (slowness of initiation of voluntary movement) At least one of the following • Muscular rigidity • 4-6Hz rest tremour • Postural instability not caused by primary visual, vestibular, cerebellar, or proprioceptive dysfunction	
Step 2	**Exclusion criteria for PD**
Parkinsonism due to identifiable causes, such as stroke, head injury, neuroleptic exposure, or brain tumour Oculogyric crises Sustained remissions Cerebellar signs Early severe autonomic insufficiency Early severe dementia Poor response to large dose of levodopa	
Step 3	**Supportive criteria for PD**
Three or more required for diagnosis of definite PD • Unilateral onset • Rest tremor present • Progressive signs and symptoms • Persistent asymmetry affecting side of onset • Excellent early response to L-dopa with persistence for = 5 years • L-dopa induced dyskinesia • Clinical course of =10 years	

After levodopa treatment for 3 to 5 years, motor complications are expected to occur in approximately 50% of patients, and after 10 years, in more than 80% of patients (Rinne, 1983). In addition to the limitations of physical function, other dimensions of

health may be impaired including mental health or social activities leading to a marked decrease in the quality of life (Shindler et al., 1995). These health (and economic) consequences, result in a considerable burden of illness to the patient and their families (Berry and Murphy, 1995).

HRQOL IN PARKINSON'S DISEASE

The syndrome of PD includes a complex variety of symptoms, including motor, autonomic, cognitive, psychological, and sensory manifestations (Koller, 1994). Parkinson's disease is a common neurodegenerative disorder, affecting 100-150 patients per 100 000 population (Martilla and Rinne, 1976, Tandberg et al., 1995). The main focus in evaluation and treatment of patients with Parkinson's disease has been on the motor disability of the disease.

During the past decade, greater emphasis has been placed on the total impact of Parkinson's disease on the patient's life, thus taking into consideration not only the physical impairment of the disease, but also various non-motor symptoms including depression, cognitive impairment, and sleep disturbances (Aarsland et al., 1996, Cummings, 1992, Lees et al., 1988). Even in the early stages of disease, mild changes in motor function result in subtle changes in quality of life, such as social embarrassment, feeling of depression, or role related impairment. Patients with early disease require special clinical assessment to ensure that they receive intervention for the more subtle, psychosocial symptoms that may not be captured in traditional clinical measures such as the Unified Parkinson's Disease Rating Scale (UPDRS) (Fahn et al., 1987) or Hoehn and Yahr Staging scale (Hoehn and Yahr, 1967). Psychosocial deterioration although less obvious than motor symptoms, can be very troubling to the patient and may imperil role related activities. Psychological changes are often extremely stressful and may be as disabling to the patient as impaired motor function.

This has been borne out in a multi-centre international survey of patients with PD which concluded that depression in PD, "satisfaction with the explanation of the condition at diagnosis" and "current feelings of optimism" have a significant impact on HRQL (Findley and Committee), 2002). These symptoms can intensify as the disease progresses, and result in a considerable burden of illness associated with

Parkinson's disease. As the medical interventions aim to produce symptomatic improvement, the concept of health related quality of life (HRQoL) should be implicit in outcome based investigation and data about the HRQoL, from the patients' perspective, are fundamental to understanding the effectiveness of care (Karlsen et al., 2000).

Thankfully there is growing awareness among the principal stakeholders in health care that health outcomes data should be used in the evaluation of medical care interventions. Health-related quality of life is a patient-reported health outcome that is gaining increasing attention. In fact, in certain diseases, quality of life may be the most important health outcome to consider in assessing treatment efficacy (Coons and Kaplan, 1992). There are a number of methodological issues in the measurement of quality of life that must be carefully considered when designing a study or evaluating existing research and these have been already been addressed.

A good quality of life scale assesses dimensions of everyday life that are missed by more narrowly designed, health specific scales. To be valid, however, quality of life scales should contain a broad range of domains relevant to the condition or treatment under study; assess recent time periods; be sensitive enough to monitor expected changes; contain sufficient range to include patient conditions; contain both positive and negative items; and contain selective subjective evaluations (Strawbridge, 1998).

Use of HRQoL in clinical trials of PD medicines

The limited available evidence of the effectiveness of many existing and emerging medical practices has led to the realization that health outcome data should be used in the evaluation of health care interventions. Interest in improving health outcome evaluation is increasing as a result of the need to balance the costs and benefits of medical technologies. Measures of health-related quality of life are receiving greater use in clinical studies and are slowly becoming common place in phase III trials in PD. Quality of life measurement provides additional data for making clinical and health care policy decisions and thus the argument for their inclusion are compelling. In addition, there is growing awareness that in certain diseases, quality of life may be the most important health outcome to consider in assessing treatment efficacy (Coons and Kaplan, 1993). However a literature search of the MEDLINE database

using the search terms quality of life, HRQoL, Parkinson's disease and clinical trials; a total of 155 published articles were found that were published between 1974 and1999. From these articles only three (2%) fulfilled the search criteria in that they referred to the measurement of quality of life (QoL) in a PD clinical trial or discussed the topic to some extent in relation to a pharmacological clinical trial. In one particular paper assessing quality of life in PD, the researchers measured QoL on a 5-point Likert scale(Koplas et al., 1999) and in the other two papers the instruments of choice were the Sickness Impact Profile in one and the other did not mention the instrument used. The two latter papers were an evaluation of surgical interventions (Arle and Alterman, 1999, Vingerhoets et al., 1999). A separate search of the key terms health related quality of life and Parkinson's disease, and clinical trials; limiting the search to clinical trials between 1999 and 2005 showed 40 entries. Of the 40 entries only 14 (35%) were of clinical trials of a therapeutic intervention or procedure such as unilateral thalamic deep brain stimulation and which specifically measured HRQoL as a (secondary) outcome. The Parkinson's disease questionnaire (PDQ-39) was cited in 8 out of the 14 papers strengthening its claim to being the most accepted disease-specific instrument of HRQoL in PD trials. The generic measures, SF-36, Sickness Impact Profile and Nottingham Health Profile were the chosen measure of HRQoL in the remaining papers. These literature searches highlight that though the use of HRQoL instruments in clinical trials is increasing it is far from being universally employed even though the rationale and justification for its inclusion have been discussed above. The possible reason for the poor uptake of the measurement of HRQoL in PD studies may be due to the plethora of generic and disease-specific instruments that have been developed or introduced in the last five to ten years and hence clinicians and trialists are perhaps confused by the benefits or advantages of one instrument over another. However, the main contributing factor is the lack of activity in the field of QoL not being communicated actively to those on the ground working with patients on a day to day basis. Nevertheless, there have been some activities in recent years to bring QoL instruments to one source to facilitate dissemination of QoL instruments such as the Compendium of Quality of Life Instruments (Salek, 1999)

Martinez-Martin (2001) conducted his own review of drug trials which measured HRQoL (Table 1.5). Nine studies on antiparkinsonian drugs applying HRQoL measures as efficacy variables have been published since 1997. Around 2750 patients have been included and the mean duration of studies was over 1 year (range, 3 weeks to 5 years). Five HRQoL instruments, two of them specific for PD, were applied. Immediate or controlled release formulations of levodopa were used in all these trials. Nevertheless, this still confirms there is still some way to go before HRQoL measures are routinely used in clinical trials of pharmacological interventions. The main applications of HRQoL in PD trials are clear. Such instruments can identify problems undetected in clinical methods; provide a comparison between different conditions e.g. dyskinesia PD patients and stable PD patients, and evaluation of clinical interventions. For example, an article about an open multi-centre study on efficacy, safety and tolerability of entacapone as adjunct therapy to levodopa by Durif and co workers (Durif et al., 2001) has been published. Evaluation was performed by application of the UPDRS, investigators global assessment and the PDQ-39 (as a secondary outcome measure). At the end of the study, significant improvement was obtained for ADL and "off" time spent during the waking period. A significant improvement was shown for all the PDQ-39 dimensions, except social support and cognition. Mobility, ADL, emotional well-being and bodily discomfort were the most improved. This highlights how the inclusion of a HRQoL instrument can complement the clinical outcome endpoints.

Recently, Damiano et al (Damiano et al., 1999) - based on a review of the literature, consultation with HRQoL experts and clinicians and interview with patients – identified 12 HRQoL dimensions particularly relevant for PD patients (Table 1.6). Self-image, social function, role function and sexual function have traditionally been ignored by clinical measures.

Table 1.5: Drug treatment of PD and HRQoL – adapted from Martinez-Martin (Martinez-Martin, 2001)

Drug treatment	Patients (n)	Study duration	HRQoL measure	Type of instrument	Main improvement	Ref
Levodopa St vs CRF	618	5 years	NHP	Generic	(CRF) Emotion Social isol	(Block et al., 1997, Koller et al., 1999)
Levodopa St ? CRF	20	3 weeks	SIP	Generic	Home manag. Mobility	(Pahwa et al., 1997)
Tolcopone adjunctive levodopa	177	1 year	SIP	Generic	(TLCP) Psychosocial Total score	(Baas et al., 1997)
Tolcopone adjunctive levodopa	202	1 year	SIP	Generic	No change	(Rajput et al., 1997)
Levodopa St ? CR	276	3 months	NHP	Generic	Emotion Mobility Pain	(Martinez-Martin et al., 1999)
Levodopa St ? CR	621	3 months	PDQ-39	Specific	All domains	(Martinez-Martin et al., 1999)
Tolcopone adjunctive levodopa	46	6 weeks	SIP	Generic	Psychosocial Total impact	(Welsh et al., 2000)
Pramipexole vs levodopa	301	2 years	PDQUALIF EuroQoL	Both	(Levodopa) Sleep Self Sexuality	(2000)
Entacapone adjunctive levodopa	489	8 weeks	PDQ-39	Specific	Emotion Mobility ADL Pain	(Durif et al., 2001)

St vs CRF: Standard versus Controlled Release Formulation. St ? CR: Conversion from standard to controlled release formulation. NHP: Nottingham Health Profile. SIP: Sickness Impact Profile. PDQUALIF: Parkinson's disease quality of life instrument. PDQ-39: Parkinson's disease questionnaire.

Table 1.6: Parkinson's disease – relevant areas of HRQoL

Physical function
Mental health and emotional state
Self-image
Social function
Health related distress
Cognitive function
Communication
Sleep and rest
Eating
Role function
Energy/fatigue
Sexual function

Despite the interest in HRQoL as a measure of outcomes in therapeutic interventions, less than one study per year using HRQoL evaluation as a variable of outcome has been published in the last decade. Additional work on the instruments for HRQoL measurement of patients with PD is absolutely necessary in order to complete aspects of validation still unexplored, to compare their performance in different settings and to obtain consensus about their application. HRQoL measurement can contribute to the knowledge of 'hidden' aspects of the disease to justify this evaluation as a complement to the clinical one.

AIMS OF THE STUDY

Improvement in health related quality of life is an important objective of therapeutic interventions used in the treatment of patients with Parkinson's disease. However, the optimal method of measuring this important outcome is far from certain and also, as highlighted in the introduction, we can see that several instruments have been used in the clinical setting. A number of instruments are available which claim to be reliable and valid, although in the absence of independent or comparative research, the selection of an appropriate instrument will be made on an ad-hoc basis.

Aims

The aim of this study is to assess and compare methods of measuring the health-related quality of life in PD patients using both generic and disease-specific instruments and to determine which instrument possesses the necessary properties to be utilised in PD research. If one can identify a measure that fulfils the required psychometric properties of a HRQoL instrument and show that it has application in the PD setting then it may be feasible to make the selection of a suitable instrument more considered. This may lead to a greater utilisation of such instruments in PD intervention trials.

Objectives

- Comparisons of generic and disease specific instruments will be made to identify their psychometric properties, establish their construct validity and comparative measurement properties.

- To examine which aspects of HRQoL are identified or omitted in the generic measure and vice versa.

- To evaluate whether generic measures such as the SF-36 and the UKSIP have utility in the measurement of HRQoL of PD patients.

- The construct validity and sensitivity to change will be measured and this will add to the existing body of knowledge of the two generic instruments.

- To evaluate which of the generic instruments might be best suited for use in the clinical setting.
- To evaluate a newly developed disease specific instrument, the PIMS, against the PDQ-39.

- To assess the performance of the PDQ-39 as arguably the most commonly used measure of PD, in a trial setting to determine whether it can differentiate between stable and dyskinesia patients and examine the relationship between poor HRQoL and dyskinesia.

- To assess the impact of caregiver burden on the carer of the PD patient.

It is hoped that the outcome of this research project will establish the most suitable instrument for use in the evaluation of health-related quality of life in patients with PD.

Chapter 2

Rationale and methodological framework of the study

RATIONALE

The clinical evaluation of PD is usually carried out using scales and simple objective tests (Martinez-Martin, 1998). The information obtained includes a variety of functional and mental aspects, in addition to data gathered in the patient's examination. However, many factors that are important to the patient may remain un-assessed. Emotional and psychosocial consequences of the disease are usually omitted from the physician's evaluation. The frustrations and restrictions caused by the disease are heavy burdens for patients and their families. Clinical rating may provide data on severity of the disease or on the degree of disability, but in no way are the medical evaluations equivalent to the quality of life (QoL) status perceived by the patient. The clinical instruments used to characterise the impact of the disease are completed by the clinicians, emphasising the rating of neurological signs such as tremor, rigidity, bradykinesia and instability of posture. Whilst such instruments are valuable in clinical studies, they do not provide a complete picture of the impact of the disease. Most importantly they fail to address the impact of the illness upon subjectively assessed quality of life of patients (Longstreth et al., 1992).

In Parkinson's disease (PD) several disease specific HRQoL instruments have become available in the past few years and there is a choice of several generic instruments that have been used in the measurement of HRQoL in PD. Investigators who want to use such an instrument are faced with the choice of several scales which differ in many respects. In the process of selecting the appropriate instrument, a comparison of the quality of these scales can be helpful. This project therefore compared and contrasted the disease specific and generic HRQoL instruments in PD and evaluated their clinimetric properties.

There are a number of instruments in the literature that measure, or purport to measure, quality of life. As part of the rationale for identifying the most suitable instrument to use in PD, the shortcomings and properties of the existing instruments must be considered. By posing certain questions and answering them in the text that follows, one can offer a clear picture of the limitations of the instruments described in the literature.

Are the instruments measuring HRQOL comprehensively?

HRQoL is a holistic concept that incorporates a number of different domains. Some of the instruments in the literature may be measuring aspects of HRQOL but are not measuring all the aspects necessary to give a comprehensive picture of a person's HRQoL. HRQoL instruments should ideally consist of items selected on the basis of interview with the patient, the clinician, nurse and even close family members.

Have all the psychometric properties been established?

Many instruments in the literature have not yet fully demonstrated all the necessary psychometric criteria. All instruments must demonstrate satisfactory reliability, validity and sensitivity in the population in which they are to be used. Sufficient evidence supporting these properties must be available. Documentation regarding developmental methodology must be available to ensure that the instrument has been correctly developed and to make sure that it is used in the correct way. It must be validated in the population in which it is to be used and any manual documenting its use should be strictly adhered to.

Does the measure have a conceptual framework?

It is important that the conceptual framework of an instrument is based on sound theory and that the constructs being measured make up a structured and complete assessment of an individual's quality of life.

Is the instrument used in accordance with its intended purpose?

Any studies using a validated instrument must be adhering to the methodology developed for its use. This means that standardised procedures for its administration and processing must be employed to ensure validity in any studies subsequent to the original validation work. It can also only be used appropriately in populations with similar characteristics to the original validation subjects. Additionally, Some HRQOL instruments designed for one purpose may not be suitable for another. For example an evaluative instrument may not be suitable for some studies that may require a discriminative instrument. Some HRQOL measures designed for pharmacoeconomic evaluation may not be suitable for assessing the outcome of drug treatment.

Is the instrument feasible and acceptable to respondents?

Simplicity and brevity are key properties for an instrument that is intended for an elderly population. Feasibility and acceptability also relates to any persons involved in administration, handling, scoring and interpretation of results. Instruments may not be suitable if they are too labour or time-intensive or if they require too much training for researchers or healthcare professionals.

Can the instrument measuring HRQOL longitudinally?

An instrument has greater application if it can be used to make comparisons between subjects at different stages of disease severity. This allows HRQOL changes across the course of the disease to be followed.

Is more than one instrument needed to measure HRQOL?

Using more than one instrument to assess HRQOL in a population increases the risk of type II error and therefore it is preferable to use one instrument to assess all the domains. However, there may be certain situations where it would be desirable to use more than one instrument e.g. in a clinical trial scenario a disease-specific and generic instrument can provide a stronger message showing that each validates the other and increases the level of confidence of the outcome.

Is there a need for a new HRQOL measure in PD or are existing measures appropriate?

At the point of initial literature review, the HRQOL instruments available for use in PD were limited in one or more of the above areas, for example; very few instruments had been tested for longitudinal sensitivity. It was because of these shortcomings that a decision was made to assess the psychometric properties of existing and recently developed HRQOL measures in head to head comparisons.

There was a clear need for such comparative studies as this would be the only way to identify which instrument was measuring areas of health considered to be of importance to the patients, which instruments were lacking in content and most importantly which instrument exhibited the best psychometric properties. It was conceptualised that the most appropriate instrument to use in an elderly population in either the trial or routine clinical setting should be acceptable and relevant to the intended population and should demonstrate satisfactory psychometric properties. It

must be suitable for assessing the impact of disease as well as drug treatment and be usable across a reasonable range of disease stages. Further, it seemed from the literature that there was a paucity of comparisons between generic and disease specific instruments in PD, an issue which will be addressed in this research.

EVALUATION OF QUALITY OF LIFE INSTRUMENTS

Since many components of quality of life cannot be observed directly, this is normally done through an instrument carefully designed to reflect the quality of life concept. Therefore, the ability to measure such a phenomenon depends on how the instrument is conceptualised (Testa and Simonson, 1996). In healthcare research, this often brings together a set of domains related to health of the individual, as previously discussed in Chapter 1. Importantly, the measurement properties which are a function of the theoretical framework of the quality of life construct and how well the scale performs, affect an instruments ability to detect meaningful results (Testa and Nackley, 1994). Thus, in addition to the conceptualisation, methodological issues for the evaluation of the quality of life instruments should be considered in order to establish the measurement properties such as applicability, practicality, comprehensiveness, reliability, validity, specificity, responsiveness and interpretability. These are described below.

Applicability

The content and emphasis of the measure used must be appropriate to the purpose being considered and acceptable to the population that will be completing it (Salek and Luscombe, 1992)

Practicality

An instrument should be feasible for use in its intended population and clinical setting. The length of the questionnaire must be appropriate for the population and place the minimum amount of additional burden on both respondents and other healthcare professional involved in the data collection (Salek and Luscombe, 1992). If the QoL measure is not practical to use then it is unlikely to receive acceptance from both the patients and healthcare professionals.

Comprehensiveness

A measure should consider the psychosocial and financial impact as well as physical. The comprehensiveness of a measure influences its accuracy and scope (Testa and Nackley, 1994).

Interpretability

HRQOL scores produced by any instrument must be meaningful in order to allow clinicians to interpret the results (Guyatt et al., 1996). For discriminative instruments it is important that a particular HRQOL score can be attributed to a particular level of impairment or functioning. For evaluative instruments, it must be possible to determine the size of improvement or deterioration represented by the magnitude of a change in HRQOL score. There are two main approaches for interpreting HRQOL data; anchor-based and distribution-based.

The **anchor-based approach** compares the HRQOL changes to other clinically meaningful outcomes (Lydick and Epstein, 1993) which can range from a global rating to life events or diagnosis(Juniper et al., 1994). A global assessment question (subject or clinician rated) is most often used as the anchor to which changes reported in an instrument are compared (Jaeschke et al., 1989).

The **distribution-based or effect-size based approach** relies on the statistical distribution of a set of results (Jacobson and Truax, 1991). **Effect size** is most commonly used and is calculated from the magnitude of change and the variability in stable subjects (Lydick and Epstein, 1993). The effect size represents the size of the HRQOL difference produced by a treatment or intervention and can be used to compare scores in different studies or determine the relative importance of a treatment effect within a study (Fletcher et al., 1992). Although the larger the effect, the greater the probability of statistical significance (Lipsey, 1990), it is difficult to interpret what this difference means in terms of impact on the patient. Although some researchers (Guyatt et al., 1987) believe that this approach at best provides an underestimate of the minimal clinically important difference, it has been used to provide evidence of clinical significance (Testa and Simonson, 1996).

Reliability

Reliability simply assesses that the instrument is measuring something in a reproducible fashion. The concept of reliability is a fundamental way to reflect the amount of error, both random and systematic, inherent in any measurement (Streiner and Norman, 1995). Reliability is a measure of the consistency and stability of the instrument. Although there is no consensus regarding critical values for reliability, most researchers refer to Nunnally (Nunnally, 1978) who considered reliability coefficients (r) of 0.70 as moderate for group comparisons. There are different types of reliability that can be measured: test-retest reliability, interrater reliability, and internal consistency reliability (Salek and Luscombe, 1992).

Test-retest reliability determines the extent to which a score can be replicated in identical or equivalent testing situations based on correlations between scores by the same person on two separate occasions. It is a measure of stability over time. The longer the interval between two assessments the greater the chance that real changes have occurred resulting in the reliability being underestimated (Hunt et al., 1981). The time period between the two occasions must be such that no true change in HRQOL has occurred. A one week interval is recommended by Meltzer (Meltzer and Hochstim, 1970).

Inter-rater reliability requires study designs that permit several raters to observe the same subjects independently (Schulzer et al., 1999) and relates to the degree of concordance shown between two or more raters in assessing HRQOL of the same sample population using the same HRQOL measure. It is only considered for interview and interview administered measures.

Internal consistency is the average degree of association or homogeneity between the items in the test (Streiner and Norman, 1995). Cronbach coefficient alphas (Cronbach, 1951) are the basic formulas for determining the reliability of instruments based on internal consistency and reveal whether items within a test are measuring different aspects of the same factor while contributing to the total score for the test.

Validity

The validity of a HRQOL instrument refers to its ability to measure what is intended and the burden of testing validity arises from a series of converging experiments (Streiner and Norman, 1995). There are three basic types of validity frequently described for health status indicators: criterion, content and construct (Salek and Luscombe D K, 1992).

Criterion validity assesses the correlation between a new measure and another that is assumed to be superior (the gold standard). Concurrent validity is established if the 'superior' measure is another HRQOL instrument that is already well established and validated in the population under study, and is completed at the same time as the new instrument. For example, establishing the concurrent validity of a new scale to measure severity of PD symptoms might require co-administration of the instrument with a gold standard such as the Hoehn & Yahr staging scale (Hoehn and Yahr, 1967). As there are currently no recognised 'superior' HRQOL measures of proven validity in PD, criterion cannot be assessed.

Construct validity is used when a gold standard for criterion validity is not available which is often the case in a relatively new area of HRQOL assessment. It must be investigated whenever no criterion is accepted as entirely adequate to define the quality of life to be measured (Schulzer et al., 1999). It is a process by which empirical evidence from several validation procedures is assembled to support the inference that a particular instrument measures what it purports to measure. Construct validity is established through a series of tests demonstrating that the variable as measured (a) does not correlate positively with measures of related but different constructs (**divergent validity**), (b) does correlate positively with related and similar constructs (**convergent validity**). The process of assessing these aspects of validity increases confidence that on observed effect pertains to the variable of interest. Construct validity can be demonstrated by correlations of items in the test measure against an external measure.

Content validity concerns the extent to which a set of items taps the content of some domain of interest i.e. is the instrument capable of measuring health status in a specific population and assessing a specified endpoint such as a successful therapeutic intervention. Moreover, this process ensures that the measure

possesses the right focus and emphasis. To demonstrate this type of validity it is necessary to possess a clear definition of HRQOL in the population under study and knowledge of the relevant area that the instrument was designed to measure.

Sensitivity (responsiveness) refers to the instruments ability to detect any clinically important changes in quality of life, even if they are small (Guyatt et al., 1989). The greater an instruments sensitivity to change, the more useful it is in a clinical trial situation, because it will be more sensitive to clinical intervention. Guyatt et al, 1992 suggests that adequate responsiveness is a prerequisite for an outcome measure, next to validity and reliability.

The **distribution based or effect size based** approach relies on the statistical distribution of a set of results (Jacobson and Truax, 1991). **Effect size** is most commonly used and is calculated from the magnitude of change and the variability in stable subjects (Lydick and Epstein, 1993). The effect size represents the size of of the HRQOL difference produced by a treatment or intervention and can be used to compare scores in different studies or determine the relative importance of a treatment effect within a study (Fletcher et al., 1992). Although the larger the effect, the greater the probability of statistical significance (Lipsey, 1990), it is difficult to interpret what this difference means in terms of impact on the patient.

Sample Size

The choice of an appropriate sample size is a crucial aspect of any study design. With an increased sample size, the standard error of an estimate will be reduced, leading to increased precision and study power.

SAMPLE SIZE

The statistical theory that underlies the choice of a sample size is governed by the power required for a given study and is critical to any study where HRQoL is an important endpoint. The power of a statistical test is determined by effect size, which is the difference between the mean value produced by the treatment, reliability of the measurement and the significance level (Lipsey, 1990). While the significance level can be chosen in advance and the reliability can be continually evaluated, the effect size produced by treatment on HRQoL of patients with PD has not been

elucidated. In addition, the responsiveness index of the quality of life instrument in this population has not been fully developed. These limitations have made the sample size calculation from a standard formula unreliable, if at all obtainable. Due to the exploratory nature of this research, sample size calculations appear irrelevant in this research study.

In exploratory research, the main consideration is the selection of study subjects with a variety of socio-demographic backgrounds and at various stages of disease severity. A population of 50 subjects may be sufficient to achieve this purpose. This number is also big enough to be subjected to most statistical tests in this research. Thus this sample size will be used for exploring the applicability of selected quality of life instruments in the study condition. When determining the implementation of the quality of life assessment in practice, however, it would be highly desirable to consider a much larger sample size.

DATA PROCESSING AND ANALYSIS

All analysis of data will be performed using SPSS for Windows. The distribution of the variables will be looked at using descriptive statistics and examined visually using normality plots. The distribution of variables in each study determines the type of statistical tests to be used. If the data is normally distributed with homogeneous variance properties then parametric statistical techniques (Brown and Hollander, 1977) (Bryman and Cramer, 1990) can be used. Such tests are reasonably robust and can cope with minor violations of normality but do not have the power to cope with heterogeneous variance properties.

HRQOL studies tend to produce asymmetric data with a number of outliers and heterogeneous variance properties and, although parametric tests may be used, it is more appropriate to use non-parametric statistical techniques (Siegal and Castellan Jr, 1988) for this data. Non-parametric statistical tests will be mostly used in analysing the data from this study. This is because non-parametric tests do not specify conditions about the parameters of the population from which the sample was drawn. Parametric tests on the other hand, assume that the observations are drawn from a normally distributed population. In most instances, non-parametric tests produce exact probabilities, regardless of the shape of the population distribution.

Non-parametric tests are available to analyse data that are inherently ranked as well as data whose scores have the strength of ranks (i.e. scores that may be higher or lower without knowing by how much), without making the unrealistic assumptions entailed by parametric methods. Non-parametric tests are also available for analysing samples taken from several different populations, again without making the unreasonable assumptions required for parametric tests.

Hypothesis testing is used for making statistical inferences about unknown parameters. The results of such tests are usually presented as a table incorporating a measure of central location, a measure of spread and a probability interval. Values known as *p*-values are computed which inform the researcher whether the hypothesis is to be rejected. Mean or median values are to be used as measures of central location while measures of spread include the standard deviation (a partner of mean with respect to spread), or inter-quartile range (a partner of the median with respect to spread). The level of statistical significance for the probability of Type I error (α) will be set at 0.05 using a two tailed test and for the probability of Type II error (β) at 0.20 giving a statistical power of 0.80 (1 - β). When comparing two independent samples, the Chi-square test can be used for categorical or nominal data and the Mann-Whitney U test for ordinal data. Where more than two independent samples are involved, the Kruskal-Wallis one-way analysis of variance can be used (Siegal and Castellan Jr., 1988).

The measure of association will also be used to evaluate the relationship of studies variables with the areas of interest. This will be analysed using the Spearman rank-order correlation coefficient (Siegal and Castellan Jr., 1988). The level of significance (α) will be set at 0.01 to decrease the probability of wrongly rejecting the null hypothesis when large numbers of correlations are calculated. For the purpose of interpreting the association, a correlation coefficient greater than 0.5 will be considered as 'high', between 0.4 – 0.5 as 'moderate' and between 0.3 – 0.4 as 'weak' (Cohen, 1977). A correlation of less than 0.3 will be considered to be negligible in most cases. In addition, with regard to a multidimensional nature of the quality of life data, a multivariate analysis will be employed where appropriate.

Reliability test – A type of reliability that can be measured for HRQOL instruments is the internal consistency reliability, which refers to the average degree of association among the items on a test. Multiple responses are obtained in most HRQOL assessments that are then summed or reduced to a single score. An implicit assumption in this approach is that the individual items measure the same thing, or that the items are consistent with each other (Spilker, 1996). The higher the inter-item correlation, the higher the multiple correlations with the factors they measure in common. Coefficient alphas are the basic formulas for determining the reliability of instruments based on internal consistency (Cronbach, 1951) and the Cronbach's alpha (Cronbach, 1951) technique will be utilised for measuring internal consistency reliability. The result indicates the level of overall correlation between items within a scale whereby a certain level of association is expected between scale items if they are measuring the same overall concept. Thus, one would expect moderate correlation amongst the items making up a domain which purported to measure physical functioning (Streiner and Norman, 1995). Alpha grows larger as the number of items and/or the average inter-item correlations increases (Schulzer et al., 1999). Cronbach's methodology (Cronbach, 1951) will be used for assessing the reliability of the HRQoL instruments of choice because it is the most commonly used. Cronbach's alpha is reported in the literature for almost all HRQOL instruments where internal consistency has been assessed. A HRQOL instrument must demonstrate reliability before it can be used to detect clinical differences between patients at a particular time-point. The most obvious means for estimating the reliability of a test is through representing the identical test on a later occasion. Measures of test reliability allow an estimate of the proportion of test variability that is due to error variance, where error variance is change in scores due to anything other than the characteristic of interest (O'Connor, 1993). Test-retest reliability is demonstrated by measuring the correlation between scores obtained by the same person on two separate occasions. The error variation corresponds to random fluctuations in performance on these two occasions. Test-retest reliability will be measured on two occasions separated by 5 or 7 days. Correlation will be measured using Spearman's rank correlation coefficients as the data will be non-parametric. Few if any standards exist for judging the minimally acceptable value for a coefficient of stability, but the commercially published individually administered aptitude test are amongst the highest. Subsets of the WAIS (Weschler Adult Intelligence Scale) have coefficients in the 0.70s, 0.80s, and low

0.90s. Crocker and Algina(Crocker and Algina, 1986) propose that well constructed test should still have test-retest coefficients in the 0.80s .

Validity test - An assessment of a HRQOL measure includes the measurement of validity. The broad definition of validity is "that an instrument measures what it is supposed to measure and nothing else" (Schulzer et al., 1999). Validation should be viewed as a continuous process that begins with the first validation study and continues with repeated use of the instrument under different conditions e.g. a cross sectional community based versus an outpatient longitudinal study. The more frequently the instrument is used, the wider the circumstances it is used in and the more often it performs as expected, then the greater the confidence in it's validity. A HRQOL instrument must demonstrate reasonable validity before it can be used in clinical situations.

The validity of an instrument is generally tested by means of correlation analysis against a variety of external measures and instruments. This technique identifies groups of items that are highly correlated with each other but exhibit low levels of agreement with other variables in the measure. There are, however, many different types of validity: criterion, content and construct are the most frequently described. In this research, construct validity of the UK-SIP will be assessed against the PDQ-39; the construct validity of the UK-SIP will be assessed against the SF-36; and the construct validity of the PDQ-39 will be assessed against the PIMS. Construct validity must be investigated whenever no criterion is accepted as entirely adequate to define the quality to be measured (Cronbach, 1955). Specifically, construct validity focuses on the assessment of whether a particular measure relates to other measures consistent with theoretically derived hypotheses concerning the concepts (or constructs) that are being measured (Zeller and Carmines, 1980). Construct validity is the main tool for the validation of quality of life questionnaires. During construct validation, evidence must be provided that demonstrates that the measured construct behaves consistently with its referent theoretical properties. The following aspects will be considered when establishing construct validity for a HRQOL measure.

1. The measured variable should correlate at a low level with measures of related but different constructs (divergent validity)

2. The measured variable should correlate positively with related and similar constructs (convergent validity)

3. The measured variable should draw information from the intended HRQOL construct

4. The measured variable should not be dominated by irrelevant factors

While strong correlations with a number of external measures support one's confidence about the construct validity of an instrument, the appropriate inference from negative evidence is more controversial. One may conclude that the measure lacks construct validity, that is, it does not measure what it purports to measure.

Various statistical tests will be used to explore these relationships including Mann Whitney U, Kruskall Wallis and Spearman's Rank. The relationship of the instruments of choice scores with scores obtained using other instruments will be examined. Multiple Correlations between the instruments will also be calculated.

Sensitivity test - The difference in HRQOL scores at baseline and six months later will be compared to see whether there is a statistically significant difference. Wilcoxon's method will be used for this as the samples are related and the data is non-parametric. Sensitivity or responsiveness is a type of validity that incorporates longitudinal information (Guyatt et al., 1992) but is mostly omitted in the testing of HRQoL instruments. Evidence of responsiveness can be established by comparing HRQoL data collected before and after a clinical intervention. This will not be possible in this research as patients will continue on their existing treatment regimen. Nevertheless, responsiveness can also be determined by examining the association between HRQoL data and an assessment of change in overall health state. A HRQoL instrument should demonstrate sensitivity to clinically significant change before it is used in a clinical trial setting. The test for responsiveness to change is carried out using two different tests. The paired t-tests compare scores at baseline and follow up for a particular group that has demonstrated change or lack of change. The unpaired t-tests are used to compare mean score differences between different

groups, e.g. between groups showing different degrees of change or between a group showing change and another group showing no change. In this study, the non-parametric equivalent of the paired samples t-test will be used; the Wilcoxon test.

Missing Data - Missing data must be dealt with appropriately. There are various types of missing data. If the data is missing for reasons unrelated to their HRQOL then the data can be said to be "missing at random". This type of missing data can make statistical analysis more complex but HRQOL estimates remain unbiased even if based only on observed HRQOL assessments. Removing entire cases where items are missing results in a loss of power and therefore where only one or two items are missing, the remaining data is included in the analyses. However, where a large number of items are missing, exclusion of the entire case should be considered. There are methods of calculating replacement values for missing data. For example, an average of the observed items can be used or the missing value can be regressed against the observed values.

Non-randomly missing data creates a much greater problem. An example would be where an assessment is missing due to disease progression or death. In these cases HRQOL would be expected to be generally poorer and therefore analyses that inappropriately assume that data is randomly missing will result in biased estimates of HRQOL. Such estimates will only reflect the group of subjects actually assessed rather than the entire population under study. One could limit the analyses to subjects with complete data but this may not be acceptable where it is necessary to compare the HRQOL assessments for all subjects. Careful selection of a method for imputing missing items for an individual who has answered most questions would in general, be preferable to deletion of the entire case or observation (Spilker, 1996). A simple method based solely on the patient's own data would use the mean of all non-missing items for the entire scale or the specific subscale. Methods based on other patients would include the mean of that item in individuals who had responded.

The introduction of possible bias should be avoided and therefore where possible, the reasons for missing assessments should be explored in order to determine whether they are related to the subject's condition or HRQOL and hence determine how the data can be interpreted. Sometimes it is not possible to determine why

assessments are missing and in these situations the possible reasons should be stated and considered when interpreting the results of subsequent analyses.

Non-parametric tests will be employed if the data generated from the questionnaires fail to meet the fundamental assumptions required for parametric analysis such as randomisation, normality and small sample numbers (Salek, 1997, Siegal and Castellan, 1988). For cross sectional evaluation within different members of the same study group, analytical techniques suitable for independent samples will be used. For example, Mann-Whitney U test to examine whether, according to chance, two independent samples have been drawn from the same population.

DESIGN OF STUDY

PART I. COMPARING A GENERIC AND DISEASE SPECIFIC QUALITY OF LIFE MEASURES

The study is aimed at evaluating the measurement properties of an established generic instrument, the UKSIP, and comparing it against a recently introduced disease specific instrument, the PDQ-39, to determine the most suitable quality of life measure for Parkinson's disease. The study will provide the researcher an insight as to the condition of patients with PD and the areas which affect them most due to the disease.

The study will be a cross-sectional postal survey of patients from the UK Parkinson's Disease Society (UKPDS). The research protocol will require an approval from a central ethics committee prior to the commencement of the study. Patients will be recruited from members of the UKPDS from the south of England and Wales region with the aim of recruiting 50 patients into the study. Patients will be considered to have PD if they are taking concurrent PD medication. The researcher will approach the selected branch members of the UKPDS and explain the nature of the study and what is required from the patients' participation. A letter inviting members who are classified as PD patients will be sent inviting their participation and if the patient agrees he or she will be asked to sign a written informed consent. Each patient will be asked to complete the quality of life questionnaires on two occasions, at an interval between 5 and 7 days. These questionnaires will be posted to the patients

with a personalised covering letter explaining what is required of them. It is hypothesised that a disease-specific measure such as the PDQ-39 will exhibit superior psychometric properties compared to the generic UKSIP as it is specifically addressing areas of health of direct relevance to the patient. The hypothesis will further test whether the UKSIP is able to identify with area or domains of quality of life considered to be relevant to the PD patient and whether it is able to identify items not considered by the PDQ-39.

PART II. COMPARING TWO GENERIC QUALITY OF LIFE MEASURES

Data from Part I will be closely examined and if the UKSIP is deemed applicable in PD it will be compared against another extensively used generic instrument, the SF-36 which has been shown to have construct validity in PD when compared against the PDQ-39. However there is no definitive answer to which instrument would be the preferred generic instrument of choice in PD as there has been no comparison between the UKSIP and SF-36.

A longitudinal assessment of the SF-36 has not been undertaken to assess the SF-36 and UKSIP's responsiveness to clinical change with patients over time. The SF-36 and UK-SIP are widely used generic Health-related Quality of Life (HRQoL) instruments but it is yet to be determined which measure is more suited to the evaluation of PD patients.

The study will be a longitudinal, prospective and single centre design. The research protocol will require an approval from the local ethics committee prior to the start of the study. The study will be carried out at an outpatient neurology clinic in Ashford, Middlesex with the aim of recruiting 50 patients into the study. Patients who meet the inclusion / exclusion criteria will be consecutively selected by the examining physician from the regular list of the outpatient clinic. The researcher will approach the selected patients and explain the nature of the study and what is required from their participation. If the patient agrees he or she will be asked to sign a written informed consent. Each patient will be asked to complete the quality of life questionnaires a few days prior to their outpatient clinic visit and then on the day of the visit the physician will assess disease severity using the Hoehn and Yahr scale, a measure of disease severity. The patient will be asked to complete the same questionnaires at

their next scheduled outpatient visit, a period no longer than 6 months. The ability of the two generic instruments to assess the impact of PD on HRQoL will be tested by comparing against available population norms for the instrument. The validation of the defined construct will be tested using a correlation coefficient between the identified quality of life domains and the overall quality of life scores. The validity of the selected questionnaires will be tested against another measure of the same or related construct. The study is aimed at comparing the construction of the selected questionnaires with other questionnaires of a related or unrelated nature. Spearman rank correlation coefficients will be used to measure the association between Hoehn & Yahr and the domains of the two generic measures, which is part of the validation process of the generic measures in the PD population.

It is anticipated that we will see a stronger correlation between the SF-36 and UK SIP (convergent validity) than between either of these HRQoL measures and the clinical Hoehn & Yahr scale (divergent validity). If construct validity is established it will confirm that the questionnaires behave in the manner hypothesised. The internal consistency reliability of identified items within each domain will also be tested. The sensitivity or responsiveness of the two instruments will be tested by comparing baseline and follow-up scores in patients who report a change in their condition as follows:

- The test for responsiveness to change will compare scores at baseline and follow-up. Differences between mean t1 (baseline) and mean t2 (follow-up) domain scores will be assessed. This approach measures the change in subjects who demonstrate a true change in HRQoL using paired Wilcoxon- tests for non-normally distributed variables to compare baseline and follow-up scores. This will provide an indication as to whether the instruments evaluated can detect change in scores of patients whose HRQoL changes over time.

- A responsiveness analysis using data from those subjects who report their HRQoL has changed from baseline, at t2 assessment will be measured. Paired Wilcoxon tests for non-normally distributed variables will be calculated to assess the significance of changes seen in HRQoL between the two assessments. The change scores for the two instruments over 6 months will be compared with all patients who report that their condition had changed and are either "**a little**

worse" or "**much worse**". All patients will complete a retrospective judgement of change questionnaire which asks "Has there been a change in the effects of your PD on your everyday life since the previous visit"; with the following answers: *much better, better, the same, a little worse, much worse*. The Wilcoxon-statistic describes the relative magnitude of the changes seen in the domains and will be taken as a comparison of the responsiveness of the questionnaire domains. The Wilcoxon-statistics for patients classed as stable will also be presented for comparative purposes. It is useful to determine whether the measure can distinguish between those subjects who remain stable and those who demonstrate a change in HRQoL using an unpaired Wilcoxon-test.

- The significance of change scores amongst patients whose symptoms worsen according to the Hoehn & Yahr assessment will be examined by means of paired t-test and standardised response mean. The standardised response mean (SRM) is calculated as the mean change in score divided by the standard deviation of individuals' change scores and provided a standardised and unit free index of sensitivity to change (Liang et al., 1990). A SRM of =0.2 is considered small, =0.5 is moderate and =0.8 is large(Timmerman et al., 2003). Those patients whose severity of symptoms worsens at the follow-up visit will be reflected by change to instrument scores indicating a worsening of HRQoL.

- The change scores for the UKSIP will be examined in relation to patterns of change in patients' SF-36 physical component, and mental component and scores between baseline and follow-up using Spearman's non-parametric correlation coefficients. This will tell us whether any changes to patients HRQoL score in one instrument is reflected by changes to the HRQoL scores of the second HRQoL instrument

- The change scores for the UKSIP and SF-36 will be examined in relation to the change scores for the clinician's ratings on the Hoehn & Yahr. This will tell us whether any changes to patients HRQoL score in one instrument is reflected by changes to the Hoehn and Yahr scores.

An important facet of any HRQoL instrument is its ability to discriminate between patients with different levels for severity (Price, 1997). The Kruskall-Wallis one-way

ANOVA will be used to test whether or not the questionnaires are able to differentiate between the categories of disease severity.

It is hypothesised that a shorter measure such as the SF-36 will exhibit lower discriminant validity as it will be compromised by the loss of items compared to the longer UKSIP. The UKSIP is a much longer instrument and may not be suitable for use in the elderly population and it is possible that higher missing items will be reflected by the UKSIP. Sensitivity to change is not routinely addressed in HRQoL research and this has not been established for either generic instrument in PD. It is hypothesised that the UKSIP will be more sensitive to change compared to the SF-36 as it contains a higher number of items within each domain and also each item is weighted on the basis of the relative severity of each disability.

PART III. **COMPARING TWO DISEASE-SPECIFIC QUALITY OF LIFE MEASURES**

Further to the introduction of the PDQ-39, an additional disease-specific instrument, the PIMS has been reported but has not undergone the same level of validation as the PDQ-39. It is a short instrument, consisting of ten items which were selected by Parkinson's disease nurses in Canada. The PDQ-39 on the other hand consists of 39 items which were arrived at by item selection after interview with PD patients and neurologists experienced in the treatment of PD patients. These instruments will be tested for their construct validity. This will establish the research grounds to confirm that the selected questionnaire behaves in the manner hypothesised, before it will be used to interpret the results in a clinical study. The validity of the selected questionnaires will be tested against other measures of the same or related construct. The study is aimed at comparing the construct of the selected questionnaires with other questionnaires of a related and unrelated nature. The study will be longitudinal, prospective design and will recruit a minimum of 50 patients from a movement disorder outpatient clinic at Romford Hospital Essex, for the comparisons. Patients will be asked to complete the questionnaires prior to their scheduled outpatient clinic appointment and then to repeat the questionnaires at their next scheduled outpatient appointment, six months later.

Though reliability and "partial" construct validity of the PIMS has been reported by the authors (Calne et al., 1996) convergent and discriminant validity have not been

reported to date. The validity of an instrument is the extent to which it measures what it claims to measure. Calne et al, 1996, validated the PIMS against the "*best external criterion for this construct*", (viz) the degree of severity of the symptoms. Calnes's patients were classified as non-fluctuators (arbitrarily assigned as having milder symptoms) and fluctuators (arbitrarily assigned as having moderate/severe symptoms). In this study construct validity will be assessed by comparing the results of the PIMS factors and global score with clinical assessments of health, i.e. the Hoehn & Yahr and UPDRS (motor subscale). Construct validity is a process by which empirical evidence from several validation procedures is assembled to support the inference that a particular instrument measures what it purports to measure (Salek and Luscombe, 1992). One often has to rely on construct validation where the best available alternative measure of health status is used. The classification of patients as non-fluctuators and fluctuators cannot be considered as the best available alternative measure of health status due to the arbitrary nature of the classification as no formal assessment was conducted to confirm the classification. The best external criterion should be a well recognised external variable. The severity of symptoms of Parkinson's disease patients can be better classified by their clinical staging, using the Hoehn & Yahr scale (Hoehn and Yahr, 1967), or by assessing the level of motor impairment as measured by the UPDRS scale (Fahn et al., 1987). In this study, construct validity will be measured against the two clinical scales which can be considered as the best available alternative measure of health status, the UPDRS (Fahn et al., 1987) and Hoehn & Yahr Scale (Hoehn and Yahr, 1967). Convergent validity of the PIMS will be assessed by correlations against a related and similar construct, the PDQ-39 (an established disease specific health care measure). Correlation coefficients will be calculated to assess linear relationships between the PIMS and the PDQ-39 and the two clinical scales. Mainly, high correlations are expected between selected questionnaire and variables of others of the same construct (convergent validity) and low correlation are expected for unrelated measures (divergent validity). Construct validity will also be tested against questionnaires of an unrelated nature where divergent validity is expected and these unrelated constructs will be used to test the discriminant validity of the disease specific instruments. It is hypothesised that the shorter instrument will have lower discriminant validity and miss items of importance to the PD patient compared to the longer instrument, particularly as the shorter PIMS did not identify items based on interview with PD patients.

Discriminant validity of the PIMS, that is, the extent to which the questionnaire scores are able to discriminate between subgroups of patients differing in clinical state, will be assessed by the method of known groups comparison. Kruskal-Wallis one-way ANOVA will be employed to test for the statistical significance of differences in PIMS scores between different categories of the Hoehn & Yahr, UPDRS and Beck Depression Inventory (Beck et al., 1961) scales.

PART IV. QUALITY OF LIFE ASSESSMENT IN CLINICAL TRIAL SETTING

Depending on the outcome of Part III of this research one disease-specific instrument will be selected to test its implementation in a clinical trial. The study is designed to to determine the difference in the levels of quality of life between dyskinetic PD patients versus non-dyskinetic patients at baseline and 6 months later and assess the impact on caregiver burden, a topic which is often neglected with respect to the PD population. The study will also evaluate the sensitivity of the chosen disease - specific instrument and evaluate the value of the quality of life information in a clinical trial setting and determine the impact of quality of life in this distinct cohort of PD patients.

The study will be a longitudinal, prospective multi-centre design and will be carried out at four outpatient neurology centers located in the USA. Consecutive patients, satisfying an inclusion / exclusion criteria, attending the clinics during a one-year period will be recruited into the study. The quality of life assessment using the selected questionnaire will be integrated into the clinical design of the study which is primarily aimed at evaluating the pharmacoeconomic aspects of treating patients with fluctuating disease. At every clinic, patients will be asked to complete the questionnaire at the waiting room before seeing their doctors and then proceed to be assessed for disease severity and functioning based on routine clinical measures. Patients will be classified as fluctuating or non-fluctuating based on their clinical assessments. The patient's caregiver will be provided with a Burden interview questionnaire to assess the level of impact the disease has on the carer and to determine whether one group experiences greater burden.

Summary statistics for each visit will be determined for each group in the study. Correlations of quality of life data with clinical variables will be used to determine

clinically meaningful differences in quality of life scores between groups (fluctuating and non-fluctuating). Paired t-statistics will be calculated to assess the significance of changes seen in HRQoL between baseline and follow-up.

It is hypothesised that patients with fluctuating disease will experience poor HRQoL compared to non-fluctuating patients and that caregivers of fluctuating patients will report higher levels of burden compared to non-fluctuating patients.

Quality of life instruments

The questionnaires selected were based on literature review and availability of a multidimensional disease-specific quality of life instrument. The PDQ-39 has recently been developed and undergone extensive testing including criterion validity and reliability. It is currently the only disease-specific measure that has undergone construct validity and comparison against the SF-36 has been reported. The UKSIP is an established generic instrument which has been used in a variety of disease conditions and its use has been reported in one cross sectional study in PD. To date no comparison of these two different instruments has been conducted. The internal consistency reliability and test-retest reliability of identified items within each domain will be tested as will convergent validity. The SF-36 has been compared to the PDQ-39 and is used extensively in HRQoL research. The SF-36 has been used as an outcome measure in a range of disease states which include, multiple sclerosis (Riazi et al., 2003), stroke (Dorman et al., 1998) and Parkinson's disease (Jenkinson, 1995). In addition, the SF-36 has performed well in cross-sectional population studies when compared with other generic instruments. Brazier et al, 1996 reported that the SF-36 questionnaire was a more sensitive instrument than the EuroQoL, especially at lower levels of health. The SF-36 is a shorter form in comparison to the UKSIP, however it is reported to be unsuitable for use in the elderly population (Hayes et al., 1995) and hence it's inclusion in this research was mandated to assess it's applicability in an elderly PD population and to compare against the UKSIP. Lastly, the PIMS is the newest of the disease-specific measures which is undergoing continues validation. It conceptually differs from the PDQ-39, is a much shorter instrument and addresses domains which are not included in the PDQ-39. A more detailed outline of these questionnaires are described below.

The Parkinson's Disease Questionnaire (PDQ-39)

The PDQ-39 consists of 39 items covering eight dimensions of health, which are separately scored; *Mobility*, *Activities of daily living* (ADL), *Emotional well-being*, *Stigma*, *Social support*, *Cognitions*, *Communication* and *Bodily discomfort* (see Appendix I). Dimension scores are coded on a scale of 0 (perfect health) to 100 (worst health) (Peto, 2001). It is possible to derive summary scales from the instrument, the PDQ summary index. Using such summary scales makes it possible to reduce the number of statistical comparisons and therefore reduce the likelihood of obtaining statistically significant differences due to chance. A users manual including the survey, its scoring algorithm and administration procedures is available from the developers (Jenkinson et al., 1998). Each of the eight scales is based on the sum of the responses to individual items within the scale and is then linearly transformed to a 0 to 100 scale. Patients respond to each question on a 5-point scale: never (0), occasionally (1), sometimes (2), often (3), and always (4).

The United Kingdom Sickness Impact Profile (UKSIP)

The Sickness Impact Profile (SIP) was designed as a measure of sickness-related behaviour for use in resource allocation, policy formation, evaluation of therapeutic outcomes and disease severity, and for use in producing population health surveys (McDowell and Newall, 1987) The UKSIP (see Appendix II) consists of 136 statements which are grouped into twelve scoring domains; *body care and movement (BCM), mobility, ambulation, emotional behaviour, social interaction, alertness behaviour, communication, sleep and rest, home management, work, recreation and pastimes and eating* (Bergner et al., 1981). Subjects are requested to read all the statements and tick only those which are applicable on the day of completion. Those items, which are deemed to be not applicable to the subject, are left un-ticked. Each of the 136 statements carries a preference weighting based on perceived importance. Each item is weighted according to the relative severity of the condition within its own category. Hence, the SIP scores are directly interpretable as reflecting the subjective evaluations of the relative degree of disability as judged by patients and clinicians and therefore greater sensitivity to treatment differences may result and the discriminative properties may be higher compared to non-weighted scales. A profile of 12 domain scores is produced and further aggregation permits calculation of *psychosocial* and *physical* dimension scores as well as an overall SIP summary score. The physical dimension score is obtained from the summation of *body care*

and movement, ambulation and *mobility*, whilst the psychosocial dimension is scored from *emotional behaviour, social interaction, alertness behaviour* and *communication*. The remaining five categories are scored independently. Questionnaire scores are presented as percentages, with 0% indicating best possible HRQOL and 100% the worst. The overall summary score can be obtained by summing the scale values of all items endorsed in the entire UKSIP, dividing that sum by the sum of the values of all the items in the UKSIP and multiplying the obtained quotient by 100. Scores for each category are calculated in a like manner (Bergner et al., 1981). The UKSIP provides an overall health question which was not present in the original SIP and this allows one to assess a change in overall health over time as perceived by the patient. The SIP has been used in a wide range of disease states, including chronic lower back pain (Follick et al., 1985), chronic pain (Jensen et al., 1992) angina pectoris (Visser et al., 1994) and Chronic Fatigue Syndrome (Anderson and Ferrans, 1997).

In this study, an Anglicised (UK) version of the SIP will be used. It was developed to allow for cultural and linguistic differences between the UK and the United States of America, where the instrument was developed. It includes a number of minor changes to the original questionnaire, and has been fully validated in a general population, as well as having been used in several disease states such as arthritis and psoriasis (Salek, 1992).

The Short-Form-36 health status questionnaire (SF-36)

The SF-36 (see Appendix III & IV) was developed following research from the Rand Corporation's health insurance experiment in the USA, which was also known as the Medical Outcomes Study (MOS). The SF-36 measures subjective health status and consists of 36 statements grouped into eight scales (domains) for scoring purposes: physical functioning (PF, 10 items), social functioning (SF, 2 items), role limitations due to physical health (RP, 4 items), role limitations because of emotional difficulties (RE, 3 items), mental health (MH, 5 items), vitality (VT, 4 items), bodily pain (BP, 2 items), and the general perception of health (GH, 5 items). In addition there is one question which asks the subject about how their health has changed in the past year, although this does not form part of the profile of domain scores. The SF-36 produces a profile of eight domains, which are all scored on a percentage scale of 0 to 100, with a score of 100 indicating the best possible functioning on each of the scales. Patients respond to items using a Likert scale. The number of response options

available varies from 2 to 6 between each section, and each domain is scored using algorithms re-produced in technical scoring manuals available from the authors (Ware and Sherbourne, 1992). None of the items in the SF-36 are weighted, except question 1 which enquires about patients' general overall health. The remainder of the items are scored on a simple scale. One of the limitations of the SF-36 is that it is not possible to compute an overall summary score. Two summary scores characterising the physical (PCS) and mental health (MCS) components are also available (Ware et al., 1994). Although, as previously stated, no overall summary score can be calculated, algorithms are now available from the authors (Ware and Sherbourne, 1992) which permits both the PCS and MCS to be calculated but their application may not be widespread. The SF-36 has performed well in cross-sectional population studies when compared with other generic instruments. Brazier et al 1996 reported that the SF-36 questionnaire was a more sensitive instrument than the EuroQoL, especially at lower levels of health.

The Parkinson's Impact Scale (PIMS)

The PIMS (Calne et al., 1996) has 10 items (*work, travel, family, community, sexuality, self, leisure, feelings, safety and financial security*) which are broadly formulated and concern domains rather than specific situations (Marinus et al., 2002). Higher scores reflect lower HRQoL. The PIMS (see Appendix V) contains four distinct, important and commonly identified factors:

- Physical (work & travel)
- Social (family, community, & sexuality)
- Psychological (patient's perception of self, leisure, feelings and safety)
- Economic (financial security)

Stable patients only score each item once, whereas patients with fluctuations judge the negative impact for both "on" and "off" periods. The scale can be self completed, but the developers recommend that patients be advised with respect to their disease state (stable or fluctuating). A global score is derived from assigning a weighting of each of the item scores providing a score in the range of 0 to 40 with 0 indicating no change and 40 indicating severe impairment.

Mini Mental State Examination (MMSE)

The MMSE (Folstein et al., 1975) is probably the most widely used measure of cognitive function. The MMSE is a quantities measure of cognitive impairment but does not include aspects of mood, abnormal mental experiences or disordered forms of thinking. The MMSE has a maximum score of 30 points, with different domains assessed: orientation to time and place (10 points), registration of three words (3 points), attention and calculation (5 points), recall of three words (3 points), language (8 points) and visual construction (1 point). A score of < 20 is indicative of cognitive impairment and a score of =15 is indicative of severe dementia.

Hoehn & Yahr Scale

The Hoehn & Yahr scale (Hoehn and Yahr, 1967) determines the stage of PD subjects. It is a five level scale used to stage symptoms of the disease based grossly upon impairments (tremor, rigidity) and disability (ability to stand or walk unassisted vs wheelchair dependant) (Koplas et al., 1999). Individuals are placed into one of five stages based upon the relative level of clinical disability: (a) unilateral involvement with minimal or no functional impairment; (b) bilateral involvement without balance impairment; (c) moderate impairment of posture; (d) severely disabling disease but patient can walk and stand unassisted; and (e) confinement to wheelchair and bed e.g.

- Stage 0 = no visible disease;
- Stage I (1) = disease that involves only one side of the body;
- Stage II (2) = disease that involves both sides of the body. but does not impair balance;
- Stage III (3) = disease that impairs balance or walking;
- Stage IV (4) = disease that markedly impairs balance or walking: and
- Stage V (5) = disease that results in complete immobility.

Stages 0-II are mild disease; Stage III is moderate disease; Stages IV and V are marked or advanced disease. There are grey areas between the successive stages. Hoehn and Yahr staging does not necessarily progress from one stage to the next but may jump more than one stage higher between assessments or visits. The Hoehn and Yahr (see Appendix VI) staging nevertheless reflects the most

appropriate global score for each patient (Hely et al., 1999) though it does not measure functional disability.

Unified Parkinson's disease Rating Scale (UPDRS)

The UPDRS (Fahn et al., 1987) (see Appendix VII) was developed to provide clinical investigators with a detailed assessment of specific areas of disability as well as a global measure of Parkinson's disease. It is an assessment of severity of specific signs and symptoms of parkinsonism. It has successfully met criteria for inter-rater reliability and validity. The UPDRS is composed of 42 items that includes four sections (I. mentation, behaviour and mood; II. activities of daily living; III. motor examination and IV. complications of therapy). The motor examination section records the observations of the examiner for a single point in time. It consists of 14 items which can be assigned a score from 0 (normal) to 4 (severe) for evaluating bradykinesia, tremor, rigidity, speech, facial expression, postural stability and gait. The UPDRS is designed to assess the severity of the cardinal motor findings in patients with Parkinson's disease (e.g., tremor, rigidity, bradykinesia, postural instability, etc.), scored for different body regions, and has a maximum (worst) score of 108.

Beck Depression Inventory (BDI)

The BDI (Beck et al., 1961) is a 21 item instrument designed to assess the severity of depression in adolescents and adults. It assesses behavioural manifestations of depression rather than symptomatology alone. The subject chooses the statement that most accurately describes his or her feelings for 21 items reflecting typical symptoms of depression including appetite, mood, and sleeping patterns. Each question is rated on a 4 point scale ranging from 0 to 3 in term of severity. A maximum score of 63 can be derived indicating severe depression. 0-9 is not depressed; 10-16 is mildly depressed; 17-29 is moderately depressed; and 30-63 corresponds to severely depressed (Koplas et al., 1999).

Dyskinesia Subjective Rating Scale

This scale, developed by Lang & Fahn (Lang and Fahn, 1999) (see Appendix VIII) was used to assess the impact of dyskinesias on five domains of activities of daily living that are generally affected by dyskinesias:

- Handwriting/drawing
- Cutting food and handling utensils
- Dressing
- Hygiene
- Walking

Each item is scored from 0 to 4 with a maximum score of 20 - higher scores indicating greater impact on ADL as a result of dyskinesia.

Carer Burden Interview

The Burden Interview (see Appendix IX) was administered via the investigator, or his/her designee, to the carer, at each visit. The Burden Interview (Zarit et al., 1987, Zarit et al., 1986) has been designed to assess the stresses experienced by family carers of elderly and disabled persons. It can be completed by carers themselves or as part of an interview. Carers are asked to respond to a series of 22 questions about the impact of the patient's disabilities on their life. For each item, carers indicate how often they have felt that way on a 5-point scale: never (0), rarely (1), sometimes (2), quite frequently (3), and nearly always (4).

The Burden Interview has been developed as a composite measure. As such, it combines different aspects or dimensions of carers' reactions to their involvement. It is useful when a single summary measure of carers' appraisal of the impact their involvement has had on their lives is needed. The psychometric validity of the Burden Interview has been documented (Anthony-Bergstone et al., 1988). An overall measure of burden is obtained by summing the responses to all 22 items. Possible scores range from 0 to 88. Higher scores indicate greater carer burden. Norms or cut-points for the Burden Interview have not been established.

POTENTIAL BENEFITS OF THE STUDY

The study is intended to bridge the knowledge gap between researchers and clinicians and enable them to work together in assessing the benefits of the quality of life measures in a clinical setting and specifically a clinical trial setting. The major perceived problem in the absence of a HRQoL measure in clinical practice or trials is the proliferation of instruments available or being developed which makes the choice

of a suitable instrument a daunting one for the healthcare practitioner. The results from the study should provide us with valuable information on the appropriate quality of life instrument for PD patients in a variety of settings. The validation of the tested instruments ought to provide a certain degree of confidence when interpreting the results and its later application in either routine practice or in clinical trials. In addition, the quality of life information generated from this study will establish a basis for the clinician, nurse or healthcare provider involved in patient care, to understand better and gain a more complete picture of their patients' well-being. Not only will the selection of the most appropriate instrument on the basis of its psychometric properties and applicability enable the clinician to evaluate the impact of their care in improving quality of life but it will also encourage the use of such instruments in clinical trials to differentiate similar treatment options on the basis of patient perceived differences.

Chapter 3

Reliability and validity of the Parkinson's Disease Questionnaire (PDQ-39) and the United Kingdom Sickness Impact Profile (UKSIP) in patients with Parkinson's disease (PD)

INTRODUCTION

A large number of disease specific clinical measures exist designed to characterise the impact of disease but there are relatively few which can provide a complete picture of the impact of the disease upon patients. Most importantly they fail to address the impact of the illness upon subjectively assessed function and well being of patients (Longstreth, 1992). An instrument to address areas salient to Parkinson's disease has been developed, the Parkinson's Disease Questionnaire (PDQ-39) (Jenkinson et al., 1997, Peto et al., 1998).

Validation of the PDQ-39 has been reported by the authors (Jenkinson, 1995) but now requires confirmation by independent researchers. As part of the validation process the PDQ-39 was compared against a well established generic health status measure, the United Kingdom Sickness Impact Profile (UKSIP – a UK version of the Sickness Impact Profile, SIP) (Salek, 1992, Salek, 1997). The UKSIP to date has not been assessed in the PD population or compared against the PDQ-39. Longstreth et al (Longstreth et al., 1992) showed that the SIP had good construct validity correlating with the Hoehn and Yahr and Columbia Rating Scale (clinical measures of PD). Internal consistency reliability and test-retest reliability were not assessed. The UKSIP has not been compared against the PDQ-39 as measures of health-related quality of life in Parkinson's patients.

In this chapter it is hoped to evaluate the measurement properties of a generic and disease-specific instrument, the UKSIP and PDQ-39 in patients with PD. Furthermore, to establish whether the UKSIP was a suitable instrument to measure health related quality of life (HRQoL) in this population.

METHODS

Both the UKSIP and PDQ-39 are self-administered questionnaires. The development and validation of the PDQ-39 has been documented (Peto et al., 1998b) and the instrument has been used widely in outcome evaluations and surveys (Gray, 2002, Harrison, 2000).

The questionnaires were administered in a postal survey designed to assess the acceptability of the UKSIP and the PDQ-39 to patients with PD, to assess which aspects of PD impacts the most and to determine the internal consistency and test-re-test reliability of the measures for this patient group.

The UKSIP and PDQ-39 were sent in a questionnaire booklet to a sample of patients with PD who had been identified from membership of local branches of the United Kingdom Parkinson's Disease Society (UKPDS) in eight areas of the UK. The UKSIP and PDQ-39 questionnaires were sent by post to a sample of 106 patients satisfying a diagnosis of idiopathic Parkinson's disease and receiving anti-Parkinsonian treatment. Patients were provided with a letter outlining the purpose of the research, what was involved and whom to contact for further information. Each participant was asked to complete consent form; the PDQ-39 and UKSIP questionnaires consecutively at the same sitting and the order in which the PDQ-39 and UKSIP were presented was randomised to avoid bias.

To measure the degree of reproducibility of the two measures the questionnaires were administered on two occasions (test-retest). Test-retest reliability is a measure of the consistency and stability of the instrument and is especially important for a self-administered instrument (Salek and Luscombe, 1992). The methodology of test-retest reliability rests on the assumption that each subject's true score has not changed over the time interval between the two occasions of measurement, so that any score differences are due purely to measurement error. It also assumes that the repeated score values are not too highly correlated due to proximity in time, and that the subject's memory of the responses on the previous occasion is not playing a part in formulating the new set of responses on the following occasion. To reduce the occurrence of overestimation or underestimation, a one week interval is recommended (Salek and Luscombe, 1992). These questionnaires were enclosed in an envelope labelled "Pack 1" and once completed by the patients were posted to the researcher in a stamped addressed envelope (Test 1). Four to seven days later, a second envelope labelled "Pack 2", containing both the PDQ-39 and UKSIP questionnaires were completed and posted to the researcher in a stamped addressed envelope (Test 2). In Test 2, in addition to the two health status instruments, patients were asked a transition question (Fitzpatrick 1993): 'Has your condition changed in any way over the last seven days, for example, have you felt

poorly, had a change of medication, had to attend for an unscheduled doctors appointment, been hospitalised?' Response categories were: 'No', or 'Yes'. All patients were asked to rate their present overall health using a global rating scale (i.e. very good, good, fair, poor, very poor) at both time points and an evaluation of correlation between the global rating scale and quality of life score were made for each time point i.e. Test 1 and Test 2.

Data Processing and Analysis

Non-parametric correlation test, Spearman rank correlation coefficient was used to examine validity (convergent and divergent) and test-retest reliability. A score of 1 indicates perfect positive association, a score of −1 indicates perfect negative association and a score of 0 indicates no association. Also, Cronbach's a correlation was carried out to determine the internal consistency reliability of the PDQ-39 and the UKSIP. Because of the multitude of comparisons, only differences with $p < 0.005$ were considered significant.

Please refer to Chapter two for more details.

Data from completed questionnaires and clinical findings were checked for completeness and accuracy. The data was then coded for computer analysis and entered onto a database and analysed using a statistical package for social scientists (SPSS version 10 for Windows).

RESULTS

Patient Characteristics

A total of 91 Test 1 and Test 2 PDQ-39 questionnaires were returned, and a total of 88 Test 1 UKSIP questionnaires and 86 Test 2 questionnaires were completed. Fifty-six patients remained 'unchanged' after 1 week and their data was analysed for test-retest reliability.

The mean age of the cohort was 70.5; range 49-89 and the gender breakdown was 70.3% male and 29.7% female. The mean disease duration was 11 years with a range of 1-42 years (Table 3.1). The majority of the patients were permanently

retired (83.5%) which reflects the age group of the cohort who were mostly over the age of retirement (87.5% were over the age of 60 years old). Education level was reported by 59.3% of the patients and from this sample, 26.4% had received school education, and 33% had received either college or university education (Table 3.1).

Table 3.1 Demographic characteristics (n=91).

Age - years	
Mean (Range)	**70.5 (49 – 89)**
Gender	
Male (%)	**64.0 (70.3)**
Female (%)	**27.0 (29.7)**
Disease duration - years	
Mean (SD)	**11.0 (7.4)**
Median (Range)	**8.0 (1 – 42)**
Educational level (%)	
School – e.g. 'O levels'	**26.4**
College – e.g. HND, City & Guilds	**16.5**
University – e.g. BSc, BA	**16.5**
Total	**59.3**

The PDQ-39 Scores

The PDQ-39 scale score at Test 1 revealed that impairment was greatest for mobility (mean = 52.4), activities of daily living (mean = 48.3) and bodily discomfort (mean = 47.0) and to a lesser extent for stigma (mean = 25.0) and social support (mean = 13.7) (Table 3.2 & Figure 3.1). Similarly, the PDQ-39 scale score at Test 2 reveal ed that impairment was greatest for mobility (mean = 49.6), activities of daily living (mean = 47.1) and bodily discomfort (mean = 44.9) and to a lesser extent for stigma (mean = 22.2) and social support (mean = 14.2) (Table 3.3 & Figure 3.2). The main difference this time was that emotional well being (mean = 47.1) shows the second joint highest level of impairment with activities of daily living.

These results suggest that PD patients experience physical functional impairment to a greater extent than psychosocial impairment. Alternatively, the data might suggest that the PDQ-39 is particularly sensitive to physical impairment in the PD population. If the PDQ-39 is more sensitive to physical domains of the disease then this may be assessed by evaluating the relationship between the PDQ-39 and a clinical measure

of PD such as the Hoehn and Yahr scale, which assesses the severity of the disease and in particular physical symptoms. Such an assessment with a clinical indicator helps demonstrate the questionnaires construct validity in PD patients. Hoehn & Yahr assessments were not performed in this study but instead the correlation between global health score as measured by the patient and PDQ-39 was performed and will be discussed later in this section.

At the significance level $p<0.05$ there was a difference detected between males and females with respect to emotional well being, cognition and communication. However, because so many tests are being carried out, the significance level is reduced to prevent problems associated with multiple testing thus reducing the significance level to 0.5% (i.e. $p<0.005$). In light of this adjustment none of the scales were statistically significant between the two groups. Differences with $p<0.05$ can be considered a trend and as such a comment can be made to suggest that emotional well being, cognition and communication affect females more so than males. In only one previous study (Kuopio et al., 2000) was there a significant difference between the sexes where women scored significantly lower on five of the eight dimensions of the Medical Outcomes Study 36-Item Short Form Health Survey (SF-36). Depression was found to be more common among parkinsonian women than men, which resonates with the difference detected in emotional well being in this study. Estimations of the occurrence of depression in PD patients vary from 20% - 90% according to the literature (Findley and Committee), 2002, Playfer, 1999, Schrag et al., 2001).

The social support dimension of the PDQ-39 which had the lowest scores at both time points (Test 1 and Test 2) measures predominantly close ties and family support e.g. "*had problems with your close personal relationships*?"; "*lacked support in the ways you need from your family or close friends*?"

Fitzpatrick et al, 1997 and Schrag et al, 2000 also found that social support scored the lowest compared to the other dimensions.

Table 3.2 Functional impairment as measured by the PDQ-39 (test 1)

PDQ-39 Category	Males n = 64	Females n = 27	Total n = 91	Mann Whitney 2 tail p-value
Overall				
Mean (S.D.)	39.7 (21.4)	38.0 (20.0)	39.2 (20.8)	0.701
Range of scores	3.9 - 94.9	3.2 - 82.1	3.2 - 94.9	
Median	39.6	36.1	38.5	
IQR	23.7 - 57.9	20.7 - 51.7	22.3 - 55.9	
Mobility				
Mean (S.D.)	49.3 (29.4)	59.9 (25.2)	52.4 (28.5)	0.157
Range of scores	0.0 - 90.0	2.5 - 90.0	0.0 - 90.0	
Median	48.3	63.1	52.5	
IQR	22.5 - 80.0	45.3 - 83.1	30.6 - 80.0	
Activities of Daily Living				
Mean (S.D.)	50.8 (28.3)	42.9 (25.8)	48.3 (27.6)	0.273
Range of scores	0.0 - 100.0	0.0 - 100.0	0.0 - 100.0	
Median	50.0	41.7	45.8	
IQR	25.0 - 75.0	25.0 - 62.2	25.0 - 69.8	
Emotional well being				
Mean (S.D.)	29.2 (22.2)	37.3 (21.5)	31.7 (22.2)	0.049*
Range of scores	0.0 - 87.5	4.2 - 75.0	0.0 - 87.5	
Median	20.8	32.3	25.0	
IQR	12.5 - 41.7	25.0 - 55.2	15.6 - 45.5	
Stigma				
Mean (S.D.)	24.4 (22.8)	26.4 (24.3)	25.0 (23.2)	0.794
Range of scores	0.0 - 100.0	0.0 - 75.0	0.0 - 100.0	
Median	18.8	18.8	18.8	
IQR	6.3 - 37.5	6.3 - 41.4	6.3 - 37.5	
Social Support				
Mean (S.D.)	15.0 (20.2)	10.4 (13.9)	13.7 (18.7)	0.353
Range of scores	0.0 - 100.0	0.0 - 41.7	0.0 - 100.0	
Median	8.3	0.0	8.3	
IQR	0.0 - 21.5	0.0 - 22.9	0.0 - 21.5	
Cognition				
Mean (S.D.)	43.3 (22.5)	32.0 (20.9)	40.0 (22.5)	0.035*
Range of scores	6.3 - 100.0	6.3 - 75.0	6.3 - 100.0	
Median	37.5	31.3	37.5	
IQR	27.6 - 56.3	14.1 - 48.4	25.0 - 56.3	
Communication				
Mean (S.D.)	40.2 (26.8)	21.5 (26.2)	34.7 (27.8)	0.004*
Range of scores	0.0 - 91.7	0.0 - 83.3	0.0 - 91.7	
Median	41.7	8.3	33.3	
IQR	16.7 - 63.2	0.0 - 42.4	8.3 - 58.3	
Bodily Discomfort				
Mean (S.D.)	44.7 (26.3)	52.4 (22.9)	47.0 (25.4)	0.211
Range of scores	0.0 - 100.0	16.7 - 91.7	0.0 - 100.0	
Median	50.0	58.3	50.0	
IQR	25.0 - 58.3	32.6 - 72.9	25.0 - 66.7	

PDQ-39 scale scores from respondents with Parkinson's disease at Test 1
Possible range of scores is 0 to 100% impairment
Increasing scores represent more impairment
*P<0.05

Figure 3.1 Summary Scores (PDQ-39 Test 1)

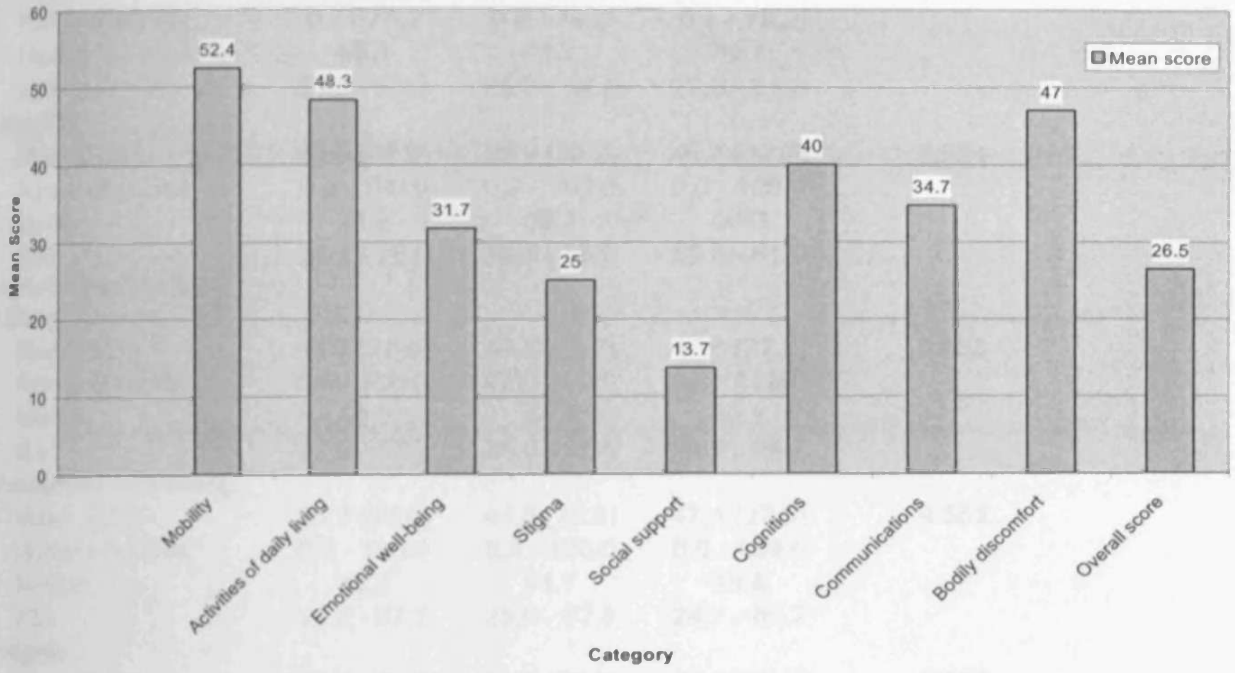
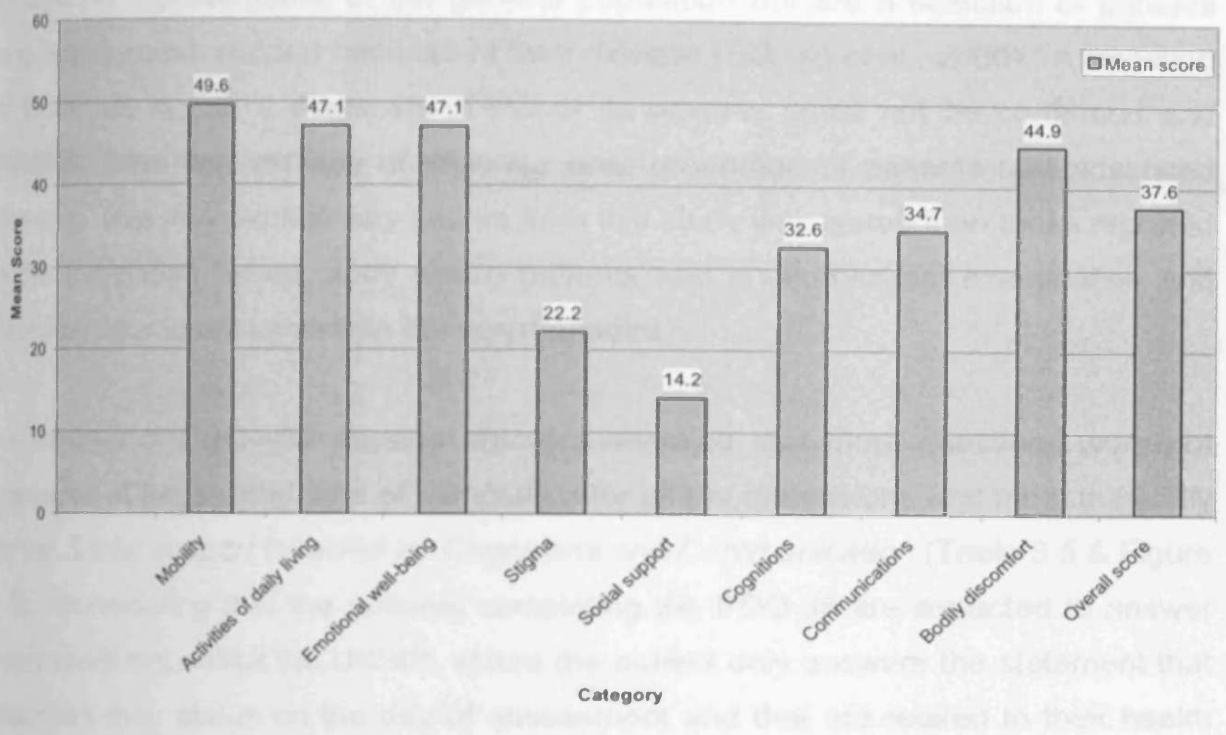


Table 3.3 Functional impairment as measured by the PDQ-39 (test 2)

PDQ-39 Category	Males n = 64	Females n = 27	Total n = 91	Mann Whitney 2 tail p-value
Overall				
Mean (S.D.)	36.3 (20.0)	40.7 (19.8)	37.6 (19.9)	0.428
Range of scores	0.7 - 78.2	9.0 - 78.2	0.7 - 78.2	
Median	39.4	41.7	39.7	
IQR	21.8 - 51.3	28.2 - 52.8	22.0 - 51.3	
Mobility				
Mean (S.D.)	45.0 (32.9)	59.9 (30.2)	49.6 (32.7)	0.064
Range of scores	0.0 - 100.0	0.0 - 100.0	0.0 - 100.0	
Median	36.9	66.3	50.0	
IQR	12.2 - 76.6	38.1 - 85.6	20.6 - 81.9	
Activities of Daily Living				
Mean (S.D.)	48.2 (28.0)	44.5 (25.8)	47.1 (27.2)	0.553
Range of scores	0.0 - 100.0	8.3 - 100.0	0.0 - 100.0	
Median	49.3	41.7	45.8	
IQR	21.2 - 67.7	25.0 - 57.8	24.7 - 66.7	
Emotional well being				
Mean (S.D.)	48.2 (28.0)	44.5 (25.8)	47.1 (27.2)	0.553
Range of scores	0.0 - 100.0	8.3 - 100.0	0.0 - 100.0	
Median	49.3	41.7	45.8	
IQR	21.2 - 67.7	25.0 - 57.8	24.7 - 66.7	
Stigma				
Mean (S.D.)	20.7 (22.8)	25.7 (26.2)	22.2 (23.9)	0.583
Range of scores	0.0 - 81.3	0.0 - 87.5	0.0 - 87.5	
Median	12.5	15.6	12.5	
IQR	3.6 - 27.6	0.0 - 47.7	1.6 - 37.5	
Social Support				
Mean (S.D.)	14.8 (22.1)	12.5 (14.7)	14.2 (20.2)	0.800
Range of scores	0.0 - 91.7	0.0 - 41.7	0.0 - 91.7	
Median	4.2	8.3	8.3	
IQR	0.0 - 25.0	0.0 - 23.6	0.0 - 25.0	
Cognition				
Mean (S.D.)	34.2 (20.0)	28.8 (18.8)	32.6 (19.7)	0.314
Range of scores	0.0 - 75.0	0.0 - 75.0	0.0 - 75.0	
Median	31.3	25.0	25.0	
IQR	18.8 - 50.0	18.8 - 37.5	18.8 - 50.0	
Communication				
Mean (S.D.)	36.1 (25.1)	31.5 (26.2)	34.7 (25.4)	0.447
Range of scores	0.0 - 75.0	0.0 - 83.3	0.0 - 83.3	
Median	33.3	26.4	33.3	
IQR	16.7 - 58.3	8.3 - 53.1	8.3 - 58.3	
Bodily Discomfort				
Mean (S.D.)	43.1 (26.3)	49.0 (26.0)	44.9 (26.2)	0.362
Range of scores	0.0 - 100.0	8.3 - 91.7	0.0 - 100.0	
Median	41.7	45.8	41.7	
IQR	16.7 - 63.2	25.0 - 72.9	25.0 - 66.7	

PDQ-39 scale scores from respondents with Parkinson's disease at Test 2
Possible range of scores is 0 to 100% impairment
Increasing scores represent more impairment
P<0.05

Figure 3.2 Summary Scores (PDQ-39 Test 2)



This may suggest that close relationships are less jeopardised by PD than are distant relationships. None of the scales at Test 2 were statistically significant between the two groups.

The scores obtained in this study of patients with PD recruited from the UKPDS was worse than those reported in a population based sample of patients from the London area (Schrag et al., 2000) but better than those reported in a population based sample of members of the UKPDS (Fitzpatrick et al., 1997) (Table 3.4). In the study by Schrag (Schrag et al., 2000) all participating patients had a semi-structured general and neurological interview and examination. In this and the Fitzpatrick (Fitzpatrick et al., 1997) study patients were not examined before inclusion into the study, but were recruited on the basis of their membership in the UKPDS. It is

therefore difficult to establish at which disease stage these patients were or how accurate the diagnosis was. Therefore, patients with other disorders, such as multiple system atrophy may have been inadvertently included in the sample. Furthermore, population based samples of active members of the UKDPS are not necessarily representative of the general population but are a selection of patients who actively seek support because of their disease (Schrag et al., 2000). A limitation of this study is that a diagnosis of PD or its severity could not be confirmed and therefore there was no way of knowing what proportion of patients had advanced disease. This may explain why scores from this study was worse than those reported in the population based study where patients had a neurological examination and interview prior to implementing the questionnaire.

An analysis of PDQ—39 missing answers revealed that more questions were not answered at the second time of completion for all the dimensions and most markedly for the *Social support* followed by *Cognitions* and *Communication* (Table 3.5 & Figure 3.3). Considering that the patients completing the PDQ-39 are expected to answer every question, unlike the UKSIP, where the patient only answers the statement that describes their status on the day of assessment and that are related to their health you might expect there to be very few missing questions. More than 10% of patients failed to complete questions in the dimensions *Cognition* and *Communication* at both time points (Test 1 and 2). The most common questions not answered at both time points were; *lacked support in the ways you need from your spouse or partner* (Test 1 = 16.28% & Test 2 = 27.5%), and *felt ignored by people* (Test 1 = 11.36% & Test 2 = 13.95%).

Female patients missed more questions than males at both time points ($p < 0.0001$). There was no relationship between duration of disease and missing items. There was a very low level of item redundancy (number of missing questions) for "felt you had to conceal your Parkinson's"; felt embarrassed in public"; "felt worried by other peoples reaction" (all items relating to *Stigma*) and "had problems writing"; (item relating to ADL). These items were affirmed in very high frequencies and may therefore not yield any information regarding differences between patients and thus detract from the discriminative properties of the questionnaire.

Table 3.4 Mean (SD) quality of life scores in patients with Parkinson's disease in this study and previous studies

	This study (Ali, 2002a, Ali, 2002b)	Schrag (Schrag et al., 2000)	Fitzpatrick et al. (Fitzpatrick et al., 1997)
	UKPDS members (Postal)	Population based study	Non clinic sample (Postal)
PDQ-39			
Mobility	52.4 (28.5)	47.0 (33.4)	64.2 (30.3)
Activities of daily living	48.3 (27.6)	36.5 (29.7)	56.3 (28.0)
Emotional well being	31.7 (22.2)	29.0 (22.5)	40.4 (23.2)
Stigma	25.0 (23.2)	23.6 (27.7)	35.1 (28.5)
Social support	13.7 (18.7)	14.8 (20.7)	20.9 (22.3)
Cognitions	40.0 (22.5)	31.1 (22.0)	48.0 (23.3)
Communication	34.7 (27.8)	20.3 (22.6)	36.1 (26.3)
Bodily discomfort	47.0 (25.4)	37.4 (22.7)	52.7 (25.7)

The UKSIP Scores

UKSIP category scores were significantly higher in PD patients than in age-gender matched controls for all categories (Table 3.6). Compared to controls PD patients suffered a great deal of functional impairment, both physical and psychosocial. This validates the applicability of the UKSIP in a PD population.

Interestingly, in the PD group the mean physical score at Test 1 (21.9%) was higher than the mean psychosocial score (15.4%). This is opposite to what Longstreth et al reported (Longstreth et al., 1992). At Test 2, mean physical and psychosocial scores were similar (21.6% and 21.2% respectively). Table 3.6 & Figure 3.4 and Table 3.7 & Figure 3.5 show the mean scores for each UKSIP category at Test 1 and Test 2. There were no differences between male and females. PD patients indicated that the greatest dysfunction was in the areas of *Home management*, *Communication* and *Recreational activities*.

Table 3.5 PDQ-39 Missing Questions at baseline (T1) and follow-up (T2)

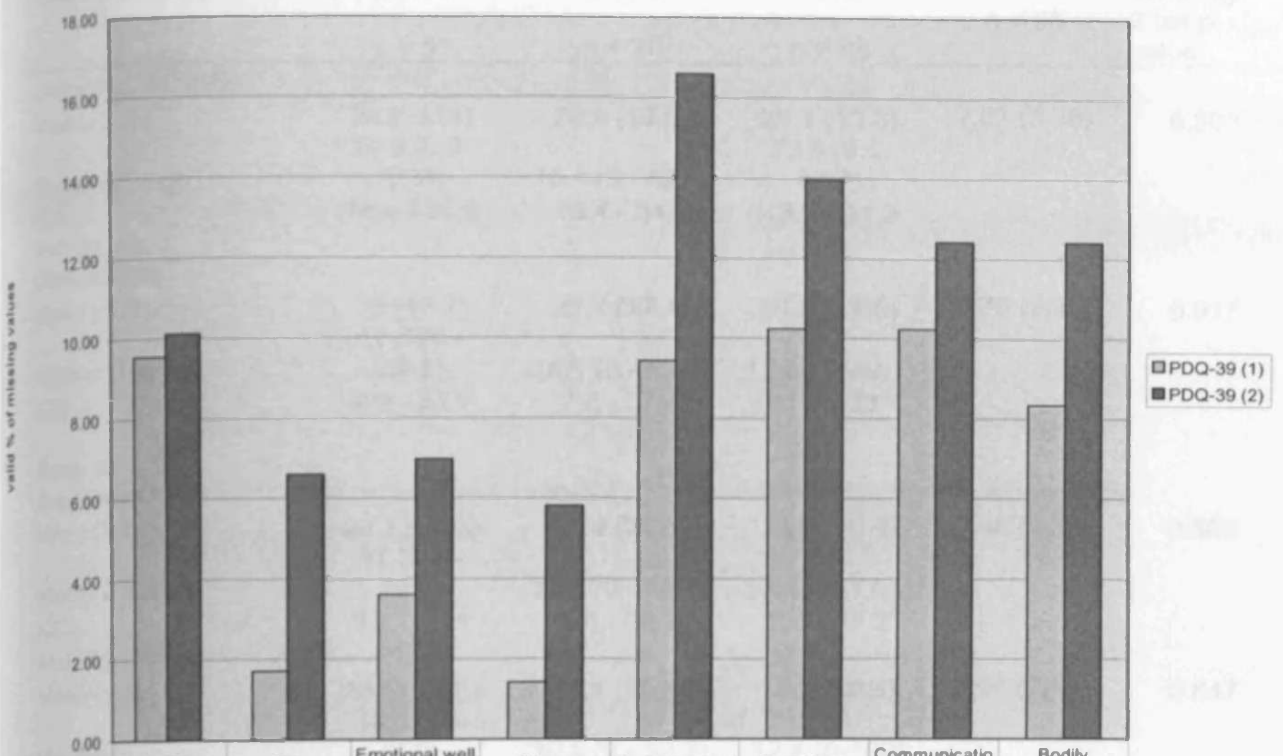
Dimension	Question no	PDQ-39 (Test 1) Non-missing questions Total	Missing questions Total	% of missing questions	PDQ-39 (Test 2) Non-missing questions Total	Missing questions Total	% of missing questions
Mobility	1	79	9	10.23	79	7	8.14
	2	77	11	12.50	78	8	9.30
	3	79	9	10.23	74	12	13.95
	4	80	8	9.09	76	10	11.63
	5	80	8	9.09	76	10	11.63
	6	79	9	10.23	77	9	10.47
	7	78	10	11.36	77	9	10.47
	8	79	9	10.23	77	9	10.47
	9	79	9	10.23	76	10	11.63
	10	86	2	2.27	83	3	3.49
Activities of daily living	11	86	2	2.27	80	6	6.98
	12	86	2	2.27	81	5	5.81
	13	86	2	2.27	81	5	5.81
	14	88	0	0.00	83	3	3.49
	15	87	1	1.14	78	8	9.30
	16	86	2	2.27	79	7	8.14
Emotional	17	84	4	4.55	81	5	5.81
	18	85	3	3.41	78	8	9.30
	19	85	3	3.41	77	9	10.47
	20	85	3	3.41	80	6	6.98
	21	85	3	3.41	83	3	3.49
	22	85	3	3.41	81	5	5.81
Stigma	23	87	1	1.14	81	5	5.81
	24	86	2	2.27	82	4	4.65
	25	87	1	1.14	82	4	4.65
	26	87	1	1.14	79	7	8.14
Social support	27	83	5	5.68	78	8	9.30
	28A	77	14	16.28	69	22	27.50
	28B	86	2	2.27	80	6	6.98
	29	85	3	3.41	80	6	6.98
Cognition	30	79	9	10.23	76	10	11.63
	31	79	9	10.23	73	13	15.12
	32	79	9	10.23	74	12	13.95
	33	79	9	10.23	73	13	15.12
Communication	34	80	8	9.09	76	10	11.63
	35	79	9	10.23	76	10	11.63
	36	78	10	11.36	74	12	13.95
Bodily discomfort	37	80	8	9.09	75	11	12.79
	38	81	7	7.95	75	11	12.79
	39	81	7	7.95	76	10	11.63

Less dysfunction was related to eating and emotional behaviour. Except for the communications category, these findings are similar to those reported for a group of patients with another chronic disease, rheumatoid arthritis (Deyo et al., 1982b). The similarities suggest that some of the dysfunction reported by PD patients may not be specific for PD but rather is seen with chronic diseases in general (Longstreth et al., 1992). The most prevalent dysfunctional behaviour in patients with PD completing the UKSIP is presented in Table 3.8.

The most commonly endorsed items were, "has trouble typing or writing" (78% of patients), "walks more slowly" (75% of patients), "takes part in fewer social activities" (64% of patients) and "sexual activity decreased" (61% of patients). The work category questions were endorsed the least with the percentage of patients ticking the work category questions ranging from 0 to 3.4%. Eighty two percent of patients were retired thus making the work section a redundant category in this patient population. Considering that PD is a chronic neurological disease affecting mostly the elderly population it is not surprising that the *Work* items are not relevant to the assessment of HRQOL in the PD population. Nevertheless, the question *"if you are retired, was your retirement related to your health?"* provides useful information on the impact of the disease on the ability to continue in employment and should be retained. Affirmative responses to ambulation, social interaction, communication and home management and recreation & pastime were more pronounced indicating that PD impaired these areas of daily functional behaviours the most.

Interestingly, sexual activity was affirmed by a large percentage of patients and this item is not addressed in the PDQ-39. The UKSIP category, Sleep and Rest score reveals 47% of patients affirmed the question "sleeps less at night" and 41% "sleeps more during the day". Sleep and Rest are not assessed in the PDQ-39. It is apparent that sexual activity and sleep disturbances are pertinent areas of health-related quality of life in PD patients and should be addressed in the PDQ-39 questionnaire. Recreational activities were highlighted as an area that was notably affected and this is a category that is not adequately addressed in the PDQ-39. The PDQ-39 has one question in the Mobility dimension; "*had difficulty doing the leisure activities which you would like to do?*" but it did not highlight this as an area of notable dysfunction.

Figure 3.3 Missing value analysis for PDQ-39



	Mobility	ADL	Emotional well being	Stigma	Social support	Cognitions	Communication	Bodily discomfort
PDQ-39 (1)	9.55	1.70	3.60	1.42	9.43	10.23	10.23	8.33
PDQ-39 (2)	10.12	6.59	6.98	5.81	16.57	13.95	12.40	12.40

Table 3.6 Functional impairment as measured by the UKSIP compared to age-gender matched controls(Salek, 1997)*(Test 1)

UKSIP Category	Males n = 61	Females n = 27	Total n = 88	Controls* n = 93	Mann Whitney 2 tail p-value
OVERALL SCORE					
Mean (S.D.)	24.6 (14)	22.8 (13.5)	24.1 (13.8)	2.07 (3.36)	0.503
Median (Range)	24.2 (0.8-84.8)	18.1 (3 - 49.4)	23.5 (0.8-84.8)		
IQR	14.4 - 31.2	13.4 - 34.6	13.9 - 31.3		
PHYSICAL DIMENSION					
Mean (S.D.)	22 (17.7)	21.7 (17.4)	21.9 (17.5)	0.76 (2.25)	0.917
Median (Range)	17.8 (0 - 89.1)	18.1 (0 - 53.8)	17.8 (0 -89.1)		
IQR	7.9 - 33.2	7.8 - 37.4	7.8 - 33		
Body care & movement					
Mean (S.D.)	24.1 (19.3)	22.4 (21.6)	23.6 (19.9)	0.48 (2.27)	0.503
Median (Range)	17.3 (0 - 77.1)	17.5 (0 - 68.7)	17.4 (0 -77.1)		
IQR	9.2 - 38.1	3.5 - 38.7	7.6 - 37.2		
Mobility					
Mean (S.D.)	18.5 (19.6)	18.1 (19.8)	18.3 (19.5)	0.94 (3.60)	0.847
Median (Range)	14.6 (0 – 100)	10.8 (0 - 52.2)	12.4 (0-100)		
IQR	0 - 33.1	0 - 40.5	0 - 33.4		
Ambulation					
Mean (S.D.)	23.5 (20.5)	24.5 (15.9)	23.8 (19.1)	1.26 (4.30)	0.447
Median (Range)	18.2 (0 - 90.1)	22.8 (0 - 58.2)	19.8 (0 -90.1)		
IQR	8.3 - 33.6	12.4 - 40.1	10.5 - 37.6		
PSYCHOSOCIAL DIMENSION					
Mean (S.D.)	16.6 (13.3)	12.9 (11.5)	15.4 (12.8)	2.90 (5.90)	0.183
Median (Range)	14.1 (0 - 75.1)	12.4 (0 - 48.2)	13.2 (0 -75.1)		
IQR	6.5 - 26.1	2.7 - 19.4	5.6 - 24.1		
Emotional behaviour					
Mean (S.D.)	11.2 (16.7)	13.9 (15.6)	12 (16.3)	3.30 (8.09)	0.368
Median (Range)	6.5 (0 – 100)	13.1 (0 - 60.6)	7.7 (0 -100)		
IQR	0 - 18.8	0 - 20.4	0 - 19.9		
Social interactions					
Mean (S.D.)	17.6 (15.2)	14.7 (12.8)	16.7 (14.5)	3.93 (7.15)	0.365
Median (Range)	15.4 (0 - 87.1)	15.1 (0 – 46.4)	15.1 (0-87.1)		
IQR	6.9 - 23.9	3.3 – 22.2	6 – 23.4		
Alertness, behaviour					
Mean (S.D.)	33.7 (31.7)	19.3 (24.2)	29.3 (30.2)	3.0 (10.17)	0.034*
Median (Range)	27 (0 – 100)	9.7 (0 - 85.5)	20.5 (0-100)		
IQR	7.6 - 53.6	0 - 39.2	0 - 46.6		
Communication					
Mean (S.D.)	37.9 (25.4)	23.4 (21.3)	33.4 (25.0)	0.35 (2.44)	0.013
Median (Range)	32.5 (0 - 91.2)	17.3 (0 - 65.2)	29.9 (0- 91.2)		

UKSIP Category	Males	Females	Total	Controls*	Mann Whitney
	n = 61	n = 27	n = 88	n = 93	2 tail p-value
IQR	19.5 - 52.4	9.7 - 35.9	9.7 - 51.9		
WORK					
Mean (S.D.)	33.2 (8.6)	35.6 (2.6)	33.9 (7.4)	4.04 (13.11)	0.213
Median (Range)	36.1 (0 - 36.1)	36.1 (22.7- 36.1)	36.1 (0 - 36.1)		
IQR	36.1 - 36.1	36.1 - 36.1	36.1 - 36.1		
RECREATION & PAST TIME					
Mean (S.D.)	34.6 (24.5)	32.4 (20.0)	33.9 (23.2)	5.33 (11.25)	0.885
Median (Range)	31.8 (0 – 100)	35.1 (0 - 66.1)	31.9 (0 – 100)		
IQR	17.8 - 48.1	18.4 - 46.3	18.1 - 46.6		
EATING					
Mean (S.D.)	7.2 (15.0)	5.7 (8.3)	6.8 (13.3)	0.78 (2.62)	0.650
Median (Range)	0 (0 - 89.1)	0 (0 - 30.2)	0 (0 - 89.1)		
IQR	0 - 10.3	0 - 13.7	0 - 11		
SLEEP & REST					
Mean (S.D.)	23.3 (20.0)	19.7 (20.3)	22.2 (20.0)	3.37 (6.30)	0.267
Median (Range)	21.9 (0 – 100)	12.2 (0 - 83.2)	21.8 (0 – 100)		
IQR	11.8 - 28.3	0.8 - 31.9	9.8 - 29.1		
HOME MANAGEMENT					
Mean (S.D.)	31.8 (24.5)	43.9 (30.5)	35.6 (26.9)	1.26 (4.38)	0.114
Median (Range)	27.5 (0 – 100)	45.8 (0 – 88.5)	30.5 (0-100)		
IQR	15.4 – 40.9	17 – 73.1	16.1 - 57		

UK-SIP scale scores from respondents with Parkinson's disease at Test 1
Possible range of scores is 0 to 100% impairment
Increasing scores represent more impairment

Table 3.7 Functional impairment as measured by the UKSIP (Test 2)

UKSIP Category	Males	Females	Total	Mann Whitney
	n = 59	n = 27	n = 86	2 tail p-value
OVERALL SCORE				
Mean (S.D.)	22.4 (14.4)	20.2 (13.6)	21.7 (14.1)	0.471
Median (Range)	22.5 (0.8 – 82)	14.8 (0.8 - 47.6)	20.2 (0.8 – 82)	
IQR	14.3 - 29.1	10 - 32.6	10.8 - 30.8	
PHYSICAL DIMENSION				
Mean (S.D.)	21.4 (17.5)	22.1 (15.6)	21.6 (16.8)	0.686
Median (Range)	18.5 (0 - 80.6)	19.1 (0 – 55)	18.4 (0 - 80.6)	
IQR	6.3 - 30.8	11.6 - 35.7	8.8 - 31.1	
Body care & movement				
Mean (S.D.)	22.5 (17.9)	22.5 (17.2)	22.5 (17.6)	0.985
Median (Range)	18.1 (0 - 76.4)	21.1 (0 - 60.3)	19.9 (0 - 76.4)	
IQR	9.5 - 32.8	10.6 - 31.1	10.1 - 32.5	
Mobility				
Mean (S.D.)	18.1 (19.1)	17.8 (18.1)	18 (18.7)	0.996
Median (Range)	15.5 (0 - 93.3)	13.3 (0 - 57.7)	14.7 (0 - 93.3)	

UKSIP Category	Males n = 59	Females n = 27	Total n = 86	Mann Whitney 2 tail p-value
IQR	12 - 23.9	0 - 22.5	10.9 - 23.7	
Ambulation				
Mean (S.D.)	23.5 (20.7)	26.1 (17.9)	24.3 (19.8)	0.297
Median (Range)	16.7 (0 - 81.1)	23.3 (0 - 65.3)	18.5 (0 - 81.1)	
IQR	5.8 - 34.5	16.4 - 40.8	10.6 - 35.3	
PSYCHOSOCIAL DIMENSION				
Mean (S.D.)	23 (17.0)	17.3 (15.8)	21.2 (16.7)	0.102
Median (Range)	20.8 (0 - 92.5)	10.7 (0 - 58.5)	20.2 (0 - 92.5)	
IQR	9.8 - 33.8	6.2 - 30.1	7.8 - 32.4	
Emotional behaviour				
Mean (S.D.)	10.4 (16.2)	13.0 (18.1)	11.2 (16.7)	0.519
Median (Range)	5.4 (0 – 100)	9 (0 - 68.7)	7.7 (0 – 100)	
IQR	0 - 17.2	0 - 19.4	0 - 17.6	
Social interaction				
Mean (S.D.)	16.7 (14.6)	13.5 (12.1)	15.7 (13.8)	0.345
Median (Range)	14.5 (0 - 79.2)	12.3 (0 - 46.6)	13.7 (0 - 79.2)	
IQR	5.4 - 23.1	3.7 - 22.5	4.7 - 22.6	
Alertness, behaviour				
Mean (S.D.)	32 (30.7)	19.8 (22.7)	28.1 (28.9)	0.108
Median (Range)	28 (0 – 100)	11.7 (0 - 85.5)	18.2 (0 – 100)	
IQR	0 - 48.6	0 - 38.7	0 - 46.4	
Communication				
Mean (S.D.)	32.9 (25.7)	23 (23.1)	29.8 (25.2)	0.090
Median (Range)	29.9 (0 - 91.2)	10.1 (0 - 85.9)	28.4 (0 - 91.2)	
IQR	10.6 - 41.9	8.9 - 44.9	9.7 - 42.2	
WORK				
Mean (S.D.)	19.3 (17.7)	17.0 (18.2)	18.6 (17.8)	0.510
Median (Range)	33.3 (0 - 36.1)	7.8 (0 - 36.1)	28.1 (0 - 36.1)	
IQR	0 - 36.1	0 - 36.1	0 - 36.1	
RECREATION & PAST TIMES				
Mean (S.D.)	36.4 (24.7)	29.1 (19.7)	34.1 (23.4)	0.226
Median (Range)	35.8 (0 – 100)	29.4 (0 - 69.9)	32.6 (0 – 100)	
IQR	18 - 53.7	9.4 - 45.3	17.7 - 49.1	
EATING				
Mean (S.D.)	6.5 (14.0)	5 (6.7)	6.0 (12.2)	0.631
Median (Range)	0 (0 - 89.1)	0 (0 - 19.3)	0 (0 - 89.1)	
IQR	0 - 10.5	0 - 9.8	0 - 9.6	
SLEEP & REST				
Mean (S.D.)	20.8 (18.7)	16.6 (18.8)	19.5 (18.7)	0.164
Median (Range)	13.4 (0 – 100)	12.2 (0 – 83.2)	12.2 (0 – 100)	
IQR	12 – 23.9	0 – 22.5	10.9 – 23.7	
HOME MANAGEMENT				
Mean (S.D.)	33.3 (27.4)	41.1 (27.7)	35.8 (27.6)	0.235
Median (Range)	28.8 (0 – 100)	41.4 (0 – 85.3)	30.5 (0 – 100)	
IQR	10.1 – 51.8	16.3 – 67.9	12.3 - 56	

UK-SIP scale scores from respondents with Parkinson's disease at Test 2
Possible range of scores is 0 to 100% impairment
Increasing scores represent more impairment

Figure 3.4 Domain Scores (UKSIP Test 1)

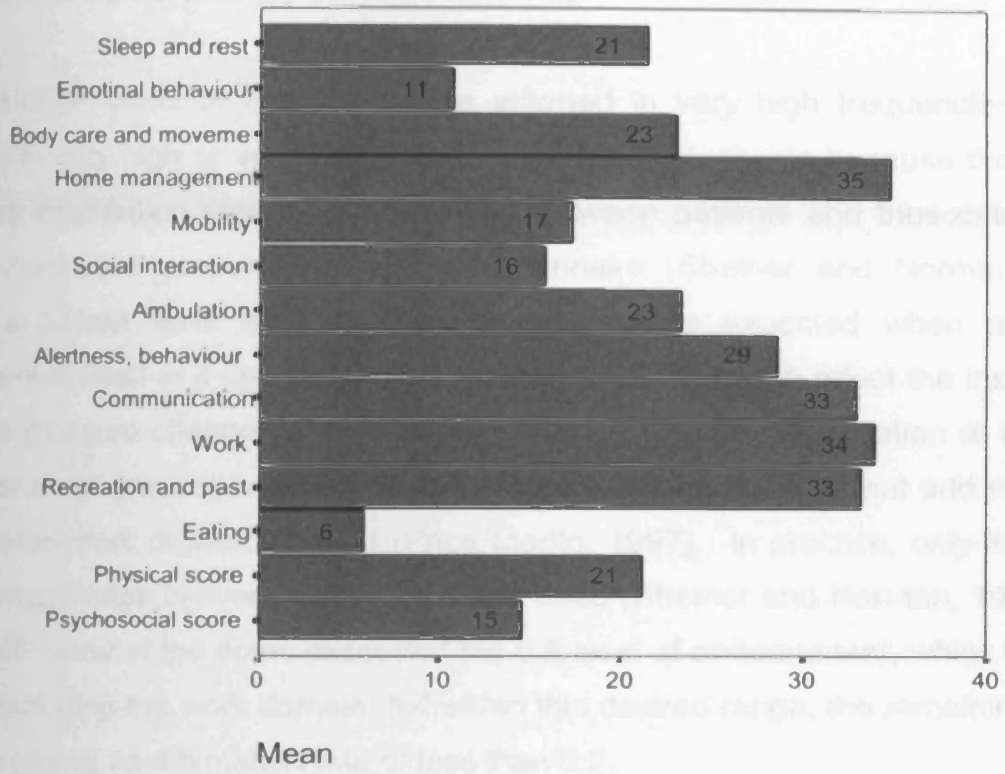
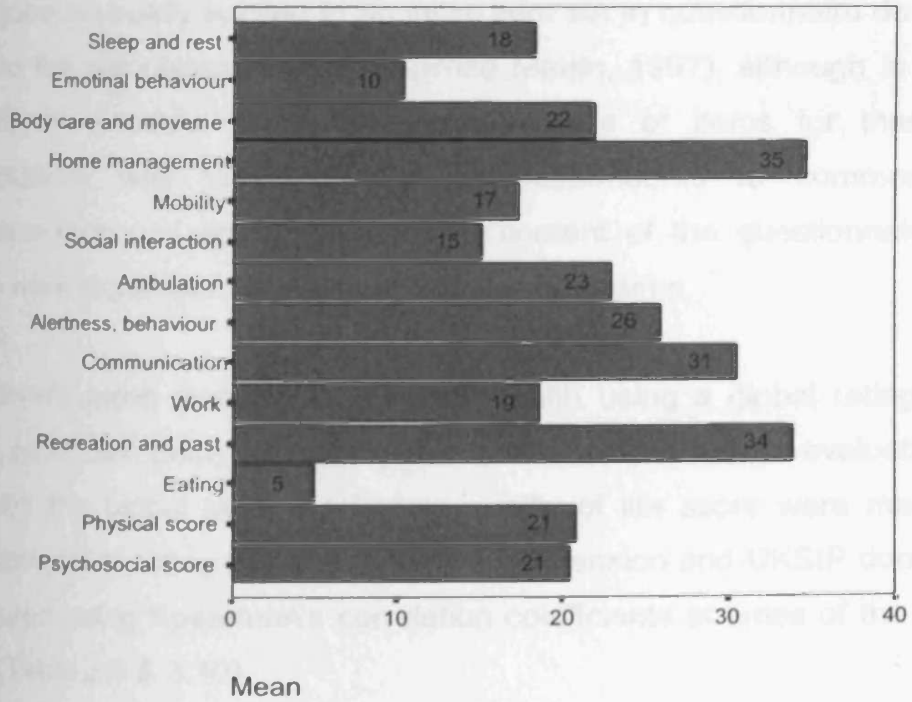


Figure 3.5 Domain Scores (UKSIP Test 2)



The UKSIP contains 8 items that directly assess recreational activities that are helpful in identifying this as an area of quality of life that is affected by PD. PD population in this study revealed that they spend less time on hobbies and recreational activities (58%) and going out to enjoy themselves (60%).

In the UKSIP, none of the items were affirmed in very high frequencies. Items affirmed in very high or very low frequencies are not desirable because they do not yield any information regarding differences between patients and thus detract from the discriminative properties of the questionnaire (Streiner and Norman, 1995). Whilst a certain level of item redundancy may be expected when a generic instrument is used in a specific disease state, a high level can affect the instruments ability to measure change over time, and can lead to patient frustration at having to complete lengthy questionnaires with high numbers of questions that add nothing to the measurement of the construct (Price Martin, 1997). In practice, only items with endorsement rates between 0.2 and 0.8 are used (Streiner and Norman, 1995). For the UKSIP none of the items exceeded the 0.8 level of endorsement, whilst 59.3% of items (excluding the work domain) fell within this desired range, the remaining 40.7% of items having an affirmation rate of less than 0.2.

This partly demonstrated the discriminative properties of the UKSIP in this group of patients. This assessment of item affirmation rates has been used as an illustrative tool rather than an in depth assessment of the item content. This is because the technique is usually applied to an initial item set in questionnaire development rather than to the completed instrument (Price Martin, 1997), although in this study it has served as a useful check on the relevance of items for these PD patients. Acceptability was tested by inviting respondents to comment on ease of understanding and appropriateness of content of the questionnaire. No adverse views were expressed for either of the questionnaires.

All patients rated their present overall health using a global rating scale (i.e. very good, good, fair, poor, very poor) at both time points and an evaluation of correlation between the global rating scale and quality of life score were made. Association between global rating score and PDQ-39 dimension and UKSIP domain scores were measured using Spearman's correlation coefficients at times of the first and second tests (Table 3.9 & 3.10).

Table 3.8 The most prevalent dysfunctional behaviour in patients with PD - UKSIP (Test 1 n = 88)

UKSIP Category	UKSIP Item	No (%) of patients ticking this item
Body care & movement	Dresses himself, but does so very slowly	**49 (56)**
	Only stand for short periods of time	**47 (53)**
	Do not keep my balance	**44 (50)**
	Kneels or bends down by holding onto something	**42 (48)**
	Has trouble putting on shoes, socks or stockings	**40 (45)**
	Make difficult movements with help	**37 (42)**
	Moves his hand or fingers with some limitation	**35 (40)**
Mobility	Only stays away from home for short periods	**38 (43)**
	Stays at home most of the time	**37 (42)**
	Does not get about in the dark unless someone to help	**30 (34)**
Ambulation	Walks more slowly	**66 (75)**
	Walks short distances or often stop for a rest	**44 (50)**
	Walk by myself but with some difficulty	**37 (42)**
Emotional behaviour	Behaves nervously or restlessly	**20 (23)**
	Says how bad or useless I am	**15 (17)**
	Keeps rubbing areas of his body that hurt	**15 (17)**
	Gets sudden frights	**15 (17)**
Social interaction	Takes part in fewer social activities	**56 (64)**
	Sexual activity is decreased	**54 (61)**
	Goes out less to visit people	**49 (56)**
	Talks less	**38 (43)**
Alertness behaviour	Has more minor accidents	**42 (48)**
	Reacts slowly to things	**35 (40)**
	Has difficulty reasoning and solving problems	**35 (40)**
Communication	Has trouble writing or typing	**69 (78)**
	Does not speak clearly when under stress	**51 (58)**
Sleep & rest	Sits for much of the day	**45 (51)**
	Sleeps less at night	**41 (47)**
	Sleeps more during the day	**36 (41)**
Home management	Is not doing any maintenance or repair work	**53 (60)**
	Is not doing heavy work around the house	**52 (59)**
	Does less of the daily household chores	**47 (53)**
Work	Retired	**72 (82)**
Recreation & pastime	Is going out to enjoy himself less often	**53 (60)**
	Spends shorter periods of time on hobbies	**51 (58)**
	Is cutting down on some of his physical recreation	**46 (52)**
Eating	Is eating much less than usual	**26 (30)**
	Is eating special or different food	**11 (12)**
	Feeds himself with help from someone else	**11 (12)**

All domains of the PDQ-39 showed an association with the global rating scale (p <0.001) at both tests with the exception of stigma, social support, communication (Test 1 only) and bodily discomfort (Test 2 only). All domains of the UKSIP showed an association with global rating scale (p <0.001) at both tests with the exception of sleep & rest, emotional behaviour, alertness behaviour, communication and work. Only emotional behaviour did not show significance at both time points. The UKSIP showed stronger correlation for physical scores (Test 1 = 0.496; Test 2 = 0.596) than psychosocial scores (Test 1 = 0.361; Test 2 = 0.541). As the global rating scale is a patient self reported scale it is feasible that patients rated their present health from the uni-dimensional perspective of physical symptoms only.

An important facet of any HRQOL instrument is its ability to discriminate between patients with different levels of severity. Only 25% (20 patients) of the sampled number reported a change in global health score (improvement or worsening) and therefore with such a small sample size we could not identify which scales could discriminate between patients who classify their health as good, fair or poor.

Further studies will be conducted to evaluate whether there is an association between measures of clinical PD severity and HRQOL.

Table 3.9 Correlation with Patient-rated Overall Health

PDQ-39

	Global health score (time 1)			Global health score (time 2)		
	Correlation Coefficient	p-value	N	Correlation Coefficient	p-value	N
Overall score	0.585	< 0.001	86	0.599	< 0.001	77
Mobility score	0.572	< 0.001	78	0.567	< 0.001	81
Activities of daily living score	0.467	< 0.001	86	0.558	< 0.001	81
Emotional well being score	0.494	< 0.001	85	0.558	< 0.001	81
Stigma score	0.254	0.019	85	0.348	< 0.001	81
Social support	0.224	0.041	83	0.313	0.006	75
Cognitions score	0.388	< 0.001	79	0.425	< 0.001	73
Communication score	0.337	0.002	79	0.475	< 0.001	74
Bodily discomfort	0.550	< 0.001	79	0.332	0.004	74

Association between overall health score and PDQ-39 domain scores measured using Spearman's correlation coefficients, at times of first and second interviews.
Statistical significance p<0.001

Table 3.10 Correlation with Patients-rated Overall Health

UKSIP

	Global health score (time 1)			Global health score (time 2)		
	Correlation Coefficient	p-value	N	Correlation Coefficient	p-value	N
Overall score	0.515	< 0.001	87	0.624	< 0.001	85
Sleep and rest	0.313	0.003	87	0.342	< 0.001	85
Emotional behaviour	0.335	0.002	87	0.279	0.010	85
Body care and movement	0.413	< 0.001	87	0.523	< 0.001	85
Home management	0.448	< 0.001	87	0.504	< 0.001	85
Mobility	0.468	< 0.001	87	0.575	< 0.001	85
Social interaction	0.378	< 0.001	87	0.516	< 0.001	85
Ambulation	0.448	< 0.001	87	0.486	< 0.001	85
Alertness, behaviour	0.280	0.009	87	0.406	< 0.001	85
Communication	0.332	0.002	87	0.477	< 0.001	85
Work	0.286	0.008	85	0.373	< 0.001	71
Recreation and past times	0.339	< 0.001	87	0.417	< 0.001	85
Eating	0.425	< 0.001	87	0.439	< 0.001	85
Physical score	0.496	< 0.001	87	0.596	< 0.001	85
Psychosocial score	0.361	< 0.001	87	0.541	< 0.001	85

Association between overall health score and UK-SIP domain scores measured using Spearman's correlation coefficients, at times of first and second interviews.
Statistical significance $p<0.001$

Reliability

Reliability refers to whether an instrument is measuring something in a reproducible and consistent fashion (Malek, 1997). Dimensions of reliability include test-retest and internal consistency.

For scales which are used as research tools to compare groups, Cronbach's alpha (to measure internal consistency) may be less than in the clinical situation, where the value of the scale for an individual is of interest. For comparing groups, alpha values of 0.7 to 0.8 are regarded as satisfactory (Bland, 1997).

The internal consistency reliability for all PDQ-39 domains were good (a 0.710-0.964); both at Test 1 and Test 2 (Table 3.11). These results imply that the PDQ-39 is a unidimensional instrument (internal consistency). Mobility at both time points and Activities of Daily Living at time point 2 were above a 0.90. This may be indicative of item redundancy within the scale and possibly loss of content validity.

The strong Cronbach's a values (> 0.70) for the UKSIP physical and psychosocial summary scores (Test 1 and Test 2) gave indication that their items were internally

consistent. All dimensions of the UKSIP had a > 0.5 (Table 3.12). The UKSIP body care & movement (T1 = 0.861; T2 = 0.824) and alertness behaviour (T1 =0.864; T2 = 0.855) showed the highest internal consistency at both test points. None of the dimensions had a > 0.90 suggesting absence of item redundancy. The work section of the UKSIP is redundant as less than 10% of respondents answered this section. Evaluation of reliability of the work dimension is meaningless as most patients in the sample were retired.

Test-retest reliability of scales was calculated from the responses of 81 individuals who completed both assessments. Correlation coefficients between scale scores for the two PDQ-39 assessments were all significant ($p<0.001$). Test-retest results supported the reliability of the PDQ-39 (r = 0.400-0.894) with strongest correlation observed for Activities of Daily Living (0.894) and Mobility (0.888) (Table 3.11). There was an outlier in that all the scales with the exception of emotional well-being scored r > 0.70. For patient reported measures, test-retest reliability should meet the 0.70 criterion to demonstrate adequate reproducibility (Damiano et al., 1999). The Emotional well being scale scored r = 0.40. This is perhaps not surprising as emotional well-being score increased between the first and second time points. Emotional well being was the second joint highest level of impairment with activities of daily living at Test 2.This finding may be because of the subjective nature of this domain or else this low r may be because the questionnaire was completed by proxy on the patients behalf.

With the exception of the *Work and Sleep* & *Rest* dimensions of the UKSIP, all test-retest values were > 0.70 (Table 3.12). The sleep & rest domain test-retest score was a satisfactory 0.550. The emotional behaviour domain of the UKSIP was 0.708 and therefore, in contrast to the PDQ-39 emotional well being scale, there was good association at both time points.

Table 3.11 Reliability of the PDQ-39

PDQ-39 domain	Test-retest Spearman's r*	Internal Consistency Cronbach's α	
		Test 1	Test 2
Overall Score	0.865	-	-
Mobility	0.888	0.960	0.964
Activities Daily Living	0.894	0.896	0.911
Emotional well being	0.400	0.900	0.889
Stigma	0.860	0.839	0.882
Social Support	0.819	0.796	0.793
Cognition	0.778	0.770	0.760
Communication	0.834	0.836	0.776
Bodily discomfort	0.788	0.710	0.847

* $p<0.001$

Table 3.12 Reliability of the UK-SIP

UK-SIP domain	Test-retest Spearman's r	Internal Consistency Cronbachs α	
		Test 1	Test 2
Overall Score	0.911	-	-
Physical Dimension	0.883	0.786	0.758
Body care & movement	0.846	0.861	0.824
Mobility	0.818	0.778	0.725
Ambulation	0.758	0.721	0.727
Psychosocial Dimension	0.819	0.774	0.777
Emotional behaviour	0.708	0.661	0.711
Social interaction	0.844	0.803	0.772
Alertness behaviour	0.762	0.864	0.855
Communication	0.804	0.767	0.772
Independent Dimensions			
Sleep & Rest	0.550	0.554	0.527
Home management	0.743	0.767	0.787
Work	-0.007	-	-
Recreation	0.815	0.665	0.645
Eating	0.799	0.737	0.690

Validity

The reliability of an instrument addresses the degree of repeatability or consistency of its measurements. If a measure fluctuates widely because of random disturbances, it will have a low reliability. Reliability is thus concerned with the amount of random error in the observed scores and makes no reference to possible systematic error in the measurements (Schulzer et al., 1999). This is because systematic errors do not detract from a measure's reliability. Systematic error can confound the measurement process by affecting the measures validity rather than it's reliability e.g. a thermometer that gives exactly the same reading for an individual on repeated occasions is reliable even if it records a temperature that is not accurate. This highlights that repeatability and consistency alone do not ensure that a measure is valid.

The broad definition of validity is that "an instrument measures exactly what it is supposed to measure and nothing else (Schulzer et al., 1999). The validity of an instrument is generally tested by means of a correlation analysis against a variety of external measures and instruments. In this study the PDQ-39 was assessed against the UKSIP to test convergent validity. Convergent Validity is one of the main tools for the validation of quality of life questionnaires. Owing to its somewhat tentative nature, this type of validity "ideally requires a pattern of consistent finding involving different researchers across a significant portion of time and with regard to a variety of diverse but theoretically relevant variables" (Zeller and Carmines, 1980).

The hypothesis tested is that the PDQ-39 domains should correlate with related variables of the UKSIP and (the other measure of this construct). The degree of association between the PDQ-39 scale scores and PDQ-39 and UKSIP overall scores were measured using Spearman's correlation coefficients.

Convergent validity was demonstrated by significant correlation of scales for the PDQ-39 with the overall scores of the UKSIP at both time points, Test 1 and Test 2 ($p < 0.001$). The UKSIP overall score correlated most strongly with the PDQ-39 Mobility (Test 1 = 0.74 & Test 2 = 0.72) and Activities of Daily Living (Test 1 = 0.67 & Test 2 = 0.73) scales on both occasions. The Stigma (Test 1 = 0.36 & Test 2 = 0.38) and Social Support (Test 1 = 0.34 & Test 2 = 0.41) scales correlated least strongly with the overall UKSIP score at both occasions (Table 3.13).

The correlation coefficients changed little from Test 1 to Test 2, demonstrating a high degree of reproducibility.

The ranking of the correlation coefficients demonstrated both a clear ordering in size (Table 3.14) and reproducibility between Test 1 and Test 2. Mobility and Activities of Daily Living had the greatest impact on this PD population and Stigma and Social Support having the least. This illustrated the relative contributions of the PDQ-39 categories in explaining patients' overall quality of life.

Table 3.13 Association between PDQ-39 categories and overall PDQ-39 and UK-SIP scores using Spearman's Correlation Coefficient (N = 74 to 88)

PDQ-39 categories*	Overall Score			
	PDQ-39 (1)	PDQ-39 (2)	UK-SIP (1)	UK-SIP (2)
Mobility	0.86	0.89	0.74	0.72
Activities of daily Living	0.83	0.84	0.67	0.73
Emotional well being	0.65	0.84	0.47	0.73
Stigma	0.56	0.54	0.36	0.39
Social Support	0.40	0.52	0.34	0.41
Cognitions	0.64	0.69	0.61	0.55
Communication	0.69	0.74	0.57	0.71
Bodily discomfort	0.60	0.54	0.40	0.41

* p-values < 0.001 for all correlation coefficients

The low association between PDQ-39 *Stigma* and *Social Support* with the overall UKSIP may suggest the (i) PDQ-39 captures aspects of the disease not measured by the UKSIP. This seems unlikely as the UKSIP category, *Social Interaction*, captures similar information to the PDQ-39 category *Social Support*. (ii) This suggests that the PDQ-39 is not adequately measuring or asking the right questions on these two

dimensions. (iii) Also, there may be too few questions present to accurately capture the impact of these dimensions on PD patients. The PDQ-39 attributes 3 questions to the *Social Support* domain whereas the UKSIP attributes 20 questions to *Social Interaction*. Finally (iv) these two dimensions may not be as important to PD patients compared to the other categories captured in the HRQoL scales.

Table 3.14 Correlation coefficients ranked in order of size (1=smallest, 8= largest)

PDQ-39 categories	Overall Score			
	PDQ-39 (1)	PDQ-39 (2)	UK-SIP (1)	UK-SIP (2)
Mobility	8	8	8	6
Activities of daily Living	7	6.5	7	7.5
Emotional well being	5	6.5	4	7.5
Stigma	2	2.5	2	1
Social Support	1	1	1	2.5
Cognitions	4	4	6	4
Communication	6	5	5	5
Bodily discomfort	3	2.5	3	2.5

DISCUSSION

The introduction of newer and similar treatments for Parkinson's disease reinforces the need to evaluate different intervention strategies from the patients' perspective by means of patient based questionnaires or health related quality of life (HRQOL) instruments. The clinical assessment of patients with PD is usually carried out by means of rating scales that vary widely in metric characteristics, length, and efficiency (Martinez-Martin, 1998). In clinical studies the predominantly used clinical rating scales such as the Hoehn & Yahr staging or the United Parkinson's Disease Rating Scale assesses from a clinical perspective important dimensions of PD such as tremor, rigidity and bradykinesia. These clinical scales do not address the broader impact of the disease on individuals' physical, psychological and social function (Fitzpatrick et al., 1992). To address the current imbalance that exists in clinical trial and treatment evaluation protocols of PD, it is necessary to propose an instrument that can compliment the currently accepted clinical measures, by introducing instruments that accounts for more than just disease symptomatology and provide a patient based opinion of the intervention that has been introduced.

If such an instrument is to be accepted its selection must be based on its psychometric attributes, which include feasibility, validity, reliability and sensitivity to change. Reliability is the extent to which a measure is free from random error in the population on interest (Guyatt et al., 1993), and refers to its internal consistency as well as its reproducibility. The reproducibility of a measure is the degree to which it yields consistent scores over time among respondents who are assumed not to have changed (test-retest reliability). Measures with poor reliability will be less efficient at distinguishing patients with different health states because differences in score may be obscured by random error (Dorman et al., 1998).

This study builds on the validation work reported by Peto et al (Jenkinson et al., 1997, Peto et al., 1998) on the development of the PDQ-39 by assessing the psychometric properties of the PDQ-39 in a community based sample of PD patients identified from the membership list of the UK Parkinson's Disease Society and comparing the instrument to the generic HRQOL measure, the UKSIP. The comparison of the PDQ-39 against the UKSIP to date has not been reported.

Reliability and validity of the PDQ-39 and UKSIP are reported here and a comparison of the two instruments is discussed.

The reliability analysis reported here indicates that the PDQ-39 is reliable both in terms of internal consistency and test-retest reliability. The questionnaire has construct validity in that scale scores are significantly associated with the overall scores of the UKSIP and respondents self-assessment of their present health (global rating). Evidence is also presented to support the overall reliability and internal consistency of the UKSIP in a PD population. For the first time the UKSIP was compared to the disease specific PDQ-39 and found to have good construct validity, particularly with respect to the *Mobility* and *ADL* domains of the PDQ-39. The UKSIP also showed good association with patients' self reported assessment of health (global rating).

The internal consistency of the PDQ-39 was good (a 0.710 – 0.964) and stable across the two completion time points (Test 1 and Test 2) indicating that the PDQ-39 is a unidimensional scale. Values above 0.5 are acceptable, although ideally scores should be in excess of 0.7 (Jenkinson et al., 1999, Nunnally, 1978). The domains of *Mobility*, *Activities of Daily Living*, *Emotional well-being* and *Stigma* all had a >0.8. The *Mobility* domain at both time points was a > 0.9. These results imply that *Mobility*, *ADL*, *Emotional well being* and *Stigma* contribute the most to the instrument in this community based population. The high internal consistency for some domains may be interpreted as a sign of redundancy of items and possibly loss of content validity. An analysis, omitting items from the PDQ-39 would reveal whether the internal consistency of the scale remained high to identify whether items could be omitted from the scale. This was not done on this occasion and it is possible that by doing so would reduce the number of items to allow us to acquire a comprehensive, overall picture of the disabilities of Parkinson's patients. This should be investigated further to identify if any items in scale are surplus to requirement (redundant) and whether there are some attributes of PD that are important to the patient but have not been adequately addressed in the instrument. Low internal consistency may not indicate unreliability, but only that domains are based on multiple constructs. None of the PDQ-39 domains had a low internal consistency.

Jenkinson et al (Jenkinson et al., 1995) reported similar internal consistency reliability values where internal consistency of the PDQ-39 were greater than 0.70 for all scales (0.72 for *Bodily discomfort* domain to 0.95 for *Mobility* domain) with the exception of *Social support* (a = 0.66) which appeared as the less consistent dimension (a < 0.70) in the two studies reported (Jenkinson et al., 1995, Peto et al., 1995).

Internal consistency reliability of the UKSIP in the PD population to date has not been reported. The UKSIP Physical and Psychosocial summary scores showed good internal reliability at both time points with Cronbach's a >0.70. *Body care & movement* and *Alertness behaviour* showed the highest internal consistency at both time points and none of the dimensions had a >0.90 suggesting absence of item redundancy. These results imply that *Body care & movement* and *Alertness behaviour* contribute the most to the UKSIP instrument in this community-based population. There are similarities with the PDQ-39 in that *Mobility* and *ADL* domains showed the highest internal consistency reliability and these domains contain items that are similar to the items that make up the *Body care & movement* domain of the UKSIP. The work section of the UKSIP is redundant in that less than 10% of respondents answered this section. The UKSIP does provide an insight into whether the respondent retired due to their health i.e. "If you are retired, was your retirement related to your health?" 39.8% of respondents stated that their retirement related to their health. This information is not captured in the PDQ-39 and should be considered, as it would provide meaningful information on the impact that the disease has on the person's ability to work.

If reliability can be accurately measured by test-retest methods, this criterion is a major guide to the performance of the instrument. However, reliability is both difficult to assess and interpret. Very high reliability is not necessarily good as it may indicate duplication and redundancy in the instrument and may be gained at the expense of responsiveness to change. Most published estimates of reliability, especially indirect estimate, need treating with caution.

Test-retest reliability was examined by means of correlation coefficients between scale scores at Test 1 and Test 2. Overall, the test-retest reliability of the PDQ-39 was high and correlation coefficients between scale scores were all significant ($p<0.001$). Strongest correlations were observed for *Activities of Daily Living* (r =

0.894) and *Mobility* (r = 0.888). Test-retest for *Emotional well-being* was low (r = 0.40) and suggests there were changes in the person or situation between test administrations. Thus low test-retest reliability does not necessarily reflect the psychometric properties of the test (Spilker, 1996). Since some behavioural or psychological attributes may lack temporal stability, Spilker (Spilker, 1996) questions the usefulness of point prevalence estimates of such attributes. The temporal stability of *Emotional well-being* should be examined to decide whether strategies are needed to increase reliability or whether alternative assessment instruments, which are more sensitive to fluctuations in mood or behaviour, are required.

All domains of the UKSIP exhibited good test-retest reliability (r >0.70) with the exception of the *Sleep & rest* domain which nevertheless was a satisfactory r = 0.550. In contrast to the PDQ-39, the UKSIP *Emotional behaviour* domain correlation was 0.708 (and r = 0.40 for the PDQ-39 *Emotional well-being* domain). The temporal stability of the UKSIP *Emotional behaviour* domain is apparently greater than that of the PDQ-39 *Emotional well-being* domain. Further exploration of the different items in each of these comparable domains may reveal why such variances in reliability exist.

Construct validity was examined by means of correlations of PDQ-39 domain scores with overall UKSIP scores and patients self-reported assessment of health (global rating). To demonstrate construct validity it should be shown not only that a test correlates highly with variables it would be theoretically expected to (convergent validity), but also that it does not correlate highly with variables with which it would be expected differ (divergent). The UKSIP and PDQ-39 origins differ whereby one is designed specifically to address attributes important to PD patients whereas the other is a generic measure of health status. It would be reasonable to expect a strong association for some of the domains of the PDQ-39 where the items are measuring similar constructs to the UKSIP. Conversely, we may expect to see a weaker association where some of the disease specific items in the PDQ-39 are measuring different variables to the UKSIP i.e. variables that are specific to PD such as difficulty with handwriting. Construct validity was also examined by means of correlations with respondents' assessment of their present health (ranging from very good to very poor). One would expect that the patients' self-assessment of health will correlate with their own PDQ-39 responses i.e. exhibit convergent validity.

Correlations were all significant for PDQ-39 scales with overall UKSIP scores ($p < 0.001$) and changed little at both time points, demonstrating a high degree of reproducibility. The UKSIP correlated most strongly with the PDQ-39 *Mobility* and *ADL* scales at both time points ($r = 0.67 - 0.74$), which is indicative of a high association. The *Stigma* ($r = 0.36$ and 0.38) and *Social support* ($r = 0.34$ and 0.41) scales correlated least strongly with the overall UKSIP score at both occasions. Low correlations may indicate that these domains of the PDQ-39 are capturing aspects of the disease not measured by the UKSIP, though this seems unlikely as the UKSIP category, *Social interaction*, captures similar information to the PDQ-39 category *Social support*. It is more likely that the PDQ-39 is not adequately measuring or asking the right questions on these two dimensions, combined with the paucity of items addressing these domains. The PDQ-39 attributes 3 questions to the *Social support* domain whereas the UKSIP attributes 20 questions to the related *Social interaction* domain. Finally the low correlations may reflect the individual nature of these two domains; for example, *Social support* is dependent on a variety of external variables at a particular point in time and can vary from the support of friends to the support of the health authority in the form of nursing/healthcare provision. These factors depend on circumstances that are independent of the patients' health status, namely the availability, affordability or access to these various forms of support.

When comparing the items that make up the *Stigma* domain of the PDQ-39, it is apparent that there is overlap with some of the *Social interaction* items of the UKSIP. For example; "I take part in fewer social activities" of the UKSIP is similar to "Avoid situations which involve eating or drinking in public" from the PDQ-39. Also, the items "Isolate myself", or "Stay alone much of the time" in the UKSIP is similar to "Felt you had to conceal your Parkinson's from people". Convergence validity of the *Stigma* and *Social support* domains of the PDQ-39 could perhaps be strengthened by combining these two related domains into one Social Function domain and requires further investigation. The similarities of these PDQ-39 items to the UKSIP items does not warrant the statement that the PDQ-39 is capturing aspects of the disease not measured by the UKSIP.

Martinez et al (Martinez-Martin et al., 1999) showed the domains of bodily discomfort, cognition and social support correlated the least with clinical measures of PD (Hoehn

& Yahr and the UPDRS). These low correlations may indicate differences in perception of disturbance and the presence of a real health problem may go unnoticed in clinical evaluation unless a HRQOL instrument is applied. This emphasises the usefulness of a HRQOL measure in the assessment of PD patients.

Association between the global rating scale (self-reported assessment of health) and the PDQ-39 domains at both time points showed moderate and significant association for all domains with the exception of *Stigma* (Test 1), *Social support* (Test 1 and 2), *Communication* (Test 1) and *Bodily discomfort* (Test 2). The lowest correlation coefficients at Test 1 were for *Social support* (r = 0.224) and *Stigma* (r = 0.254); and at Test 2 *Social support* (r = 0.313) and *Bodily discomfort* (r = 0.332). This data implies that patients are not associating, in particular, *Social support* as a factor that impacts significantly on their well-being. The items in the Social support domain refer to close relationships and therefore close relationships may be less jeopardised by PD than distant relationships. The *Social support* dimension of the PDQ-39 had the lowest scores amongst all domains at both time points (Test 1 and Test 2), which supports this suggestion. Stigma at Test 1 had the second lowest correlation to self-reported assessment of health and could suggest that self-image is not a major problem in PD or else in an elderly population. The PD patient may be less concerned about their disability in public places or else the reaction of others towards their impairment and are more troubled by their impaired mobility and resulting limitations on their ability to undertake routine tasks. It is therefore not surprising that the highest correlations were for the *Mobility*, and *Activities of daily living* domains.

Association between the global rating scale (self-reported assessment of health) and the UKSIP domains at both time points showed moderate and significant association for *Body care & movement*, *Home management*, *Mobility*, *Social interaction*, *Ambulation*, *Recreation & Past times*, *Eating*, Physical score and Psychosocial score. The UKSIP showed stronger correlation for physical scores than psychosocial scores and provides tentative evidence that patients rated their present health from the perspective of physical symptoms only. This data also underpins the importance of incorporating a HRQOL measure in the assessment of PD patients. Though the patient in this sample associates their physical symptoms more strongly to their own perception of health, the UKSIP was able to identify close associations between non-

physical domains such as *Social interaction* which showed significant moderate and good correlation at both time points respectively. The reason for patients showing a stronger awareness of their physical impairment can possibly explained by the fact that patients in this study were active members of the UKDPS who are not necessarily representative of the general population but are a selection of patients who actively seek support because of their disease (Schrag et al., 2000). A limitation of this study is that a diagnosis of PD or its severity could not be confirmed and therefore there was no way of knowing what proportion of patients had advanced disease or in other words severe physical impairment. This may explain why there was a stronger correlation between physical score in the UKSIP and self reported perception of health.

Examination of the association between HRQOL instruments and established clinical indicators of disease is a commonly used approach that is adopted in construct validation procedures (Price Martin, 1997). As clinical outcomes were not evaluated in this study, the self reported patients' assessment of health was adopted as an indirect correlate of disease severity. Construct validity is an on-going process that accumulates evidence over time by examining relationships built on *a priori* hypothetical constructs. The high correlations between questionnaire scores and patient reported outcomes forms part of the evidence for construct validity of both the PDQ-39 and UKSIP. There were stronger overall association between domains related to patients' physical functioning, than those related to psychosocial a domain which is perhaps not too surprising. Evidence for construct validity was further backed up by correlations observed between the PDQ-39 domains and overall UKSIP score at both time points.

The PDQ-39 had good convergent validity in this community dwelling population of patients with Parkinson's disease. It correlated strongly with the generic UKSIP overall score particularly for *Mobility* and *Activities of daily living*, and to a lesser degree for *Stigma* and *Social support*.

The UKSIP's reliability and validity in a community-based sample of PD patients was good and this study highlights that this generic HRQOL measure is relevant to the measurement of Quality of Life in Parkinson's disease.

Missing data has been found to be strongly correlated with older age and reduced health (Wagner et al., 1993). The PDQ-39 was analysed for the number of items (questions) that were not answered bearing in mind that each item in the PDQ-39 questionnaire requires an answer based on five possible responses ranging from "Never" to "Always". Some items were visibly missing i.e. did not receive a response at both time points which may point towards a lack of relevance. A limit of 10% was arbitrarily set to highlight items that were not considered to be relevant based on non-completion by the patient.

An analysis of missing questions from the PDQ-39 shows that more questions (above 10% limit) were not answered for *Cognition*, *Communication* and *Social support* (the percentage of missing items for *Social support* at Test 1 was 9.43% but rose to 16.57% at Test 2). More questions were not answered at the second time of completion for all domains, which may be due to frustration at repeating the same questionnaire within a short interval period. 16.28% (Test 1) and 27.5% (Test 2) of respondents did not answer the question "Lacked support in the ways you need support from your spouse or partner?" There are two possible reasons for this. 1. This item is irrelevant and patients feel that they are supported from their partners who are likely to be carers as well; 2. The patient felt uncomfortable answering this question in the presence of their partner/spouse and preferred to avoid the answer totally. This raises the question of how reliable such a self reported questionnaire is where the respondent is sensitive to answering certain questions honestly for fear of offending a close family member or else apprehensive to reveal their true feelings. This is not within the remit of this research but should be borne in mind. Surprisingly there were 7 items from the Mobility domain at Test 1, which were above the 10% limit. Impaired mobility is a major symptom of PD but when you look at the nature of questions that comprise the Mobility dimension it is possible to understand why there are missing items. The Mobility dimension contains questions such as; "Had difficulty doing the leisure activities which you would like to do"; "Had difficulty looking after your home, e.g. DIY, housework, cooking"; "Had problems walking half a mile". The mean age of the PD population was 70.5 years old and one would not be surprised to see healthy people in their 4^{th} or 5^{th} decade of life having difficulty in pursuing leisure activities if leisure activities include sport, or walking; nor is it uncommon with today's prevailing sedentary life-style, for people to not walk half a mile, or to participate in DIY activities, or cook and do housework. It would then not be surprising to expect

an elderly PD patient to find these types of questions irrelevant. Mobility is of major importance and relevant to PD patients but perhaps questions that reflect and are more in tune with the age range associated with PD would be more appropriate.

Such a level of missing data is undesirable in a HRQOL instrument and strongly indicates that the questionnaire contains non-relevant items.

More than 10% of items were missing at both time points for the Cognition domain and may indicate that this cohort of patients did not have any serious or noticeable level of cognitive impairment. Cognitive impairment was not assessed in this study.

The most prevalent dysfunctional behaviour (which was identified as the most commonly endorsed items on the UKSIP) in patients with PD were items related to writing, walking, taking part in social activities and interestingly sexual activity. Studies have indicated that the need for intimacy and sexual expression are important dimensions of QoL (Welsh et al., 1997). Although disturbed sexual functions are a common complaint of patients with PD, it is not assessed by the PDQ-39 (Moore et al., 2002). The quality of sexual life has shown to be correlated with disease severity in PD, as well as patient's general satisfaction from life. It is of significance that sexual interest often remains intact in older age(Moore et al., 2002) and therefore old age cannot be attributed to a loss of interest in sexual activity. This study has highlighted that sexual activity was decreased in 61% of respondents thus confirming its importance in the life's of PD patients and also revealing that the generic measure, the UKSIP is able to identify this pertinent dimension of HRQOL. This should lead to a recommendation to address this important area of HRQOL in the PDQ-39 instrument.

Sleep problems are also common in PD and were recognised by Parkinson in his original description (Factor et al., 1990). Insomnia in PD may result from difficulty with sleep initiation, or with sleep maintenance and may be worsened by motor limitations or tremor (Stacy, 1999). Larsen et al (Larsen et al., 2000) showed that PD patients had more sleep disturbances and fatigue than patients with diabetes mellitus and healthy elderly people. The UKSIP confirmed these previous findings as nearly half of the respondents from this study confirmed that they "sleep less at night" or "sleep more during the day". This was not detected in the PDQ-39, which does not

specifically measure sleep related problems. Fifty one percent of respondents reported on the UKSIP that they "sit for much of the day" and this aspect of fatigue is not addressed in the PDQ-39; in fact the PDQ-39 does not include questions on energy problems. Fatigue has been identified as risk factor for poor HRQOL in PD patients (Herlofson and Larsen, 2003) and should be addressed in any future disease specific HRQOL instrument.

Non-motor problems like sleep disturbance and fatigue should be actively diagnosed and treated when possible and the UKSIP is an instrument that can be used to identify patients with such non-motor impairment problems to allow the medical practitioner to implement suitable treatment strategies such as the administration of tricyclic antidepressants to aid night time sleep or else the timing of the evening dose of anti-PD medication to ensure symptoms are adequately controlled during the night. Also, in a clinical trial setting the UKSIP can be utilised to identify or distinguish between which dopamine agonists precipitate sudden daytime sleep episodes. Reports that dopaminergic medication, in particular the D_1 and D_2 receptor agonists pramipexole and ropinirole, precipitate daytime sleepiness with sudden sleep episodes in PD patients has received widespread attention (Happe, 2001).

Analysis of the most frequently endorsed items of the UKSIP shows that PD patients will experience some degree of social isolation, impaired sleep, limitations in activities of daily living and exclusion due to non-participation in recreational activities.

The PDQ-39 reveals that impairment is poorer for *Mobility*, *Activities of daily living* and *Bodily discomfort* and to a lesser extent for *Stigma* and *Social support*. Peto et al, 1995 also reported greater impairment for *Mobility*, *Activities of daily living* and *Bodily discomfort* whereas *Social support* and *Communication* was to a lesser extent impaired. These results suggest that PD patients experience physical function impairment to a greater extent than psychosocial impairment or else that the PDQ-39 is more sensitive to physical impairment in the PD. A comparison of SIP scores from elderly controls taken from a study assessing the SIP in PD patients (Longstreth et al., 1992) showed that PD patients from this study had poorer UKSIP scores. The greatest differences from the general population norms seen were seen in domains for *Communication*, *Home management* and *Alertness behaviour*. The UKSIP completed by respondents from this study indicated greatest dysfunction in the areas

of *Home management*, *Communication* and *Recreational activities* which is identical to that reported by Longstreth et al, 1992 who evaluated patients from a US neurology clinic and consistent with the analysis of most frequently endorsed items in this study. This is also similar to those reported for a group of patients with another chronic disease, rheumatoid arthritis (Deyo et al., 1982) and thus these similarities suggest that some of the dysfunction reported by PD patients may not be specific for PD but rather with chronic diseases in general. The main difference in this study was that the physical dimension impairment score was higher than the mean psychosocial dimension score. This may be explained by the mean age of the patients recruited from the UKPDS (70.5 years), which was noticeably higher than the mean age of patients recruited from the Longstreth study (66.3 years) and that these patients were recruited from a neurology clinic. Increasing age may result in worse physical scores whereas a younger population may be more aware of their psychosocial dimensions. Less dysfunction was related to *Eating* and *Emotional behaviour*, which indicates that eating is not a problem associated with PD patients and that these patients are not emotionally disturbed in that they do not feel they are a burden to others or fearful of the future. Emotional behaviour is very dependent on a patients circumstances and health at a single point in time and should therefore be evaluated in more depth and detail than is in either of the HRQOL instruments. The UKSIP asks patients to describe how they are on the day of completing the questionnaire and therefore does not take into account the temporal nature of emotional behaviour and may therefore miss that a patient was worried about the future or nervous in the preceding week or month. The PDQ-39 asks patients how they were "during the last month" but addresses *Emotional well being* in 4 questions as apposed to the UKSIP *Emotional behaviour domain* which has attributed 9 items. The PDQ-39 could attribute further items to *Emotional well being* domain to capture whether the patients are irritable, blames themselves, gets sudden frights, or frustration. These items are addressed in the UKSIP but avoided in the PDQ-39. Adding items that identify different areas of the *Emotional well being* domain of the PDQ-39 will strengthen the comprehensiveness of the instrument, particularly as "current feelings of optimism" has been shown to have a significant impact on HRQOL (Findley and Committee, 2002). The coping strategies adopted by patients with chronic diseases are widely recognised as influencing their ability to maintain reasonable levels of emotional well-being (Felton and Revenson, 1984). It is feasible that patients with PD can adapt to their situation by lowering their expectations regarding their own physical capacity



and therefore it is vital to ensure that the true emotional well being of the patient is captured. The UKSIP identifies non-physical dimensions of health as having the greatest impairment in HRQOL in PD patients whereas the PDQ-39 identifies physical dimensions as having the greatest impairment in HRQOL in PD patients. This is not surprising as the PDQ-39 is a disease specific measure, which focuses on the physical aspects of PD such as *Mobility* and *Bodily discomfort* whereas the UKSIP covers in more depth the psychosocial and independent domains of HRQOL. This would highlight the need to accompany the PDQ-39 with the generic UKSIP to ensure the non-physical features of PD are adequately addressed in assessment and monitoring of the patient.

It is unlikely that a single HRQOL measure will address all issues relevant to a particular therapy. For example, new therapies may introduce effects that were not previously associated with other PD therapies and, therefore, not included in existing PD specific HRQOL measures (Damiano et al., 1999). This can be further explained by the incidence of somnolence reported by patient taking the second-generation dopamine agonists, which might lead to future PD HRQOL instruments comprehensively capturing these relevant areas. It would be more practical to use a generic measure such as the UKSIP which adequately captures items relating to sleep and rest in association with the disease specific PDQ-39. Including a general HRQOL measure such as the UKSIP may capture further unanticipated effects, such as sexual function, which is not addressed in the PDQ-39.

SUMMARY

- We can conclude that the PDQ-39 and UKSIP are both reliable, internally consistent and valid instrument for the assessment of the disabilities on patients with Parkinson's disease from a community based sample of PD patients recruited from the membership list of the UK Parkinson's Disease Society.

- The UKSIP is a robust and reliable measure that has demonstrated strong evidence of construct validity in the PD population. The UKSIP correlated most strongly with the PDQ-39 *Mobility* and *Activities of Daily Living* (ADL) categories.

- This study indicates that the UKSIP is useful for identifying sickness related dysfunction amongst patients with PD, and in particular non-physical dysfunction.

- HRQOL scores assessed by the PDQ-39 have demonstrated strong evidence of convergent validity and confirm the findings of the originators of the PDQ-39 that the instrument has good test-retest reliability and internal consistency.

- Much of the dysfunction reported by PD patients was related to non-physical difficulties rather than physical problems according to the UKSIP, and vice versa for the PDQ-39. This shows that the UKSIP should be used together with the PDQ-39 to identify non-physical difficulties in the treatment and evaluation of PD patients.

- Ranking correlation coefficients demonstrated both a clear ordering in size and reproducibility between both time points. *Mobility* and *ADL* had the greatest impact on this PD population. *Stigma* and *Social support* had the least impact. This illustrates the relative contributions of the PDQ-39 categories in explaining patients' overall quality of life.

- Impairment is greatest for *Mobility, ADL and Bodily discomfort* categories of the PDQ-39. PDQ-39 suggests that physical functional impairment is experienced to a greater extent than psychosocial impairment. Similarly, the mean physical score of the UKSIP was higher than the mean psychosocial score but the greatest areas of dysfunction were in the independent categories of *Home management* and

Recreational activities, along with *Communication*, which is a psychosocial category.

- *Recreational activities* such as time spent on hobbies and recreation are not directly addressed in the PDQ-39 and should be further investigated.

- The *Social support* dimension of the PDQ-39 had the lowest scores and the most common question not answered at both time points was "lacked support in the ways you need from your spouse or partner". The *Social support* domain is weighted towards close family relations and spouse and hence the data may be revealing that PD patients do not experience any problems with close family members. Items describing dysfunctional behaviour in more generalised social interactions should be considered.

- Sexual activity was highlighted in the UKSIP as an area of concern to patients who affirmed that this had an affect on their health. This is not addressed in the PDQ-39 and should be as it has been reported previously that there is a strong relationship between decreased sexual activity and affection with poor HRQOL.

- The *Work* category of the UKSIP is a redundant item with the exception of the question "if you are retired, was your retirement related to your health?" This should be retained and also addressed in the PDQ-39 as it provides valuable information highlighting that PD has a direct impact on a person's ability to undertake and continue their employment.

Chapter 4

Comparison of psychometric properties of the UK-SIP and SF-36 in idiopathic Parkinson's disease

INTRODUCTION

Preliminary research using the disease-specific PDQ-39 yielded useful data from a community based sample of Parkinson's disease (PD) patients recruited from the membership list of the UK Parkinson's Disease Society. This research also revealed the utility of a generic measure, the UK version of the Sickness Impact Profile (SIP), in the PD population. The work revealed that the UKSIP was useful for identifying sickness related dysfunction amongst patients with PD, and in particular non-physical dysfunction. It seems sensible to test the hypothesis of whether these findings remain true for a PD population recruited from a hospital clinic and compare the results against another generic measure, the Medical outcomes Study 36-Item Short Form Health Survey (SF-36). The literature shows that no comparison of the UKSIP against the SF-36 in PD has been conducted and this merits further investigation.

The clinical evaluation of PD is usually carried out using scales and simple objective tests (Martinez-Martin, 1998). The information obtained includes a variety of functional and mental aspects, in addition to data gathered in the patient's examination. However, many factors that are important to the patient may remain un-assessed. Emotional and psychosocial consequences of the disease are usually omitted from the physician's evaluation. The frustrations and restrictions caused by the disease are heavy burdens for patients and their families. Clinical rating may provide data on severity of the disease or on the degree of disability, but in no way are the medical evaluations equivalent to the Quality of Life (QoL) status perceived by the patient.

A large number of disease-specific clinical measures exist designed to characterise the impact of the disease but there are comparatively few disease-specific health-related quality of life measures in Parkinson's disease. Clinical disease based scales for PD focus on rating mental deterioration and physical signs such as tremor, rigidity, bradykinesia and postural instability. The clinical measures, however, do not fully capture the outcomes of PD. A variety of instruments exist which measure the multiple dimensions of health-related quality of life (HRQoL). These HRQoL measures use patient self-report to assess from the patients perspective, their physical, mental and social function and satisfaction with health. There are numerous generic health-related quality of life measures, which may be suitable for

this population. The selection of an appropriate instrument is crucial. These measures must be practical to administer, comprehensive in their content, reliable and valid in the specific population under study and contribute pertinent information without being unnecessarily redundant (Rubenstein et al., 1998). In addition, acceptance of an instrument would be aided if it were able to discriminate between patients with different levels of HRQoL and were sensitive to change.

Generic instruments address multiple aspects of quality of life across a range of different patient or disease groups. Thus they focus on general issues of health (or ill health) rather than specific features of a particular disease: the role of disease-specific instruments (Thompson and Yu, 2003). A number of generic instruments are commonly used in research and clinical evaluation in populations with chronic diseases. The two most commonly used are the Sickness Impact Profile (Bergner et al., 1981) and Medical outcomes Study 36-Item Short Form Health Survey (Ware and Sherbourne, 1992).

Comparison studies are needed of the validity, reliability, and especially the responsiveness of these generic measures in the same population to improve the precision of the dimensions included in existing instruments. Only then can a HRQoL instrument be used with confidence and be accepted by the medical community who rightly demand that such a measure should contribute information beyond that obtained by the clinical rating scales used by neurologists such as the Hoehn & Yahr (Hoehn and Yahr, 2001) and Unified Parkinson's Disease Rating Scale (UPDRS) (Fahn et al., 1987).

These two established generic quality of life measures, the UK version of the Sickness Impact Profile (UK SIP) (Salek, 1997) and the Short Form Health Survey (SF-36) (Ware and Sherbourne, 1992) were compared in a PD population attending a neurology put-patient clinic. These are commonly used in a wide range of neurological indications but as far as the literature shows, not routinely in Parkinson's disease. It would therefore be advantageous to be able use either the UK-SIP or SF-36 in Parkinson's disease to measure quality of life changes as these questionnaires have been validated and extensively tested and also, to compare quality of life parameters between different neurological disorders. This would mean that a validated and relevant generic measure could be utilised by neurologists, PD

specialist nurses and of course the patient, alongside other therapeutic disciplines, and health care professionals.

The aims of this study were to establish the longitudinal measurement properties of the SF-36 and UKSIP and to assess the practical use of this questionnaire in a neurology clinic setting. This study will also aim to investigate the impact that PD has on patients' general quality of life. The objectives of the study are outlined below.

Objectives

- To determine whether the SF-36 and UKSIP are suitable generic instruments for use in discriminative PD studies

- To establish the discriminative and convergent construct validity of two generic HRQOL assessments, the SF-36 and UKSIP in PD patients attending a Parkinson's out-patient clinic

- To explore the relationships between HRQoL patient scores and various patient characteristics

- To explore the internal consistency of the SF-36 and UKSIP domains

- Assess the applicability and practicality of the UK-SIP and SF-36 in PD patients by assessing missing items in the SF-36 and prevalence of UKSIP items

- Assess the sensitivity to change of the two generic measures

The Hoehn and Yahr staging scale (Hoehn and Yahr, 2001) was implemented and completed by a consultant neurology physician to allow a comparison of clinical status to quality of life as measured by the two generic health-related quality of life measures. Data from this research will part help users to decide which of these instruments should be recommended as the generic instrument of choice for use in discriminative PD studies.

METHODS

Study Instruments

Both the UKSIP and SF-36 are self-administered questionnaires. The development and validation of the SF-36 has been documented and the instrument has been used widely in outcome evaluations and surveys (Brazier et al., 1992, Hayes, 1995, Jenkinson, 1995). The UKSIP (Salek, 1992) has been validated for use in disability outcomes research and in the assessment of PD patients (Andresen and Meyers, 2000, Bergner et al., 1981, Longstreth et al., 1992) and is reported in more detail in chapters two and three.

Patient Recruitment

Patients attending the neurology out patient clinic of Ashford Hospital, Middlesex, and satisfying a clinical diagnosis of idiopathic Parkinson's disease and satisfying the following inclusion and exclusion criteria were invited to participate in the study:

Inclusion criteria

- Men and women aged 45 years and older
- Diagnosis of idiopathic PD as determined by a neurologist
- Hoehn & Yahr stage 1-5 during an 'on' period
- Stable PD (i.e. patients were not required to return to the clinic for evaluation within 1 month)
- Written informed consent

Exclusion criteria

- Dementia
- Other neurological complications e.g. epilepsy
- Patients who in the opinion of the investigator was not suitable for participation to the study e.g. other chronic disease

Study Design

The study was designed as a comparative, longitudinal study to measure quality of life in Parkinson's disease patients comparing two different measuring instruments. Patients with impaired mental status and/or physical disability were excluded from entering the study. The neurologist selected suitable patients from an appointment list at Ashford hospital, Middlesex, and invited suitable patients to participate in the study in writing in advance of their clinic visit. The eligible patients were asked to sign a written-informed consent form before entering the study.

At baseline, the first patient assessment visit, the neurologist examined the patients in the clinic and rated them on the Hoehn and Yahr scale. Prior to their assessment by the neurologist, patients were asked to complete the SF-36 and UK-SIP questionnaires (preferably the day before or else on the day of the clinic appointment). After a period of no longer than six months, the same procedure was repeated on the same patient population.

In addition to the above information, demographic data was collected i.e. age, gender, level of education, level of income, work status and time since initial diagnosis (years). Ethics committee approval was sought prior to undertaking this research from the North West Surrey local research ethics committee.

Data Processing and Analysis

All data collected during the study was checked for completeness and accuracy and then coded and entered onto a computer database. The data was analysed using SPSS for windows statistical package. Non-parametric statistical test were employed in the analysis of questionnaire data, since normal distribution of the data could not be guaranteed. Statistical significance was set at the $p<0.05$ value and adjustment for multiple testing (Bonferroni adjustment) was not incorporated in to the analysis as this study is considered to be preliminary research and hypothesis generating. Differences with $p < 0.05$ can be considered a trend that justifies further investigation with larger sample sizes.

As this study is hypothesis generating rather than hypothesis testing a sample size calculation is irrelevant. This is a 'practicality' study and there is no 'target difference'

involved, therefore, 30 subjects will secure an acceptable power and confidence interval. A total of 36 subjects were recruited to this study.

The appropriate statistical methods were used for data analysis to determine the applicability and reliability of the quality of life instruments. Comparison of the measurement properties of the two instruments will form the basis for the selection of the single most suitable measure in assessing patient outcome.

RESULTS

The results generated in this study are overwhelming and therefore for the purpose of clarity will be presented here in five different parts:

Part I: HRQoL instruments domain coverage and scores
Part II: Applicability and specificity
Part III: Reliability
Part IV: Construct validity (convergent, divergent and discriminative)
Part V: Responsiveness (sensitivity) to change

Study Patient's Demographic Characteristics

A total of 53 patients were screened for entry to the study, from which 36 (67.9%) satisfied the inclusion/exclusion criteria and 17 (32.1%) were excluded. The patients were excluded from the study for the following reasons: cognitive impairment, co-morbidity such as arthritis, multiple system atrophy, psychiatric complication, non-idiopathic PD and lost to follow-up.

The mean age of the cohort was 68.9; range 51-83 and the gender breakdown was 66.7% male and 33.3% female. The mean disease duration was 6 years with a range of 1-22 years (Table 4.1). The majority of the subjects were permanently retired (83.5%) which reflects the age group of the cohort who were mostly over the age of retirement (87.5% were over the age of 60 years old). Education level was reported by 59.3% of the subjects and from this sample, 26.4% had received school education, and 33% had received either college or university education.

Twelve patients (37.5%) were diagnosed by the neurologist as having early stage PD at baseline compared to 20 patients (62.5%) with Advanced PD (Table 4.2). Of the 20 Advanced PD patients, 8 (40%) were assessed as stable and 12 (60%) as having insufficient control of motor fluctuations.

Table 4.1: Demographic characteristics (n= 36).

Age - years	
Mean (Range):	**68.9 (51 – 83)**
Gender	
Male (%):	**24.0 (66.7)**
Female (%):	**12.0 (33.3)**
Disease duration – years	
Mean (SD):	**6.0 (5.35)**
Median (Range):	**8.0 (1 – 22)**
Stage of Parkinson's at enrolment	(n = 32)
Early stage:	**12**
Advanced:	**20**
Insufficient control	10
Motor fluctuation	2
Stable	8

The majority of patients at baseline were staged at Hoehn and Yahr I and II (64.7%) and the remaining at Hoehn & Yahr stage III and IV (35.3%). None of the patients in this study had Hoehn & Yahr stage V (Table 4.3). At the follow-up visit, 56% were diagnosed at stage I and II and 44% at stage III and IV showing a worsening in the severity of PD in this population over a 6 month period.

All patients who completed their second visit answered a change in condition questionnaire (Table 4.4.) to rate their overall health in comparison to their baseline visit (i.e. much better, better, the same, a little worse, and much worse). The majority of patients reported they were "a little worse" compared to the baseline visit (44.4%); the same (33.3%); better (18.5%) and much worse (3.7%). This corresponds with

the clinician rating of disease severity and shows that the PD patients in this study reported their health was deteriorating.

Table 4.2: No of patients at each visit classified Early or Advanced PD

	Baseline		Follow-up	
	n	%	n	%
Early stage patient	12	37.5	6	24.0
Advanced-Stable	8	25.0	4	16.0
Advanced-Motor fluctuations	2	6.3	2	8.0
Advanced-Insufficient control	10	31.3	13	52.0
Total	32		25	
Missing	4		11	

Table 4.3: No of patients at each visit classified according to PD Hoehn & Yahr staging

	Baseline		Follow-up	
	n	%	n	%
Stage 1	14	41.2	8	32.0
Stage 2	8	23.5	6	24.0
Stage 3	9	26.5	10	40.0
Stage 4	3	8.8	1	4.0
Stage 5	0	0.0	0	0.0
Total	34		25	
Missing	2		11	

Table 4.4: Transition Question at follow-up visit

Change in condition		
	n	%
Better	5	18.5
The same	9	33.3
A little worse	12	44.4
Much worse	1	3.7
Total	27	
Missing	9	

Part I: HRQoL instruments domain coverage and scores

SF-36 and UKSIP Domain Coverage

The HRQoL dimensions suggested by Patrick and Erickson 1993 were used; (Table 4.5 and 4.6) to contrast the Hoehn & Yahr, SF-36 and UKSIP. The administrative and scoring considerations for each instrument are also included in the table. The Hoehn & Yahr focuses on physical functional status and pain symptoms. It includes questions about signs and symptoms such as tremor, postural instability and mobility. The SF-36 and UKSIP contribute additional information about patients' health outcomes which are not solely limited to physical signs and symptoms such as activities related to independent living such as working around the house, participation in recreational or pastime activities. For this very reason a HRQoL instrument is an important measure of the impact of disease, effect of treatment and other variables affecting people's lives.

The SF-36 and UKSIP both measure psychological aspects of health derived from questions pertaining to mood and social activity derived from questions pertaining to attending or avoiding social events. Both instruments provide information about the effects of physical problems by including items that ask about limitations in kind of activity, decrease in amount of work accomplished (from SF-36), difficulty with movement, and not walking (from UKSIP).

In contrast to the SF-36 the UKSIP provides information about sexual satisfaction, thought disorders such as (concentration and memory problems) and intellectual impairment. The UKSIP has a domain exclusively addressing social interactions (such as going out to visit people, and attending social activities).

SF-36 Domain Scores

Lower scores on the SF-36 reflect poorer health status. The SF-36 summary score at baseline and follow-up visits revealed that there was greater impairment for the physical component (baseline ms = 36.8, follow-up ms = 35.3) compared to the mental component (baseline ms = 45.3, follow-up ms = 44.4). There was no statistically significant difference detected between males and females. The minimal

Table 4.5: Domain coverage of the SF-36 and UKSIP:

Instrument	Scored Dimensions	Number of items
SF-36		36
Scales:	Physical functioning	
	Role functioning/physical	
	Role functioning/emotional	
	Vitality (energy)	
	Emotional well-being	
	Social functioning	
	Pain	
	General Health	
Components:	Physical	
	Mental	
UKSIP		136
Categories:	Ambulation	
	Body care and movement	
	Mobility	
	Home management	
	Recreation and pastime	
	Social interaction	
	Emotional behaviour	
	Alertness behaviour	
	Sleep and rest	
	Eating	
	Communication	
	Work	
Dimensions:	Physical	
	Psychosocial	

fluctuation in scores between the baseline and follow-up suggest that HRQoL in this population remained stable (Table 4.7 and 4.8). The category or dimension scores of the SF-36 (Table 4.7) add up to reflect a component score for physical and mental components. The physical role limitation category which includes such items (or questions) as *limitations in kind of activity, decreases in amount of work accomplished* and *difficulty in performing work or activities* are the most impaired of all the dimensions as reflected in the lower mean scores (ms1 = 38.7, ms2 = 40.4). Hence the physical component score is lower than the mental component score where lower scores indicate poorer HRQoL. The second highest level of impairment was vitality, which measures a respondent's level of energy while performing daily tasks (ms1 = 41.8, ms2 = 42.2). Fatigue is a common but poorly understood symptom in non-depressed Parkinson's disease(Friedman and Friedman, 1993). These highlight that the patients' ability to undertake activities of daily living are

impaired (Figure 4.1).The scales that had the least impairment were psychosocial domains such as mental health (ms1 = 65.0, ms2 = 64.0), role limitations because of emotional difficulties (ms1 = 63.2, ms2 = 57.3), and social functioning (ms1 = 61.3, ms2 = 57.8). The mental health category assesses how the patient is feeling and it appears that this dimension is not adversely affected in comparison to the other dimensions. This may be due to a number of reasons such as an improved state of mind resulting from participation in a study and greater than before access to a physician; patients mean duration of disease (6 years) may imply that the disease has not progressed to a stage where it may begin to have adverse effects on the patients feeling and state of mind.

Table 4.6: Dimensions of health, mode of administration and scoring considerations for the SF-36 and UKSIP

Dimensions/Measures	SF-36	UK-SIP	Hoehn & Yahr
Health perceptions	X	X	
General Health perceptions	X	X	
Satisfaction with health			
Functional status	X	X	
Physical function	X	X	X
Activities of daily living	X	X	X
Fitness	X	X	
Social function	X	X	
Limitation in usual roles:	X	X	
Work / regular activity	X	X	
Other social activities	X	X	
Bed or restricted time		X	
Sexual satisfaction		X	
Interaction and contact		X	

Dimensions/Measures		SF-36	UK-SIP	Hoehn & Yahr
Psychological function		X	X	
Affective:	distress	X	X	
	happiness	X	X	
	depression		X	
	motivation			
	vitality	X	X	
Cognitive:			X	
	intellectual impairment		X	
	thought disorder		X	
	Limitations in activity due to emotional problems	X		
Impairment				
Symptoms				
	Sensations			
	Pain	X	X	X
	Signs			
	Self-reported disease			
Self-administered		X	X	
Observer administered				X
Summary scores		X	X	
Subscales/dimension scores		X	X	
Population norms		X	X	

Table 4.7: SF-36 scores at baseline visit (n = 36)

Category/Dimension	Males	Females	Total	Mann Whitney 2 tail p-value
Physical Component				
Mean (S.D.)	35.6 (10.7)	38.8 (12.4)	36.8 (11.2)	0.493
Median (IQR)	32.8 (28.2– 45.3)	40.7 (28.9–48.6)	35.1 (28.7–45)	
Mental Component				
Mean (S.D.)	45.4 (12.6)	45.1 (11.4)	45.3 (11.9)	1.000
Median (IQR)	47.8 (34.3-57.6)	45.7 (37.9–56.5)	45.7 (36.3–56.5)	
Physical Functioning				
Mean (S.D.)	51.4 (30.4)	56.5 (25.6)	53.1 (28.6)	0.811
Median (IQR)	44.2 (25–82.9)	62.6 (32.5–74.6)	60.8 (25–75.6)	
Role Physical				
Mean (S.D.)	32.5 (41.4)	50 (41.8)	38.7 (41.8)	0.221
Median (IQR)	0 (0–77.1)	54.2 (2.1–97.9)	25 (0–81.3)	
Role Emotional				
Mean (S.D.)	59.6 (49.2)	70 (42.9)	63.2 (46.6)	0.657
Median (IQR)	100 (0–100)	100 (29.2–100)	100 (0–100)	
Social Functioning				
Mean (S.D.)	58.8 (27.4)	65.9 (21.9)	61.3 (25.5)	0.519
Median (IQR)	65.8 (32.5–90)	73.3 (45–89.2)	67.5 (44–90)	
Mental Health				
Mean (S.D.)	67.2 (18.6)	61.1 (18.5)	65 (18.5)	0.456
Median (IQR)	67 (51.7–84.3)	65.3 (37.7–79.3)	65 (51–81)	
Vitality				
Mean (S.D.)	42 (20.0)	41.4 (19.5)	41.8 (19.5)	0.901
Median (IQR)	42.5 (24.6-60.6)	45 (31.3–51.9)	45 (25.4–51.9)	
Bodily Pain				
Mean (S.D.)	57 (26.7)	64.6 (24.9)	59.6 (26)	0.485
Median (IQR)	53.7 (31.6-79.8)	54.3 (48.5–97.7)	52 (40.3–81.5)	
General Health				
Mean (S.D.)	47.5 (22.5)	47.3 (12.9)	47.4 (19.3)	0.829
Median (IQR)	47.8 (28.8-63.9)	50.3 (40.6-52)	50 (30-62)	

Lower scores indicate poorer health and high scoring indicates better health

Figure 4.1
Mean SF36 Domain Scores in PD patients

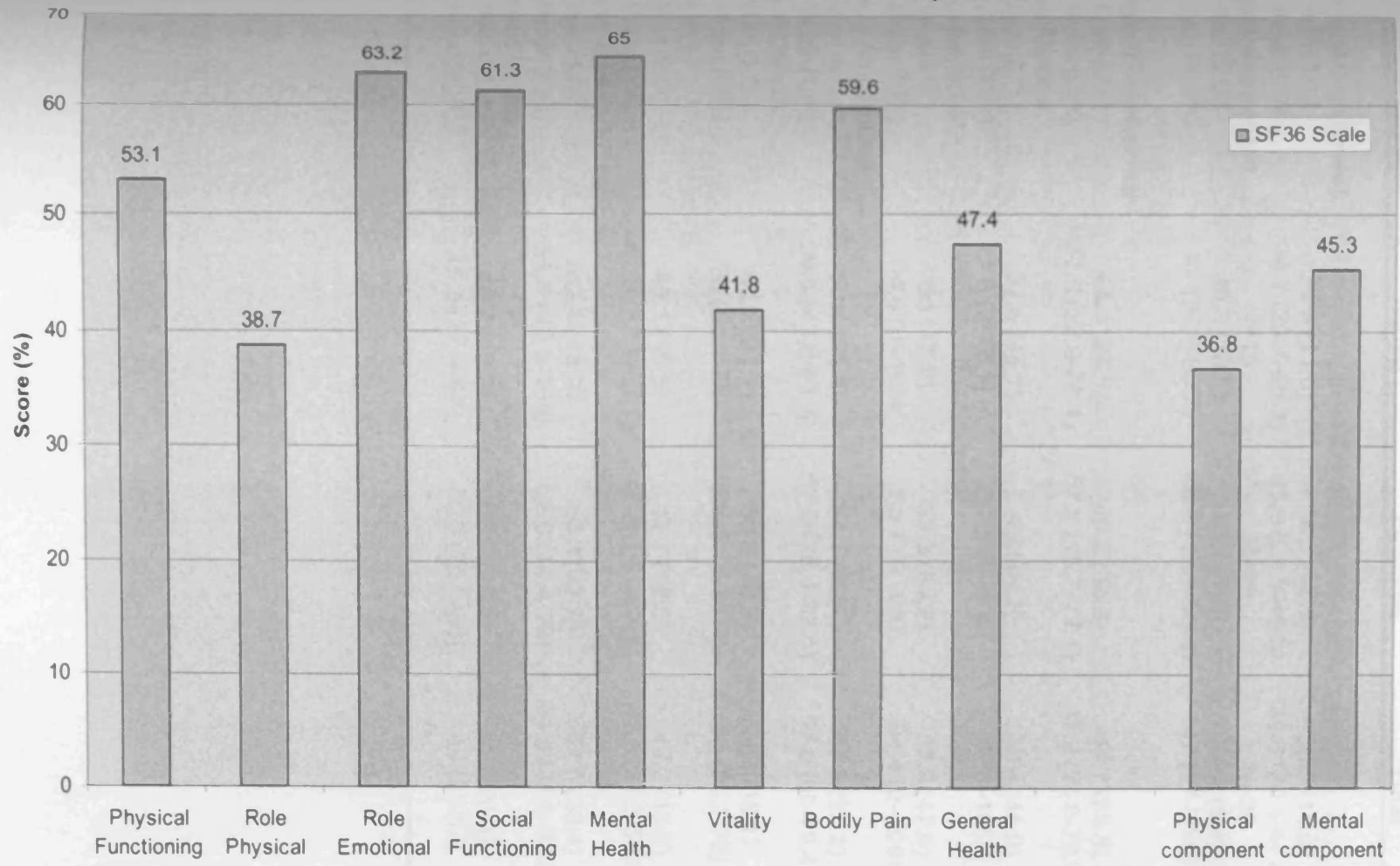


Table 4.8: SF-36 scores at follow-up visit (n = 36)

Category/Dimension	Males	Females	Total	Mann Whitney 2 tail p-value
Physical Component				
Mean (S.D.)	35.8 (11.6)	34.7 (12.3)	35.3 (11.6)	0.868
Median (IQR)	34.7 (29.7-46.8)	35.3 (22.8-46.3)	34.6 (29.4-44.8)	
Mental Component	N=15	N=10	N=25	
Mean (S.D.)	45.3 (10.5)	43.1 (7.8)	44.4 (9.4)	0.782
Median (IQR)	47 (36.9-55.2)	45.8 (37.1-50.4)	47 (37-50.6)	
Physical Functioning				
Mean (S.D.)	44.5 (28.5)	45.2 (30.8)	44.7 (28.8)	0.885
Median (IQR)	37.9 (22.9-78.1)	57.5 (16.7-73.1)	37.9 (24.6-73.8)	
Role Physical				
Mean (S.D.)	37.5 (44.7)	45 (49.7)	40.4 (45.9)	0.581
Median (IQR)	0 (0-77.1)	37.5 (0–100)	0 (0–100)	
Role Emotional				
Mean (S.D.)	60 (47.5)	53.3 (50.2)	57.3 (47.6)	0.782
Median (IQR)	94.4 (0–100)	66.7 (0–100)	94.4 (0–100)	
Social Functioning				
Mean (S.D.)	60.1 (24.1)	54.1 (27.7)	57.8 (25.2)	0.541
Median (IQR)	59.4 (43.8-81.3)	57.5 (29.4-82.5)	57.5 (35-79.4)	
Mental Health				
Mean (S.D.)	66.6 (21.1)	59.2 (10.3)	64 (18.1)	0.134
Median (IQR)	70 (54.6-76)	62 (51.5-67.3)	64 (53.1-75)	
Vitality				
Mean (S.D.)	43.1 (16.5)	40.7 (18.5)	42.2 (16.9)	0.904
Median (IQR)	42.5 (32.9-53.1)	43.8 (25. -58.1)	42.5 (31.3-53.8)	
Bodily Pain				
Mean (S.D.)	62.1 (28.9)	50.3 (32.7)	57.6 (30.4)	0.308
Median (IQR)	51.8 (39.9-100)	52.8 (14.4-79)	51.3 (31.6-93.3)	
General Health				
Mean (S.D.)	45.7 (19.2)	38.5 (23.4)	43 (20.8)	0.321
Median (IQR)	45.3 (27.9-67)	36.3 (20.6-56)	40 (25-57)	

UKSIP Domain Scores

Higher scores on the UKSIP reflect poorer health status. The mean physical score at baseline and follow-up visits (ms1 = 20.6, ms2 = 20.4) were higher than the mean psychosocial score (ms1 = 14.7, ms2 = 12.7). This is consistent with the findings of the previous chapter where the UKSIP was administered in a postal survey to members of the UK Parkinson's disease society (postal survey). Tables 4.9 and 4.10 show the mean scores for each UKSIP category at baseline and follow-up visit. There were no significant differences between scores for male and females. PD patients indicated that the greatest dysfunction was in the areas of *Home management*, *Alertness, behaviour*, *Recreational activities and Communication* (Figure 4.2). Less dysfunction was related to *Eating* and *Emotional behaviour*. Again these findings are consistent with the previous study (chapter three) and are also similar to those reported for a group of patients with another chronic disease, rheumatoid arthritis (Dey et al., 1982).

Figure 4.2 Mean UKSIP Domain Scores

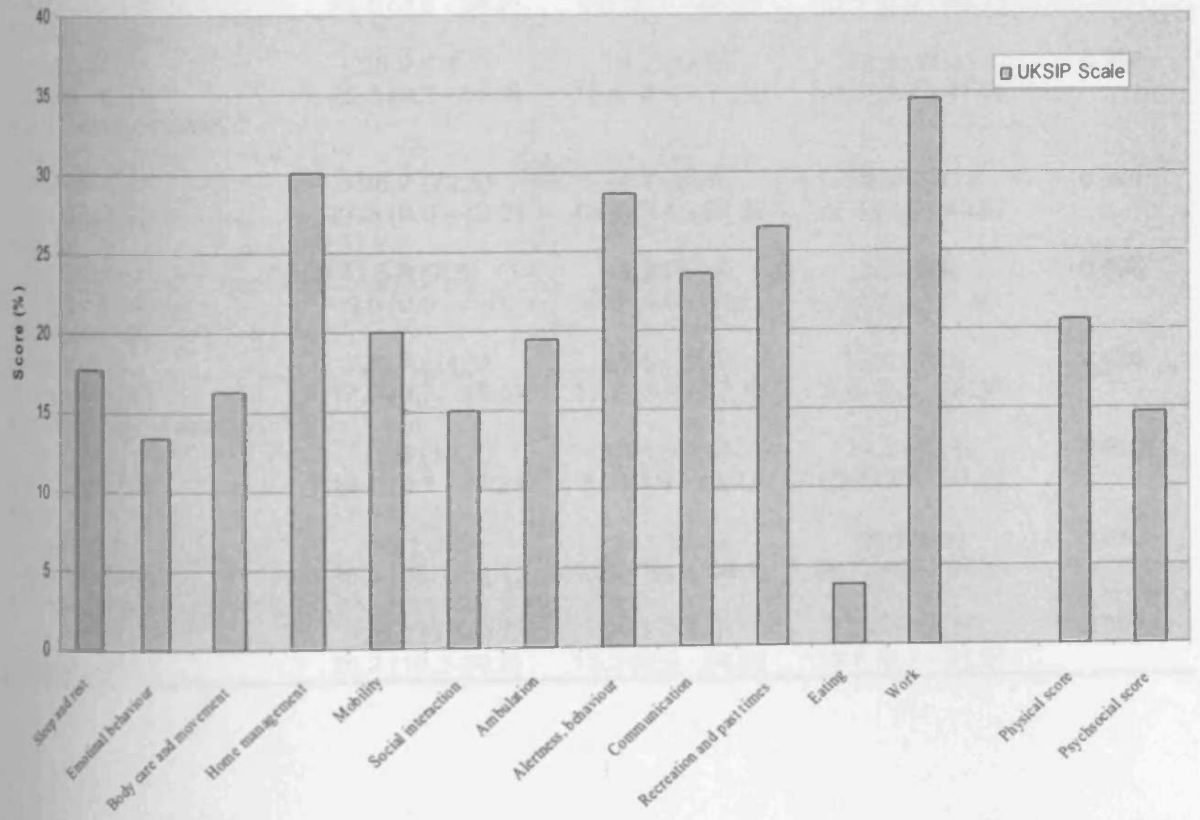


Table 4.9: UKSIP (T1) Baseline Scores (n = 36)

Category/Dimension	Males	Females	Total	Mann-Whitney p-value
Sleep and rest				
Mean (S.D.)	13.3 (10.1)	26.1 (28.3)	17.7 (19.0)	0.066
Median (IQR)	12.0 (9.8-12.2)	12.2 (12.2-31.6)	12.2 (9.8-21.8)	
Emotional behaviour				
Mean (S.D.)	14.3 (16.9)	11.8 (16.6)	13.4 (16.6)	0.731
Median (IQR)	8.8 (0.0-32.5)	7.7 (0.0-17.4)	8.8 (0.0-25.8)	
Body care and movement				
Mean (S.D.)	16.5 (12.7)	15.9 (17.5)	16.3 (14.3)	0.601
Median (IQR)	15.2 (5.7-29.9)	9.3 (3.7-24.3)	14.2 (5.1-26.7)	
Home management				
Mean (S.D.)	30.0 (25.4)	30.1 (32.2)	30.1 (27.4)	0.766
Median (IQR)	29.5 (8.1-56.9)	17.2 (1.6-62.6)	28.4 (6.6-56.9)	
Mobility				
Mean (S.D.)	15.7 (15.2)	28.3 (45.8)	20.0 (29.4)	0.957
Median (IQR)	15.3 (0.0-22.7)	5.0 (0.0 -52.5)	14.9 (0.0-33.4)	
Social interaction				
Mean (S.D.)	16.4 (11.2)	12.4 (11.4)	15.0 (11.3)	0.313
Median (IQR)	12.9 (9.0-27.0)	11.0 (0.9 - 19.7)	12.6 (5.8 - 25.0)	
Ambulation				
Mean (S.D.)	20.1 (17.8)	18.3 (17.9)	19.5 (17.6)	0.944
Median (IQR)	17.1 (4.2 - 33.6)	12.1 (4.8 - 28.9)	12.1 (4.2 - 31.0)	
Alertness, behaviour				
Mean (S.D.)	29.7 (34.2)	26.8 (37.7)	28.7 (34.9)	0.661
Median (IQR)	20.5 (0.0 - 54.2)	5.0 (0.0 - 39.3)	18.5 (0.0 - 46.7)	
Communication				
Mean (S.D.)	28.2 (24.5)	14.7 (12.0)	23.6 (21.8)	0.256
Median (IQR)	20.3 (9.2 - 51.9)	15.4 (2.4 - 21.1)	19.3 (9.2 - 41.0)	
Recreation and past times				
Mean (S.D.)	25.7 (22.5)	28.1 (25.5)	26.5 (23.2)	0.827
Median (IQR)	27.5 (0.0 - 42.0)	18.6 (3.5 - 57.2)	21.3 (0.0 - 43.8)	
Eating				
Mean (S.D.)	3.8 (7.3)	3.9 (6.4)	3.8 (6.9)	0.966
Median (IQR)	0.0 (0.0 - 5.1)	0.0 (0.0 - 9.5)	0.0 (0.0 - 5.2)	
Physical score				
Mean (S.D.)	20.7 (14.3)	20.5 (19.6)	20.6 (16.0)	0.626
Median (IQR)	17.0 (9.7 - 35.5)	13.2 (6.4 - 37.1)	16.5 (9.1 - 35.5)	
Psychosocial score				
Mean (S.D.)	15.6 (12.7)	13.1 (13.3)	14.7 (12.8)	0.520
Median (IQR)	14.2 (5.1 - 29.2)	8.2 (2.0 - 23.1)	10.2 (3.6 - 24.2)	
Work				
Mean (S.D.)	36.1 (0.0)	32.1 (9.6)	34.6 (6.1)	0.070
Median (IQR)	36.1 (36.1-36.1)	36.1 (36.1- 36.1)	36.1 (36.1- 36.1)	
Overall score				
Mean (S.D.)	20.2 (12.0)	20.7 (17.9)	20.4 (14.1)	0.702
Median (IQR)	20.2 (10.3-30.2)	13.3 (8.0 - 34.5)	19.8 (9.7 - 31.5)	

Table 4.10 UKSIP (T2) Follow-up Scores (n = 23)

Category/Dimension	Males	Females	Total	Mann-Whitney p-value
Sleep and rest				
Mean (S.D.)	21.2 (19.3)	30.9 (37.1)	25.4 (28.1)	0.707
Median (IQR)	12.2 (10.7-28.7)	12.2 (11.9-43.2)	12.2 (11.6-24.2)	
Emotional behaviour				
Mean (S.D.)	11.4 (15.7)	10.6 (16.1)	11.1 (15.5)	0.897
Median (IQR)	6.5 (0.0 - 22.5)	7.7 (0.0 - 12.3)	6.5 (0.0 - 16.2)	
Body care and movement				
Mean (S.D.)	18.2 (19.8)	14.6 (16.8)	16.6 (18.2)	0.641
Median (IQR)	8.5 (4.4 - 35.4)	7.6 (2.4 - 23.7)	7.8 (3.2 - 24.6)	
Home management				
Mean (S.D.)	25.5 (24.4)	28.9 (34.0)	27.0 (28.3)	0.826
Median (IQR)	15.9 (4.6 - 39.8)	23.1 (0.0 - 48.4)	15.9 (0.0 - 39.1)	
Mobility				
Mean (S.D.)	15.6 (15.3)	20.4 (24.2)	17.7 (19.3)	0.950
Median (IQR)	11.3 (0.0 - 26.0)	10.5 (0.0 - 41.3)	11.3 (0.0 - 28.7)	
Social interaction				
Mean (S.D.)	16.4 (12.2)	10.2 (8.3)	13.7 (10.9)	0.171
Median (IQR)	13.5 (4.5 - 27.9)	10.6 (0.0 - 17.0)	11.9 (3.5 - 22.6)	
Ambulation				
Mean (S.D.)	21.5 (14.2)	19.4 (21.7)	20.6 (17.4)	0.455
Median (IQR)	22.9 (7.0 - 37.4)	14.3 (0.0 - 28.9)	17.1 (4.2 - 37.4)	
Alertness, behaviour				
Mean (S.D.)	20.2 (26.4)	26.9 (34.6)	23.1 (29.7)	0.698
Median (IQR)	9.7 (0.0 - 41.5)	15.6 (0.0 - 46.0)	9.7 (0.0 - 37.1)	
Communication				
Mean (S.D.)	24.5 (25.7)	9.4 (17.0)	17.9 (23.2)	0.036
Median (IQR)	11.4 (9.7 - 36.8)	0.0 (0.0 - 12.5)	9.7 (0.0 - 31.7)	
Recreation and past times				
Mean (S.D.)	31.1 (25.9)	31.2 (24.6)	31.1 (24.8)	0.755
Median (IQR)	40.0 (0.0 - 48.9)	25.0 (8.5 - 56.2)	31.8 (8.5 - 50.0)	
Eating				
Mean (S.D.)	8.9 (17.0)	5.8 (7.7)	7.5 (13.6)	0.784
Median (IQR)	0.0 (0.0 - 11.1)	2.6 (0.0 - 11.6)	0.0 (0.0 - 10.9)	
Physical score				
Mean (S.D.)	21.1 (16.9)	19.5 (20.2)	20.4 (18.0)	0.577
Median (IQR)	17.8 (6.2 - 36.6)	15.0 (1.3 - 33.5)	15.7 (2.8 - 32.0)	
Psychosocial score				
Mean (S.D.)	13.3 (12.9)	11.8 (11.7)	12.7 (12.1)	0.975
Median (IQR)	9.8 (3.7 - 21.9)	5.9 (4.2 - 18.4)	6.2 (4.5 - 16.8)	
Work				
Mean (S.D.)	33.8 (8.0)	32.4 (7.7)	33.2 (7.7)	0.448
Median (IQR)	36.1 (36.1- 36.1)	36.1 (30.0-36.1)	36.1 (36.1-36.1)	
Overall score				
Mean (S.D.)	20.5 (13.5)	19.8 (17.9)	20.2 (15.1)	0.756
Median (IQR)	19.4 (9.7 - 30.7)	11.6 (7.3 - 30.8)	18.8 (8.4 - 28.2)	

Part II: Applicability and specificity

Applicability and Specificity of the SF-36 – Comparison against non PD controls

For comparison, SF-36 scores from this study were compared against age-matched controls (Brazier et al., 1992) and PD patients reported from a population based study (given in 10 year age bands) (Schrag et al., 2000a) – (Table 4.11).

In a comparison of SF-36 scores with the general UK population Schrag et al 2000 showed that PD patients identified from General Practice lists in the London area had poorer SF-36 scores than did the general population in the same age group. The greatest differences from the general population norms were seen in the domains of physical and social functioning, physical role limitations, and general health perceptions, particularly in the younger age group (45-54 years old). In contrast, pain, emotional role limitations, vitality and mental health were only slightly worse for the patients with PD than for the age-matched UK sample (Figures 4.3, 4.4 and 4.5).

In this study a similar trend was observed with a few exceptions. PD patients from this study were patients who routinely attended an out-patient neurology clinic and had poorer SF-36 scores in all domains (excluding physical role limitations in the youngest age band) across all the age groups, compared to the general population norms. Social functioning was considerably impaired in the out-patient group particularly in the younger and older age bands compared to controls. Physical functioning was more impaired in the younger age band compared to controls and the other age bands (Figures 4.3, 4.4 and 4.5). In the Schrag population based prevalence study, PD patients reported better health status in the youngest age band for social functioning, role emotional, mental health and bodily pain. The notable difference here is that in this study, social functioning in the younger age band was considerably impaired compared to controls and in the Schrag study the opposite was observed.

Table 4.11: SF-36 PD Patients versus Controls (Mean (SD) SF-36 scores by age in comparison to the general UK population)

	45 – 54 yrs			55 – 64 yrs			65 – 74 yrs		
	UK POPULATION CONTROLS(BRAZIER J.E ET AL., 1992)	PD PATIENTS(SCHRAG ET AL., 2000A)	**ASHFORD PD PATIENTS N = 3**	UK POPULATION CONTROLS(BRAZIER ET AL., 1992)	PD PATIENTS(SCHRAG ET AL., 2000A)	**ASHFORD PD PATIENTS N = 5**	UK POPULATION CONTROLS(BRAZIER ET AL., 1992)	PD PATIENTS(SCHRAG ET AL., 2000A)	**ASHFORD PD PATIENTS N = 14**
Physical Functioning	84.0	68.7 (10.3)	**61.7 (18.9)**	74.0	62.9 (24.0)	**68.0 (25.9)**	60.0	50.8 (32.3)	**56.4 (30.7)**
Social Functioning	87.0	87.5 (17.7)	**51.7 (23.2)**	84.0	65.0 (25.6)	**72.5 (16.5)**	80.0	59.9 (27.4)	**60.0 (29.5)**
Role Physical	83.0	62.5 (32.3)	**91.7 (14.4)**	72.0	62.9 (24.0)	**40.0 (45.4)**	59.0	35.3 (54.7)	**48.2 (41.0)**
Role Emotional	82.0	83.3 (19.2)	**50.0 (70.7)**	80.0	79.0 (52.5)	**66.7 (47.1)**	73.0	63.5 (52.8)	**69.0 (46.2)**
Mental Health	72.0	81.0 (17.1)	**45.3 (10.1)**	74.0	62.0 (23.1)	**73.6 (19.1)**	73.0	62.6 (20.4)	**65.2 (17.5)**
Vitality	59.0	55.0 (26.8)	**40.0 (18.0)**	59.0	50.0 (18.7)	**37.0 (23.6)**	57.0	41.3 (22.1)	**41.9 (23.5)**
Bodily Pain	77.0	81.0 (21.9)	**58.0 (39.3)**	73.0	61.4 (32.8)	**60.4 (19.8)**	67.0	62.2 (28.7)	**63.7 (21.2)**
General Health	70.0	49.0 (24.0)	**44.0 (24.8)**	65.0	48.5 (19.5)	**38.4 (21.3)**	58.0	45.8 (19.9)	**51.4 (19.1)**

Figure 4.3: Mean SF-36 Summary Scores in comparison to the general UK population (45-54 year old band)

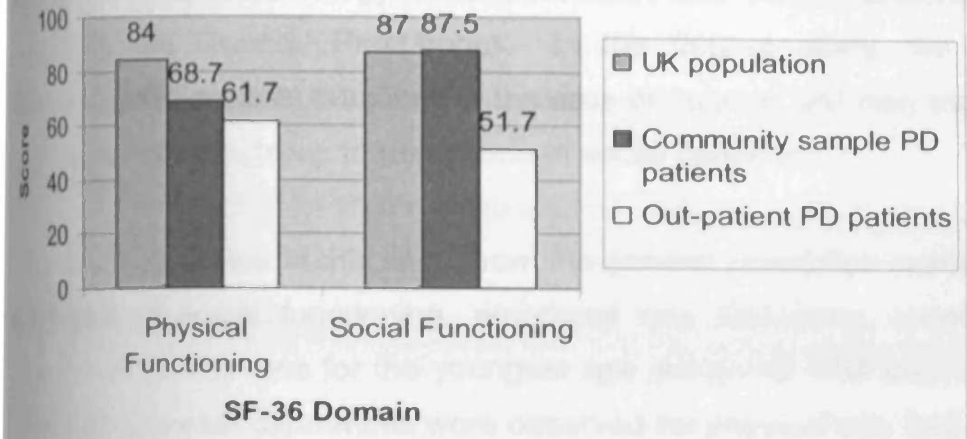


Figure 4.4: Mean SF-36 Summary Scores in comparison to the general UK population (55-64 year old band)

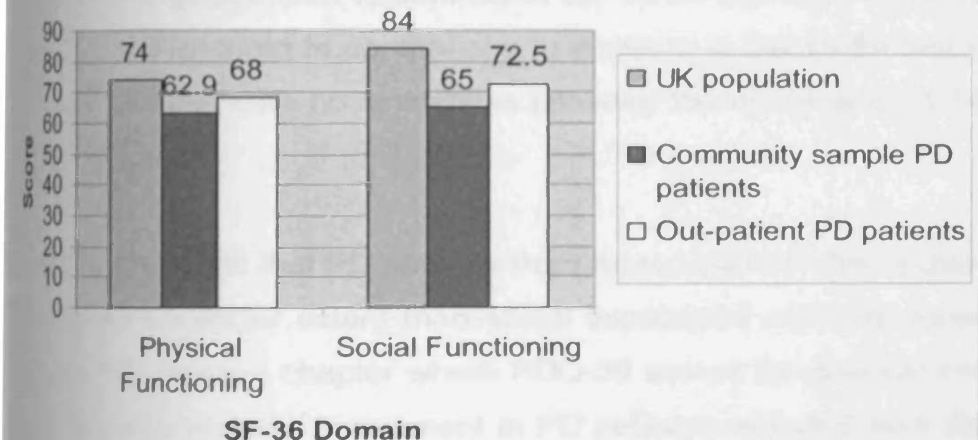
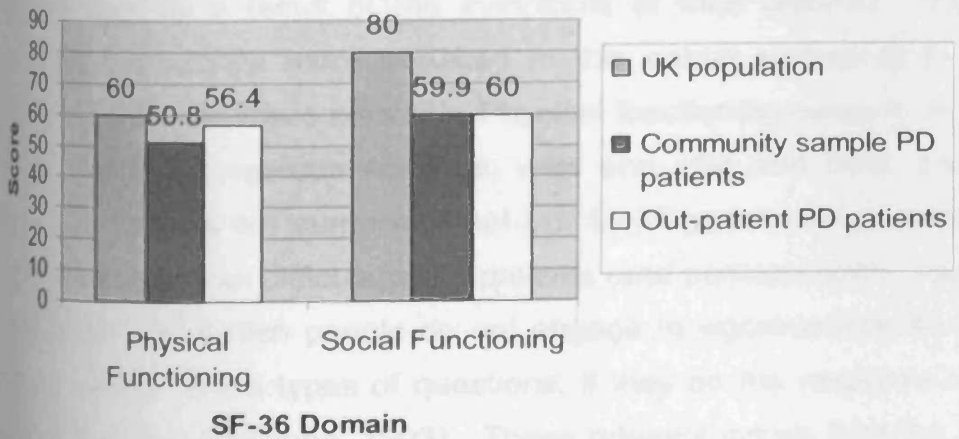


Figure 4.5: Mean SF-36 Summary Scores in comparison to the general UK population (65-74 year old band)



The key differentiation in the two populations is that in this study the patients were routinely attending a neurology out-patient clinic and were therefore referred to a specialist by the General Practitioner. In the Schrag study, the patients were recruited from the general practices in the area of London and may therefore contain a more mobile or able group to participate in social activities.

The greatest differences in this study from the general population norms were seen in the domains of social functioning, emotional role limitations, mental health and general health perceptions for the youngest age group (45 – 54 yrs). In the 55 – 64 yrs group the greatest differences were observed for physical role limitations, general health perceptions and vitality. In the oldest group with comparable controls (65 – 74 yrs) the greatest differences from the general population norms were seen in the domains of social functioning, vitality and physical role limitations. Social functioning was impaired in comparison to controls in the 45-54 and 65-74 year old age groups and vitality was impaired in comparison to controls in the 55-64 and 65-74 year old age groups. There were no similarities between the 45-54 and 55-74 year old age groups.

These results suggest that PD patients from an out-patient clinic experience physical impairment to a greater extent than social impairment and are consistent with the findings of the previous chapter where PDQ-39 scores for physical impairment were greater than psychosocial impairment in PD patients recruited from the membership list of the UK Parkinson's Disease Society. Interestingly, it would appear that the PD patients in this study experienced a higher level of social functioning impairment in contrast to physical functioning when compared to age matched controls. This highlights that PD has a considerable impact on a patient's ability to participate in social activities as a result of the symptoms of their disease. Though physical functioning was slightly more impaired in this cohort compared to age matched controls, the difference was minimal. Physical functioning refers to the respondent's ability to undertake vigorous activities, walk one mile and climb several flights of stairs. It is perhaps not surprising that the 55-74 years old group of controls had found such activities as difficult as PD patients (and perhaps some younger controls). A vast majority of British people do not engage in vigorous activities and are thus unable to answer these types of questions; if they do the response would have no meaning(Hunt and McKenna, 1993). These relevant issues limit the applicability of

the SF-36 in PD even though the instrument was able to distinguish between healthy and PD population. If the questions in an instrument are deemed irrelevant by the patient then there is a high likelihood that the question will not be answered.

SF-36 *Missing Data*

Missing data can compromise the quality of trials having a HRQoL component. Missing data can be subdivided into *missing forms* (the whole questionnaire is missing) and *missing items* (a subset of items are not filled in by the patient). Missing forms may be due to staff related factors e.g. the study coordinator at site forgot to hand out or mail the questionnaire. Missing items may be problematic if the patients skip some items in the questionnaire because they perceive these items as too intrusive (Groenvold, 1999) (e.g. items on sexuality), because the questionnaire was felt to be too long, or due to mistakes (unclear lay out or complicated completion instructions). In general, missing forms are the main reason for missing data in clinical trials: patients typically fill in a high percentage of items, unless there are specific problems with upsetting or intrusive items or the questionnaire is too long (Ocoba and Zee, 1998). An analysis of SF-36 missing answers revealed that more questions were not answered at the second time of completion for all the dimensions (with the exception of General health) and most markedly for the *Role-emotional* and *Role-physical* (Table 4.12 & Figure 4.6).

The SF-36 asks the respondent to "answer every question", and to give the "best answer" when they are unsure about how to answer; one would therefore expect very few missing questions. More than 10% of subjects failed to complete questions in the dimensions *Role Physical* and *Role Emotional* at the follow-up visit. The most common questions not answered at both time points were; "vigorous activities" (baseline = 9.38%, follow-up = 10.34%), "cut down on amount of time spent at work" as result of physical health (baseline = 12.5%, follow-up = 10.34%), "cut down on amount of time spent at work" as result of emotional problems (baseline = 9.38%, follow-up = 13.79%). The most common questions not answered at the second time of completion (follow-up) were; "limited in the kind of work or other activities" due to physical health (13.79%), "accomplished less than you would like" due to emotional problems (13.79%), "limited in kind of work or activities" due to physical health (13.79%). There was a very low level of item redundancy (number of missing questions) for the remaining questions in the SF-36. Questions 3f (health limiting

ability to bend, kneel or stoop) and 7 (bodily pain during last 4 weeks) were affirmed in very high frequencies and may therefore not yield any information regarding differences between patients and thus detract from the discriminative properties of the questionnaire. Overall the SF-36 produced a low level of missing items with notable exceptions outlined above.

Applicability and specificity of the UKSIP – Comparison against non PD controls

The mean scores for the PD and control groups are presented in Table 4.13 and Figure 4.7. The UKSIP scores for controls were extracted from a study by Salek(Salek, 1997). These controls were "healthy" elderly age-sex matched for rheumatoid arthritis (RA) and osteoarthritis (OA) patients. There were 93 controls, 51 Male: 42 Female with median age 62 years (range 24-82). The results show that there was a marked deficit in HRQoL for patients who had PD, compared with non PD controls. This deficit was seen across all twelve domains of the UKSIP. The differences between the two groups for all domains were prominent, with PD patients scoring considerably higher, indicating worst possible HRQoL, than that for non PD patients. The PD group of patients recorded significantly higher scores for the domains of *work, home management, alertness & behaviour and communication* compared to the non-PD controls (Figure 4.7).

Compared to controls, PD patients in the study suffered a great deal of functional impairment, both physical and psychosocial. This validates the applicability of the UKSIP in a PD population.

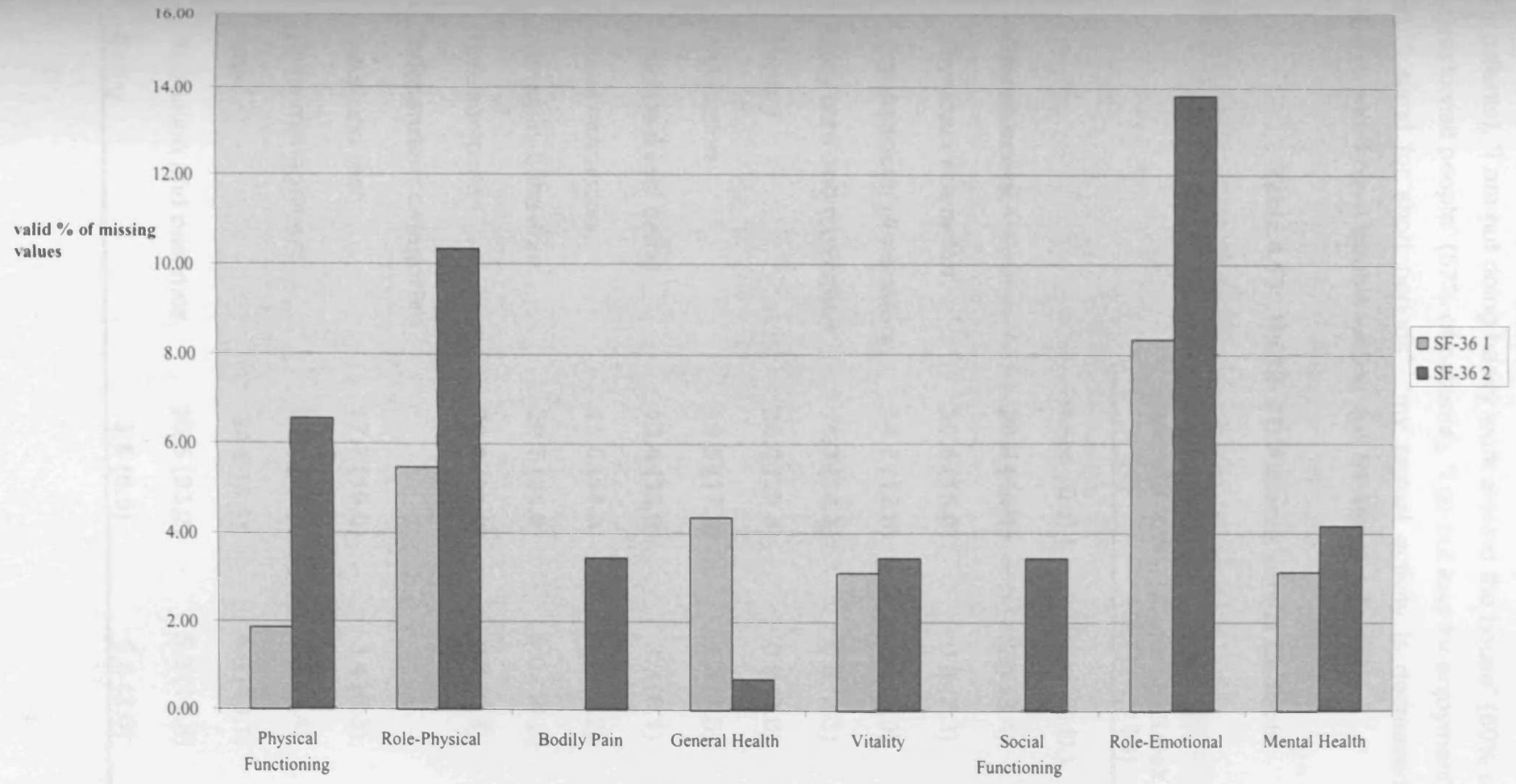
UKSIP – Affirmative responses

The most prevalent dysfunctional behaviour, categorised as more than half of the respondents, in patients with PD as measured by the UKSIP is presented in Table 4.14.

Table 4.12 Missing items analysis on SF-36

SF36 Category	Question no	Baseline % missing	Follow-up % missing	Baseline Scale % missing	Follow-up Scale % missing
Physical Functioning	qn3a	9.38	10.34	**1.88**	**6.55**
	qn3b	3.13	6.90		
	qn3c	0.00	3.45		
	qn3d	0.00	10.34		
	qn3e	0.00	6.90		
	qn3f	0.00	0.00		
	qn3g	3.13	3.45		
	qn3h	3.13	10.34		
	qn3i	0.00	10.34		
	qn3j	0.00	3.45		
Role Physical	qn4a	12.50	10.34	**5.47**	**10.34**
	qn4b	3.13	6.90		
	qn4c	3.13	13.79		
	qn4d	3.13	10.34		
Role Emotional	qn5a	9.38	13.79	**8.33**	**13.79**
	qn5b	6.25	13.79		
	qn5c	9.38	13.79		
Social Functioning	qn6	0.00	3.45	**0.00**	**3.45**
	qn10	0.00	3.45		
Bodily Pain	qn7	0.00	0.00	**0.00**	**3.45**
	qn8	0.00	6.90		
Vitality	qn9a	0.00	3.45	**3.13**	**3.45**
	qn9e	6.25	6.90		
	qn9g	3.13	3.45		
	qn9i	3.13	0.00		
Mental Health	qn9b	3.13	6.90	**3.15**	**4.16**
	qn9c	3.13	3.45		
	qn9d	3.13	3.45		
	qn9f	3.23	3.57		
	qn9h	3.13	3.45		
General Health	qn1	3.13	0.00	**4.38**	**0.69**
	qn11a	6.25	3.45		
	qn11b	3.13	0.00		
	qn11c	3.13	0.00		
	qn11d	6.25	0.00		

Figure 4.6: Missing value analysis for SF-36 Scales

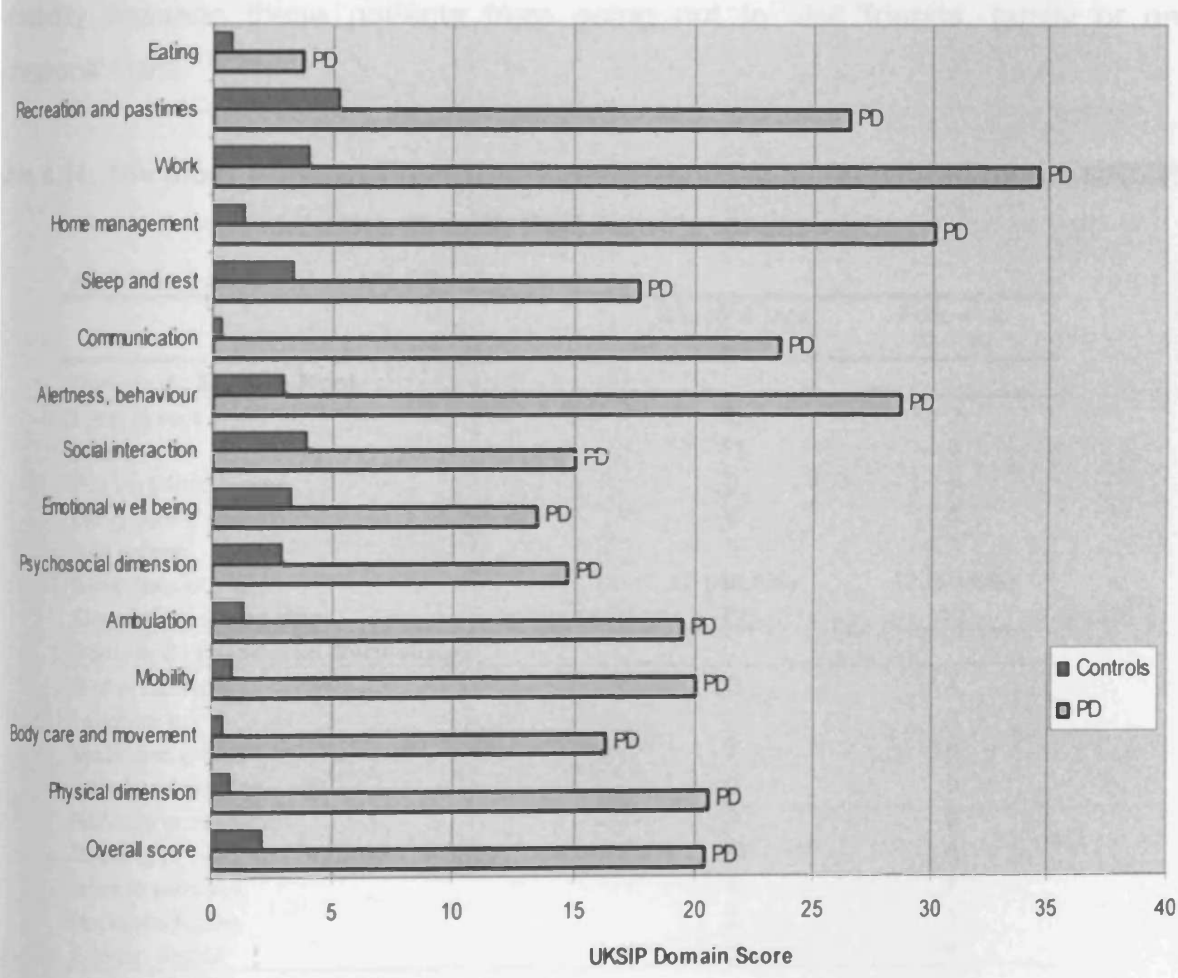


The most commonly endorsed items at the baseline visit were; "I walk more slowly" (77.1% of patients), "I am not doing heavy work around the house" (60% of patients), "I go out less to visit people" (57% of patients), "I go out less for enjoyment" (54.3% of patients"), "I stand for short periods", "my sexual activity is decreased", "I dress myself slowly", and "I have trouble writing" (all 51.4% of patients).

Table 4.13: UKSIP PD Patients versus Controls

	Ashford PD patients (n = 35)	**Controls(Salek, 1997) (n = 93)**
	mean (S.D.)	mean (S.D.)
Overall score	**20.4 (14.1)**	2.1 (3.4)
Physical dimension	**20.6 (16.0)**	0.8 (2.3)
Psychosocial dimension	**14.7 (12.8)**	2.9 (5.9)
Body care and movement	**16.3 (14.3)**	0.5 (2.3)
Mobility	**20.0 (29.4)**	0.9 (3.6)
Ambulation	**19.5 (17.6)**	1.3 (4.3)
Emotional well being	**13.4 (16.6)**	3.3 (8.1)
Social interaction	**15.0 (11.3)**	3.9 (7.2)
Alertness, behaviour	**28.7 (34.9)**	3.0 (10.2)
Communication	**23.6 (21.8)**	0.4 (2.4)
Independent categories		
Sleep and rest	**17.7 (19.0)**	3.4 (6.3)
Home management	**30.1 (27.4)**	1.3 (4.4)
Work	**34.6 (6.1)**	4.0 (13.1)
Recreation and pastimes	**26.5 (23.2)**	5.3 (11.3)
Eating	**3.8 (6.9)**	0.8 (2.6)

Figure 4.7: Mean Domain Scores for Parkinson's Disease (PD) and Control Groups



The most commonly endorsed items at the follow-up visit (up to 6 months after the baseline visit) were; "I walk more slowly" (82.7% of patients), "I go out less to visit people" (72.4% of patients), "I go out less for enjoyment" (65.5% of patients), "my sexual activity is decreased" (62.1% of patients), "I walk short distances" (58.6% of patients), and "I do no maintenance or repair work" (55.2% of patients). Ambulation and social interaction domains are the most commonly endorsed items at both time points which reflects how the physical aspect of this chronic disorder impedes the person's social situation. For example, in this study population, more than half reported that they walk more slowly, and walk shorter distances which in all probability dissuade these patients from going out to visit friends, family or on recreational visits.

Table 4.14: The most prevalent dysfunctional behaviour as measured by the UKSIP in patients with Parkinson's disease

No. of items endorsed on UK-SIP

	Baseline visit N = 35	Follow-up N = 29
Domain A: Sleep & Rest		
Lying to rest	2	3
Sitting	11	7
Dozing 24hrs a day	2	4
Lying during day	4	8
Half asleep	3	3
Sleep less at night	**17 (48.6%)**	**17 (58.6%)**
Sleep more during day	11	12
Domain B: Emotional Behaviour		
Bad or useless	5	4
Laugh or cry	2	4
Moan and groan	7	5
Attempted suicide	0	0
Nervous or restless	8	9
Rubbing or holding hurt areas of body	13	4
Irritable with self	4	7
Hopeless future	2	1
Sudden frights	8	6
Domain C: Body Care & Movement		
Difficult movements	9	10
Need help to move	3	5
Stand for short periods	**19 (54.3%)**	11
No balance	13	8
Limited hand movement	9	7
Stand with help	1	2
Bend down holding onto something	16	13
Restricted position	0	1
Clumsy	7	5
Use stick or support to get in and out of bed	12	7
Lying mostly	2	2
Change position frequently	9	7

No. of items endorsed on UK-SIP

	Baseline visit N = 35	Follow-up N = 29
Hold onto something to move in bed	8	7
Help with bathing	10	8
Can not bathe one self	1	1
Use a bed pan	0	0
Difficulty putting on socks etc.	12	11
No control of bladder	1	1
Difficulty in fastening clothes	6	6
Mostly partly dressed or in pyjamas	1	0
No control of bowels	0	0
Dresses oneself	**18 (51.4%)**	12
Dress with help	3	5
Domain D: Home Management		
Housework with breaks	15	9
Less daily chores	16	11
No daily chores	3	2
No usual maintenance jobs	15	**16 (55.2%)**
No shopping	10	8
No cleaning	7	4
Difficulty using hands	11	7
No washing	5	6
No heavy house work	**21 (60%)**	**15 (51.7%)**
No paying bills etc.	12	10
Domain E: Mobility		
Confined to building only	7	6
Confined to room only	1	1
Staying in bed more	2	2
Confined to bed	0	0
No use of public transport	8	9
Stay at home	14	13
Out if toilet is nearby	5	5
Don't go to town	7	5
Away from home for short periods	13	14
Help in the dark	13	9
Domain F: Social Interaction		
Out to visit less	**20 (57.1%)**	**21 (72.4%)**
Don't visit at all	5	2
Less interest in others problems	5	4
Irritable with those around	7	6
Less affection shown	2	2
Fewer social activities	**15 (42.8%)**	**15 (51.7%)**
Shorter stays at friends	9	8
Avoid having visitors	7	4
Decreased sexual activity	**19 (54.3%)**	**18 (62.1%)**
Concern over health	10	9
Talk less with those around	9	8
Demands on other people	3	3
Stay alone	7	7
Disagreeable with family	3	4
Frequently angry with family	1	1
Isolated	1	1
Pay less attention to the children	2	2
Refuse contact with family	0	0
Don't take care of family	2	1

No. of items endorsed on UK-SIP

	Baseline visit N = 35	Follow-up N = 29
Don't joke	2	3
Domain G: Ambulation		
Walk short distances	16	**17 (58.6%)**
Don't walk up and down hills	15	**15 (51.7%)**
Use stairs with physical aid	14	14
Use stairs with assistance from somebody	3	3
Get about in wheelchair	1	0
Don't walk	0	1
Walk with difficulty	15	12
Walk with help	1	2
Stairs tackled slowly	13	14
Don't use stairs	0	1
Use a walking frame	4	5
Walk more slowly	**27 (77.1%)**	**24 (82.7%)**
Domain H: Alertness, Behaviour		
Confused, more than one task at once	6	7
Minor accidents	11	13
Slower reactions	14	7
Don't finish	10	6
Difficulty reasoning	12	11
Confused, not knowing who people are	7	6
Forgetful	15	12
Short attention span	6	6
More mistakes	11	6
Difficulty in thought provoking jobs	12	8
Domain I: Communication		
Trouble writing	**18 (51.4%)**	14
Communicate non-verbally	1	1
Speech not generally understood	2	4
Lose control of voice	14	9
Only write to sign name	12	9
Sit close to carry a conversation	3	3
Speak with difficulty	11	6
Understood with difficulty	4	5
Do not speak clearly under stress	14	8
Domain K: Work		
Not working	2	1
Job from home	0	0
Not as much work done as usual	0	0
Irritable with work mates	0	0
Work shorter hours	2	2
Light work	1	0
Work for short periods	0	1
Use special aids or job swap	1	1
Jobs not so carefully and accurately done	0	1
Domain L: Recreation		
Less hobbies	**17 (48.6%)**	**15 (51.7%)**
Go out less for enjoyment	**19 (54.3%)**	**19 (65.5%)**
Less inactive pastimes	8	10
No inactive pastimes	0	2
More inactive pastimes	9	7
Fewer community activities	11	10
Less physical recreation	11	9

No. of items endorsed on UK-SIP

	Baseline visit N = 35	Follow-up N = 29
No physical recreation	9	10
Domain M: Eating		
Eat less	6	5
Special utensils for feeding	2	6
Eat no food only liquids	1	3
Eat special foods	0	0
Pick or nibble at food	1	1
Drink less	4	4
Help with feeding	3	3
Do not feed one self	0	0
Eat not food at all	0	0

In the study reported in chapter three, the most commonly endorsed items were "I have trouble writing" (78%), "I walk more slowly" (75%), "I take part in fewer social activities" (64%) and "my sexual activity is decreased" (61%). Having trouble with writing, (which usually results from tremor and/or insufficient control), is not as prominent in this study population and may be explained by the fact that the majority of patients were early stage or advanced stable PD patients. These patients were probably controlling their symptoms with medication and by regular clinic visits or in the case of early stage patients the symptoms were not yet prominent. The chapter three study patients mean duration of illness was longer and the mean age higher which may also in part explain this difference.

The UKSIP addresses items relevant to PD which are corroborated by the prevalence analysis. Sexual activity has shown to be affected and a concern to PD patients in this study (and the study reported in chapter three) but is not addressed in the SF-36. Preoccupation with health problems interfering with sex and dissatisfaction with body appearance are more prevalent in parkinsonian women compared to age and sex matched controls (Welsh et al., 1997).

Micrographia (small cramped hand writing) is a common motor symptom associated with PD, manifesting in hand writing problems. This is highlighted in the UKSIP as an item that impedes communication and is not addressed in the SF-36. Patients in this age bracket are likely to be on benefits and expected to signature a variety of documents to ensure social security payments are forthcoming. An inability to communicate due to hand writing difficulties is more likely to add to the psychological

distress experienced by the patient. The UKSIP addresses this and other communication issues such as difficulty with speech. Forty percent of patients at baseline reported that they often lose control of their voice. Speech becomes fainter and may be unintelligible; this is a common characteristic of PD. Communication is not adequately dealt with in the SF-36 where there are no items relating to a communication dimension.

The UKSIP and SF-36 both show that impairment is greater for physical compared to psychosocial dimensions in PD patients. Nevertheless the UKSIP reveals that >70% of patients "walk more slowly" which is not surprising as one of the major characteristics of PD is slowness of movement (akinesia). This is a very relevant item for PD patients and is absent in the SF-36, which focuses on the patient's ability to "walk 1 mile", walk or "climb several flights of stairs". These items seem to be irrelevant as a vast majority of the British population do not engage in vigorous activities and are thus unable to answer such questions. Hunt and McKenna, 1993 have commented in a similar vain with respect to the adequacy of these SF-36 items with respect to an elderly population and state that "some people would not attempt to walk a mile even in the best of health for reasons of safety, climate or indolence". The social interaction item "goes out less to visit people" featured prominently in UKSIP responders. The UKSIP extensively addresses social interaction, which is a dimension of disease which can greatly impact HRQoL. The UKSIP contributes 20 questions to this dimension compared the 2 in the SF-36 (social-extent and social-time).

Body care and movement items; "stand for short periods" (54.3%) and "dresses very slowly" (51.4%) were the 4th and 5th most prevalent problems reported at the baseline visit. The clinical features of PD include tremor, rigidity, and slowness of movement and loss of postural reflexes. It is not surprising that these physical impairments will have an obstructive impact on the patient's ability to dress or stand for long periods. These and other equally important items such as ability to bathe oneself get in and out of bed, clumsiness, etc are included in the UKSIP *body care & movement* domain but are not adequately covered in the SF-36. The SF-36 dedicates one item to such matters; "Bathing or dressing yourself" as part of the PF dimension compared to 23 items on the UKSIP relating to *body care & movement*.

More than half of the UK respondents reported that they were not doing heavy house work around the house and more than 40% of patients reported that they were not doing any of the maintenance or repair work that they would usually do. This highlights how PD impairs a person's ability to undertake usual household chores and contributes to a poorer quality of life. PD patients indicated that the greatest dysfunction was in the areas of *home management*. This clearly identifies *home management* as a contributory factor to poor health-related quality of life in PD patients and is a dimension that should be considered in greater depth when evaluating PD patients HRQoL. The SF-36 contributes only one question to address the impact of disease on the patient's ability to carry out normal home related activities i.e. "lifting or carrying groceries" and there is no dedicated *home management* component.

Another key dimension that is included in the UKSIP but not the SF-36 is *sleep & rest*. Patients showed a marked sleep disturbance during the night with 48.6% reporting that the "sleep less at night" at baseline, increasing to 58.6% at the follow-up assessment. Sleep problems are an under-emphasised cause of disability in Parkinson's disease (PD) and may be seen independently of PD, associated with primary PD pathology, or as a result of anti-parkinsonian medications. Common sleep disorders include excessive daytime sleepiness, rapid eye movement (REM) sleep behavior disorder, night-time wakefulness and restless legs syndrome(Stacy M, 2002) Sleep disturbances can lead to a reduced quality of life in PD patients and should be incorporated in a HRQoL instrument.

Comparison of UKSIP and SF-36 Scores

Both questionnaires are scored on a scale between 0 and 100, although in different directions. A score of zero on the UKSIP indicates that patients are enjoying the best possible HRQoL in that domain of functioning whilst on the SF-36 it indicates the worst possible score for the scale. A summary of the scoring achieved at baseline for the two questionnaires in the study group of PD patients is presented in Tables 4.7 and 4.11 and Figures 4.1 & 6 and the results have been remarked upon above. A wide range of scores is reported for the various subscales of both questionnaires suggesting heterogeneity in the population sampled. The mean and median scores in both instruments exhibit parity which suggests the absence of outliers in the data set.

Direct graphical comparisons of mean scores between the two questionnaires are not straightforward because they have scales, which measure HRQoL in opposing directions. To provide some comparison of the mean scores, the scaling of the SF-36 was reversed by subtracting 100 from each domain score (hence a perfect score would be zero). This was done for illustrative purposes only and cannot be quantitatively analysed with any degree of true confidence. The results of this manipulation are shown in Figure 4.8. The SF-36 scores were reversed because it is ultimately scored on a relatively simple percentage scale, whilst the UKSIP has external preference weights attached to each item. Figure 4.8 shows that patients reported lower levels of dysfunction on the UKSIP scales than the SF-36, though it must be reiterated that direct comparison is not feasible. The UKSIP mean domain scores were mainly around the 10 – 25% level, whilst the adjusted SF-36 reported 30 - 50% levels of dysfunction across the domains.

Part III: Reliability

Subjective health status measures should produce the same results with repeated use under the same conditions. Although it is possible to assess reliability using the test-retest approach, it may be difficult to distinguish a change in health status from any measurement error (Staniszewska, 1998). Reliability can be assessed by examining the internal consistency of items with the aim of assessing the degree of agreement between items. Internal consistency reliability was measured using Cronbach's alpha coefficient. The closer the coefficient is to 1, the greater the homogeneity between the items and therefore, the greater the confidence that can be attributed that items relate to the domain under investigation. However, caution should be noted as alpha coefficients of >0.95 can mean that several of the items are in fact measuring the same thing (Thompson and Yu, 2003).

Internal Consistency of the SF-36

The internal consistency reliability (Table 4.15) for all SF-36 domains were good (a 0.722 – 0.946 at baseline and 0.760 – 0.942 at follow-up visit). The SF-36 items within the scales are measuring similar areas of functioning in PD patients, which adds validity to the scoring profile produced by the questionnaire. These results suggest that the SF-36 has good internal consistency reliability in PD patients.

Figure 4.8 Comparison of the Mean Impact of PD on HRQoL as Measured by the UKSIP and (Adjusted) SF-36

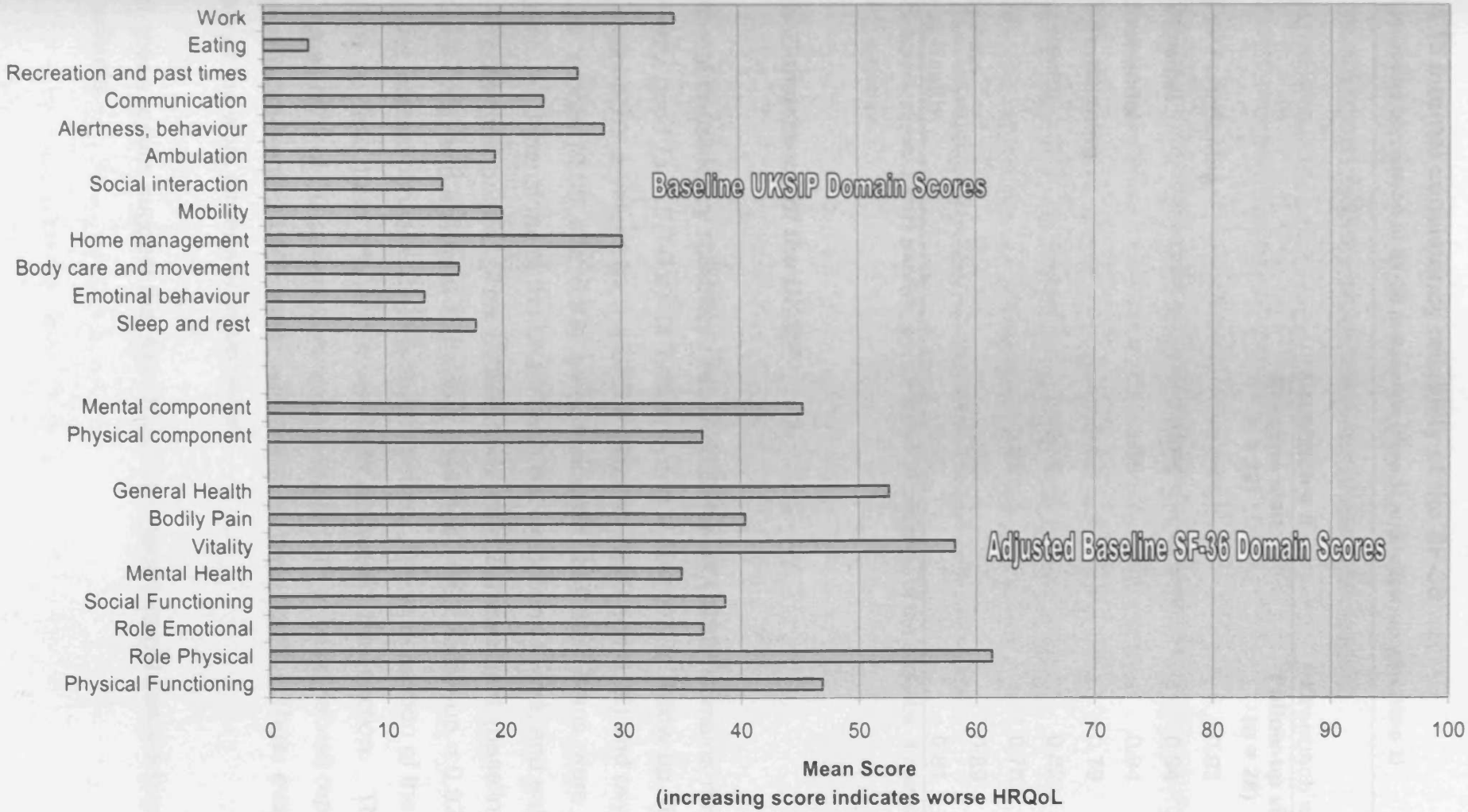


Table 4.15 Internal consistency reliability of the SF-36

Internal reliability correlation of SF-36 at baseline (time 1) and follow-up visit (time 2)

Scale	Cronbach's a Baseline visit (n = 32)	Cronbach's a Follow-up visit (n = 28)
Physical Functioning	0.91	0.93
Role Physical	0.86	0.94
Role Emotional	0.95	0.94
Social Functioning	0.87	0.79
Mental Health	0.78	0.82
Vitality	0.85	0.76
Bodily Pain	0.86	0.89
General Health	0.72	0.81

High Cronbach's alphas (>0.7) indicate good internal consistency of the scale i.e. it measures what it purports to measure

Internal Consistency of the UKSIP

The internal consistency reliability (Table 4.16) for all UKSIP domains ranged from low to very good (a 0.37–0.93 at baseline and 0.16–0.93 at follow-up visit). The strong Cronbach's a values for the UKSIP physical (0.90 and 0.92) and psychosocial summary scores (0.90 and 0.93) gave indication that their items were internally consistent. All dimensions of the UKSIP with the exception of *work*, and *eating* had a > 0.6 at both time points. The UKSIP body care & movement (baseline = 0.83; follow-up = 0.88) and alertness behaviour (baseline =0.93; follow-up = 0.92) showed the highest internal consistency at both test points. The work section of the UKSIP is redundant as less than 2% of respondents answered this section. This is not surprising as 80% of PD patients completing the UKSIP at baseline visit reported that they were retired and 28.6% were retired due to their health. Thus, evaluation of reliability of the work dimension is pointless.

Overall, these results suggest that the UKSIP has good internal consistency reliability in PD patients.

Table 4.16: Internal consistency reliability of the UKSIP

UKSIP Category	Cronbach's a Baseline visit (n = 35)	Cronbach's a Follow-up visit (n = 29)
Overall UKSIP	0.77	0.95
Physical dimension	0.90	0.92
Psychosocial dimension	0.94	0.93
Sleep and rest	0.53	0.70
Emotional behaviour	0.69	0.66
Body care & movement	0.83	0.88
Home management	0.83	0.84
Mobility	0.70	0.76
Social interaction	0.62	0.71
Ambulation	0.81	0.79
Alertness behaviour	0.93	0.92
Communication	0.72	0.84
Work	0.29	0.21
Recreation and past times	0.75	0.74
Eating	0.37	0.16

High Cronbach's alphas (>0.7) indicate good internal consistency of the scale i.e. it measures what it purports to measure

Part IV: Construct validity

Construct validity must be investigated whenever no criterion is accepted as entirely adequate to define the quality of life to be measured (Cronbach, 1955). Specifically, construct validity focuses on the assessment of whether a particular measure relates to other measures consistent with theoretically derived hypotheses concerning the concepts (or constructs) that are being measured (Zeller and Carmines, 1980). During construct validation, evidence must be provided demonstrating that the measured construct behaves consistently with its referent theoretical properties. The following aspects should be considered when establishing construct validity for a HRQOL measure.

1. The measured variable should correlate at a low level with measures of related but different constructs (divergent validity).
2. The measured variable should correlate positively with related and similar constructs (convergent validity).
3. The measured variable should draw information from the intended HRQOL construct.

Relationship between Hoehn & Yahr staging and disease classification

The severity of symptoms of Parkinson's disease patients can be better classified by their clinical staging, using the Hoehn & Yahr scale which is based grossly upon impairment (tremor and rigidity) and disability (ability to stand or walk unassisted). Correlations between Hoehn & Yahr staging and disease classification (1. early stage PD; 2a. advanced stable PD, 2b. advanced with motor fluctuations, 2c. advanced with insufficient control of symptoms) are presented in Table 4.17. Hoehn and Yahr staging and disease classification are both determined by the clinician. There is a strong and statistically significant association between staging of disease and the patient's disease classification (r_s = 0.774; p<0.001). This highlights that as staging increases (from stage I to IV) the patient is more likely to be classified as deteriorating or progressing in disease state (from early stage to advanced PD). This supports the use of the Hoehn & Yahr as a suitable clinical construct to correlate against the two generic instruments as part of the construct validity assessment.

Table 4.17: Correlations* between Hoehn & Yahr staging and disease classification

Baseline visit	Hoehn and Yahr Stage					
	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Total
Early stage patient	9	2				11
Advanced-Stable	3	3	2			8
Advanced-Motor fluctuations			2			2
Advanced-Insufficient control		3	4	3		10
Total	12	8	8	3	0	31

*Spearman's ccrrelation coefficient = 0.774 (p-value <0.001)

Relationships between duration of disease and Hoehn & Yahr staging and disease classification

Correlations between disease duration and Hoehn & Yahr staging and disease classification are presented in Table 4.18. There is strong association between disease duration and both disease classification (r_s 0.796, p<0.001) and Hoehn & Yahr staging (r_s 0.641, p<0.001). These results indicate that as duration of disease increases the severity of symptoms worsen.

Table 4.18 Correlations between duration of disease and Hoehn & Yahr staging/disease classification (n = 31)

Baseline	Spearman rank r_s	Correlation p-value
Disease classification	0.796	0.001
Hoehn & Yahr staging	0.641	0.001

Relationships between SF-36 and Hoehn & Yahr (indicator of disease severity)

Hoehn and Yahr assessments were conducted by the clinician at both baseline and follow-up visits, on the day of their outpatient appointment. Spearman's rank correlation coefficients for the association between Hoehn and Yahr staging and SF-36 domain scores are presented in Table 4.19. Negative Correlation coefficients (r_s) were found to be statistically significant at p=0.05, p=0.01, and p=0.005 for the domains *physical functioning*, *role physical*, *role emotional*, and *mental health* and showed that increasing disease severity was associated with both reduced physical and psychological aspects of quality of life. At the baseline visit, correlations were highest with the dimensions measuring physical functioning and social functioning.

In the SF-36 the social dimension contains items asking respondents to what extent their health, including physical function, has limited their ability to engage in social activities. Hence this dimension partly taps aspects of physical, as well as social well being (Jenkinson et al., 1995).

Table 4.19: Correlations between the Hoehn & Yahr and SF-36 Domain Scores (n = 36)

	Spearman rank correlation			
	Baseline		Follow-up	
SF-36 Domain	rs	p-value	rs	p-value
Physical Functioning	-0.581	0.001	-0.520	0.011
Role Physical	-0.428	0.020	-0.604	0.002
Role Emotional	-0.414	0.0317	-0.431	0.045
Social Functioning	-0.544	0.002	-0.386	0.062
Mental Health	-0.611	0.00	-0.455	0.029
Vitality	-0.422	0.023	-0.289	0.179
Bodily Pain	-0.358	0.052	-0.205	0.336
General Health	-0.306	0.113	-0.061	0.777
Reported Health Transition	0.066	0.732	0.245	0.248
Physical Component Score	-0.332	0.112	-0.306	0.166
Mental Component Score	-0.648	0.001	-0.450	0.019

High negative correlations for the mental domain at baseline which have not been previously reported in the literature would suggest that severity of disease has a negative impact on the person's state of mind and anxiety levels. This data may be revealing that the subject is demoralised, i.e. feeling sad as opposed to depressed. Feeling of sadness is often time limited and secondary to adverse psychosocial circumstances such as stress or disappointment in failing to live up to ones own expectations.

There was a negative relationship between all the remaining domains and the Hoehn and Yahr staging indicating a trend towards a reduced HRQoL although none reached a level of statistical significance at both visits.

Relationships between UKSIP and Hoehn & Yahr (indicator of disease severity)

Spearman's rank correlation coefficients for the association between Hoehn and Yahr staging and UKSIP domain scores are presented in Table 4.20. Correlation coefficients (r_s) were found to be statistically significant at p=0.05, p=0.01, and

p=0.005 for the domains *body care and movement*, *social interaction*, and *ambulation* at both time visits. Disease severity has a negative impact on *social interaction* which is not surprising as PD patients symptoms lead to physical impairment such as tremor and involuntary body movement can lead to a sense of self-isolation and a fear of social contact. The strongest correlations at baseline were observed for *body care and movement*. The strong association with *body care and movement* is not unexpected as the Hoehn and Yahr in part measures disability such as the ability to stand or walk unassisted which is similar to the body care and movement domain which addresses the patient's ability to stand, keep their balance and stand with someone's help. The lower correlations between clinical measure and the UKSIP were observed for *sleep and rest*, *recreation and pastimes* and *eating* and might suggest that the UKSIP assesses aspects of well being which is not captured by the clinical measure.

Statistically significant associations were found between the two generic measures and the clinical indicator of disease severity used in the study, suggesting construct validity in PD. Weak correlations in both measures were present for some domains suggesting that the HRQoL instrument taps into areas of well-being not covered in the clinical instrument. The UKSIP showed that *social interaction*, and *ambulation* correlated significantly with the Hoehn and Yahr and these two domains were also the most commonly endorsed items at both time points which most likely explain the strong correlations observed here.

Convergent Validity

In this study the UKSIP was assessed against the SF-36 to assess convergent validity. Convergent Validity is one of the main tools for the validation of quality of life questionnaires. The hypothesis tested is that the UKSIP domains should correlate with related variables of the SF-36. The degree of association between the UKSIP overall and dimension scores and SF-36 component scores were measured using Spearman's correlation coefficients and are presented in Table 4.21 and 4.22.

Table 4.20: Spearman correlations between the Hoehn & Yahr and UKSIP Domain Scores (n = 36)

	Spearman rank correlation			
	Baseline		Follow-up	
UKSIP Domains	r_s	p-value	r_s	p-value
Sleep and rest	0.269	0.130	0.109	0.648
Emotional well being	0.482	0.005	0.439	0.053
Body care and movement	0.513	0.002	0.531	0.016
Home management	0.312	0.078	0.649	0.002
Mobility	0.237	0.184	0.292	0.211
Social interaction	0.390	0.025	0.474	0.035
Ambulation	0.455	0.008	0.665	0.001
Alertness, behaviour	0.424	0.014	0.183	0.440
Communication	0.454	0.008	0.401	0.080
Recreation and pastimes	0.205	0.260	0.153	0.519
Eating	0.100	0.586	0.168	0.479
Physical dimension	0.488	0.004	0.625	0.003
Psychosocial dimension	0.556	0.001	0.418	0.067
Work	0.202	0.294	0.432	0.073
Overall UKSIP	0.458	0.007	0.406	0.076

It would be expected that domains aiming to measure similar areas of function would be correlated with one another.

Convergent validity was demonstrated by significant correlation of summary scales for the UKSIP with the overall scores of the UKSIP at both time points, Baseline and follow-up visit ($p < 0.0001$). The UKSIP overall score correlated most strongly with the SF-36 physical component (Baseline = -0.767 & Follow-up = -0.606) compared to the SF-36 mental component (Baseline = -0.392 & Follow-up = -0.389) scales on both occasions. The correlation coefficients changed little from baseline to follow-up, demonstrating a high degree of reproducibility.

Table 4.21 Relationships† between the SF-36 component and UKSIP Summary Scores – Baseline visit (n = 36)

	UK-SIP Physical dimension	UK-SIP Psychosocial dimension	UK-SIP Overall score	M0 SF-36 Physical Component Score	M0 SF-36 Mental Component Score
UK-SIP Physical dimension		0.747**	0.914**	-0.752**	-0.341*
UK-SIP Psychosocial dimension	0.747**	-	0.878**	-0.613**	-0.482*
UK-SIP Overall score	0.914**	0.878**	-	-0.767**	-0.392*
M0 SF-36 Physical Component Score	-	-	-	-	-0.026
M0 SF-36 Mental Component Score	-	-	-	-0.026	-

†Spearman rank correlation coefficient
*** p=0.05; **p=0.001**

Table 4.22 Relationships† between the SF-36 component and UKSIP Summary Scores – Follow-up visit (n = 25)

	UK-SIP Physical dimension	UK-SIP Psychosocial dimension	UK-SIP Overall score	M0 SF-36 Physical Component Score	M0 SF-36 Mental Component Score
UK-SIP Physical dimension	-	0.782**	0.878**	-0.672**	-0.246
UK-SIP Psychosocial dimension	0.782**	-	0.907**	-0.557*	-0.595*
UK-SIP Overall score	0.878**	0.907**	-	-0.606*	-0.389
M0 SF-36 Physical Component Score	-	-	-	-	0.075
M0 SF-36 Mental Component Score	-	-	-	0.075	-

†Spearman rank correlation coefficient
*** p=0.05; **p=0.001**

The UKSIP physical dimension score correlated significantly with the SF-36 physical component but not the mental component scores at both time points. The UKSIP psychosocial dimension correlated significantly with the SF-36 physical and mental components at both time points. It is surprising that the UKSIP psychosocial dimension correlated with the SF-36 physical component but this may indicate that the physical component partly taps aspects of social well being and interactions by asking question such as "cut down on amount of time you spent on work or other activities" .

The UKSIP and SF-36 domains were correlated against the UKSIP overall scores at both time points. The ranking of the SF-36 domain correlation coefficients demonstrated both a clear ordering in size and reproducibility between baseline and the follow-up visit (Table 4.23). Social functioning, role-physical and physical functioning had the greatest impact on this PD population and bodily pain and mental health having the least. This illustrated the relative contributions of the SF-36 domains in explaining patients' overall quality of life. The ranking of the UKSIP domain correlation coefficients also demonstrated both a clear ordering in and reproducibility between baseline and the follow-up visit (Table 4.24). Body care and movement, home management and recreation and past times had the greatest impact on this PD population and work and emotional well being having the least. This illustrated the relative contributions of the UKSIP domains in explaining patients' overall quality of life.

There are similarities in the ranking of coefficients in that the similar domains of mental heath (SF-36) and emotional well-being (UKSIP) contributed the least to overall quality of life in this sample of PD patients. Social functioning (SF-36) and recreational past-times (UKSIP) are related in that they both address social activities and hence the high ranking of these domain suggest that the ability to participate in social activities is impaired by PD resulting in poorer HRQoL. Bodily pain has been reported to reduce HRQoL(Karlsen et al., 2000) in a four year follow-up study but it seems to have had little impact in this study population. The reason for this could be that the follow-up in this study was limited to 6 months and disease had not progressed in that time to adversely affect matters relating to pain. The work domain in the UKSIP is redundant as a significant number of patients in this study were retired or not working.

Table 4.23 Ranking* of correlation coefficients for relationships between SF-36 domain scores and UKSIP Overall Score (n = 25)

	UKSIP Overall Score		Ranking of coefficients	
SF-36 Domain	Time 1 r_s (p-value)	Time 2 r_s (p-value)	Time 1	Time 2
Physical Functioning	**0.788 (0.0001)**	**0.730 (0.0001)**	10	8
Role Physical	**0.786 (0.0001)**	**0.788 (0.0001)**	9	10
Role Emotional	**0.765 (0.0001)**	**0.657 (0.0001)**	8	6
Social Functioning	**0.864 (0.0001)**	**0.894 (0.0001)**	11	11
Mental Health	**0.644 (0.001)**	**0.642 (0.001)**	3	5
Vitality	**0.708 (0.0001)**	**0.462 (0.05)**	7	2
Bodily Pain	**0.530 (0.005)**	**0.611 (0.001)**	2	3
General Health	**0.705 (0.0001)**	**0.636 (0.001)**	6	4
Reported Health Transition	**-0.406 (0.05)**	**-0.311 (0.130)**	1	1
Physical Component Score	**0.683 (0.0001)**	**0.751 (0.0001)**	5	9
Mental Component Score	**0.682 (0.0001)**	**0.674 (0.0001)**	4	7

***Correlation coefficients ranked in order of size (1=smallest, 11= largest)**

Table 4.24 Ranking* of correlation coefficients for relationship between UKSIP domain scores and UKSIP Overall Score (n = 31)

	UKSIP Overall score		Ranking of coefficients	
UKSIP Domain	Time 1 r_s (p-value)	Time 2 r_s (p-value)	Time 1	Time 2
Sleep and rest	**0.316 (0.084)**	**0.674 (0.001)**	1	6
Emotional well being	**0.489 (0.005)**	**0.558 (0.009)**	3	2
Body care and movement	**0.818 (0.0001)**	**0.956 (0.0001)**	10	14
Home management	**0.826 (0.0001)**	**0.865 (0.0001)**	11	11
Mobility	**0.721 (0.0001)**	**0.655 (0.001)**	5	4.5
Social interaction	**0.773 (0.0001)**	**0.849 (0.0001)**	6	10
Ambulation	**0.816 (0.0001)**	**0.764 (0.0001)**	9	7
Alertness, behaviour	**0.789 (0.0001)**	**0.852 (0.0001)**	8	9
Communication	**0.783 (0.0001)**	**0.655 (0.001)**	7	4.5
Recreation and pastimes	**0.837 (0.0001)**	**0.822 (0.0001)**	12	8
Eating	**0.584 (0.001)**	**0.560 (0.008)**	4	3
Physical dimension	**0.926 (0.0001)**	**0.935 (0.0001)**	14	13
Psychosocial dimension	**0.874 (0.0001)**	**0.927 (0.0001)**	13	12
Work	**0.369 (0.041)**	**0.364 (0.105)**	2	1

***Correlation coefficients ranked in order of size (1=smallest, 11= largest)**

Discriminative Validity of the SF-36 and UKSIP

Discriminative validity (i.e. the extent to which the questionnaire scores were able to discriminate between subgroups of patients differing in clinical state) was assessed by evaluating the SF-36 and UKSIP ability to discriminate between:

- Stage I, II, and III/IV Hoehn & Yahr staging. The Hoehn and Yahr is a 5 point category scale of disease severity and is a useful scale to categorise the severity of patients PD. Higher categories or stages indicate deterioration in clinical status and therefore health with stage I indicating mild disease and stage IV/V indicating severe disease.

It was anticipated that quality of life scores would worsen as symptom severity increased. The Kruskall-Wallis one-way ANOVA was used to test whether or not the questionnaires were able to differentiate between the three categories of disease severity (stage III and IV were grouped together due to the small sample size). Results are presented in Table 4.25 and in Figures 4.9, 4.10, 4.11 and 4.12.

SF-36 ability to discriminate between H&Y stages:

Clinicians rating of the severity of disease symptoms using the H&Y showed a statistically significant pattern of lower scores (indicating worse health) for *physical functioning, role physical, role emotional, social functioning*, and *mental health* domains as well as the mental component ($p<0.05$). The other SF-36 scales; *vitality, bodily pain* and physical component score showed a pattern of lower scores but were not statistically significant. These results show that differences exist among the means but do not indicate differences between which means e.g. stage 1 and stage 2 or stage 1 and stage 3/4. A post hoc analysis was not employed in this case. Once you have determined that differences exist among the means, post hoc range tests and pair wise multiple comparisons can determine which means differ. Range tests identify homogeneous subsets of means that are not different from each other. Pair wise multiple comparisons test the difference between each pair of means, and yield a matrix where asterisks indicate significantly different group means at an alpha level of 0.05.

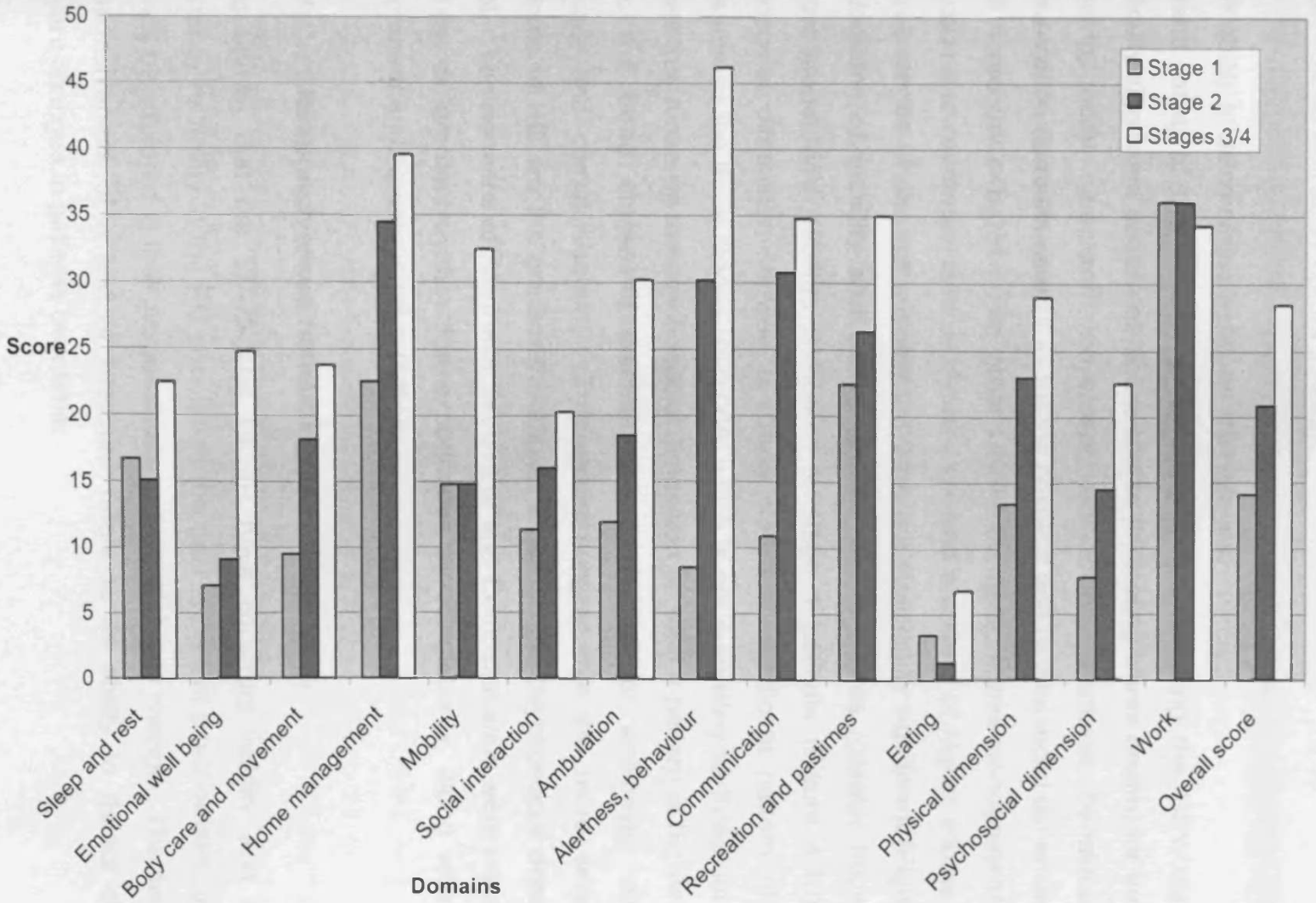
Overall the SF-36 demonstrated a level of discriminative sensitivity and was able to distinguish between patients with different disease severities, and in particular

between Stage I and stage III/IV patients (Figure 4.11), with the exception of *General Health*, and *Reported Health Transition*. The biggest differences exist between stage 1 and stage 3/4 patients (Figure 4.12) for the domains of *role physical*, *role emotional* and *social functioning* which highlights that the more severely impaired patients are hindered in their ability to work or undertake other regular activities due to physical and emotional problems which interferes with their social interactions.

What is interesting in these findings is that *physical role limitations* was found to be the most impaired of all the dimensions as reflected in the lower mean scores in this study population and this is echoed by a clear indication that physical health is impaired with patients with greater disease severity. The contrary is found for the scales that had the least impairment; *role limitations because of emotional difficulties* and *social functioning*. The discriminative analysis showed that these domains had much lower scores (indicative of worse health) in the severe patients group compared to the "mild", stage 1 group. The key observation from this is that within group differences may vary considerably in the various dimensions of HRQoL and in this population the more "severe" patients experience greater disability in psychosocial domains than the whole study population grouped together.

What was surprising is that the Physical component score did not show a statistically significant pattern of lower scores between the disease severity staging; considering that the domains that form part of the "physical" component of the SF-36 (physical functioning, and role physical) reached statistical significance. The cause of this may relate to the very low physical score for stage II patients on the *role physical* domain (mean score 9.4) indicating patient's worse health. Stage I and II means scores for this domain was 65.9 and 25.0 respectively. This might suggest that in this study population the severity of disease in patients that has progressed from "mild " severity to "moderate" severity has a considerable impact on their ability to continue satisfactorily with daily activities. The most severe PD patients having higher scores in comparison to the moderate severity group may relate to their lower expectations to undertake such daily activities such as work.

Figure 4.9 HRQoL Scores versus Severity of PD: UKSIP



The General Health scale for stage 1 patients was elevated indicating better HRQoL compared to stage II, and III/IV. Stage III/IV patients reported slightly better General health score compared to stage II patients which could be attributed to more severely impaired patients having a lower expectation of their well-being. Future evaluation of the discriminative properties of the SF-36 should include more stable measures of disease such as the UPDRS which addresses the functional problems associated with disease as well as symptom severity.

UKSIP ability to discriminate between H&Y stages:

Clinicians rating of the severity of disease symptoms using the H&Y showed a statistically significant pattern of higher scores (indicating worse health) for *emotional well-being, body care and movement, ambulation, alertness behaviour,* and *communication* domains as well as for the physical and psychosocial dimensions and overall score (all $p<0.05$). The other UKSIP domains; *home management, social interaction and recreation and past-times* showed a pattern of higher scores across the three stages of disease severity but were not statistically significant (Figure 4.9). The domains of mobility and eating showed higher scores (greater impairment) amongst stage III/IV patients compared to stage 1 patients (Figure 4.10). The psychosocial dimension showed a strong statistical significant pattern of higher scores across the three stages ($p<0.005$) which is not surprising as three out of four domains that make up the psychosocial dimension showed a pattern of higher scores across the three stages of disease severity (*emotional well-being, alertness behaviour* and *communication*). This would suggest that the more severe the symptoms of PD are the greater the impact it has on the psychosocial domains of HRQoL. The domains of *Alertness, behaviour* and *Communication* were reported as the areas of greatest dysfunction experienced by PD patients, along with *Home management* and *Recreational activities.*

Part V: Responsiveness (sensitivity) to change

Having shown that the SF-36 and UKSIP have construct validity and internal consistency reliability in the PD population the next important psychometric property that must be explored is their responsiveness (sensitivity) to change. The sensitivity or responsiveness of a test or measure refers to its ability to detect clinically significant changes in patients over time.

A comparison of the SF-36 and UKSIP scores at baseline and follow-up were analysed using Wilcoxon matched pair tests for non-parametric variables to provide an indication as the whether the two generic instruments were able to detect change in scores of patients whose HRQoL had changed over a 6 month period. Tables 4.26 and 4.27 and Figures 4.13 and 4.14 show that the difference in domain scores between baseline and follow-up were not statistically significant though there were marginal improvements and deteriorations in each of the scale scores.

Role emotional and role physical showed the greatest magnitude of difference between baseline and follow-up in the SF-36, the trend being for improvement in functioning (Figure 4.13). In the UKSIP, alertness behaviour and communication showed the greatest magnitude of difference, again the trend being for improvement (Figure 4.14). The non-significance of these results illustrates that over the 6 month period the magnitude of improvement or deterioration in health status was not sufficiently dramatic and that this cohort of patients had in effect stable disease. The change between baseline and follow-up assessments are not statistically significant. These initial sets of result indicate that the analysis of responsiveness to change will be difficult to determine. It must be noted that small sample size could have also contributed to the instruments lack of responsiveness in detecting change over time.

The patient population was divided into groups for testing depending on the degree of change demonstrated during the 6 month period. These groups were those that showed no change and those that showed deterioration based on the i) self-reported change in condition question (Tables 4.28 and 4.29), and ii) Hoehn & Yahr scale (Tables 4.30 and 4.31).

Baseline data (t1) includes all first completions of the SF-36 and UKSIP while follow-up data (t2) consists of all questionnaires completed after a time interval of 6 months. The "p" values for each test demonstrate the individual responsiveness of each of the SF-36 and UKSIP categories. The ability of both instruments to detect change was tested for all those who showed any degree of deterioration as well as for those who showed no change.

Figure 4.10 HRQoL Scores vs Severity of PD: UKSIP Comparison of Satge 1 vs Stage 3/4 Patients

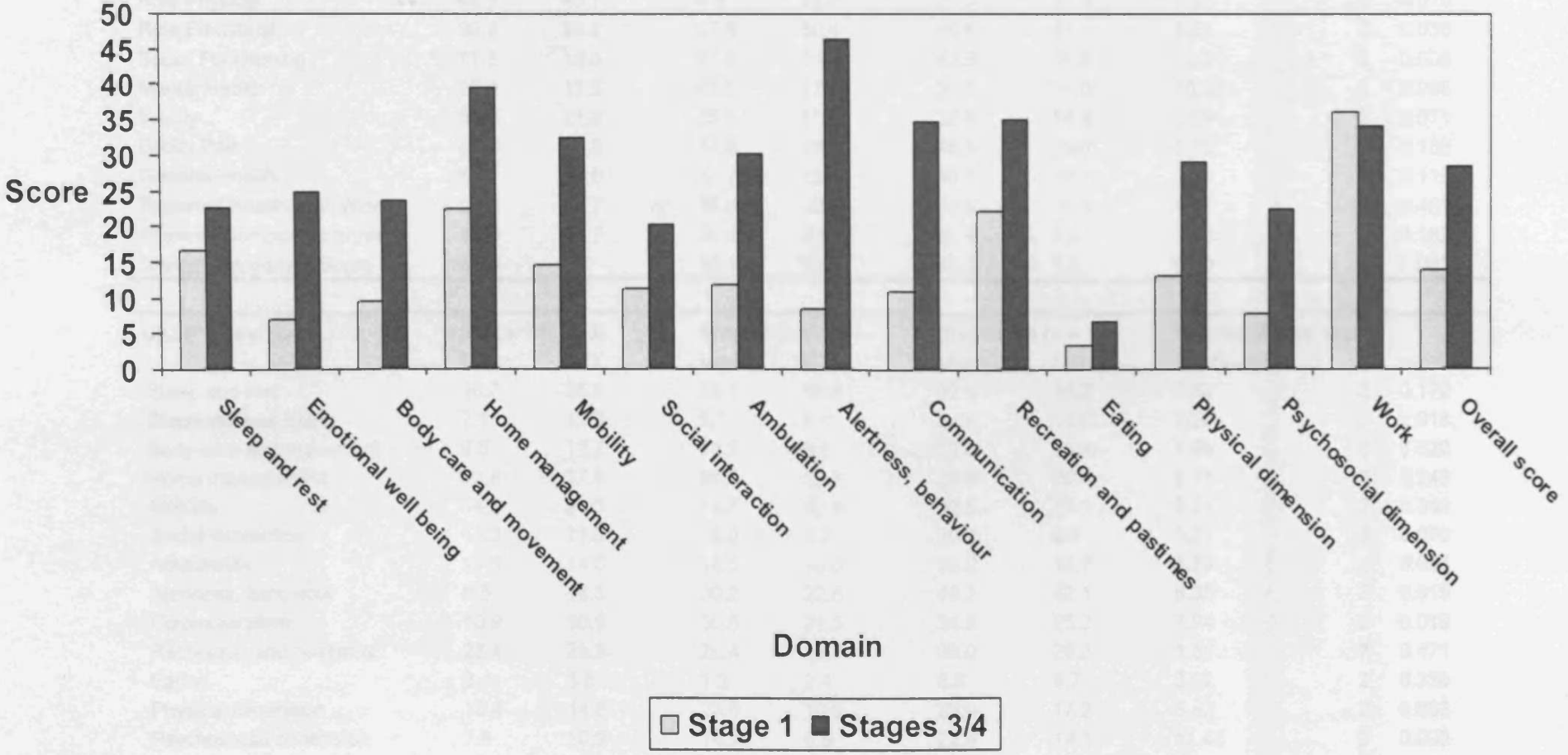


Table 4.25: DISCRIMINATIVE ANALYIS – Comparisons between disease severity and quality of life scores

SF-36 (Baseline)	Stage 1 (n = 13)		Stage 2 (n = 8)		Stages 3/4 (n = 12)		Kruskal Wallis test		
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Chi-Square	df	p-value
Physical Functioning	69.8	26.2	47.5	25.9	35.2	25.7	9.31	2	0.010
Role Physical	65.9	40.7	9.4	18.6	25.0	37.3	9.29	2	0.010
Role Emotional	90.9	30.2	47.6	50.4	44.4	47.1	6.64	2	0.036
Social Functioning	77.5	16.0	57.8	24.5	43.3	24.9	10.20	2	0.006
Mental Health	77.0	17.3	65.0	12.0	51.6	14.5	10.55	2	0.005
Vitality	50.0	21.5	35.0	15.1	32.8	14.4	5.29	2	0.071
Bodily Pain	65.5	18.9	57.0	28.5	46.6	25.5	3.72	2	0.156
General Health	56.1	18.5	37.7	19.4	40.1	16.2	4.26	2	0.119
Reported Health Transition	56.8	11.7	65.6	22.9	57.5	16.9	1.55	2	0.461
Physical Component Score	40.3	11.7	31.4	9.6	31.6	5.2	3.40	2	0.182
Mental Component Score	52.5	8.0	45.4	11.2	35.5	7.9	9.79	2	0.007

UKSIP Baseline	Stage 1 (n =13)		Stage 2 (n = 8)		Stages 3/4 (n = 12)		Kruskal Wallis test		
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Chi-Square	df	p-value
Sleep and rest	16.7	25.8	15.1	10.5	22.5	16.2	3.52	2	0.172
Emotional well being	7.1	13.8	9.1	9.4	24.8	19.0	7.99	2	0.018
Body care and movement	9.5	12.7	18.2	9.8	23.8	15.8	7.66	2	0.022
Home management	22.6	27.8	34.5	19.8	39.6	30.4	2.78	2	0.249
Mobility	14.8	21.5	14.8	12.8	32.5	42.1	2.11	2	0.349
Social interaction	11.3	12.9	16.0	8.2	20.3	9.8	5.31	2	0.070
Ambulation	11.9	14.0	18.5	16.0	30.2	18.7	6.79	2	0.034
Alertness, behaviour	8.5	16.5	30.2	22.8	46.2	42.1	8.35	2	0.015
Communication	10.9	10.6	30.8	21.5	34.8	25.3	7.94	2	0.019
Recreation and pastimes	22.4	25.2	26.4	16.7	35.0	25.3	1.51	2	0.471
Eating	3.4	5.8	1.3	2.4	6.8	9.7	2.22	2	0.330
Physical dimension	13.4	14.8	23.0	10.9	29.0	17.2	6.82	2	0.033
Psychosocial dimension	7.8	10.9	14.5	6.5	22.6	14.1	11.43	2	0.003
Work	36.1	0.0	36.1	0.0	34.3	5.7	1.90	2	0.387
Overall score	14.1	13.3	20.9	7.1	28.5	15.4	7.15	2	0.028

Figure 4. 11 HRQoL Score versus Severity of PD: SF-36

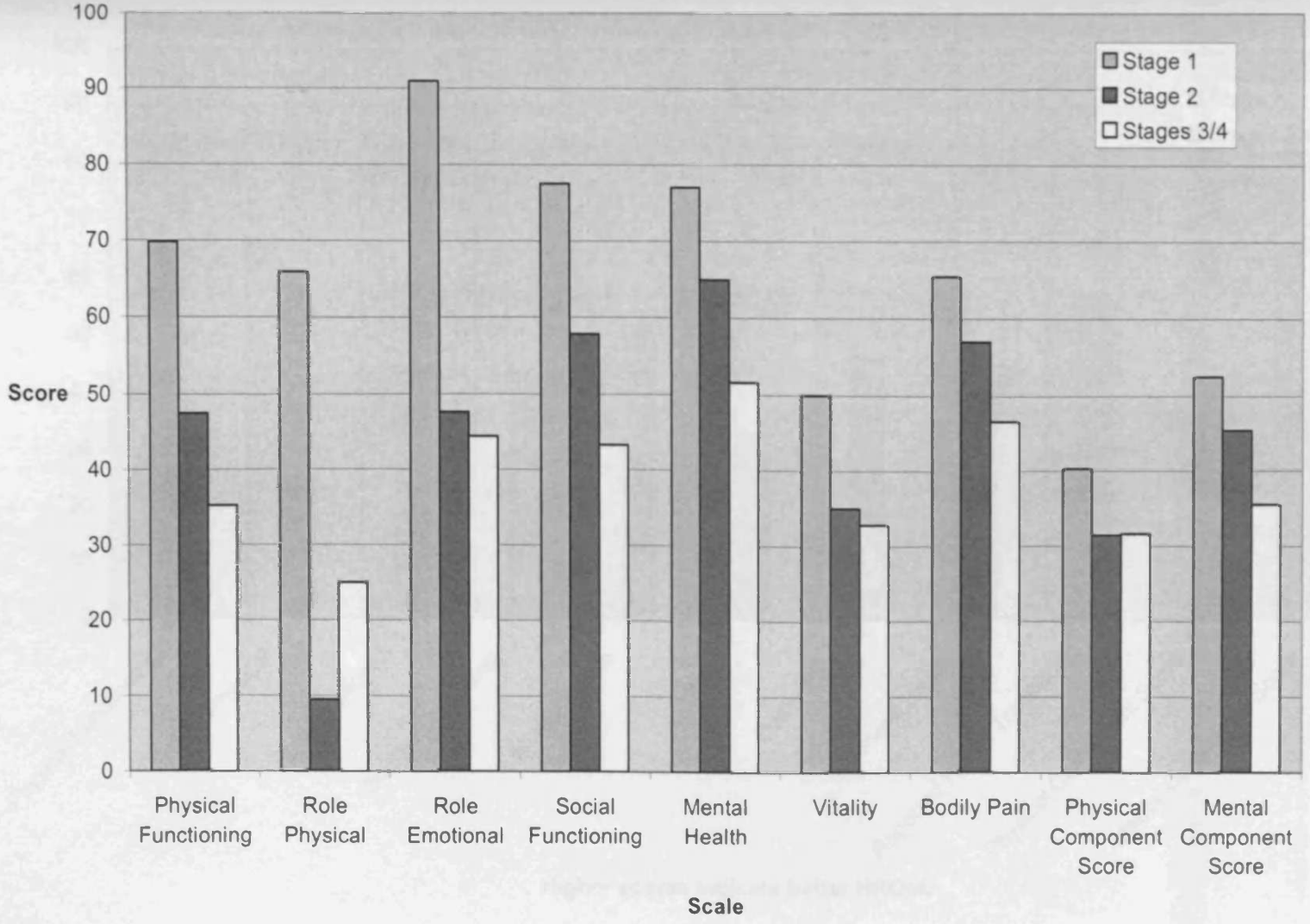
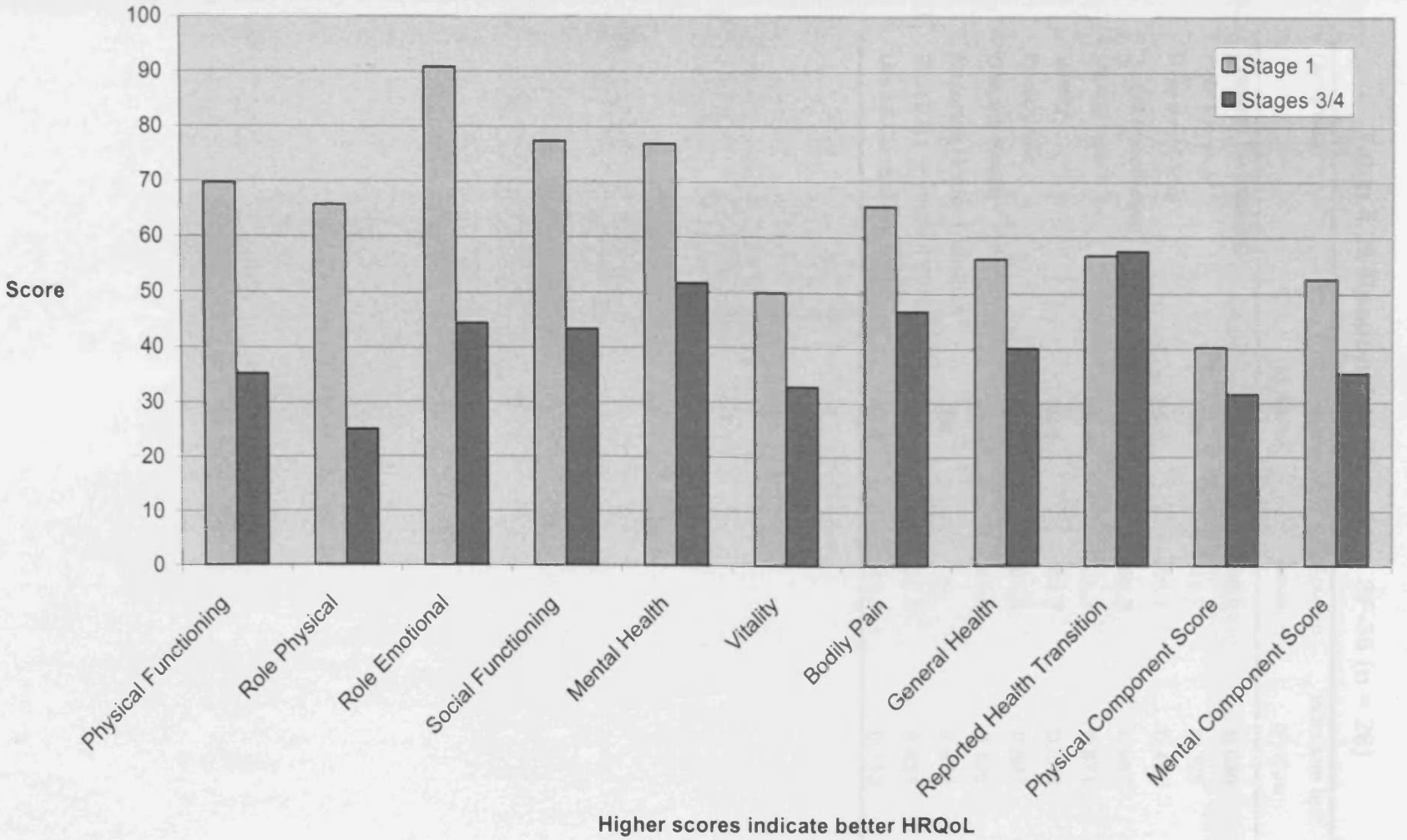


Figure 4.12 HRQoL Scores vs Severity of PD: SF-36
comparison of Stage 1 versus Stage 3/4

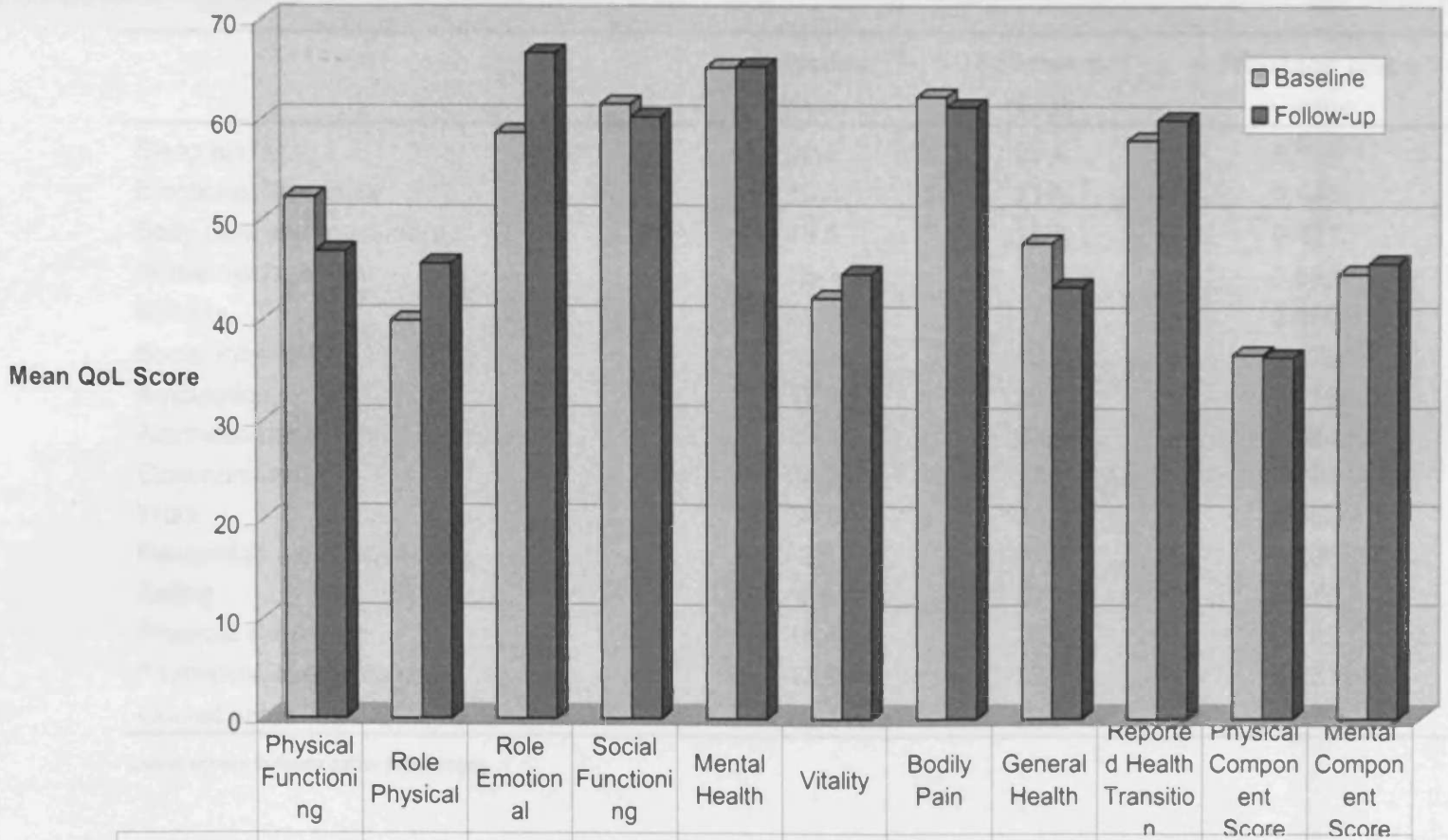


Comparison of SF-36 and UK-SIP scores at baseline and follow-up.

Table 4.26 Responsiveness of the SF-36 (n = 26)

Dimension	Baseline Mean	Follow-up Mean	Wilcoxon test p-value
Physical Functioning	52.3	46.9	0.099
Role Physical	40	45.7	0.325
Role Emotional	58.7	66.7	0.435
Social Functioning	61.6	60.3	0.987
Mental Health	65.2	65.3	0.871
Vitality	42.3	44.7	0.555
Bodily Pain	62.3	61.3	0.897
General Health	47.8	43.5	0.126
Reported Health Transition	58	60	0.672
Physical Component Score	36.8	36.5	0.931
Mental Component Score	44.8	45.8	0.715

Figure 4.13 Mean HRQoL Score as measured by the SF-36



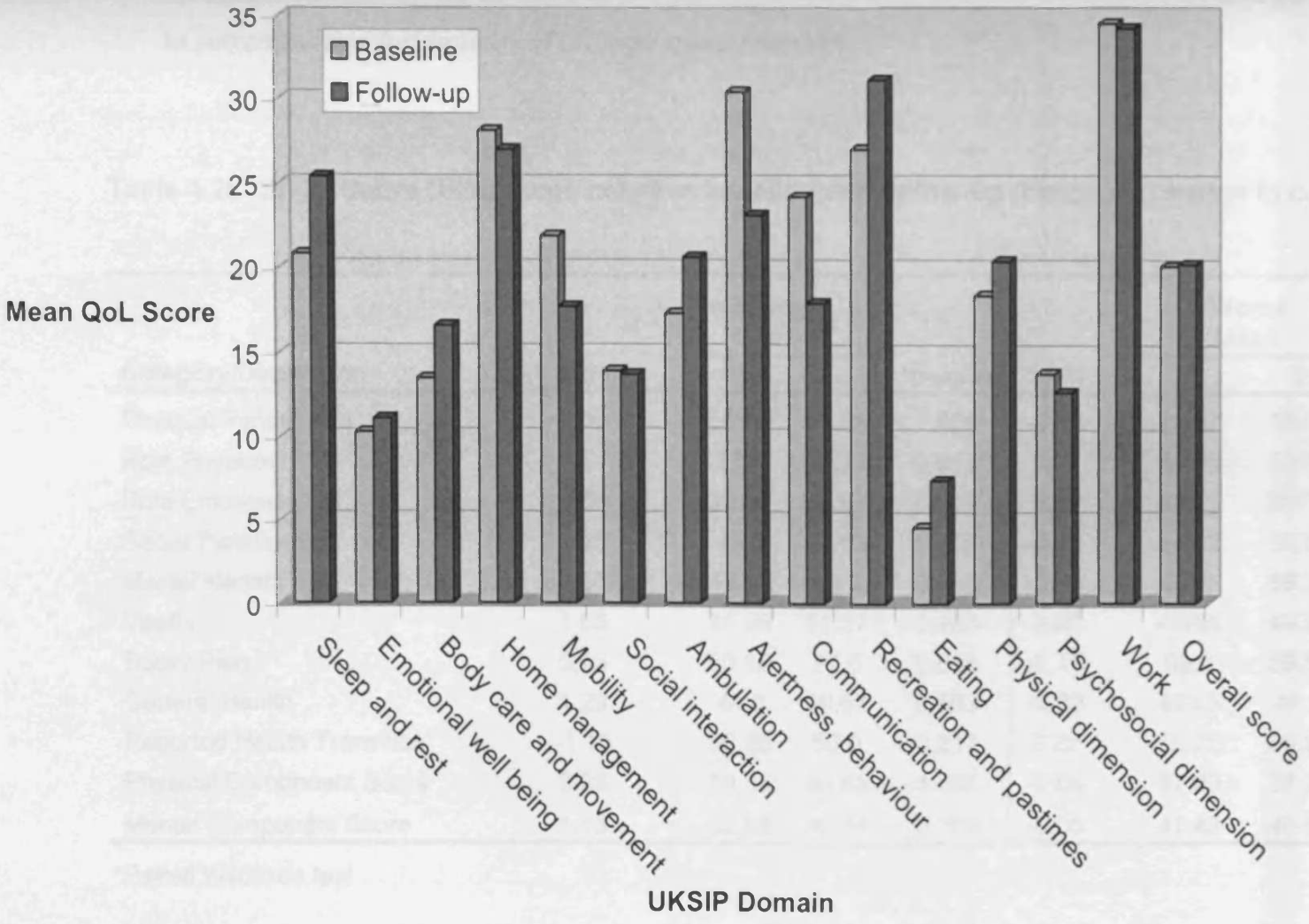
	Physical Functioning	Role Physical	Role Emotional	Social Functioning	Mental Health	Vitality	Bodily Pain	General Health	Reported Health Transition	Physical Component Score	Mental Component Score
Baseline	52.3	40	58.7	61.6	65.2	42.3	62.3	47.8	58	36.8	44.8
Follow-up	46.9	45.7	66.7	60.3	65.3	44.7	61.3	43.5	60	36.5	45.8

Table 4.27 Responsiveness of the UKSIP (n = 26)

	Baseline Mean	Follow-up Mean	Wilcoxon test p-value
Sleep and rest	20.8	25.4	0.154
Emotional behaviour	10.3	11.1	0.443
Body care and movement	13.5	16.6	0.121
Home management	28.1	27	0.887
Mobility	21.9	17.7	0.569
Social interaction	13.9	13.7	0.879
Ambulation	17.3	20.6	0.116
Alertness behaviour	30.4	23.1	0.263
Communication	24.2	17.9	0.209
Work	34.5	34.2	0.655
Recreation and pastimes	27	31.1	0.296
Eating	4.6	7.4	0.176
Physical dimension	18.4	20.4	0.192
Psychosocial dimension	13.8	12.7	0.931
Overall score	20.1	20.2	0.523

Lower scores indicate better functioning

Figure 4.14 Mean HRQoL Score as measured by the UKSIP



Responsiveness of the SF-36 and UK-SIP in patients that showed deterioration and no change in answer to retrospective judgement of change questionnaire.

Table 4.28: SF-36 Score Differences between baseline and follow-up (based on change in condition question)

		No change Mean				Worse Mean		
Category/Dimension	SRM	t1	t2	*p-value	SRM	t1	t2	*p-value
Physical Functioning	0.00	56.66	56.66	1.000	-0.14	52.37	48.18	0.285
Role Physical	0.19	25.0	33.33	0.655	0.23	43.05	52.77	0.310
Role Emotional	0.58	33.33	66.67	0.317	0.00	68.75	68.75	1.000
Social Functioning	1.65	40.0	65.83	0.109	-0.15	64.62	60.62	0.623
Mental Health	0.55	49.33	64.0	0.285	-0.08	68.0	66.38	0.569
Vitality	1.06	31.66	51.67	0.285	0.06	43.61	44.81	0.583
Bodily Pain	0.46	60.66	77.0	0.655	-0.16	63.2	59.55	0.844
General Health	1.29	44.0	59.67	0.593	-0.39	48.15	41.6	0.058
Reported Health Transition	-1.15	66.66	50.0	0.317	0.22	55.26	59.21	0.438
Physical Component Score	0.28	39.16	40.63	1.000	-0.04	37.69	37.25	0.959
Mental Component Score	1.13	33.95	46.54	0.109	-0.06	47.42	46.63	0.642

*Paired Wilcoxon test

Table 4.29: UKSIP Score Differences between baseline and follow-up (based on change in condition question)

		No Change				Worse		
		Mean				Mean		
Category/Dimension	SRM	t1	t2	*p-value	SRM	t1	t2	*p-value
Sleep and rest	0.23	17.79	20.08	0.588	0.15	21.81	25.75	0.332
Emotional well being	-0.32	16.36	10.09	0.715	0.07	8.48	9.59	0.284
Body care and movement	0.14	18.47	20.37	1.000	0.19	11.56	14.40	0.158
Home management	-0.16	39.34	37.09	0.686	-0.08	24.58	22.29	0.689
Mobility	0.49	22.36	29.87	0.068	-0.21	23.41	15.03	0.158
Social interaction	-0.39	25.48	22.16	0.686	-0.02	10.97	10.75	0.753
Ambulation	0.29	23.37	27.83	0.655	0.18	15.26	18.26	0.130
Alertness behaviour	-0.54	48.03	28.28	0.144	-0.17	25.82	19.28	0.362
Communication	-0.81	36.0	17.10	0.080	-0.19	19.56	15.62	0.575
Recreation and pastimes	-0.21	41.82	37.67	1.000	0.20	25.17	30.24	0.294
Work	0.00	36.1	36.1	1.000	-0.04	33.83	33.53	0.655
Eating	-0.33	7.51	5.46	0.180	0.50	4.095	8.43	0.080
Physical dimension	0.23	24.41	26.69	0.686	0.08	16.28	17.64	0.433
Psychosocial dimension	-0.53	23.46	16.74	0.500	-0.09	11.47	10.34	0.955
Overall score	-0.21	26.29	24.34	0.893	-0.03	18.56	18.12	0.836

*Paired Wilcoxon test

Cohorts that showed a large amount of deterioration over time might be expected to produce more marked changes in HRQoL. Also, certain categories of the two instruments might be more sensitive to change than others. The sample size in this study was relatively small as this is in essence an exploratory study but nevertheless it is feasible that some changes may not achieve significance that would be expected to in a larger sample size.

In the SF-36 and UKSIP, groups showing deterioration based on the patients own retrospective judgement of change none of the categories were statistically significant between baseline and follow-up. In the group where no change had occurred, no significant differences were found between baseline and follow-up for any of the categories of the SF-36 and UK-SIP.

The standardised response mean (D/SD (SRM)) was calculated as the mean change in score divided by the standard deviation of individuals' change scores and provided a standardised and unit free index of sensitivity to change(Liang M H et al., 1990). A SRM of =0.2 is considered small, =0.5 is moderate and =0.8 is large(Timmerman et al., 2003). Those patients whose severity of symptoms worsened at the follow-up visit should be reflected by change to instrument scores indicating a worsening of HRQoL. The SRM scores were equivocal due to the absence of statistical significance between baseline and follow-up scores and included negative values, thus reflecting both the uncertainty in the estimates and the small study size.

In the group showing a deterioration (based on the Hoehn & Yahr measure of disease severity) none of the SF-36 categories were found to be responsive to change with the exception of vitality (p =0.034). For those categories of the UKSIP showing deterioration, *recreation and past-times* and *psychosocial dimension* were significantly different between baseline and follow-up (p = 0.046 and p = 0.028 respectively). In the group where no change had occurred, no significant differences were found between baseline and follow-up for any of the categories of the UKSIP apart from social interaction (p =0.036). The standardised response mean (SRM) for the two statistically significant categories, *recreation and pastimes* and *psychosocial dimension* were 0.81 and 0.66 respectively and hence in the moderate to large range using Cohen's criteria (Cohen, 1977).

Responsiveness of the SF-36 and UK-SIP in patients that showed deterioration and no change according to clinician's assessment of PD severity using the Hoehn & Yahr scale.

Table 4.30: SF-36 Score Differences between baseline and follow-up (based on change in Hoehn & Yahr stage)

		Stable				Worse		
		Mean				Mean		
Category/Dimension	SRM	t1	t2	*p-value	SRM	t1	t2	*p-value
Physical Functioning	-0.05	52.5	50.97	0.964	-0.40	43.33	32.31	0.058
Role Physical	0.37	35.41	52.08	0.221	0.20	25.0	33.33	0.317
Role Emotional	0.13	54.54	60.60	0.683	0.28	66.66	80.0	0.317
Social Functioning	0.35	53.33	63.33	0.312	-0.80	68.92	50.35	0.116
Mental Health	0.05	61.33	62.25	0.878	-0.42	75.33	67.33	0.249
Vitality	0.48	35.0	45.13	0.067	-0.50	45.83	37.5	0.034
Bodily Pain	0.43	55.16	63.08	0.236	-0.06	51.42	49.85	0.581
General Health	0.06	39.66	40.66	0.755	-0.58	45.85	37.71	0.176
Reported Health Transition	0.00	59.09	59.09	1.000	0.54	60.71	71.42	0.257
Physical Component Score	0.29	36.18	39.01	0.248	-0.12	28.93	27.63	0.686
Mental Component Score	0.17	42.43	44.45	0.534	-0.27	52.48	49.42	0.225

*Paired Wilcoxon test

Table 4.31: UKSIP Score Differences between baseline and follow-up (based on change in Hoehn & Yahr stage)

	Stable				Worse			
		Mean				Mean		
Category/Dimension	SRM	t1	t2	*p-value	SRM	t1	t2	*p-value
Sleep and rest	0.08	27.19	29.53	0.674	0.83	17.66	26.71	0.235
Emotional well being	-0.38	14.30	6.83	0.225	0.67	10.33	19.25	0.207
Body care and movement	0.07	14.29	15.37	0.838	0.58	15.11	23.95	0.063
Home management	-0.11	33.68	30.14	0.398	0.28	26.73	31.80	0.237
Mobility	-0.21	32.12	22.75	0.441	-0.07	16.63	15.73	0.686
Social interaction	-0.40	17.51	12.30	0.036	0.66	11.76	18.62	0.034
Ambulation	0.09	17.37	18.92	1.000	0.48	18.74	25.55	0.068
Alertness, behaviour	-0.40	38.02	21.04	0.116	0.30	20.68	28.40	0.138
Communication	-0.50	29.95	17.05	0.093	0.22	25.32	30.69	0.109
Recreation and pastimes	-0.20	33.81	28.17	0.292	0.81	22.81	37.57	0.046
Eating	0.06	6.53	7.21	0.461	1.31	2.75	9.78	0.655
Physical dimension	-0.01	19.93	19.68	0.959	0.58	19.22	26.86	0.063
Psychosocial dimension	-0.45	17.78	10.66	0.022	0.66	10.73	17.75	0.028
Work	0.00	36.1	36.1	1.000	0.00	36.1	36.1	1.000
Overall score	-0.23	24.225	19.75	0.110	0.67	18.72	25.34	0.063

*Paired Wilcoxon test

For the remaining categories the SRM scores were equivocal due to the absence of statistical significance between baseline and follow-up scores and included negative values, thus again reflecting both the uncertainty in the estimates and the small study size.

The change scores for the UKSIP were examined in relation to patterns of change in patients' SF-36 physical component, and mental component for baseline and follow-up.

Change in the UKSIP psychosocial and physical dimension scores at follow-up were reflected in a changed overall score at the follow-up visit. The correlations were strong and statistically significant indicating that any change in the UKSIP overall score is influenced by changes to the dimensions scores. This is not unanticipated and provides evidence for the robustness of the instrument (Table 4.32).

The SF-36 mental component did not correlate with any of the UKSIP change scores or the physical component of the SF-36. The SF-36 physical component showed moderate and statistically significant correlation with the UKSIP psychosocial dimension which was not expected. There is not sufficient evidence to suggest that the SF-36 and UKSIP are responsive to changes in corresponding HRQoL domains and nor is this a weakness as both are generic instruments made up of differing items and domains that are addressing a wide range of physical and psychosocial aspects of well being. It is perhaps more relevant that such instruments can reflect change in clinical parameters of disease before they can be considered useful.

The change scores for the UKSIP and SF-36 were examined in relation to the change scores for the clinician's ratings on the H & Y.

Changes scores for the physical and psychosocial dimensions and overall score of the UKSIP correlated positively and moderately (r_s 0.592 – 0.662) with change in Hoehn and Yahr staging which signifies that when a patients severity of symptoms changes from the baseline grading it is associated with a corresponding change to the UKSIP dimension and overall scores (Table 4.33). This provides evidence that the UKSIP is responsive to change in patient's clinical status over time.

Table 4.32: Relationships† between change scores for UKSIP and SF-36 dimensions (between follow-up (t2) and baseline (t1)). Sensitivity to change between follow-up and baseline visit (n = 23).

	UK-SIP t2-t1 Psychosocial dimension	UK-SIP t2-t1 Overall score	SF-36 t2-t1 Physical Component Score	SF-36 t2-t1 Mental Component Score
UK-SIP t2-t1 Physical dimension	0.480*	0.704**	-0.104	-0.150
UK-SIP t2-t1 Psychosocial dimension		0.798**	-0.568*	-0.026
UK-SIP t2-t1 Overall score			-0.461	-0.044
			17	17
SF-36 t2-t1 Physical Component Score				-0.017

*p<0.05, **p<0.001

†Spearman rank correlation

The physical and mental component change scores of the SF-36 did not correlate with change in Hoehn and Yahr staging with weak correlations for both components (r_s physical component = -0.318 and mental component = -0.401). This shows that a change in patient symptom severity is not associated with a corresponding change to the SF-36 components.

Table 4.33: Relationship† between change scores for UK-SIP and SF-36 dimensions (between follow-up (t2) and baseline (t1)) and change scores for clinicians' ratings of H&Y

	Change in Hoehn & Yahr scale
UK-SIP t2-t1 Physical dimension	0.592*
UK-SIP t2-t1 Psychosocial dimension	0.662**
UK-SIP t2-t1 Overall score	0.623**
M0 SF-36 t2-t1 Physical Component Score	-0.318
M0 SF-36 t2-t1 Mental Component Score	-0.401

*p<0.05, **p<0.005

†Spearman rank correlation

The significance of change scores amongst patients whose symptoms worsened according to the Hoehn & Yahr.

The significance values for each test demonstrate the individual responsiveness of each UKSIP and SF-36 dimension and summary measure as well as the overall sensitivity of the UKSIP (Table 4.34). The ability of the UKSIP and SF-36 to detect change was tested for all those who showed any degree of deterioration (as measured by the Hoehn and Yahr).

Certain categories of the two instruments might be more sensitive to change than others. It is important to point out that the sample size used here is smaller than what one would require for a longitudinal assessment of sensitivity and therefore some changes may not achieve significance that would be expected to in a larger study population.

Prospective change in quality of life on the UKSIP dimensions and overall score were high (standardized response mean >0.8) but only the physical dimension was significantly different between baseline and follow-up. This supports the commonly held view that physical related domains are more responsive to change than those concerned with psychosocial areas of HRQoL(Hyland et al., 1994). Neither the physical nor the mental component scores of the SF-36 were responsive to change and exhibited small SRM.

The analysis provides some evidence to support the responsiveness of the UKSIP patient scores to change over time but the responsiveness of the SF-36 has not been demonstrated.

The change scores for the UKSIP and SF-36 over 6 months were compared with all patients who reported that their condition had changed and was either "a little worse" or "much worse".

The significance values for each test demonstrate the individual responsiveness of each UKSIP and SF-36 dimension and summary measure as well as the overall sensitivity of the UKSIP (table 4.35). The ability of the UKSIP and SF-36 to detect

Table 4.34: Sensitivity of SF-36 and UKSIP in patients that showed any degree of deterioration based on Hoehn & Yahr staging

SF-36 (n = 7)	Baseline	Follow-up		Paired t-test
	Mean	Mean	**SRM**	**p-value**
Physical Component Score	29.05	29.19	-0.22	0.647
Mental Component Score	53.71	45.33	-0.68	0.203
UK-SIP (n = 8)				
	Mean	Mean	**SRM**	**p-value**
Physical dimension	21.26	26.87	0.94	0.048*
Psychosocial dimension	10.67	17.75	0.84	0.068
Overall score	19.36	25.35	0.84	0.067

***p<0.05**

SRM = standardised response mean where a value of 0.2 is small, 0.5 is moderate and 0.8 or greater is large.

change was tested for all those who showed any degree of deterioration (as measured by a patient reported transitional questionnaire).

Change in UKSIP dimension and overall scores ranged from low to moderate with only the physical dimension significantly different between baseline and follow-up. The psychosocial and overall score did show a trend to deterioration at follow-up. This supports the commonly held view that physical related domains are more responsive to change than those concerned with psychosocial areas of HRQoL(Hyland et al., 1994). Neither the physical nor the mental component scores of the SF-36 were responsive to change and exhibited small SRM.

These results in conjunction with the responsiveness data from the Hoehn and Yahr analysis above provides confidence to state that the SF-36 is not responsive to change in the PD population.

Table 4.35: Patients for whom change in condition between baseline and follow-up was either 'a little worse' or 'much worse'

SF-36 (n = 12)	Baseline	Follow-up		Paired t-test
	Mean	Mean	**SRM**	**p-value**
Physical Component Score	35.90	32.48	-0.39	0.248
Mental Component Score	49.38	43.40	-0.44	0.198
UK-SIP (n = 13)				
	Mean	Mean	**SRM**	**p-value**
Physical dimension	19.82	22.11	0.64	0.050*
Psychosocial dimension	13.28	15.45	0.39	0.200
Overall score	20.75	23.10	0.51	0.103

*P=0.05

SRM = standardised response mean where a value of 0.2 is small, 0.5 is moderate and 0.8 or greater is large.

Sample sizes for UKSIP

The sample size required to detect a 20% improvement in QoL (using the overall score of the UKSIP at baseline) was assessed and the figures below (Table 4.36) are given for a two-sided test at the 5% and 1% significance level and 90, 80, 70 and 60% power to detect the stated change.

Table 4.36: Sample Size - UKSIP Overall Score at baseline

alpha	0.05	0.05	0.05	0.05	0.01	0.01	0.01	0.01
power	0.9	0.8	0.7	0.6	0.9	0.8	0.7	0.6
Observed baseline mean	20.4	20.4	20.4	20.4	20.4	20.4	20.4	20.4
% increase	20	20	20	20	20	20	20	20
delta	4.08	4.08	4.08	4.08	4.08	4.08	4.08	4.08
Observed standard deviation of change - SD (t1 - t2)	9	10	11	12	9	10	11	12
Sample size required to detect 20% mean change	**53.0**	**49.1**	**46.8**	**44.3**	**75.7**	**73.5**	**73.2**	**72.6**

DISCUSSION

As health-related quality of life instruments become more common place in the evaluation of the impact of treatment regimens and the use of such information (i.e. patient-reported outcomes) in clinical decision making, it is necessary to conduct comparative research to determine the relative merits of various instruments. Head to head comparisons of such instruments are limited and as far as the literature shows a comparison of the UKSIP and SF-36 in the PD population has not been undertaken.

Patients from a neurology out-patient clinic completed both instruments, the UKSIP and SF-36, to reveal areas that affect health status in the PD population. The mean age of the cohort was 68.9 years (51 – 83) with median disease duration of 8 years (1 – 22). Though more than half of the sample population had advanced disease 64.7% were graded as stage I and II on the Hoehn & Yahr scale indicating mild disease severity.

The minimum and maximum scores achieved for each of the subscales of both questionnaires show a wide range of score suggesting that these instruments are sensitive to the impact of PD on the various HRQoL domains. However on closer examination, despite the wide range of scores achieved the mean scores were considerably lower that the reported maximum scores. For example, the mean baseline UKSIP domain score for sleep and rest is 17.7 with a maximum reported score of 100. This would suggest that some of the patients reported a considerable degree of impairment in some of the domains thus skewing the data. This observation is confirmed by the comparison of the median with the mean. For the majority of the UKSIP subscale scores there is a disparity between the mean and median scores and to a lesser extent in the case for the SF-36. The UKSIP subscale, eating had a median score of zero at baseline meaning that over half the patients in this group could not show improved HRQoL in this area during the follow-up visit. This was true as at the follow-up visit the median not surprisingly remained as zero. A similar observation was noted for role emotional in the SF-36. Eating is a known complication of PD as patients have difficulty chewing and swallowing in the later stages of the disease. Both instruments showed a good range of scores though there was disparity between the mean and median scores, particularly for the

UKSIP. Apart from eating in the UKSIP and role emotional in the SF-36, the majority of patients could show improvements in HRQoL in all the domains thus making these instruments suitable for longitudinal studies i.e. there is an absence of a ceiling effect. A direct comparison of the mean scores between the two instruments showed that patients reported lower levels of dysfunction on the UKSIP scales than the SF-36.

It was in the alertness behaviour category that the first major difference between the two instruments was revealed. Though both the SF-36 and UKSIP show that there is greater impairment in the physical dimensions of health-related quality of life compared to psychosocial dimensions, the UKSIP highlights that *alertness behaviour* is greatly impaired in this study population. This item addresses the subject's feelings e.g. "I am confused, I react slowly, I do not keep my attention". Though physical impairment is a major handicap it is apparent that a cognitive decline is also observed in the PD cohort which was not identified in the SF-36. These are not addressed in the SF-36 and therefore would be overlooked and hence based on the content of the items of both instruments the UKSIP would be recommended over the SF-36, particularly as cognitive impairment is a major characteristic of the elderly and a debilitating symptom of PD. Neuropsychological investigations of patients with Parkinson's disease have shown specific impairments even in the early stages of the disease (Dubois and Pillon, 1997). Depression, disability, postural instability, and cognitive impairment have the greatest influence on QoL in Parkinson's disease. The improvement of these features should therefore become an important target in the treatment of the disease(Schrag et al., 2000) . The UKSIP also highlights that communication was affected and this is a domain which has a real and direct impact on the PD sufferer. The communication domain on the UKSIP addresses items such as "I am having trouble writing or typing", my speech is understood by only a few people" and "I do not speak clearly"; items which are not dealt with in the SF-36. Speech problems occur in more than 70% of Parkinson's patients (Schulz, 2002), by some estimates. Motor abnormalities that limit action in the hand may develop in late stages. Handwriting, for instance, often becomes diminutive (micrographia) and this is one of the secondary symptoms of PD. Speech difficulty can be caused by rigidity of the facial muscles, loss of motor control, and impaired breath control. Tone can become monotonous, words may be repeated over and over, or the rate of speech may even be very fast. Furthermore, the SF-36 does not contain questions referring to sleep. This is a serious omission as changes in the pattern and quality of sleep

are commonly associated with ill health (Hunt and McKenna, 1993). Sleep problems are an under-emphasised cause of disability in Parkinson's disease (PD) and may be seen independently of PD, associated with primary PD pathology, or as a result of anti-parkinsonian medications. Common sleep disorders include excessive daytime sleepiness, rapid eye movement (REM) sleep behaviour disorder, night-time wakefulness and restless legs syndrome (Stacy, 2002). The UKSIP addresses Sleep & Rest with questions such as "I am sleeping or dozing most of the time".

Impairment in both instruments was seen in the areas related to physical and social functioning, and none of the areas of HRQoL were similar in prevalence in patients with PD and the general population. The impairment of HRQoL was seen in all age groups. This clearly reveals that PD patients experience considerable impairment in a variety of quality of life domains and highlights the impact of the disease compared to the general population.

Missing data is one of the most important problems compromising the quality of trials having a HRQoL component. Missing items may be problematic if patients skip some items in the questionnaire because they perceive these items as too intrusive, because the questionnaire was felt to be too long or due to mistakes (unclear lay-out) (Groenvold, 1999). Overall the SF-36 produced a low level of missing items and more questions were not answered at the second time of completion. Missing data at the second time of completion did not follow any pattern and may be considered as random. Missing data has been found to be strongly correlated with older age and reduced health (Wagner et al., 1993). The SF-36 was analysed for the number of items (questions) that were not answered bearing in mind that each item in the SF-36 questionnaire requires an answer based on a Likert scale or simple dichotomous yes or no answer. More than 10% of subjects failed to complete questions in the dimensions of *Role Physical* and *Role Emotional.* A limit of 10% was arbitrarily set to highlight items that were not considered to be relevant based on non-completion by the patient. The most common questions not answered at both time points were "vigorous activities, such as running..." (question 3a), "cut down on amount of time spent at work" as a result of physical health (question 4a) and "cut down on amount of time spent at work" as a result of emotional problems (question 5a). These are questions that are perhaps redundant in a population of mean age 68.9 years who are beyond retirement age and very unlikely to be in work or participate in any

vigorous activity. It could be argued that a younger cohort would find questions on vigorous activity redundant particularly as more people find themselves leading a sedentary lifestyle. The question stem for questions 4 and 5, relating to strenuous physical and work related activities would not surprisingly be omitted by elderly subjects. In a UK study assessing the SF-36, subjects were recruited from an outpatient site and a day centre for the elderly. Results revealed that 91% of respondents regarded all or most of the questions in the SF-36 as clear and easy to understand. However, problems of applicability related particularly to questions asking about work or other regular activities (questions 4 and 5). Although these questions include both work or other regular daily activities, respondents repeatedly perceived the question in terms of paid employment and replied that they were retired and no longer worked (Hayes et al., 1995). These findings are consistent with what was found in the PD cohort of this study. An amended version of the Short Form 36 health survey questionnaire (SF-36), which has been suggested as being more suitable for self-completion in older adults, was evaluated among a group of elderly subjects with Parkinson's disease. However, the modified questionnaire still did not overcome the problem of missing responses from elderly subjects. A total of 24% of respondents missed one or more of the 36 statements and 80% of missing responses were concentrated in the three statements which had been modified (Hobson and Meara, 1997). These results would suggest that the SF-36 in its original form would need to be combined with a disease-specific measure to adequately evaluate the health status of older adults with PD or else some items must be revised to reflect the pertinent aspects of HRQoL in an elderly population. Nevertheless, there was a low level of missing of item redundancy for the remaining questions as observed in the study of elderly (Hayes et al., 1995) and in fact some questions were affirmed in high frequencies such as question 3f (health limiting ability to bend, kneel or stoop) and question 7 (bodily pain during last four weeks). These questions may not yield any information regarding differences between patients and thus detract from the discriminative properties of the questionnaire.

The UKSIP asks patients to "tick only those questions that you are sure describe you today and are related to your health". Inspection of the most frequently endorsed items revealed that PD patients are likely to suffer from reduced mobility, social isolation, limitations in household activities and recreational activities as well as decreased sexual activity. Some of the frequently endorsed items can be anticipated

based on disease characteristics such as slowness of movement or postural instability. A decrease in sexual activity was reported in 54.3% of respondents at baseline and though surprising, it has been reported that the complaint may be just as common among patients with chronic diseases that do not affect the nervous system (Longstreth et al., 1992). Satisfaction from sexual life has been shown to decrease with age and advanced disease (Moore et al., 2002) and both these variables apply to the PD population making the inclusion of a question on sexual activity in a HRQoL instrument mandatory. Such a question is missing from the SF-36. This study has highlighted that sexual activity was decreased in more than half of respondents thus confirming its importance in the life's of PD patients and also revealing that the generic measure, the UKSIP is able to identify this pertinent dimension of HRQOL.

To highlight the applicability of an instrument in the pertinent population is best highlighted by the UKSIP where 77.1% of respondents at baseline endorsed the item "walking more slowly". This is not surprising as decreased mobility is a key facet of the symptomatology of PD but nevertheless it is of great value that such a question exists in a generic measure such as the UKSIP. The Ambulation domain in the UKSIP refers to questions that would be applicable to the PD population such as "I do not walk at all", "I walk short distances", and "I walk more slowly". Compare this to the SF-36 which asks whether the subject's health is limited in such activities as; "walking half a mile", "walking more than a mile", and "climbing several flights of stairs". These are not activities that would be relevant or easy to accomplish in the elderly population never mind an elderly population with a chronic disease that significantly affects mobility. There is not an ambulation domain in the SF-36 but this would not make it irrelevant but however, the omission of questions on mobility that are of relevance to the population studied would make sense.

The domain of the UKSIP where there was a very low item endorsement was in Work. This domain is only required to be completed if the subject works (apart from looking after their home) and hence in a retired elderly cohort this domain would seem redundant. What is helpful though that there exist questions to ascertain whether the person is no longer working because of their health. In this study 96.5 % of participants who answered the question reported that they were retired of whom 36% retired due to their health.

These findings relating to the UKSIP are consistent with that of the postal survey study reported in Chapter three and highlight that there are no differences in impairment between PD patients attending an out-patient clinic at a hospital and those PD patients who are in the community and served by the General Practitioner.

Overall the SF-36 items within the scales are measuring similar areas of functioning and hence the results suggest that the SF-36 has good reliability in PD patients though an assessment of test-retest reliability would be a necessary next step. The Cronbach's alpha values at baseline and at follow-up remained similar showing stability of the instrument in a longitudinal setting. These results are consistent with a previously published cross sectional study (Rubenstein et al., 1998) where the a coefficients for all domains were >0.7. Overall the UKSIP items within the scales are measuring similar areas of functioning and hence the results suggest that the UKSIP has good reliability in PD patients especially as good test-retest reliability has been previously reported(Ali and Salek, 2002).

Steiner and Norman (Streiner and Norman, 1995) have proposed that because HRQoL is an abstract concept which is not directly comparable, one must set up hypothetical constructs which will relate the measures. The hypothetical construct used in this study was the clinical severity of PD as measured by the Hoehn & Yahr (Hoehn and Yahr, 1967) and it is assumed that disease severity will be related to some aspect of HRQoL, for example, one could reasonably hypothesis that there would be a correlation between disease severity and HRQoL domain of physical functioning (PF), as was the case in this study. At baseline, correlations were highest with the dimensions measuring PF and social functioning (SF). The high association with SF relates to the questions attributed to this domain. These questions refer to what extent the patients health, including physical function, has limited their ability to engage in social functions. Hence the dimension partly taps aspects of physical as well as social well-being. More interestingly is the high correlation with the MH domain and this would suggest that as disease progresses the patient experiences greater levels of psychological distress and becomes more nervous and sad, items that the mental health (MH) domain addresses. These variable have been shown by Findley and colleagues (Findley on behalf of the Global Parkinson's Disease Steering Committee, 2002) as factors which impact on HRQoL in PD patients. They reported that depression and "current feelings of optimism" have a significant impact on

HRQL. It has also been reported that clinical stage and depression have a significant association to poor HRQoL as measured by the SF-36(Kuopio et al., 2000), findings which have been replicated here. There was low correlation between bodily pain and disease severity and this highlights that the SF-36 measures aspects of well-being which are not adequately covered by the clinical measure. Karlsen et al (Karlsen et al., 2000) reported in a 4-year follow-up period using the Nottingham Health Profile (NHP) that there was a significant increase in NHP scores, reflecting a decreased HRQoL, in the dimensions of emotional reactions, pain, and social isolation, dimensions which are not evaluable using a clinical measure of disease such as the Hoehn & Yahr.

Significant associations were found between the UKSIP and Hoehn & Yahr at both baseline and follow-up visit for the domains of body care and movement, ambulation and social interaction. One would anticipate a strong correlation with the Physical domains of the UKSIP and the measure of disease severity which assesses tremor, rigidity and movement; hence the strong correlations with body care and movement and ambulation. A possible explanation to the strong correlation to social interaction is that as the disease progresses the subject is less likely to involve themselves in social activities such a going out or interacting with others. This can lead to isolation and subsequently depression both factors which can have severe impact on the patients HRQoL. Also, the strong correlation of the UKSIP domains of social interaction and ambulation against the Hoehn and Yahr could be further accounted for by the observation that these two domains were also the most commonly endorsed items at both time points which explains the strong correlations observed here. At baseline there was a strong correlation between emotional well being and disease severity which reinforces the belief that as disease progresses or deteriorates there is an increase in isolation, social interaction and subsequently the presence of depressive symptoms, a pattern that is reflected in other chronic diseases. Lower correlations were observed for sleep and rest, recreation and pastimes and eating which suggest that the UKSIP assesses areas of relevance to well -being which are not captured by the clinical measure. It is reassuring that there was not perfect agreement between all domains of the UKSIP and Hoehn and Yahr as this is not a favourable property of a HRQoL instrument; otherwise the HRQoL instrument would appear to measure the same underlying characteristics as the clinical measure.

Both the SF-36 and UKSIP have shown that disease severity does impact on physical, social and psychological domain and exhibit construct validity in the PD.

The UKSIP and SF-36 have not been evaluated against one another in the PD population and this study reveals convergent validity of the instruments. Though it is difficult to conclusively identify a clear difference between the two instruments based on construct and convergent validity analysis it is suggested that an instrument with an overall score and individual domain score would be more suited to a research and clinical environment. Questionnaires with both a profile score of domains and also an index score (overall score) are likely to be more useful in evaluative studies such as clinical trials. The summary score provides a means of assessing overall change which allows the clinician to easily interpret whether the patients HRQoL has improved, whilst the domain scores provide a useful descriptive assessment of the various areas being assessed allowing clinicians to focus most of their efforts on improving impaired aspects of a patient's daily activity.

The relative contributions of the two instruments in explaining patients overall quality of life were examined by ranking of correlation coefficients arising from the relationship between category scores and overall score. There were similarities in the ranking of coefficients in that the similar domains of mental health (SF-36) and emotional well-being (UKSIP) contributed the least to overall quality of life in this sample of PD patients. Social functioning (SF-36) and recreational past times (UKSIP) were related in that they both address social activities and hence the high ranking of these domain suggest that the ability to participate in social activities is impaired by PD resulting in poorer HRQoL.

Bodily pain has been reported to reduce HRQoL (Karlsen et al., 2000) in a four year follow-up study but it seems to have had little impact on this study population. The reason for this could be that the follow-up in this study was limited to six months and disease had not progressed in that time to adversely affect matters relating to pain. There are numerous reasons for the decrease in HRQoL suffered by PD patients: restrictions in mobility, falls, emotional disorders, social embarrassment, isolation, sleep disturbances, dyskinesias, and fluctuations. Many aspects of these disorders go unnoticed in clinical evaluation, and only QoL assessment allows them to be rated and this analysis highlights two instruments that have a role to play in the

measurement of HRQoL in PD. Both the SF-36 and UKSIP highlight problems in the areas of leisure, home care and access to public areas that might be expected from the characteristics of the disease (Fitzpatrick et al., 1997, Longstreth et al., 1992).

Discriminative analysis results show that the SF-36 is able to discriminate between PD patients according to their disease severity as measured by the Hoehn and Yahr a finding which has been confirmed by a previously published study by Rubenstein et al (Rubenstein et al., 1998). The results also show that there is a strong link between clinical disease severity and HRQoL measurement. The clinician's disease severity rating was found to correspond to a wide range of SF-36 domains, with only bodily pain (BP), general health (GH) and vitality (VT) domains failing to discriminate between the three levels of disease severity (stage I, stage II and stage III/IV). Figure 4.12 highlights visually how the SF-36 was able to distinguish between patients with different disease severities and in particular between stage I and stage III/IV. The most apparent differences between stage I and stage III/IV subjects are in the domains of role-physical, role-emotional, and social-functioning which demonstrates that increasing severity of disease has a proportional effect in that it results in deterioration in physical, emotional and social domains of well-being. Whilst patients with mild disease (stage I) had higher scores for role-emotional and social-functioning the opposite was observed in the more severe cohort (stage III/IV) where patients had much lower scores (indicative of worse health). This highlights a very important observation that within group differences exist and very different patterns of scoring may be achieved for the various HRQoL domains based on disease stage or severity of the cohort studied. One particularly interesting finding relates to the role-physical score of stage II patients in comparison to stage I and stage III/IV patients. Stage II patients mean role-physical score was 9.4 (lower scores indicating worse health) compared to 65.9 and 25.0 for the stage I and stage III/IV cohort respectively. This is unexpected as you would expect the stage II mean score to be higher than the clinically rated more severe stage III/IV cohort and the magnitude of difference between the stage I and stage II scores to be less severe. What might be happening here is that the most severe PD patients have higher scores (indicating better health) in comparison to the moderate severity group (stage II) because of the lower expectations to undertake such daily activities such as work. Also the magnitude of difference between stage I and stage II score for role-physical is due to the more noticeable disability affecting the subject's ability to continue

satisfactorily with daily activities. This may in part explain why statistical significance was not achieved for the physical component score. Stage III/IV patients also reported better general health scores compared to stage II subjects which suggest that more severely impaired patients have a lower expectation of their well-being or else they are benefiting most from treatment. Future evaluation of the discriminative properties of the SF-36 should include a more stringent measure of disease such as the UPDRS which addresses the functional problems associated with disease as well as symptom severity. Overall, the results show that the SF-36 was able to discriminate between PD patients with different levels of disease severity.

The ability of the UKSIP to discriminate between PD patients according to disease severity also shows that it is able to discriminate between different levels of disease severity. A statistical significant pattern of higher scores was observed for all domains with the exception of; home management, social interaction, and recreation and pastimes which showed a pattern of higher scores across the three stages of disease severity, but were not statistically significant. Mobility and eating domains showed higher scores (greater impairment) amongst stage III/IV patients compared to those in stage I. The statistical significance of the results for the SF-36 were higher than for the UKSIP with 5 out of 10 subscales showing a significant difference at $p=0.01$ compared to one of the UKSIP domains showing a statistical difference at $p=0.01$. Overall, the UKSIP and SF-36 were found to measure HRQoL in a similar manner, with both questionnaires able to discriminate between patients according to their disease severity. Both instruments ability to discriminate between patients by disease severity suggests that these profiles can enhance the description of health status and provide estimates of the magnitude of diminished HRQoL as (Hoehn and Yahr) stage increases. Also, the comparisons with general population mean scores indicate that individuals with disease have considerably diminished health. This indicates the potential for a quantifiable estimate of the attributable reduction in HRQoL caused by PD.

This analysis compares for the first time the discriminative properties of both the SF-36 and UKSIP in the PD population and adds to the evidence in support of the construct validity of both questionnaires such that they may have sufficient discriminative properties for use in an out-patient population of PD patients.

Although quality of life in patients with Parkinson's disease has been increasingly explored and described, there is very limited information on the longitudinal evolution of distress and well-being in patients with PD. Such information could indicate which patients need a closer follow-up in future management and provides insight into change over time of HRQoL in a group of patients with a traditional motor controlled focused treatment (Karlsen et al., 2000). To achieve such an assessment the instruments used must exhibit responsiveness (sensitivity) to change. The sensitivity of a test or measure refers to its ability to detect clinically meaningful changes in patients over time. The sensitivity to change of the two generic instruments were explored in a variety of ways as outlined in the "methods" section of this chapter.

Differences between baseline and follow-up mean domain scores were assessed to measure the change in patients who demonstrate a true change in HRQoL. This provides an indication as to whether the instruments evaluated can detect change in scores of patients whose HRQoL changes over time. The results show that there was neither a trend for improvement nor deterioration over time for both instruments. The non-significance of these results illustrates that over a six month period the magnitude of change in health status was not dramatic and that the study cohort perhaps represents a relatively stable group of patients with respect to disease status and severity. This result suggests that in the absence of a significant and large enough trend to change over time, the analysis of responsiveness to change will be difficult to determine. This may also be a reflection of the small sample size, notably the low numbers in the group for those patients demonstrating no change.

More encouragingly, for those categories of the UKSIP showing deterioration in the Hoehn and Yahr category, the domains of recreation and past times and psychosocial dimension showed statistical significance ($p=0.05$) whilst the domains of body care and movement, ambulation, physical dimension and overall score showed a *trend* towards significance when compared to baseline. The direction towards significance was more notable for the UKSIP domains compared to the SF-36 and if the sample size had been larger we may have observed more statistically significant results. Additionally, the standardised response mean (SRM) for the UKSIP domains of recreation and past times and psychosocial dimension were in the moderate to

large range using Cohen's criteria (Liang et al., 1990) and therefore suggest that the magnitude of change identified was clinically meaningful. In clinical trials, where HRQoL instruments are being increasingly used as primary outcome measures, it is simple to determine the statistical significance of a change in HRQoL, but placing the magnitude of these changes in a context that is meaningful for health professionals and patients has not been so easy. Ascertaining the magnitude of change that corresponds to a minimal important difference would help address this problem (Juniper et al., 1994). So when determining an important change standard the perspective can influence the assessment approach and the way in which an important difference is determined. The minimal important difference (MID), from the patient perspective, can be defined as "*the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management*" (Jaeschke et al., 1989). Thus, individual change standards are needed to provide meaningful interpretation of HRQoL intervention and treatment effects and to classify patients based on this standard as improved, stable or declined. To date two broad strategies have been used to interpret differences or changes in HRQoL following treatment: distribution based approaches – the effect size (ES); and anchor-based measures – the minimum clinically important difference (MCID). Distribution based approaches rely on relating the difference between treatment and control groups to some measure of variability. The most popular approach uses Cohen's standardised effect size, the mean change divided by the standard deviation to serve as an "effect size index", that is suitable for sample size estimation. Cohen suggested that standardised effect sizes of 0.2 to 0.5 should be regarded as "small", 0.5 to 0.8 as "moderate" and those above 0.8 as "large" (Liang et al., 1990).

The responsiveness data so far, could be interpreted to suggest that the UKSIP is a more valid instrument in the measurement of change and the instrument of choice in longitudinal studies as it shows statistical significance in more domains and trends to significance in domains of interest. Responsiveness to change was further investigated by looking at change scores between the two instruments and in relation to the change scores for the clinician's rating on the Hoehn and Yahr at the two assessments.

In the analysis using change scores for the clinician's rating on the Hoehn and Yahr positive correlations were observed for UKSIP physical and psychosocial dimensions and overall score ($p<0.005$). This shows that the UKSIP exhibit a corresponding change in score with change in severity of symptoms. The SF-36 physical and mental component scores did not correlate with change scores of the Hoehn and Yahr. This is the first evidence to suggest that the UKSIP is responsive to change in patients' clinical status over time and the strong correlation with a clinical measure contrasts with the disease-specific PDQ-39 where none of the changes in the PDQ-39 were significantly related to changes in clinical scores (Fitzpatrick R et al., 1997). Fitzpatrick and colleagues have explained that the lack of correlation with the PDQ-39 is due to the distinct dimensions assessed by the two different instruments (the PDQ-39 and Hoehn and Yahr). What this analysis reveals is that a clinical scale which focuses more specifically on impairment and physical function (and is completed by the clinician) can be compared with the UKSIP which focuses on patients' perceptions of health-related quality of life (and is completed by the patient). The UKSIP appears to be sensitive to changes that matter to patients and are closely associated with the characteristics of the disease which are routinely measured by a clinical scale. This would make the instrument an important addition to outcome measures in clinical trials.

The test for responsiveness was further tested by comparing scores at baseline and follow-up for the group that had demonstrated change (or deterioration) based on Hoehn and Yahr assessment. In the group showing deterioration or worsening on condition according to the Hoehn and Yahr, physical dimension score was significantly different between baseline and follow-up whilst there were trends towards significance on psychosocial dimension and overall score of the UKSIP. For the SF-36 no significant differences were found between baseline and follow-up. Fitzpatrick et al (Fitzpatrick et al., 1997) in a similar analysis, showed that the change scores of the PDQ-39 (a disease-specific measure of HRQoL) were significantly correlated with the retrospective judgment of change for mobility and ADL, but no other dimension of the PDQ-39. This study differed in two important area, 1) larger sample size ($n = 51$) and a longer follow-up. This may in part explain why a correlation was observed for two of the PDQ-39 domains of interest (mobility and ADL) but a more important reason is that 2) a disease-specific measure is more likely to be more sensitive to change versus a generic measure due to the relevance of the

questions and domains which makeup such an instrument. With this in mind it is quite exceptional that the generic instrument, the UKSIP exhibited signs of sensitivity to change and the number of statistically significant findings may have increased with a larger sample size and longer duration of follow-up. The same analysis, but on this occasion, by comparing scores at baseline and follow-up for the group that had demonstrated change (or deterioration) based on the patient reported transitional questionnaire was completed. This revealed that the UKSIP physical dimension score was significantly different between baseline and follow-up where as none of the SF-36 summary categories reached significance. The psychosocial summary and overall scores of the UKSIP did show a trend towards deterioration at follow-up but did not reach significance. It has been reported previously that physical related domains are more responsive to change than those concerned with psychosocial areas of HRQoL. Also, the rating of overall patient deterioration may also be too crude to detect change as such a measure does not take into account the various factors that may contribute to a patient's perceived change in condition such a depressive symptoms and cognitive status.

Based on this comprehensive review of responsiveness it is concluded that the SF-36 is not responsive to change in the PD population whereas the UKSIP exhibits responsiveness in some domains and a strong trend to responsiveness in other domains of importance to the PD patient.

Further research using a larger population would be useful in the case of this chapter in order to build on the evidence of sensitivity demonstrated by the analysis in this chapter and to more firmly establish the responsiveness of the UKSIP and SF-36 to longitudinal change. To evaluate the number of patients required to detect an arbitrary 20% mean change in the HRQoL score at a range of powers from 60% to 90% shows that at the 5% significance level as many as 53 patients would be required (at 90% power, $p<0.05$) or 76 patients if the significance was increased (90% power, $p<0.01$). Though 20% change in HRQoL score is arbitrary it is based loosely on a 4 year longitudinal study where a 30% increase in HRQoL using the Nottingham Health Profile over the four year period was detected in the majority of patients(Karlsen et al., 2000). As the duration of follow-up was considerably less in this analysis the power calculation was based on a magnitude of change in HRQoL score that was lower than 30%. To recruit such numbers at single centre is not

practical and would require a multi-centre set up which for the purpose of this chapter was not feasible nor economical to do so.

SUMMARY

- Both the SF-36 and UKSIP were able to detect impairment in key domains of relevance to the PD population, with the greatest impairment being in physical dimensions of HRQoL when compared to data from the general population

- The UKSIP detected impairment in cognitive symptoms, an item which is not addressed in the SF-36 and also the SF-36 omits any items referencing sleep disturbances which are a facet of PD impairment

- Items in the SF-36 that have little bearing on elderly PD patients were identified and should be considered redundant, e.g. "vigorous activities such as running" and "time spent at work"

- A decrease in sexual activity was reported by more than half of the UKSIP respondents a domain which is not addressed in the SF-36

- Internal consistency reliability for all SF-36 domains were good whereas in the UKSIP the internal consistency reliability ranged from low to very good across all domains

- Both the SF-36 and UKSIP exhibited good construct and convergent validity in the PD population studied

- The SF-36 and UKSIP are able to discriminate between patients according to disease severity and can be considered to be a useful discriminators of PD severity

- The analysis in this chapter has provided evidence supporting the responsiveness of the UKSIP patient scores to change over time

- Differences between UKSIP baseline and follow up scores were significant or showed trends towards significance in domains of interest in the cohort who reported deterioration

- The analysis has produced evidence suggesting the possibility of different degrees of responsiveness for different UKSIP domains

- The evidence from this analysis does not support the responsiveness of the SF-36 in PD nor suggest a trend towards responsiveness

- Further research in a larger population is indicated to build on the evidence provided by the analysis in this chapter and to firmly establish the responsiveness of the UKSIP

- The duration of follow-up should be considerably longer than the six month follow-up in this study to ensure a meaningful change in disease severity is observed

- Preliminary work has indicated that the UKSIP is a valid measure of HRQoL for use in longitudinal clinical studies

- It is concluded that the most appropriate method of assessing HRQoL of PD patients appears to be to use the UKSIP and that such a generic instrument may assess important general effects of PD on HRQoL that could be missed by disease-specific instruments covering a narrower range of domains.

Chapter 5

An evaluation of the reliability and validity of the Parkinson's Impact Scale (PIMS) in patients with idiopathic Parkinson's disease (PD)

INTRODUCTION

Following preliminary research into the psychometric properties of the PDQ-39 and further evaluations that have been extensively reported in the literature (Jenkinson, 1999, Jenkinson, 1997, Jenkinson, 1995) it was evident that this disease specific measure of HRQoL was reliable, internally consistent and valid for the assessment of PD (Ali and Salek, 2002). Nevertheless a new disease specific HRQoL instrument has recently been introduced, the Parkinson's Impact Scale (PIMS) (Calne et al., 1996) which was developed primarily for use with patients with PD in a clinical setting. The development of a new HRQoL instrument should not be undertaken unless there are compelling reasons to do so and as Streiner (Streiner and Norman, 1995) indicates there is a daunting array of available scales. Proliferation of the number of scales available impedes research, since there are problems in generalising from one set of findings to another. Instrument development is justifiable if there is none available which meets the needs of the proposed study or if the instruments already developed are inadequate on the basis of published research or lack of practicality as a result of their length, complexity or mode of administration. It is necessary to evaluate the PIMS in comparison to the PDQ-39 which is becoming the gold standard disease-specific HRQoL instrument in PD to ascertain whether an additional disease-specific instrument adds to the knowledge already in existence and collated from the established PDQ-39. Continued evaluation of the suitability of existing instruments such as the PDQ-39 for use as therapeutic outcome measures in PD management, and clinical studies is mandated and the goal of this study was to establish the disease specific questionnaire of choice for the assessment of HRQoL in PD patients by clinicians, nurses and other health care professionals.

The PIMS (Calne et al., 1996) has 10 items (*work, travel, family, community, sexuality, self, leisure, feelings, safety and financial security*) which are broadly formulated and concern domains rather than specific situations (Marinus J et al., 2002). Higher scores reflect lower HRQoL. The PIMS contains four distinct, important and commonly identified factors:

- Physical (work & travel)
- Social (family, community, & sexuality)

- Psychological (patient's perception of self, leisure, feelings and safety)
- Economic (financial security)

Stable patients only score each item once, whereas patients with fluctuations judge the negative impact for both "on" and "off" periods. The scale can be self completed, but the developers recommend that patients be advised with respect to their disease state (stable or fluctuating). A global score is derived from assigning a weighting of each of the item scores providing a score in the range of 0 to 40 with 0 indicating no change and 40 indicating severe impairment.

AIMS

- To explore the construct and convergent validity of a disease specific HRQOL assessment, the PIMS in PD patients attending a Parkinson's out-patient clinic. For more detail about the PIMS see chapter two
- To explore the relationships between PIMS patient scores and various patient characteristics
- To explore the discriminative validity of the PIMS patient scores using a measure of depression (BDI), clinical severity (Hoehn & Yahr) and motor impairment (UPDRS). For more detail about the BDI, Hoehn and Yahr and UPDRS see chapter two
- To explore the internal consistency of the PIMS factors

METHODS

Fifty six patients meeting the brain bank criteria (Hughes A, 1992) for PD were randomly screened from a Movement Disorder clinic register at Harold Wood Hospital, Romford, Essex by the consultant neurologist and Parkinson's Clinical Nurse Specialist during a 6 month period.

Each patient had explained to them the purpose of the study, and what was involved before being asked to sign an informed consent form. The Parkinson's Clinical Nurse Specialist at the hospital explained the PIMS and PDQ-39 to each patient in a standard scripted fashion. Each patient self completed the Parkinson's Impact Scale (PIMS), Parkinson's disease Quality of Life Questionnaire (PDQ-39), and Beck

Depression Inventory (BDI) prior to their scheduled clinical assessment. The Clinical Nurse Specialist or Consultant Neurologist performed a clinical assessment during each visit including administration of the Mini Mental State Examination (MMSE).

Disease severity of motor symptoms was measured by the UPDRS, (motor section), clinical staging by the Hoehn & Yahr Index, cognitive impairment by the Mini Mental State Examination and depressive symptomatology by the self-report Beck Depression Inventory. Additional information was collected about treatment at baseline and change in treatment at each subsequent visit. Basic demographic data was also collected at baseline.

There were a total of 3 visits over a 6 month duration. At visit 1 (month 0), visit 2 (month 3) and visit 3 (month 6) the patient was asked to complete the PIMS, PDQ-39, MMSE, BDI, and treatment information. At visits 2 and 3 the patient was asked a transition question to assess any change in the effects of their Parkinson's disease on their everyday life since their last visit. The Clinical Nurse Specialist or Consultant Neurologist completed a UPDRS (motor section) and Hoehn and Yahr assessment at all of the visits (see figure 5.1 below). Ethics approval was obtained from the Barking & Havering Local Research Ethics Committee (ref: LREC B&H 19).

Data processing and analysis

The summary scores and global scores of the PIMS and PDQ-39 were calculated according to the respective scoring algorithms. Non-parametric tests for comparisons were selected because of the ordinal nature of the scales and the small sample size (Petrie A and Sabin C, 2000). Group differences were analysed by Wilcoxon signed ranks test (paired observations), Wilcoxon Two Sample test (independent groups of observations) and paired *t*-test to test the null hypothesis that the mean of a set of differences of paired observations is equal to zero. Kruskal-Wallis one-way ANOVA was employed to test for the statistical significance of differences in PIMS scores between different categories of the Hoehn & Yahr, UPDRS and Beck Depression Inventory scales. Associations between scales were measured with Spearman rank correlation coefficients.

The validity of the PIMS was examined in three ways:

- **Construct validity**: by measuring the association between the PIMS global score, the factor (domain) scores between stable, fluctuating and non-fluctuating patients with the clinical ratings of disease severity in Parkinson's disease (Hoehn & Yahr and UPDRS);
- **Convergent validity**: by examining the PIMS correlation with the disease specific (PDQ-39) measure of QoL;
- **Discriminant validity**: determining whether the PIMS successfully differentiated between patients with and without features of advanced disease e.g. disease severity, motor complications and depression.

Statistical significance was set at the $p<0.05$ value and adjustment for multiple testing (Bonferroni adjustment) was not incorporated in to the analysis as this study is considered to be preliminary research and hypothesis generating. Differences with $p<0.05$ can be considered a trend that justifies further investigation with larger sample sizes.

Data from completed questionnaires and clinical findings were checked for completeness and accuracy. The data was then coded for computer analysis and entered onto a database and analysed using a statistical package for social scientists (SPSS version 10 for Windows).

RESULTS

Demographic Characteristics of the Study Patients

Of the 56 patients screened, a total of 47 patients completed quality of life questionnaires at visit 1 (month 0), a response rate of 83.9%; 35 subjects completed questionnaires at visit 2 (month 3) and 27 subjects completed questionnaires at visit 3 (month 6). The mean age of the cohort was 65.4 years; range 34-88 and the gender breakdown was 53.2% male and 46.8% female at visit 1; 57.1% male and 42.9% female at visit 2; and 59.3% male and 40.7% female at visit 3. The mean disease duration was 6.7 years with range of <1 - 21 years (Table 5.1).

Figure 5.1 Parkinson's Impact Scale Study - Methods

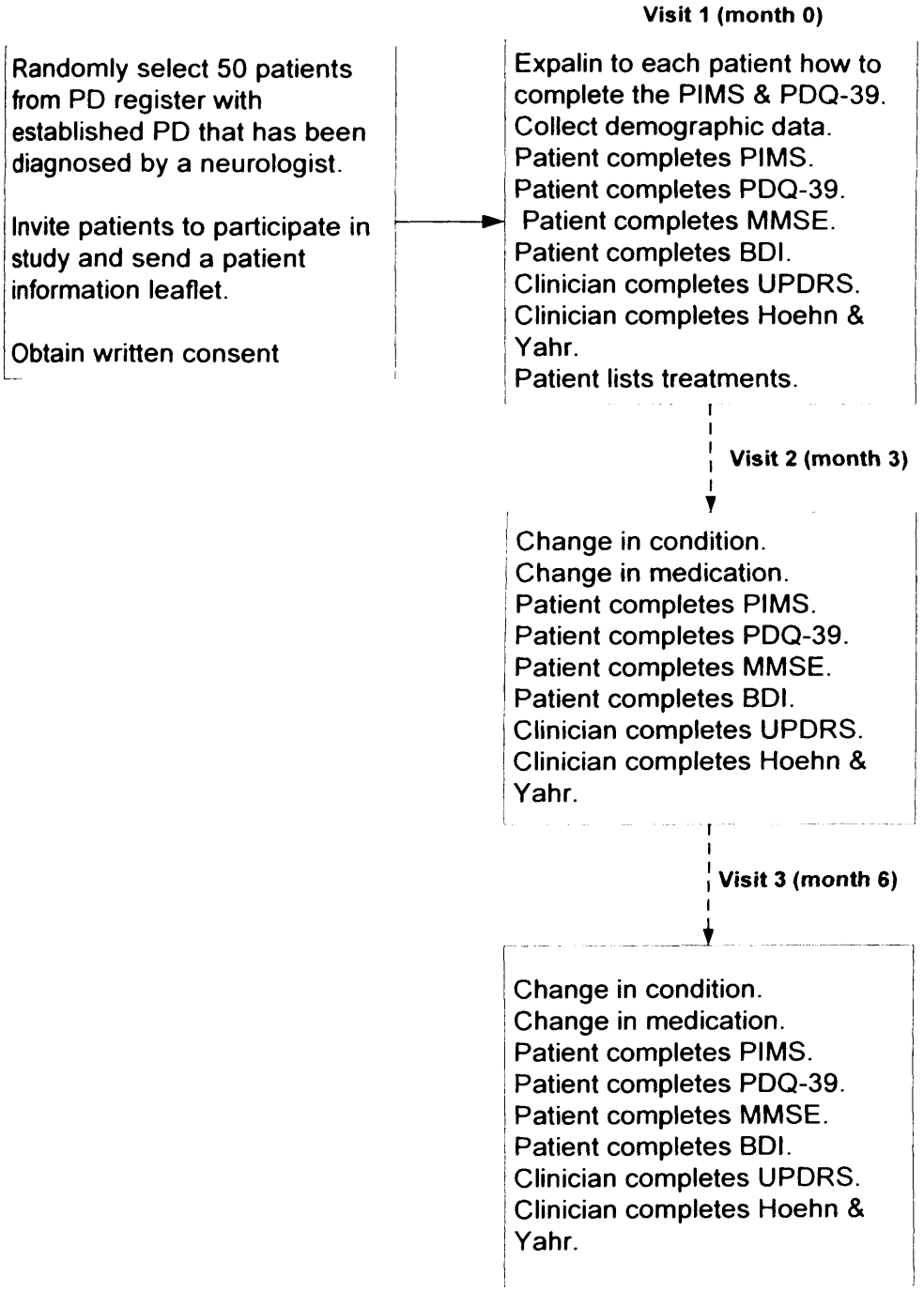


Table 5.1: Demographic characteristics of the study patients (n=47)

Age - years	
Mean (Range)	**65.4 (34 – 88)**
Gender	
Male (%)	**25.0 (53.2)**
Female (%)	**22.0 (46.8)**
Disease duration - years	
Mean (SD)	**6.7 (6.2)**
Median (Range)	**5.0 (<1 – 21)**

Patient's level of performance

Patients were asked to report their level of performance based on the severity of their perceived PD symptoms. The patient was asked to tick the box that most resembled their PD state at the time of the visit and these questions were similar to the Hoehn & Yahr staging questions but simplified for the purpose of self reporting. The level of performance questions were as follows:

Stage 1: Symptoms on one side of the body only. Daily activities only mildly impaired

Stage 2: Symptoms on both sides of the body only. Daily activities only mildly impaired

Stage 3: Balance and co-ordination are now affected and falls often occur. Daily activities moderately impaired

Stage 4: Unable to carry out daily activities

Stage 5: Wheelchair required, or unable to move from bed

Five patients (10.6%) reported a level of performance which were self reported as mild (stage 1); 24 patients (51%) reported a level of performance which was mild to moderate (stage 2 and 3); and 13 patients (27.6%) reported a level of performance which was severe (stage 4 and 5). Five patients (10.6%) did not record their level of self reported performance (Table 5.2). More that half the sample of patients reported

a level of performance that was mild to moderate where daily activities were only mildly impaired.

Patient's disease severity

Patients self reported levels of performance were similar to clinicians staging of Parkinsonism as measured by the Hoehn & Yahr scale. The sample mostly comprised individuals with less severe scores according to the Hoehn & Yahr scale which was completed by the clinician or nurse (Table 5.3). 33.3% of subjects were assessed as mild symptoms (Stage I); 57.2% as mild to moderate (Stage II-III) and 9.5% as severe (Stage IV). Here 33.3% of patients were classified as having minimal or no functional impairment and the majority as bilateral symptoms without impairment of balance; i.e. a stable PD sample. Modern treatment strategies have shifted the time to reaching later stages of PD significantly to the right, so that it is common to see patients with PD, still at stage III after more than 15 years(Dewey Richard B, 2000). It is therefore unlikely to expect marked deterioration of clinical staging in the duration of this study.

Table 5.2 Level of performance (as reported by patient)

	Visit 1	
	N	%
Number of subjects with data	47	(100.0%)
Stage 1: Symptoms on one side of the body only. Daily activities only mildly impaired.	5	(10.6%)
Stage 2: Symptoms on both sides of the body only. Daily activities only mildly impaired.	15	(31.9%)
Stage 3: Balance and co-ordination are now affected and falls often occur. Daily activities moderately impaired.	9	(19.1%)
Stage 4: Unable to carry out daily activities	9	(19.1%)
Stage 5: Wheelchair required, or unable to move from bed	4	(8.5%)
Stage not recorded	5	(10.6%)

Table 5.3 Hoehn and Yahr stage (Completed by clinician or nurse)

	Visit 1	
	N	%
Number of evaluable patients	42	(100.0%)
Stage 1	14	(33.3%)
Stage 2	13	(31.0%)
Stage 3	11	(26.2%)
Stage 4	4	(9.5%)

Stage 1 – mild symptoms of PD

Stage 2 & 3 – mild to moderate symptoms of PD

Stage 4 and 5 – severe symptoms of PD

However, as Table 5.4 shows, when patients were asked to indicate their impression of global change compared to the baseline visit nearly half of those sampled indicated that they were "a little" or else "much worse". This highlights how patients own evaluation of their health status can differ from the clinician's assessment.

Table 5.4 Change since previous visit

	Visit			
	2		3	
	N	%	N	%
Number of evaluable patients	34	(100.0%)	26	(100.0%)
Much better	1	(2.9%)		
Better	3	(8.8%)	7	(26.9%)
The same	13	(38.2%)	4	(15.4%)
A little worse	15	(44.1%)	13	(50.0%)
Much worse	2	(5.9%)	2	(7.7%)

Cognitive impairment in PD

Neuropsychological investigations of patients with Parkinson's disease have shown cognitive impairments even in the early stages of the disease (Dubois and Pillon, 1997). Cognitive impairment erodes ones ability to think, memorize and maintain mood and thus undoubtedly contributes to poor QoL. PD patients with cognitive impairment tend to have more severe motor impairment and also higher levels of depression than those who are not demented (Hobson and Meara, 1999). Therefore, patient's level of cognitive function as measured by the MMSE was assessed. In this cohort of patients, none of the subjects were assessed as having cognitive impairment (all scored > 20 on the MMSE scale where a score of < 20 is indicative of cognitive impairment and a score of = 15 is indicative of severe dementia). The mean MMSE score at visit 1 and 2 was 27.8 and at visit 3, 27.2 (range 21 – 30) (Table 5.5). The absence of cognitive impairment is probably reflected by the fact that this was a hospital-based study. Referral bias would tend to reduce the prevalence of cognitive impairment and dementia in hospital based clinic populations. Also, only 9.5% of patients were assessed as severe with respect to clinical staging which reflects that the majority of patient's symptoms and associated symptoms of PD were not advanced. The mean age of patients from a community sample study of cognitive function in elderly patients with PD(Hobson and Meara, 1999) was 76.6 years reflecting that dementia in PD is strongly linked to age. In this study the mean age of patients was 65.4 which reflect a slightly younger population in comparison.

Depression in PD

Depression is a common problem in patients with Parkinson's disease, but its mechanism is poorly understood (Schrag et al., 2001). According to published studies, the prevalence of depression ranges between 20 to 90% (Dogu Okan et al., 2002). The average rate of depression in PD was found to be 40%. In this sample of patients depression measured by the BDI ranged from 0 – 32. The mean BDI score at visit 1 was 12.4; visit 2, 13.5; and visit 3, 13.4 (Table 5.6). This highlights that patients from this sample were mildly depressed as measured by the BDI. Depressive symptoms are related to both disease severity and disease duration (Playfer, 1999).

Table 5.5 MMSE mean total scores for Patients with PD

	Visit			Wilcoxon Matched Pairs Signed Ranks test	Wilcoxon Matched Pairs Signed Ranks test
	1	2	3	V2 & V1	V3 & V1
Number of evaluable patients	46	28	26		
Mean	27.8	27.8	27.2	0.586	0.472
Std dev	1.8	1.8	2.6		
Minimum	23.0	23.0	21.0		
25%ile	27.0	27.0	26.0		
Median	28.0	28.0	28.0		
75%ile	29.0	29.0	29.0		
Maximum	30.0	30.0	30.0		

Maximum total score is 30

Score <20 is indicative of cognitive impairment

Score <15 is indicative of severe dementia

The mild severity of depressive symptoms in this cohort of PD patients is reflected by a mean disease duration of 6.7 years which when compared to a community based study (Ali and Salek, 2002) reported in this thesis was 11 years. A large international randomised survey of PD patients shows that there is a strong correlation between depression and quality of life (Findley and Committee), 2002).

The following behavioural traits of depression were the most common across the three visits; difficulty with sleeping, ability to do any activity or work, tiredness, difficulty making decisions and loss of interest in sex (Table 5.7).

Motor function

The level of motor impairment was assessed by the Part III of the UPDRS which evaluates bradykinesia, tremor, rigidity, speech, facial expression, postural stability and gait; all of which are the cardinal features of PD.

Table 5.6 Beck Depression inventory mean total scores in patients with PD

	Visit			Wilcoxon Matched Pairs Signed Ranks test	Wilcoxon Matched Pairs Signed Ranks test
	1	2	3	V2 & V1	V3 & V1
Number of evaluable patients	47	35	25		
Mean	12.4	13.5	13.4	0.489	1.000
Std dev	6.9	7.8	8.3		
Minimum	0.0	1.0	2.0		
25%ile	7.0	7.0	6.0		
Median	11.0	13.0	11.0		
75%ile	18.0	20.0	21.0		
Maximum	25.0	32.0	29.0		

Possible range of scores is 0 to 63

Increasing scores represent more depression

10-16 = mild; 17-29 = moderate; 30-63 = severe depression

The range of scores was 4.0 – 40 at visit 1 with a mean of 17.6 (Table 5.8) which reflects a mild level of motor impairment. UPDRS scores improved at visit 2 (mean = 13.6) and at visit 3 (mean = 13.0). The improvement in motor impairment between visit 1 and visit 3 was statistically significant ($p < 0.0005$ Wilcoxon Matched Pairs Signed Rank test). In clinical trials of therapeutic interventions the UPDRS is primarily used as the primary outcome measure. The outcome measure in a multi-centre, double blind, randomised 3 month trial of pergolide (a dopamine agonist) (Barone P et al., 1999) showed a =30% decrease in UPDRS part III score at end point compared to placebo. Though this study was not a randomised, controlled clinical trial it is interesting to reflect that the patients showed a 26.13% decrease in UPDRS part III score after 6 months (visit 3).

This is likely to be a consequence of symptomatic relief and control due to PD medication and probably also a consequence of dedicated follow up of the sampled patients by the Parkinson's nurse through the duration of this study.

Table 5.7: Responses to the 3rd or 4th items of each BDI statement

Statement	N visit 1	%	N visit 2	%	N Visit 3	%
Feeling sad/unhappy	0	**0**	3	**8.6**	1	**4**
Feeling hopeless	7	**15.2**	8	**22.9**	5	**20**
Failure	2	**4.3**	4	**11.5**	0	**0**
Satisfaction	4	**8.5**	2	**5.8**	2	**8**
Feel guilty	9	**19.6**	7	**20.6**	4	**16**
Punished	1	**2.2**	0	**0**	1	**4**
Feel disappointed	0	**0**	1	**2.9**	0	**0**
Blame myself	4	**8.7**	2	**5.8**	0	**0**
Suicidal thoughts	1	**2.2**	0	**0**	1	**4.3**
Crying	3	**6.6**	1	**2.9**	2	**8.4**
Irritated	1	**2.1**	3	**8.6**	2	**8**
Lost interest	1	**2.1**	0	**0**	2	**8.3**
Difficulty making decisions	11	**23.4**	11	**31.4**	7	**28**
Look ugly	2	**4.3**	7	**20**	6	**24**
Cant do any work	17	**36.2**	8	**22.9**	8	**32**
Cant sleep	17	**36.2**	16	**45.7**	11	**44**
Too tired	13	**27.7**	12	**34.3**	7	**28**
No appetite	2	**4.3**	2	**5.7**	1	**4**
Lost weight	2	**4.8**	0	**0**	2	**9**
Physical problems	5	**10.6**	5	**14.7**	2	**8.7**
Lost interest in sex	11	**25.6**	7	**21.9**	9	**36**

3rd and 4th items of each BDI question indicate worst possible scores

Table 5.8 UPDRS mean total scores in patients with PD

	Visit (V)			Wilcoxon Matched Pairs Signed Ranks test	Wilcoxon Matched Pairs Signed Ranks test
	1	**2**	**3**	**V2 & V1**	**V3 & V1**
Number of evaluable patients	40	13	23		
Mean	17.6	13.6	12.0	0.722	0.0005
Std dev	8.2	8.2	5.7		
Minimum	4.0	3.0	2.0		
25%ile	11.0	6.0	8.0		
Median	14.5	14.0	12.0		
75%ile	22.5	17.0	14.0		
Maximum	40.0	32.0	22.0		

Possible range of scores is 0 to 56

Increasing scores represent worsening motor function

=14 = slight; 15-28 = mild; 29-42 = moderate; 43-56 = severe motor impairment

Patients were classified as early-stage (patients who present with the first symptoms and signs of PD – minor motor manifestations) or Advanced PD patients with symptoms > 1 year). Advanced PD patients were further categorised as those with insufficient control of motor symptoms and stable. At baseline (Visit 1) 46% were early stage PD patients, 54% advance patients of which 32.4% had insufficient control of motor symptoms and 21.6% were stable (Table 5.9).

Parkinson's Impact Scale (PIMS)

On the PIMS, patients assigned a score to each question in column 1 if their symptoms were stable over time and were not subject to random fluctuations ("on/off" attacks). Patients with fluctuations in mobility assigned scores to two other columns, 2a "best" and 2b "worst".

At baseline (visit 1) 31 patients (67.4%) reported stable symptoms compared to 15 (32.6%) who reported fluctuating symptoms (Table 5.10).

Table 5.9 PD staging of the study patients as classified by UPDRS

Early stage or advanced PD:	Visit					
	1		2		3	
	N	%	N	%	N	%
Number of subjects with data	37	(100.0%)	10	(100.0%)	22	(100.0%)
Early stage PD patient	17	(46.0%)	6	(60.0%)	13	(59.1%)
Advanced PD: insufficient control of motor symptoms	12	(32.4%)	2	(20.0%)	7	(31.8%)
Advanced PD: stable	8	(21.6%)	2	(20.0%)	2	(9.1%)

Table 5.10 PD patients completing column 1 (indicating stable disease) and column 2 (indicating fluctuating disease)

Subject break down according to visit:			
PIMS	**Visit 1 (n)**	**Visit 2 (n)**	**Visit 3 (n)**
Symptoms stable (Column 1)	31 (67.4%)	20 (58.8%)	15 (55.6%)
Symptoms fluctuate (Column 2a & 2b)	15 (32.6%)	14 (41.2%)	12 (44.4%)
Total	46 (100%)	34 (100%)	27 (100%)

Specific problems most commonly reported by PD patients:

Problems most commonly reported via individual questionnaire items are listed in Table 5.11. Patients were particularly likely to experience problems of physical function such as pursuing leisure activities (68.1%), looking after the home (63.8%), getting around in public places (55.3%) and being confined to the house (51.1%). Fitzpatrick et al (Fitzpatrick et al., 1997) reported a similar pattern in a study of outpatient clinic attendees with the exception of the items " being confined to the house". The most common problem of all was pursuing leisure activities (68.1%). Problems' writing clearly (63.8%) was the second joint highest problem reported from this sample of clinic attendees.

Frequency of missing items in PIMS:

An analysis of missing items from questionnaires completed by stable patients (PIMS 1 – stable), fluctuating patients in their best state (PIMS 2a – best) and, worst state (PIMS 2b – worst), reveals that the social scales have the highest number of missing items followed by the economic scale (Table 5.12). Fluctuating best patients missed less items compared to fluctuating worst patients of whom 40% did not answer community, family relationships and financial security items. The frequency of missing items for stable patients was highest for the work (25.8%) and sexuality items (25.8%). Such a short instrument (10 items) should not have so many missing items which reveal that there is either a difficulty in understanding how to complete the questionnaire or else the items have no relevance to the PD group sampled.

The PIMS Scores

Scores across the three visits (Table 5.13) were similar with a mean global score (stable) of 12.8 at visit 1 (quartile 7.7-19.9), 14.1 at visit 2 (quartile 9.0-21.2) and 13.6 at visit 3 (quartile 5.8-17.6). There were no significant differences between baseline and visit 2 and 3 scores. Global scores for fluctuating patients (best) were also similar with scores of 22.0 (quartile 19.2-24.1) at visit 1; 16.8 (quartile 13.2-20.0) at visit 2; 21.8 (quartile 8.4-30) at visit 3. There were no differences between baseline and visit 2 and 3 scores (p = NS).

Global scores for fluctuating patients (worst) were similar with scores of 30.4 (quartile 27.6-33.0) at visit 1; 30.0 (quartile 22.1-40.0) at visit 2; 28.5 (quartile 20.2-37.3)) at visit 3. There were no differences between baseline and visit 2 and 3 scores (p = NS). Fluctuating (best) respondents scored higher scores than stable patients indicating lower HRQoL and fluctuating (worst) respondents scored higher scores than fluctuating (best) and stable respondents. This indicates that fluctuating patients experiencing both on and off periods judge their symptoms to have a negative impact on their quality of life.

Analysis of the factor scores (Table 5.15) shows that the physical domain of the PIMS across all three clinical states (stable", fluctuating best and worst) is the most impacted with the level of impairment greater in the fluctuating groups and in particular when in the "off" state which is defined as "fluctuating – worst". The next domain which the PD has the greatest impact as measured by the PIMS is the

psychological domain. The domain with the lowest score indicating least change to HRQoL was the economic domain which refers to the subject's ability to support them-selves, family and pay medical costs.

Table 5.11: Problems reported from the PDQ-39 as experienced sometimes, often, or always in the last month by more than half the clinic sample

Visit 1	Reporting a problem	
PDQ-39 Item (scale)	n	%
Difficulty with leisure activities (mobility)	32	68.1
Difficulty looking after home (mobility)	30	63.8%
Difficulty getting around in public places (mobility)	26	55.3
Confined to the house (mobility)	24	51.1
Problems writing clearly (activities of daily living)	30	63.8
Aches and pains in the body (bodily discomfort)	28	59.6
Unexpectedly fallen asleep during the day (cognitive impairment)	26	55.3
Felt depressed (emotional well being)	24	51.1

Table 5.12: Frequency of missing items in PIMS at Visit 1

		Missing items	
	Item (scale)	**n**	**%**
PIMS 1: Stable (N = 31)	**Community relationships** (social)	**5**	**16.1**
	Family relationships (social)	**5**	**16.1**
	Financial security (economic)	**7**	**22.6**
	Leisure (psychological)	**5**	**16.1**
	Safety (psychological)	**5**	**16.1**
	Self (negative) - (psychological)	**6**	**19.4**
	Self (positive) - (psychological)	**5**	**16.1**
	Sexuality (social)	**8**	**25.8**
	Travel (physical)	**6**	**19.4**
	Work (physical)	**8**	**25.8**
PIMS 2a: Fluctuating – best (N = 15)	**Community relationships** (social)	**0**	**0.0**
	Family relationships (social)	**0**	**0.0**
	Financial security (economic)	**3**	**20.0**
	Leisure (psychological)	**2**	**13.3**
	Safety (psychological)	**2**	**13.3**
	Self (negative) - (psychological)	**2**	**13.3**
	Self (positive) - (psychological)	**3**	**20.0**
	Sexuality (social)	**2**	**13.3**
	Travel (physical)	**2**	**13.3**
	Work (physical)	**2**	**13.3**

Table 5.12: Frequency of missing items in PIMS at Visit 1- cont.

PIMS 2b: Fluctuating – worst (N = 15)	**Community relationships** (social)	**6**	**40.0**
	Family relationships (social)	**6**	**40.0**
	Financial security (economic)	**6**	**40.0**
	Leisure (psychological)	**3**	**20.0**
	Safety (psychological)	**5**	**33.3**
	Self (negative) - (psychological)	**2**	**13.3**
	Self (positive) - (psychological)	**3**	**20.0**
	Sexuality (social)	**4**	**26.7**
	Travel (physical)	**4**	**26.7**
	Work (physical)	**3**	**20.0**

Table 5.13: Functional Impairment in out-patient study participants as measured by the PIMS (Global Scores)

Scale	Visit 1 (V1) n = 47	Visit 2 (V2) n = 35	Visit 3 (V3) n = 27	Wilcoxon Matched Pairs Signed Ranks test V2 & V1	Wilcoxon Matched Pairs Signed Ranks test V3 & V 1
Global Score *(stable)*					
Mean (S.D.)	12.8 (6.5)	14.1 (7.4)	13.6 (9.2)	0.878	0.500
Median (Range)	12.5 (0.0–23.0)	15.7 (1.8-24.4)	15.2 (0.6-30.6)		
25%ile – 75%ile	7.7 – 19.9	9.0 – 21.2	5.8 – 17.6		
Global Score *(fluctuate best)*					
Mean (S.D.)	22.0 (6.4)	16.8 (7.1)	21.8 (8.4)	0.144	0.108
Median (Range)	23.1 (11.4-30.7)	16.7 (6.2-30.7)	24.0 (8.4-30.0)		
25%ile – 75%ile	19.2 – 24.1	13.2 – 20.0	19.9 – 26.7		
Global Score *(fluctuate worst)*					
Mean (S.D.)	30.4 (6.2)	30.0 (7.6)	28.5 (8.3)	1.000	0.654
Median (Range)	31.1 (19.5-40.0)	28.5 (22.1-40.0)	28.2 (20.2-37.3)		
25%ile – 75%ile	27.6 – 33.0	23.2 – 37.6	21.5 – 35.5		

PIMS scale scores from out-patient clinic respondents with Parkinson's disease
Possible range of scores is 0 to 40
Increasing scores represent more impairment
P<0.001

Table 5.14: Functional Impairment in out-patient study participants as measured by the PIMS (factors scores):

Scale	Visit 1 (V1) n= 23	Visit 2 (V2) n = 19	Visit 3 (V3) n = 14	Max possible score	% of total score V1, V2, V3
Stable patients *(column 1)*:					
Physical 1 (stable)				5.2	36.5, 40.4, 34.6
Mean (S.D.)	1.9 (1.5)	2.1 (1.4)	1.8 (1.7)		
Median (Range)	1.4 (0.0-5.2)	2.1 (0–5.2)	1.8 (0–5.2)		
Social1 (stable)				12.0	26.7, 23.3, 23.3
Mean (S.D.)	3.2 (2.4)	2.8 (2.5)	2.8 (2.2)		
Median (Range)	3.0 (0.0-8.3)	3.2 (0-7.2)	2.7 (0-6.1)		
Psycholog 1 (stable)				20.8	37.5, 37.5, 34.1
Mean (S.D.)	7.8 (4.5)	7.8 (4.5)	7.1 (5.5)		
Median (Range)	7.0 (0-20.8)	8.0 (0-14.5)	6.3 (0-18.4)		
Economic 1 (stable)				2.0	30, 20, 30
Mean (S.D.)	0.6 (0.7)	0.4 (0.6)	0.6 (0.8)		
Median (Range)	0.5 (0-2.0)	0 (0-2.0)	0.2 (0-2.0)		
Fluctuating patients -best *(column 2a)*:					
Physical 2a (best)				5.2	63.5, 57.8, 50.0
Mean (S.D.)	3.3 (1.0)	3.0 (1.0)	2.6 (1.5)		
Median (Range)	3.6 (1.3-4.6)	2.6 (1.3-4.5)	3.2 (0.6-3.9)		
Social2a (best)				12.0	44.2, 41.7, 50.8
Mean (S.D.)	5.3 (2.4)	5.0 (2.6)	6.1 (2.9)		
Median (Range)	5.2 (0-9.0)	5.4 (1.2-9.6)	7.2 (1.3-9.0)		
Psycholog 2a (best)				20.8	52.9, 40.9, 58.2
Mean (S.D.)	11.0 (4.1)	8.5 (4.2)	12.1 (3.4)		
Median (Range)	13.0 (4.0-15.6)	9.3 (2.4-15.2)	13.1 (6.5-15.6)		
Economic 2a (best)				2.0	30.0, 30.0, 60.0
Mean (S.D.)	0.6 (0.6)	0.6 (0.7)	1.2 (0.6)		
Median (Range)	0.5 (0-1.5)	0.5 (0-2.0)	1.0 (0-2.0)		
Fluctuating patients -worst *(column 2b)*:					
Physical 2b (worst)				5.2	84.6, 88.5, 69.2
Mean (S.D.)	4.4 (0.8)	4.6 (0.6)	3.6 (1.5)		
Median (Range)	4.6 (2.5-5.2)	4.5 (3.9-5.2)	3.6 (1.3-5.2)		
Social2b (worst)				12.0	64.2, 67.5, 70.0
Mean (S.D.)	7.7 (2.7)	8.1 (3.0)	8.4 (3.3)		
Median (Range)	8.5 (3.6-12.0)	8.4 (2.4-12.0)	8.9 (4.3-11.4)		
Psycholog 2b (worst)				20.8	76.9, 78.8, 76.9
Mean (S.D.)	16.0 (3.6)	16.4 (3.0)	16.0 (3.6)		
Median (Range)	16.9 (9.8-20.8)	15.6 (11.7-20.8)	16.3 (11.6-19.7)		
Economic 2b (worst)				2.0	55.0, 40.0, 60.0
Mean (S.D.)	1.1 (0.8)	0.8 (1.0)	1.2 (0.7)		
Median (Range)	1.0 (0-2.0)	0 (0-2.0)	1.2 (0-2.0)		

Increasing scores represent more impairment

The group with fluctuations-worst reported greater levels of impairment in all domains compared to the other groups confirming that fluctuating patients in an "off" state experience more disability with respect to HRQoL and this is quite important to realise and the choice of initial therapy may prove fundamental to a treatment strategy that maximizes symptomatic control while minimizing the chances for long-term complications such as motor fluctuations and dyskinesias (Stern, 2000). Such a finding would allow the health care practitioner to modify the treatment approach to alleviate the possible side-effects of the existing treatment. As stated all factors had an increase in score (indicating worsening of HRQoL) in fluctuating patients in the "off" or worst state (column 2b) relative to the non-fluctuating patients. All factors reflected an increase in values from the "on" state (column 2a) to the "off state (column 2b). Construct validity was thus established with respect to the external criterion of level of symptoms.

Internal Consistency Reliability

Reliability is an important concept to consider in the assessment of a HRQoL instrument and refers to the consistency, stability or the reproducibility of scores obtained during different administrations of the scales when conditions remain relatively unchanged (Coons and Kaplan, 1992). *Internal consistency reliability* is based on the average inter-item correlation among the individual questions in the scale (Damiano et al., 1999) and hence is a measure of the homogeneity of items within a scale, whereby a certain level of association is expected between scale items if they are measuring the same overall concept. Hence one would expect moderate correlations amongst the items making up a domain which purports to measure physical, psychological or social well-being.

The desirable level of internal consistency chosen in this study was 0.7 as it is often cited in the literature (Nunnally, 1978). The results show that the PIMS exhibits varying degrees of internal consistency and mostly below a reliability coefficient of 0.7 (Tables 5.15a, b and c). The coefficient values varied from each visit and more so at visit 3 where the psychological domain exhibited good internal consistency compared to visit 1 and 2 coefficients alpha. Internal consistency was not stable between baseline and the last patient visit. The reliability at visit 3 is likely to improve

Table 5.15a: Internal Consistency Reliability (Cronbach a) for PIMS

Visit 1 (baseline)	PIMS 1	PIMS 2a	PIMS 2b
	N = 21	N = 9	N = 8
Scale			
Psychological	0.665	0.365	0.661
Social	0.680	0.284	0.623
Physical	0.640	0.693	0.597
Economic	0.747	0.731	0.658

Table 5.15b:

Visit 2 (month 3)	PIMS 1	PIMS 2a	PIMS 2b
	N = 14	N = 9	N = 6
Scale			
Psychological	0.544	0.302	0.610
Social	0.413	0.447	0.654
Physical	0.627	0.525	0.754
Economic	0.741	0.558	0.662

Table 5.15c:

Visit 3 (month 6)	PIMS 1	PIMS 2a	PIMS 2b
	N = 11	N = 9	N = 4
Scale			
Psychological	0.757	0.792	0.750
Social	0.561	0.712	0.570
Physical	0.537	0.808	0.674
Economic	0.702	0.904	0.883

because patients are beginning to complete the questionnaires more consistently and with a better appreciation of what is being asked. The authors of the PIMS (Calne et al., 1996) reported that internal consistency of all 10 items was high (0.89) as measured by Cronbach's alpha on a sample size of 109 in a cross-sectional study. In this analysis internal consistency was measured using the 4 identified factors (dimensions) i.e. psychological, social, physical and economic dimensions and the sample size at baseline was smaller by comparison. The economic factor in general exhibited good internal consistency across all visits and for both fluctuating and non-fluctuating groups but this is not surprising as only one item makes up the economic domain. One would expect higher correlation coefficients to be reported in this analysis as the domains constitute a range of items not exceeding 4.

The PDQ-39 Scores

Scores across the three visits were similar with a mean overall score of 30.8 at visit 1 (quartile 13.6-42.8), 29.5 at visit 2 (quartile 12.8-45.0) and 28.0 at visit 3 (quartile 11.5-44.2). There were no differences between baseline and visit 2 and 3 scores (p = NS). The domains of bodily discomfort, mobility and activities of daily living revealed the highest scores, indicating a lower HRQoL across the three visits. Communication and social support had the lowest scores across the three visits indicating that these domains had the least impact, in comparison, on HRQoL (Table 5.16).

The internal consistency of the PDQ-39 has been previously reported and commented in this thesis (Ali and Salek, 2002) but nevertheless for comparison the PDQ-39 results show that all scales (domains) have good internal consistency and are all well above a reliability coefficient of 0.7 (Table 5.17). Overall the PDQ-39 items within the scales are measuring similar areas of functioning in PD patients, which adds validity to the scoring profile produced by the questionnaire. The PIMS on the other hand exhibited varying degrees of internal consistency between the visits (Figures 5.2, 5.3 and 5.4) and also between the fluctuating and non-fluctuating cohort suggesting further testing in larger sample size in probably necessary before concluding on the internal consistency reliability of the PIMS.

Construct validity

Construct validity reflects the logical relationship that should exist between scores on the instrument when given to patients with differing disease severity (discriminant validity) and the degree of correlation between scores on the new instrument and existing validated measures covering similar domains (convergent validity). Construct vaiidity has not been thoroughly established for the PIMS as only known group comparisons were performed, demonstrating significant differences between stable and fluctuating patients in their off situations. Construct validity involves comparison between measures and examines the logical relations that should exist between a measure and characteristics of patients and patient groups (Lydick et al., 1997). A good instrument should distinguish between groups with and without disease related impact on symptoms, function, health status and emotional correlates.

Table 5.16: Functional Impairment in subjects from an out-patient clinic using the PDQ-39

PDQ-39 Category	Visit 1 (V1) n = 47	Visit 2 (V2) n = 35	Visit 3 (V3) n = 27	Wilcoxon Matched Pairs Signed Ranks test V2 & V1	Wilcoxon Matched Pairs Signed Ranks test V3 & V 1
Overall					
Mean (S.D.)	30.8 (15.7)	29.5 (17.9)	28.0 (18.6)	0.890	0.793
Median (Range)	31.6 (6.0–62.9)	26.5 (8.6–66.1)	23.0 (5.3–61.6)		
25%ile – 75%ile	13.6 – 42.8	12.8 – 45.0	11.5 – 44.2		
Mobility					
Mean (S.D.)	42.6 (28.7)	41.1 (35.7)	41.2 (35.9)	0.860	0.602
Median (Range)	40.0 (0.0-97.5)	35.0 (0.0–100)	32.5 (0.0–95.0)		
25%ile – 75%ile	17.5 – 67.5	5.0 – 75.0	7.5 – 77.5		
Activities of Daily Living					
Mean (S.D.)	36.9 (26.5)	33.5 (27.0)	37.0 (29.3)	0.954	0.435
Median (Range)	29.2 (0.0–91.7)	29.2 (0.0–95.8)	29.2 (0.0–95.8)		
25%ile – 75%ile	16.7 – 54.2	12.5 – 50.0	8.3 – 66.7		
Emotional well being					
Mean (S.D.)	31.1 (22.0)	34.5 (24.2)	33.2 (24.5)	0.665	0.951
Median (Range)	29.2 (0.0–75.0)	29.2 (0.0–87.5)	25.0 (0.0–79.2)		
25%ile – 75%ile	12.5 - 50.0	16.7 – 54.2	16.7 – 50.0		
Stigma					
Mean (S.D.)	23.1 (24.6)	27.6 (24.7)	24.0 (22.9)	0.249	0.736
Median (Range)	18.8 (0.0–87.5)	18.8 (0.0–81.3)	18.8 (0.0–62.5)		
25%ile – 75%ile	0.0 – 37.5	6.3 – 81.3	0.0 – 62.5		
Social Support					
Mean (S.D.)	11.2 (15.0)	11.9 (16.5)	11.3 (15.8)	0.575	0.944
Median (Range)	8.3 (0.0–66.7)	0.0 (0.0–62.5)	0.0 (0.0–50.0)		
25%ile – 75%ile	0.0 – 16.7	0.0 – 16.7	0.0 – 16.7		
Cognition					
Mean (S.D.)	34.6 (21.4)	33.2 (19.5)	32.1 (20.9)	0.290	0.616
Median (Range)	37.5 (0.0–87.5)	31.3 (6.3–75.0)	37.5 (6.3–75.0)		
25%ile – 75%ile	18.8 – 50.0	18.8 – 50.0	12.5 – 50.0		
Communication					
Mean (S.D.)	22.2 (21.9)	15.5 (17.3)	13.5 (13.8)	0.070	0.074
Median (Range)	16.7 (0.0–91.7)	16.7 (0.0–58.3)	8.3 (0.0–50.0)		
25%ile – 75%ile	8.3 – 33.3	0.0 – 25.0	0.0 – 25.0		
Bodily Discomfort					
Mean (S.D.)	42.4 (22.8)	44.8 (26.9)	41.7 (26.2)	0.665	0.493
Median (Range)	50.0 (8.3–100)	58.3 (0.0–100)	50.0 (8.3–91.7)		
25%ile – 75%ile	25.0 - 58.3	25.0 – 75.0	25.0 - 66.7		

PDQ-39 scale scores from out-patient clinic respondents with Parkinson's Disease
Possible range of scores is 0 to 100% impairment
Increasing scores represent more impairment
P<0.001

Construct validity is assessed here to establish whether the PIMS correlates positively with the PDQ-39 (related and similar construct) to exhibit convergent validity; and does not correlate positively with measures of UPDRS and Hoehn and Yahr (related but different constructs) to exhibit divergent validity.

Table 5.17: Internal Consistency Reliability (Cronbach a) for PDQ-39

Visit 1 (baseline)	**PDQ-39**
	N = 43
Scale	
Mobility	0.792
Activities of daily living	0.821
Emotional well being	0.826
Stigma	0.819
Social support	0.836
Cognitive impairment	0.828
Communication	0.832
Bodily discomfort	0.821

Convergent and Divergent validity

Agreement between the PIMS and the PDQ-39

The correlations between domains of both questionnaires were assessed to examine whether related domains measure HRQoL in a similar way. It would be expected that domains aiming to measure similar areas of functioning would be correlated with one another for example the PIMS psychological domain would be expected to correlate strongly with the emotional well being domain of the PDQ-39.

There were widespread strong correlations between the domains of the PDQ-39 and PIMS for patients whose symptoms were stable i.e. non-fluctuating and hence had completed only column 1 of the PIMS. Particularly strong and statistically significant correlations were observed for PIMS physical domain and PDQ-39 mobility and activities of daily living (ADL) domains; the PIMS social domain and PDQ-39 stigma and social support domains; the PIMS psychological domain and PDQ-39 mobility and ADL domains (Table 5.18).

The cohort classified as fluctuating (that is patients whose symptoms were not stable) and who completed columns 2a and 2b revealed a very different picture. None of the PIMS domains for the fluctuating group correlated, even moderately or were statistically significant. The PIMS domain for the fluctuating-worst group exhibited

very low correlations which were not significant with the PDQ-39 domains of mobility and ADL (Table 5.18). This conflicts for the same domains of the PIMS for the non-fluctuating cohort.

The results show that none of the domains of the PIMS from the fluctuating cohort correlated with the PDQ-39 domains, even for domains where an association would be expected such as PIMS physical and PDQ-39 mobility or PIMS social and PDQ-39 social support.

This may suggests that the columns 2a (best) and 2b (worst), that were completed by the patients with fluctuations in symptoms, is redundant, as there is a widespread paucity of correlation between domains. An alternative explanation which is perhaps more plausible is that patients completing columns 2a and 2b are making a distinction between the two possible states they are being asked to assess, the state of when their symptoms are adequately controlled by treatments (column 2a) and the state where their symptoms are not controlled by treatment (column 2b). The patients are probably focussing on the worse of the two possible states i.e. when they are fluctuating and this determines how they answer the questions in the PIMS. The patient is more likely to report how they expect they would grade the items whilst in a fluctuating state rather than how they really are and this may provide for either over-grading or under-grading of the items that make up the PIMS. The poor association with the PDQ-39 in the fluctuating group may also be a reflection of the fact that the PDQ-39 doesn't differentiate between stable and unstable cohorts and hence patients grade the questions on the instrument based on their present state of health in the same way they do with column 1 (symptoms stable) and where an association exists. The precision of the PIMS to detect change in fluctuating patients is not adequate on the basis of these results.

The analysis indicates that the PIMS is likely to have convergent validity only for stable patients, since measures of physical, social and psychological function correlated significantly with conceptually related subscales of the PDQ-39. The PIMS does not appear to be suitable for use in patients who experience fluctuations of symptoms as none of the measures correlated adequately or significantly with conceptually related subscales of the PDQ-39. Also, none of the PIMS domain for both fluctuating and non-fluctuating cohorts correlated with the PDQ-39 cognitions domain which implies that cognition is not an item or factor addressed in the PIMS.

In its entirety, the PIMS does not exhibit adequate convergent validity in PD but also it does highlight that the PDQ-39 may not be able to detect important changes to HRQoL domains in a cohort of fluctuating patients as it does not differentiate between fluctuating and stable cohorts. There is however, a cautionary note with respect to the absence of association between PIMS and PDQ-39 domains, and particularly for fluctuating patients; it is that patient numbers for column 2a and 2b were half of that for non-fluctuating patients.

Agreement between the PIMS Scores and Clinical Indicators of Disease Severity

Two indicators of disease severity chosen for comparison with the PIMS were the UPDRS and the Hoehn and Yahr scale which were completed by the clinician or else the nurse. These indicators were chosen because they are commonly used indicators of disease severity. Spearman's rank correlation coefficients were calculated to measure the association between UPDRS and Hoehn and Yahr scores and the four factors (domains) including the global score of the PIMS (Table 5.19).

The Hoehn and Yahr scores were obtained on the day of the scheduled visit and completed by the clinician or nurse. Scores were not available for all patients, and therefore a reduced sample was used for statistical analysis. None of the correlation coefficients (r_s) were found to be statistically significant at $p<0.05$ between the PIMS domains and the Hoehn and Yahr indicating a lack of association between this clinical outcome measure and the domains of HRQoL measured by the PIMS. However there were statistically significant correlations between the domain of PIMS physical ($p<0.05$), PIMS psychological ($p<0.005$) and the UPDRS total score in the stable cohort (i.e. non-fluctuating group of patients) indicating that change in motor symptoms of PD as measured by the UPDRS was associated with changes to physical and psychological domains. The association between the UPDRS which measures tremor, rigidity, bradykinesia and postural stability and the physical domain of the PIMS which is made up of items which refer to how parkinsonism has had an effect on ones job, running of the home, ability to support the family (Work items) and ability to reach ones destination e.g. work (Travel items) is not surprising. The relationship between the psychological domain and UPDRS on the other hand is less expected as you would not anticipate an association between a clinical instrument which measures physical symptoms and HRQoL domain which considers psychological aspects of well being such as how one feels about themselves with

respect to stress, anxiety and depression (Self items) and ability to enjoy activities (Leisure items).

However if one considers that an impairment or reduction in a patients physical symptoms such as tremor and rigidity is very likely to impact adversely on the patients ability to enjoy activities and limit their functionality and ability to undertake activities of daily living this is likely to cause increased levels of anxiety and stress and possibly depression. This probably explains why a change in UPDRS score is likely to be reflected in changes to the psychological domain of the PIMS.

Negative relationship between all domains of PIMS and the UPDRS total score in the non-stable (i.e. fluctuating group of patients) was observed (Table 5.19). This finding was not anticipated because the fluctuating group of patients by definition would be experiencing fluctuation of symptoms such a bradykinesia, tremor and rigidity, and these are symptoms which are measured by the UPDRS clinical scale and are suggestive of deterioration in disease state. Hence a strong and statistically significant correlation between the related PIMS domains for fluctuating patients and UPDRS would have been anticipated.

A possible explanation for this difference is that the clinician assessment using the UPDRS records the observations of the examiner for a single point in time and does not differentiate patients based on whether they were fluctuating or non-fluctuating patients. In other words the UPDRS did not differentiate between stable and unstable cohorts and hence was not sufficiently sensitive to correspond with deterioration of PIMS domains of the fluctuating cohort. Similar findings were experienced when assessing the relationship of the PIMS against the PDQ-39. Further, it should be noted that the number of patients completing columns 2a and 2b were very low in comparison to those completing column 1 and this is likely to have influenced the results observed. A pattern was emerging of fluctuating patients not completing the relevant columns 2a and 2b which would have suggested that there was ambiguity on how to complete this instrument. The uncertainty on how to complete the two columns has led to a poor completion of the instrument and patients may be confused by having to assign a score to each question at a single

Figure 5.2: PIMS 1 Internal Consistency Reliability

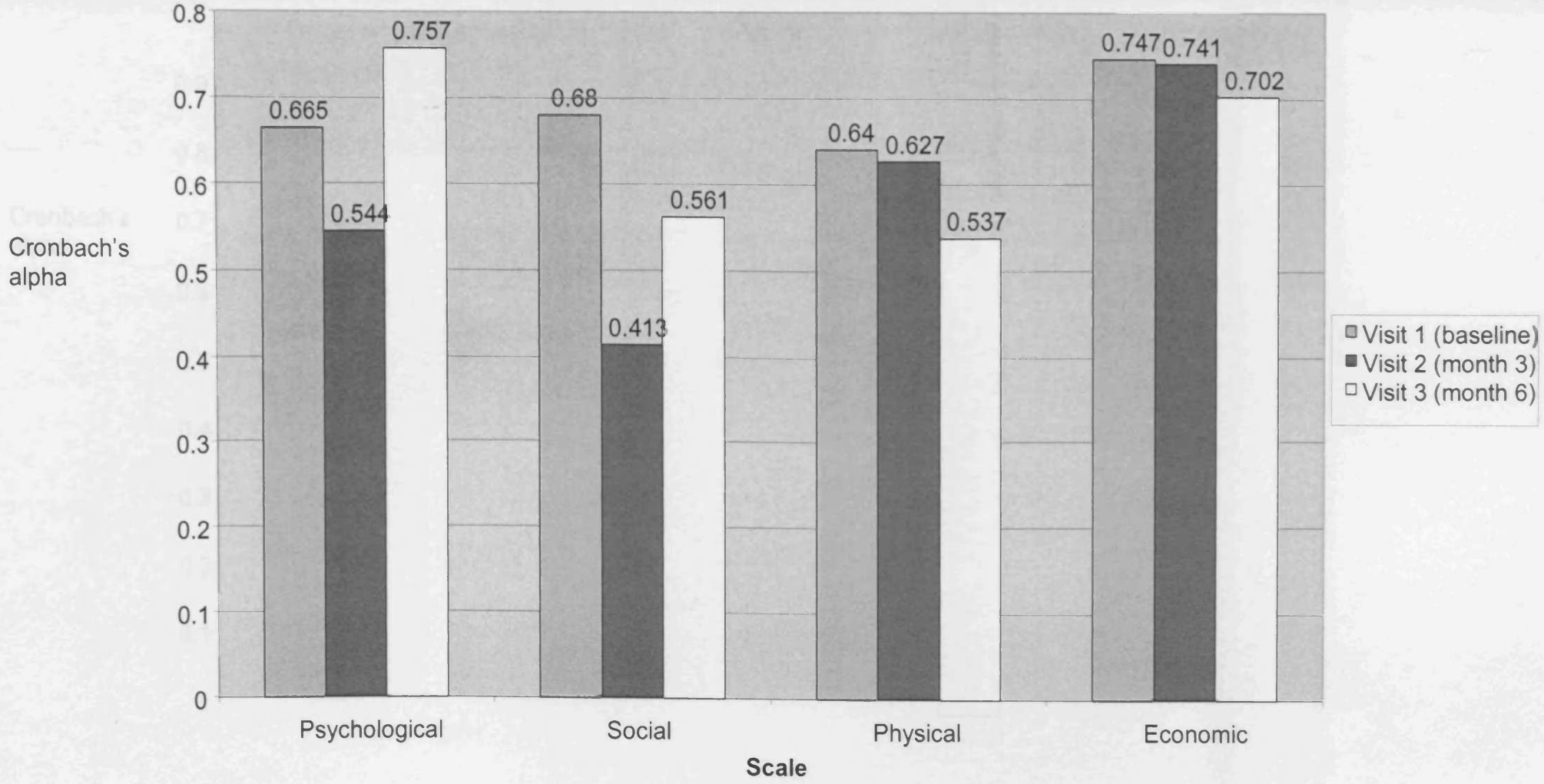


Figure 5.3: PIMS 2a Internal Consistency Reliability

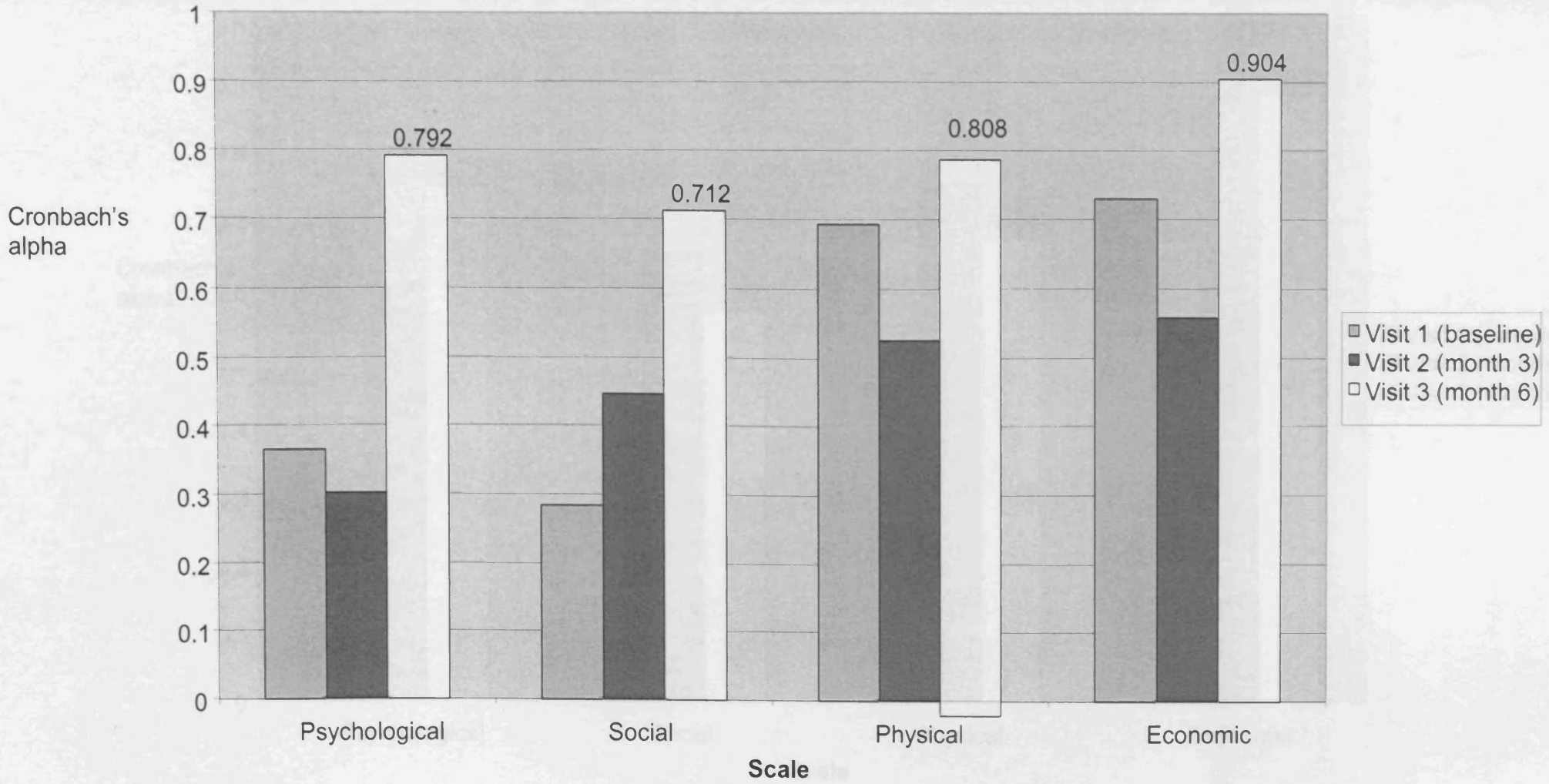


Figure 5.4: PIMS 2b Internal Consistency Reliability

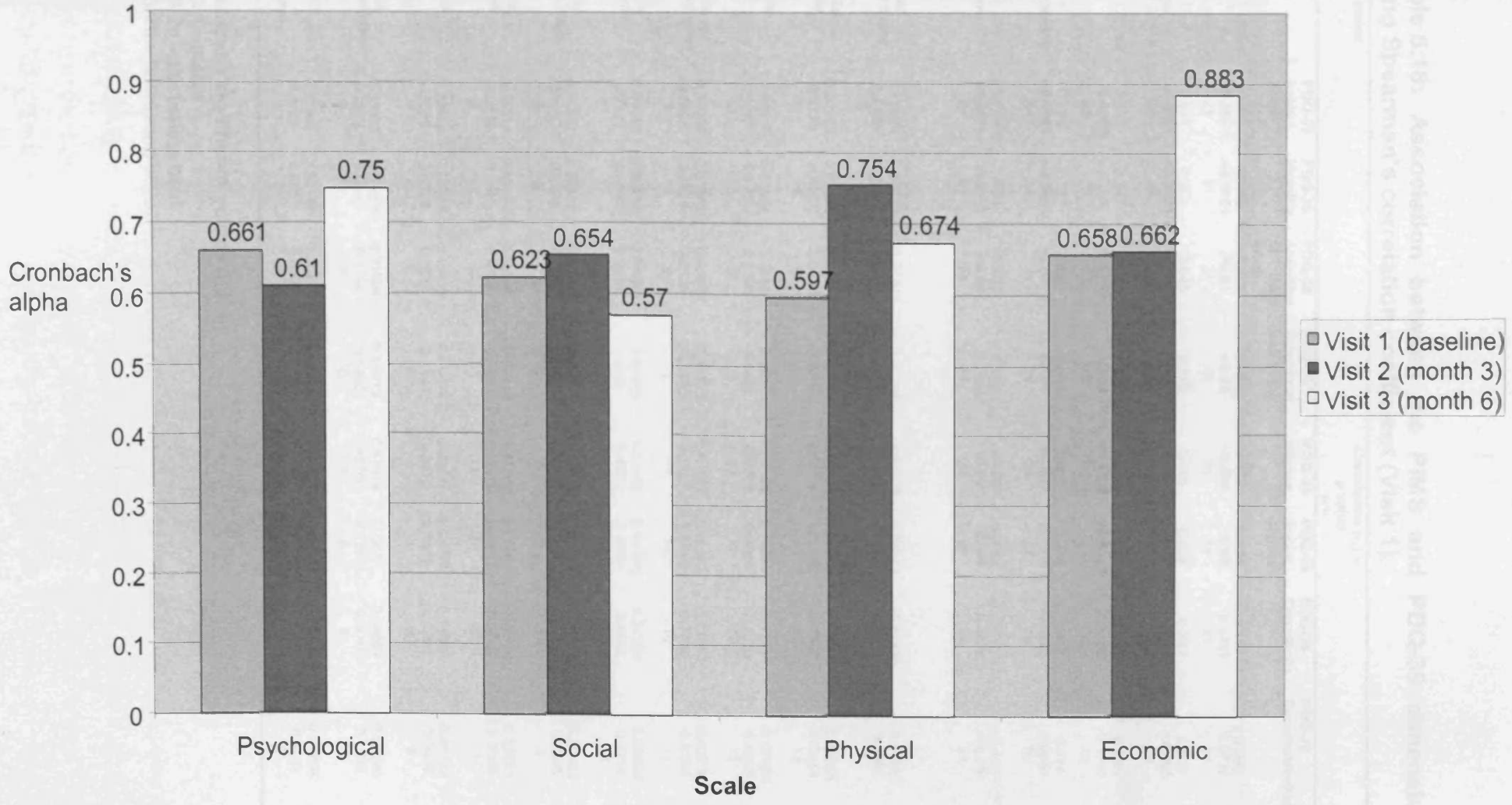


Table 5.18: Association between the PIMS and PDQ-39 dimension scores using Spearman's correlation coefficient (Visit 1).

PIMS scales	Correlation (r_s) = p value n =								
	PDQ-39 Index Score	**PDQ-39 Mobility**	**PDQ-39 Activities of daily living**	**PDQ-39 Emotional well being**	**PDQ-39 Stigma**	**PDQ-39 Social support**	**PDQ-39 Cognitions**	**PDQ-39 Communication**	**PDQ-39 Bodily discomfort**
PIMS 1 Global Score	0.866 **<0.0001** 21	0.776 **<0.0001** 21	0.579 **<0.01** 21	0.468 **<0.05** 21	0.614 **<0.005** 21	0.494 **<0.05** 21	0.177 0.4412 21	0.393 0.0772 21	0.504 **<0.05** 21
PIMS 1 Physical	0.814 **<0.0001** 23	0.802 **<0.0001** 23	0.648 **<0.0005** 23	0.279 0.1964 23	0.495 **<0.05** 23	0.336 0.1169 23	0.233 0.2827 23	0.535 **<0.005** 23	0.504 **<0.05** 23
PIMS 1 Social	0.643 **<0.0005** 23	0.547 **<0.05** 23	0.335 0.1170 23	0.553 **<0.05** 23	0.669 **<0.0005** 23	0.674 **<0.0005** 23	0.136 0.5356 23	0.285 0.1859 23	0.335 0.1172 23
PIMS 1 Psychological	0.799 **<0.0001** 24	0.787 **<0.0001** 24	0.627 **<0.0005** 25	0.451 **<0.05** 25	0.630 **<0.001** 25	0.329 0.1079 25	0.319 0.1201 25	0.374 0.0651 25	0.518 **<0.01** 25
PIMS 1 Economic	0.722 **<0.0001** 24	0.750 **<0.0001** 24	0.493 **<0.05** 24	0.263 0.2126 24	0.472 **<0.05** 24	0.420 **<0.05** 24	0.150 0.4815 24	0.412 **<0.05** 24	0.471 **<0.05** 24
PIMS 2a Global Score	0.59524 0.1195 8	0.36136 0.3393 9	-0.67475 0.0664 8	0.21281 0.5825 9	0.06723 0.8636 9	0.39347 0.2948 9	-0.67475 0.0664 8	-0.54857 0.1262 9	-0.05130 0.8957 9
PIMS 2a Physical	0.00963 0.9776 11	0.21298 0.4647 14	0.06978 0.8294 12	0.18641 0.5234 14	0.46811 0.0914 14	0.07059 0.8188 13	0.41848 0.1547 13	0.28055 0.3313 14	0.07826 0.7903 14
PIMS 2a Social	0.43836 0.1775 11	0.38674 0.1917 13	-0.36684 0.2408 12	0.18724 0.5402 13	0.15139 0.6215 13	0.18990 0.5544 12	-0.11111 0.7310 12	-0.07650 0.8038 13	0.07325 0.8120 13
PIMS 2a Psychological	0.35929 0.3821 8	0.09382 0.7838 11	-0.45992 0.2129 9	0.47928 0.1358 11	0.44267 0.1727 11	0.10101 0.7813 10	0.11963 0.7420 10	-0.02759 0.9358 11	0.17363 0.6097 11
PIMS 2a Economic	0.09013 0.7921 11	0.03283 0.9152 13	0.04060 0.9057 11	0.63063 <0.05 13	0.23682 0.4360 13	0.20766 0.4960 13	0.24324 0.4462 12	0.21929 0.4716 13	0.06495 0.8330 13
PIMS 2b Global Score	0.17857 0.7017 7	0.59881 0.1168 8	0.25459 0.5817 7	0.22893 0.5855 8	0.33737 0.4138 8	0.20294 0.6298 8	0.01802 0.9694 7	0.20483 0.6265 8	-0.55241 0.1557 8
PIMS 2b Physical	0.45314 0.2206 9	0.46488 0.1497 11	0.28936 0.4174 10	-0.00714 0.9834 11	0.37392 0.2573 11	0.27261 0.4460 10	0.22154 0.5385 10	0.33741 0.3102 11	-0.34146 0.3041 11
PIMS 2b Social	0.03604 0.9389 7	0.19831 0.6090 9	0.13415 0.7515 8	0.56356 0.1141 9	0.36017 0.3410 9	0.03190 0.9402 8	0.11585 0.7847 8	0.59149 0.0934 9	-0.65096 0.0576 9
PIMS 2b Psychological	0.38095 0.3518 8	0.19453 0.5902 10	0.31094 0.4154 9	0.19513 0.5890 10	0.67074 <0.05 10	0.03651 0.9257 9	0.38657 0.3041 9	0.51990 0.1235 10	-0.39891 0.2535 10
PIMS 2b Economic	0.33945 0.4107 8	0.20176 0.6027 9	0.20246 0.6306 8	0.62009 0.0748 9	0.40612 0.2781 9	0.40631 0.2778 9	0.09938 0.8149 8	0.27708 0.4704 9	-0.04386 0.9108 9

Statistically significant values in bold
PIMS 1 - stable
PIMS 2a – fluctuating best
PIMS 2b – fluctuating worst

time-point, based on how they might be when in a fluctuating best state (column 2a) versus a fluctuating worse state (column 2b). The number of missing answers would strongly indicate that this instrument is not ideally suited for self-completion by the patient and that it should be administered by the health professional. Also, a more explicit set of instructions on how to complete the columns 2a and 2b seems to be mandated.

The PDQ-39 which like the UPDRS in this study, does not differentiate between fluctuating and non-fluctuating patients exhibited strong correlation ($p<0005$) against the UPDRS score and also against the Hoehn and Yahr indicating construct validity in PD. This is consistent with the pattern of results obtained in a study of the PDQ-39 (Fitzpatrick et al., 1997).

Comparison of PIMS scores between each of the states (stable – column 1, fluctuating best – column 2a and fluctuating worse – column 2b).

It is evident from the analysis so far that there are differences in the PD patients recruited to this study in that some of those sampled have stable disease which for the purposes of this analysis means they do not experience a fluctuation in their symptoms. The other group are the sampled patients who experience fluctuations of symptoms and were asked to grade the PIMS based on a best and worse case scenarios i.e. how they would respond during an "on" and "off" attack. The analysis has shown that there was a difference in the scores between fluctuating and non-fluctuating patients and therefore the next logical step seemed to be to compare whether there was a difference between these assumingly distinct groups of PD patients at each of the visits. The Wilcoxon rank sum test (a nonparametric alternative to the two-sample *t*-test) was used to examine differences between the two groups. Where significant values are reported in Table 5.20 there is evidence to reject the null hypothesis that the distributions of item scores are the same in the two groups.

The key observation from this analysis is that all items with the exception of financial were significantly different when comparing stable patient scores against fluctuating patient scores, particularly for the fluctuating worst group at all visits. This confirms that the distribution of items scores differ significantly between the two group. Statistically significant differences were strongest for self positive, self negative,

family relationship, leisure, travel and work (all $p<0.001$) at baseline. At visit 2 and visit 3 the difference in item scores between the two groups was less powerful but nevertheless still statistically significant ($p<0.05$ and $p<0.01$).

The financial and sexual items were not significant at both visit 2 and visit 3 and this may be a reflection of the smaller sample size completing all follow-up visit questionnaires and also a reflection of the number of respondents who did not complete these items. As reported previously in this chapter, 26% of respondents did not complete the sexual item question and 23% of respondents did not complete the financial item question. This part of the analysis confirms that there are true and meaningful differences which were statistically significant in the scores of patients with fluctuation of symptoms compared to stable patients and this is reflected also by the higher item scores observed indicating decreased HRQoL.

Discriminative validity

The PIMS appears to have discriminative validity as evidenced by its ability to define groups of PD patients with differing degrees of disease severity (Table 5.20).

In addition, discriminant validity (i.e. the extent to which the questionnaire scores were able to discriminate between subgroups of patients differing in clinical state) was assessed by evaluating the PIMS and PDQ-39 ability to discriminate between:

- Stage I, II, III and IV/V Hoehn & Yahr staging (Table 5.21)

- Slight (score =14), mild (score 15-28) and moderate (score 29-42) motor impairment as measured by the UPDRS (Table 5.22)

- Minimal (score =10), mild (score 10-16) and moderate (score 17-29) depression as measured by the BDI (Table 5.23)

Table 5.19: Association between PIMS dimensions and Clinical Measures of PD (UPDRS & H&Y) - r_s = Spearman's correlation coefficient; n = number of observations

PIMS / PDQ-39 SCALES	UPDRS Total Score	Hoehn & Yahr
PIMS 1 Global Score	r_s = 0.334	0.225
	p = 0.189	0.353
	n = 17	19
PIMS 1 Physical	0.457	0.390
	<0.05	0.08
	19	20
PIMS 1 Social	0.146	0.237
	0.55	0.30
	19	21
PIMS 1 Psychological	0.642	0.168
	<0.005	0.452
	21	22
PIMS 1 Economic	0.191	0.407
	0.419	0.06
	20	21
PIMS 2a Global Score	-0.800	0.0
	0.104	1.0
	5	7
PIMS 2a Physical	0.119	0.540
	0.741	0.069
	10	12
PIMS 2a Social	-0.288	-0.006
	0.452	0.985
	9	10
PIMS 2a Psychological	-0.541	0.149
	0.209	0.700
	7	9
PIMS 2a Economic	-0.463	-0.132
	0.209	0.696
	9	11
PIMS 2b Global Score	-0.800	0.414
	0.2	0.414
	4	6
PIMS 2b Physical	0,00	0.458
	1.00	0.252
	7	8
PIMS 2b Social	-0.564	-0.072
	0.321	0.876
	5	7
PIMS 2b Psychological	-0.028	0.144
	0.956	0.757
	6	7
PIMS 2b Economic	-0.359	0.151
	0.552	0.746
	5	7
PDQ-39		
PDQ-39 Index Score	0.550	0.576
	<0.0005	**<0.0005**
	36	37

Statistically significant values in bold

PIMS 1 = Stable symptoms; PIMS 2a = Fluctuating best; PIMS 2b = Fluctuating worst

Table 5.20: Comparison of PIMS mean ranks between columns 1 (Stable), 2a (Fluctuate best) and 2b (Fluctuate worst) using Wilcoxon rank sum (two-sample) test:

PIMS Categories	Stable (A)	Fluctuate best (B)	Fluctuate worst (C)	Change 2-sided p-value
VISIT 1				
Self positive	17.6	24.8	28.8	AC***
Self negative	19	20.5	28.2	AC***
Family relationship	18.4	26.5	27.4	AB*, AC***
Community	19.3	25.1	25.7	AC**
Safety	16.6	27.8	30.3	AB**, AC***
Leisure	18.0	24.1	27.9	AC***
Travel	16.7	26.0	28.2	AB*, AC***
Work	15.3	25.1	26.3	AB**, AC***
Financial	18.3	20.3	21	NS
Sexual	17.7	19.9	24.1	AC**
VISIT 2				
Self positive	16.2	15.8	21.2	AC**
Self negative	15.8	15.1	19.6	AC*
Family relationship	15.0	20.2	20.4	AC*
Community	14.5	20.9	21.8	AC**
Safety	14.6	15.7	21.3	AC**
Leisure	13.8	20.0	22.4	AC**
Travel	13.7	20.2	23.1	AB*, AC***
Work	12.9	16.3	20.8	AC***
Financial	15.1	17.5	14.9	NS
Sexual	13.1	18.1	17.4	NS
VISIT 3				
Self positive	9.9	17.6	17.1	AB**, AC**
Self negative	8.9	13.5	16.0	AC**
Family relationship	9.1	20.1	16.8	AB***, AC**
Community	9.6	18.9	16.0	AB**, AC*
Safety	10.0	18.3	16.0	AB**, AC*
Leisure	9.6	17.3	17.7	AB**, AC**
Travel	10.5	16.8	16.2	AB*, AC*
Work	9.5	11.4	14.4	NS
Financial	10.3	14.7	13.0	NS
Sexual	10.1	11.5	14.3	NS

*p<0.05; **p<0.01; ***p<0.001, NS = not significant

Patients completing the PIMS were sub-divided into two groups determined by the completion of column 1 and column 2a/2b. Those patients completing column 1 were classified as stable patients. Those patients completing columns 2a and 2b were

classified as patients with fluctuations in their "on/off" state. It was anticipated that quality of life scores would worsen as symptom severity increased.

The PIMS ability to discriminate between H&Y stages

The clinician's rating of the severity of disease symptoms using the H&Y showed a consistent but not statistically significant pattern of elevated scores (indicating worse health) for physical, social and psychological scales of the PIMS for stable patients as symptom severity increased. The social scale did not discriminate beyond stage III where only one patient was classified at stage IV. Similarly, the global PIMS score showed a consistent but not statistically significant pattern of elevated scores up to stage III.

An assessment of the PIMS to discriminate between stages I to IV for fluctuating patients was not possible because of the low number of patients with fluctuating symptoms who had an H&Y evaluation. There were no patients with fluctuating (worst) symptoms with a stage I and IV or else these patients had not had a clinical staging performed.

The PDQ-39 ability to discriminate between H&Y stages

The clinician's rating of the severity of disease symptoms using the H&Y showed a consistent and statistically significant pattern of elevated scores (indicating worse health) for mobility, activities of daily living and stigma scales as well as the PDQ-39 Single Index score. The other PDQ-39 scales; emotional well being, social support, and communication showed a pattern of elevated scores but did not reach statistical significance.

The PIMS ability to discriminate between differing levels of motor impairment

The clinician's rating of the severity of motor impairment using the UPDRS showed a consistent but not statistically significant pattern of elevated scores (indicating worse health) for physical, social and psychological scales of the PIMS for stable patients as symptom severity increased. The same pattern was observed for the PIMS global score. The noted exception to this is seen with the economic subscale.

An assessment of the PIMS to discriminate between three defined levels of motor impairment for fluctuating patients was not possible because of the low number of patients who had had an UPDRS evaluation or else due to the high number of

missing items from column 2 (the column to be completed by fluctuating patients). One noticeable difference was that patients completing column 2b, indicating that they are fluctuating and were in an "off" state scored higher on all the PIMS scales (indicating worse health) compared to scores on column 1 which was completed by stable patients i.e. those that did not experience an "on/off" state and column 2a, patients who fluctuated but were in the "on" state.

The PDQ-39 ability to discriminate between differing levels of motor impairment

Clinician's rating of the severity of motor impairment using the UPDRS showed a consistent and statistically significant pattern of elevated scores (indicating worse health) for mobility, activities of daily living and stigma scales as well as the PDQ-39 Single Index score. This is a similar pattern to that shown for the Hoehn & Yahr staging. The other PDQ-39 scales; cognitive impairment and bodily discomfort showed a pattern of elevated scores but were not statistically significant. The social support, emotional well being and bodily discomfort scales did not show a pattern of elevated scores.

The PIMS ability to discriminate between differing depressive states

The respondents' rating of the severity of depression using the BDI showed a consistent but not statistically significant pattern of elevated scores (indicating worse health) for all scales of the PIMS for stable patients as symptom severity increased. The same pattern was also observed for the PIMS global score. Fluctuating patients showed a pattern of elevated scores for minimal, mild and moderate depression when compared to stable patients. Fluctuating "off" patients had higher scores for all scales with the exception of the economic sub-scale when compared to fluctuating "on" patients.

The PDQ-39 ability to discriminate between differing depressive states

Respondents rating of the severity of depression using the BDI showed a consistent and statistically significant pattern of elevated scores (indicating worse health) for mobility, activities of daily living (ADL) and stigma scales as well as the PDQ-39 Single Index score. This is the same for the Hoehn & Yahr and UPDRS revealing that the PDQ-39 is able to discriminate between subgroups of patients differing in clinical state with respect to mobility ADL, stigma and overall health.

The other PDQ-39 scales; emotional well being, social support, cognitive impairment, communication and bodily discomfort showed a pattern of elevated scores but did not reach statistical significance.

The Kruskall- Wallis one-way ANOVA was used to test whether or not the PIMS and PDQ-39 were able to differentiate between the categories of disease severity outlined above. Overall the PDQ-39 demonstrated a high level of discriminative sensitivity and was able to distinguish between patients with different severities for the domains of mobility ($p<0.001$), activities of daily living ($p<0.01$), stigma ($p<0.05$) and the global index score ($p<0.05$) indicating reduced HRQoL with increased disease severity. Two of the three PDQ-39 domains which reached statistical significance were related to physical aspects of well being (mobility and ADL) which could be expected to have a strong association with clinical disease severity.

Table 5.21: Mean scale scores for the PIMS and PDQ-39 by clinical staging (Hoehn & Yahr staging)

PIMS/PDQ-39 Dimension	Hoehn and Yahr stage – mean (s.d)				'p' value*
	Stage I (n =13)	**Stage II (n =13)**	**Stage III (n =11)**	**Stage IV/V (n =4)**	
PIMS (Stable)					
PIMS 1 Global Score	10.9 (4.5)	12.4 (8.3)	10.7	19.9	0.65
PIMS 1 Physical	1.3 (1.1)	1.9 (1.6)	3.3 (2.7)	3.9	0.28
PIMS 1 Social	2.4 (1.9)	2.8 (2.7)	6.3 (2.9)	3.4	0.35
PIMS 1 Psychological	6.8 (2.3)	7.1 (4.3)	10.3 (8.4)	11.1	0.74
PIMS 1 Economic	0.4 (0.5)	0.4 (0.5)	1.5 (0.7)	1.5	0.094
PDQ-39					
Single Index Score	19.3 (9.4)	26.2 (13.7)	41.5 (12.0)	40.3 (11.4)	**0.004**
Mobility	20.8 (15.6)	30.0 (19.9)	69.3 (23.9)	71.9 (8.3)	**<0.0001**
Activities of Daily Living (ADL)	17.0 (12.3)	33.7 (25.9)	50.8 (22.4)	41.7 (23.8)	**0.008**
Emotional Well-being	26.6 (22.6)	27.9 (25.0)	33.7 (21.9)	34.4 (13.8)	0.71
Stigma	11.1 (18.2)	20.7 (23.0)	22.7 (24.7)	50.0 (16.9)	**0.020**
Social Support	6.4 (11.9)	8.3 (11.8)	17.5 (21.0)	22.9 (20.8)	0.18
Cognitive Impairment (Cognitions)	26.0 (21.6)	33.3 (16.5)	39.2 (20.7)	32.8 (19.3)	0.35
Communication	12.2 (13.9)	22.4 (23.9)	25.8 (18.4)	33.3 (18.0)	0.12
Bodily Discomfort	34.6 (15.5)	42.3 (28.8)	47.0 (21.5)	35.4 (17.2)	0.55

*p-values obtained with the Kruskal-Wallis statistics test **(bold indicating significant results).**

Though none of the domains of the PIMS reached statistical significance for any of the measures of disease severity there was a trend towards reduced HRQoL as disease severity increased thus indicating the existence of a weak relationship.

Table 5.22: Mean scale scores for the PIMS and PDQ-39 by severity of motor impairment (UPDRS)

PIMS/PDQ-39 Dimension	UPDRS total score – mean (s.d.) Slight motor impairment (n =19)	Mild motor impairment (n =17)	Moderate motor impairment (n =3)	'p'-value*
PIMS (Stable)				
PIMS 1 Global Score	12.2 (6.3)	13.6 (9.9)	23.0	0.65
PIMS 1 Physical	1.7 (1.4)	1.9 (1.8)	4.3 (1.3)	0.28
PIMS 1 Social	2.6 (1.9)	4.0 (3.4)	4.5	0.35
PIMS 1 Psychological	6.5 (3.6)	9.9 (4.4)	17.0 (5.4)	0.74
PIMS 1 Economic	0.6 0.6	0.2 (0.3)	2.0 (0.0)	0.094
PIMS (fluctuating – best)				
PIMS 2a Global Score	30.0	21.9 (9.8)	21.7	1.00
PIMS 2a Physical	3.0 (1.5)	3.5 (1.2)	3.3	0.31
PIMS 2a Social	6.2 (3.0)	4.8 (3.3)	4.8	0.52
PIMS 2a Psychological	15.6	9.7 (4.9)	13.1	0.90
PIMS 2a Economic	1.0 (0.5)	0.7 (0.8)	0.5	0.38
PIMS (fluctuating – worse)				
PIMS 2b Global Score	40.0	33.0 (1.0)	31.1	0.35
PIMS 2b Physical	5.2	4.7 (0.5)	4.6	0.22
PIMS 2b Social	12.0	7.0 (3.0)	7.2	0.86
PIMS 2b Psychological	20.8	15.3 (4.0)	18.3	0.72
PIMS 2b Economic	2.0	1.3 (0.8)	1.0	0.71
PDQ-39				
Single Index Score	24.3 (12.7)	36.6 (15.5)	46.1 (15.3)	**0.004**
Mobility	29.2 (21.4)	54.6 (32.5)	73.3 (6.3)	**<0.001**
Activities of Daily Living (ADL)	22.4 (20.1)	46.4 (24.7)	75.0 (14.4)	**0.008**
Emotional Well-being	26.8 (22.0)	33.6 (19.7)	33.3 (21.7)	0.71
Stigma	15.8 (22.1)	26.5 (27.4)	45.8 (29.5)	**0.020**
Social Support	7.9 (12.6)	14.6 (19.1)	8.3 (8.3)	0.18
Cognitive Impairment (Cognitions)	32.2 (21.3)	36.8 (18.5)	50.0 (34.8)	0.35
Communication	18.9 (15.4)	25.0 (21.2)	16.7 (16.7)	0.12
Bodily Discomfort	35.5 (21.5)	43.1 (20.9)	66.7 (30.0)	0.55

*p-values obtained with the Kruskal-Wallis statistics test (**bold** indicating significant results).

Table 5.23: Mean scale scores for the PIMS and PDQ-39 by severity of depression (BDI)

PIMS/PDQ-39 Dimension	BDI total score – mean (s.d.) Minimal depression (n = 21)		Mild depression (n = 12)		Moderate depression (n = 14)		'p'-value*
PIMS (Stable)							
PIMS 1 Global Score	10.6	(5.8)	14.6	(7.7)	18.4	(2.7)	0.65
PIMS 1 Physical	1.5	(1.2)	2.1	(2.0)	3.3	(0.7)	0.28
PIMS 1 Social	2.0	(1.7)	4.2	(2.7)	5.4	(2.3)	0.35
PIMS 1 Psychological	6.1	(3.5)	9.2	(5.9)	10.4	(2.6)	0.74
PIMS 1 Economic	0.4	(0.6)	0.7	(0.9)	1.2	(0.3)	0.094
PIMS (fluctuating – best)							
PIMS 2a Global Score	24.1		21.1	(13.6)	22.0	(5.2)	1.00
PIMS 2a Physical	2.2	(1.5)	3.7	(1.0)	3.6	(0.6)	0.31
PIMS 2a Social	5.1	(3.0)	4.7	(3.8)	5.8	(1.6)	0.52
PIMS 2a Psychological	8.5	(6.4)	10.2	(4.7)	12.3	(3.4)	0.90
PIMS 2a Economic dimension	0.2	(0.3)	1.0	(0.7)	0.8	(0.6)	0.38
PIMS (fluctuating – worse)							
PIMS 2b Global Score	19.5		33.0	(1.0)	31.5	(5.6)	0.35
PIMS 2b Physical	2.5		4.6	(0.5)	4.6	(0.6)	0.22
PIMS 2b Social	7.2		7.0	(3.0)	8.2	(3.1)	0.86
PIMS 2b Psychological	9.8		15.3	(4.0)	17.7	(2.0)	0.72
PIMS 2b Economic	0.0		1.0	(0.7)	1.3	(0.9)	0.71
PDQ-39							
Single Index Score				(12.7)	45.7	(10.8)	**0.004**
Mobility	20.0	(11.0)	32.9	(30.0)	58.4	(22.0)	**<0.001**
Activities of Daily Living (ADL)	27.5	(25.2)	51.1	(28.9)	48.4	(25.3)	**0.008**
Emotional Well-being	24.0	(20.6)	45.8	(21.6)	51.5	(12.6)	0.71
Stigma	17.3	(16.3)	31.6	(19.7)	49.6	(19.8)	**0.020**
Social Support	7.4	(12.9)	19.8	(7.9)	23.8	(16.3)	0.18
Cognitive Impairment	4.8	(12.0)	7.6		47.1		
(Cognitions)	26.5	(20.6)	35.4	(17.3)	35.7	(21.0)	0.35
Communication	15.9	(17.9)	17.4	(22.0)	14	(22.5)	0.12
Bodily Discomfort	33.7	(20.8)	39.6	(20.8)	57.7	(20.8)	0.55

*p-values obtained with the Kruskal-Wallis statistics test (**bold** indicating significant results).

Sample Size

The sample size required to detect a 20% change in QoL (using the overall score of the summary or global scores at baseline) was assessed for a two-sided test at the 5% significance level and 90% power to detect the stated change (Table 5.24). The sample size required was based on the observed values for PDQ-39 single index score, PIMS global scores for PIMS 1, 2a and 2b at Visits 1 and 3: The number of patients required would be n = 146 (the sum of the sample size required from each of the instruments).

Table 5.24 Sample size estimation based on the observed values for PDQ-39 single index score, PIMS global scores for PIMS 1, 2a and 2b at Visits 1 and 3:

PDQ-39/PIMS Score	Observed baseline mean	Observed standard deviation of changes	Sample size required to detect 20% mean change
PDQ-39 single index score	30.8	11.0	34
PIMS 1 global score	12.8	6.1	60
PIMS 2a global score	22.0	3.8	8
PIMS 2b global score	30.4	12.4	44

As it is feasible to obtain two global scores for fluctuating patients completing the PIMS, Table 5.24 provides a sample size for each of the instruments global scores. However, it should be noted that due to the high number of missing questionnaires, particularly from patients completing the PIMS columns 2a and 2b; sample size estimates were performed on a limited number of observed values and hence cannot be considered accurate. On the other hand, the PDQ-39, where more patients completed this instrument (i.e. there were considerably less missing questionnaires) one can with some certainty perform a sample size estimation and determine that the number of patients required would be 34.

DISCUSSION

The PIMS was not as well comprehended by PD patients compared to the PDQ-39 as shown by the relatively high rate of missing data and some of the comments of the patients whose most common report was that they did not understand whether they

ought to complete columns 1, 2a and 2b at the same time or in other cases only one of columns 2a or 2b. This highlights the requirement for clearer instructions which can be understood by such a group of patients and the importance of the interactions between the patient and the questionnaire administrator explaining how such an instrument ought to be completed. A sensible approach would be that as a first step the clinician or administrator of the questionnaire ought to inform the patient under which category they fall at each visit i.e. stable or fluctuating patients and then advise them of which column to complete i.e. column 2a or 2b. This should be preceded with an explanation in layman terms of the difference between stable and fluctuating conditions and further the difference between fluctuating best and fluctuating worse conditions. Others mentioned that they either did not find the "financial security" item relevant to them as they were retired or else they did not wish to divulge their financial situation. Admittedly this item is optional and 22% of respondents did not complete this question. In the UK, most PD patients receive invalidity or state benefit and also free medical health care under the NHS and hence their financial security is not as affected as say, in North America, where free NHS like provision is not as accessible as it is in the UK. The two questions most frequently missed were that of sexuality and work (both 25.8% at baseline visit) and all 10 items had more than 16% of respondents not answering the questions which are quite high considering the brevity of the PIMS and requirement for all items to be answered in this particular case. The fluctuating patients missed more items on column 2b (fluctuating worst) compared to column 2a (fluctuating best). A significant number of fluctuating patients (40%) did not complete column 2b (fluctuating worst) for the items of community, family relationship and financial security; a third of patients did not complete the safety item; and a quarter of patients sampled did not complete the items of sexuality and travel. This may be highlighting that patients who were in a fluctuating state were not aware that they were to complete both columns 2a and 2b and not just the column which reflected their state at the time of completion of the questionnaire whether it was as a fluctuating patient in a stable state or else a fluctuating patient in a worse state. The analysis reveals the need for a more comprehensive explanation on how to complete the PIMS or in the case of the financial security item that it is redundant in the UK population.

The potential impact of missing responses is often ignored in health related quality of life studies (Fairclough et al., 1998b). Missing data from study participants can cause

bias in parameter estimates of models predicting HRQoL outcomes (Fairclough et al., 1998a). Regression models are frequently interpreted with the assumption that the available data are representative of the entire study population which is of course not always the case. One approach to minimise the impact of potential bias arising from missing data is to compare the clinical characteristics of respondents with and without missing surveys and then to attempt to assess the impact of these differences on the regression model parameter estimates and ultimately on the results of the study (Sales et al., 2004). This method was not utilised in this assessment as the problem was not missing data from patients who had withdrawn from the trial or had not completed a follow-up question but rather in this case missing questions arose from subjects who had not withdrawn from the study but had nevertheless not completed the questionnaire in it's entirety thereby resulting in part completed questionnaires. Nevertheless, assumptions cannot be based unless one can demonstrate that the data missing from a study are truly missing at random, making them ignorable, which is rarely the case. In fact it is evident that missing questions from the PIMS are not random and that these missing items arise from a lack of understanding on how to complete the instrument accurately.

Newer statistical techniques have been developed to assess and correct for the bias resulting from missing HRQoL data. One technique is the Heckman two step method (Heckman & MaCurdy, 1981). The Heckman method was developed by an economist, James Heckman, to address problems of self-selection among women participating in the labour force. This method makes it possible to assess whether selection bias is present, identify factors contributing to the selection bias, and to control for the bias in estimating the outcomes of interest. The Heckman method attempts to control for the effect of non-random selection by incorporating both the observed and unobserved factors that affect non-response (Sales et al., 2004) This can be used when whole questionnaires are missing due to patient withdrawal or when questionnaires are only partly completed thus making it relevant to this analysis. However, due to the specialist models and algorithms required this was not utilised in this analysis.

In the absence of such specialist models or know-how there is an alternative method for dealing with missing data which should be considered for future analysis i.e., the imputation method. Multiple imputation methods can be employed to address

missing HRQoL data (Fairclough, 2002). This method derives missing values from existing data in the dataset, thereby creating a 'complete' dataset and eliminating the need to drop patients from analysis. Imputation thereby eliminates bias from missing data per se (i.e., there is no longer missing data), but is highly dependent on the validity of deriving the missing HRQoL data from the existing dataset. Multiple imputation replaces each missing value with a set of plausible values that represent the uncertainty about the right value to impute (Shih, 2002). For longitudinal data the conventional last observation-carried forward (LOCF) approach is a common practice of another simple imputation which assumes that the missing data after the patients withdrawal are the same as the last value observed for that patient and requires the strong assumption stability and hence the more it imputes the more bias it introduces if the assumption of stability does not hold.

In summary there are a number of possible strategies to implement to minimise the bias introduced as a result of missing data and each has its merits and requires further investigation but it is probably an area which requires a multidisciplinary approach due to the requirement for the formulation of the appropriate regression model, statistical technique to be implemented and computer modelling required to analyse such missing data. By far the most common approach is to simply omit those cases with missing data and to run the analyses on what remains. Thus if some subjects have missing scores on one or more variables, one would simply omit those individuals from the analysis. This approach is usually called listwise deletion, but has the disadvantage that it reduces the analytical sample size, lowering the power of any test carried out. The issue of what to do about missing data caused by dropouts (and partially complete questionnaires) in clinical trials is a research topic that is still under development in statistical literature (Shih, 2002). As has been noted in the ICH-E9 guideline (ICH, 1998), no universally applicable method of handling missing values can be recommended. The issue of handling missing data is intrinsically difficult because it requires a large proportion of missing data to investigate a method.

One of the key differences between the PIMS and the PDQ-39 is that the former instrument allows discrimination between stable and fluctuating patients. In PD, adequate treatment with levodopa is able to reduce most parkinsonian symptoms; however, long term levodopa use is associated with a gradual decline in efficacy and

the development of complications such as motor fluctuations where patients fluctuate between "on" responses, in which they experience a good antiparkinsonian effect, and "off" responses, in which levodopa does not adequately treat the parkinsonian features. When "on", the patient is mobile and symptoms such as akinesia, rigidity or rest tremors are mild or absent. The "off" state results in a return of parkinsonism. Hence it would be reasonable to anticipate a difference in HRQoL between stable patients and fluctuating patients particularly amongst those in the "off" state. In fact Schrag et al 2000 investigated the contributing factors to quality of life by using the PDQ-39 and the EQ-5D instrument in a population-based study. They found that the "on"-"off" fluctuations had a significant impact on HRQoL when evaluated by the EQ-5D but not the PDQ-39. Indeed in the study reported in this chapter it was found that fluctuating patients completing the PIMS scored higher scores indicating lower HRQoL compared to stable patients. Also, fluctuating worst patients ("off") scored higher scores compared to fluctuating best ("on") patients. This indicates that fluctuating patients experiencing both "on" and "off" periods judge their symptoms to have a negative impact on their HRQoL. The PDQ-39 is completed by patients irrespective of their stable or fluctuating status and therefore does not allow one to differentiate scores between these two populations and thus leads to an amalgamation of scores which may not truly reflect the real differences amongst stable and fluctuating cohorts within a study sample. This is an issue that ought to be addressed in the PDQ-39 probably with the addition of an additional question which asks the fluctuating patient to score the items in the PDQ-39 according to their "on" and "off" state. The PIMS by allowing the respondent to score according to their true clinical state reveals some interesting findings which would not otherwise be evident by completion of the PDQ-39. For example, the results of the analysis of the factor scores across the three visits reveals that the physical domain is the most impacted with the level of impairment greater in the fluctuating groups and in particular when in the "off" state. The group with fluctuations in the "off" state reported greater levels of impairment in all domains compared to the other groups confirming that these patients experience more disability and hence such information could impact on the choice of initial therapy or treatment strategy e.g. delay the administration of levodopa and replace with a dopamine agonist. The PIMS reveals useful information in this respect such that it highlights greater level impairment in a sub population of the subject's samples in this analysis and therefore allows the medical practitioner to modify the treatment approach to alleviate the possible side effects of the existing

treatment. Motor fluctuations have a major impact on patients with PD in regard to mobility and perception of quality of life. A treatment that could prevent the occurrence of motor complications would have a major impact on each of these factors (Dodel et al., 2001).

It is generally accepted that items on a scale are chosen for their homogeneity (how well items correlate with one another and the overall score). Using a Cronbach's a of 0.7 as the benchmark, the results showed that the PIMS exhibits varying degrees of internal consistency and mostly below a reliability coefficient of 0.7 for most of the domains, though for the stable cohort the domains of psychological, social and physical were very close to 0.7 (0.66, 0.68, and 0.64 respectively). Internal consistency of the PIMS was not stable across the three visits whereas the PDQ-39 reported that all domains had good internal consistency above the reliability coefficient of 0.7. One really would have anticipated higher correlation coefficients to be reported as the authors of the PIMS (Calne et al., 1996) reported the internal consistency reliability of the 10 items (which make up the domain scores) to be 89.8%. Considering the limited number of items that make up the PIMS it is surprising to see such a high alpha coefficient reported. Alpha coefficient depends on the size of the questionnaire or the size of the subscales, as well as on the average correlation among the items within it. Alpha grows larger as the number of items and/or the average inter-item correlation increases (Schulzer M et al., 1999). These properties of alpha coefficient should be borne in mind when the internal consistency reliability of a questionnaire is estimated. A longer questionnaire will automatically record a higher value of alpha reliability than will a shorter one with the same average degree of association among its items. It is feasible to argue that whilst the average inter-item correlation of the PIMS may be similar to the PDQ-39, by the very nature of the size of the questionnaire its alpha will always deemed to be lower in comparison to longer questionnaires. On the other hand, from the point of view of the burden placed on the patient who is required to complete the questionnaire, practical and ethical considerations advocate economy in the size of the instrument, and distinct advantage if the PIMS. We can hence conclude that the PDQ-39 items within the scales are measuring similar areas of functioning in PD patients, which adds validity to the scoring profiles produced by the instrument. Conversely, the PIMS seems to indicate the opposite with lower alpha coefficients which are not stable across each of the visits but this may in part be a reflection of

the size of the instrument as previously discussed. Further reliability testing of the PIMS such as test-retest is recommended to conclude the reliability of the instrument and probably in a larger sample size as there was considerable variability in results between the fluctuating best and fluctuating worst groups which was probably a consequence of the low sample size in these two sub-groups.

The results of the analysis of convergent construct validity show that there was widespread strong correlations between the domains of the PDQ-39 and PIMS for patients whose symptoms were stable and particularly strong correlations were identified for the PIMS physical domain and PDQ-39 mobility and activities of daily living domains; the PIMS social domain and PDQ-39 stigma and social support domains and perhaps unexpectedly, the PIMS psychological domain and the PDQ-39 mobility and ADL domains. Though the correlation between the PIMS psychological domain and the PDQ-39 mobility and ADL domains was not expected as they are purporting to be measuring different aspects of well being it is feasible to infer that reduced psychological functioning is a consequence of physical deterioration. As anticipated there were stronger associations between domains which measured similar areas of HRQoL than between domains which assessed different aspects. This indicates that there are some important differences in the domains of HRQoL.

The key finding was that none of the PIMS domains for fluctuating patients correlated with the PDQ-39 domains even for domains where an association would be expected such as PIMS physical and PDQ-39 mobility. This is a particularly relevant finding that shows that patients who are in a fluctuating state report different perception of well-being when measured by the PIMS but that these perceptions are not adequately detected by the PDQ-39. The PIMS asks patients to score each item based on their best and worst state if they are classified as a fluctuating patient and hence the patient is making a distinction between the two possible states whereas such a mechanism is not available in the PDQ-39. Patients' perception of quality of life is mediated by the process of coping which can be linked to various factors such as the perceived threat of each symptom or problem to the patient and the success or otherwise of coping strategies employed to maintain control. The lack of association between the PIMS columns completed by fluctuating patients and the whole group who complete the PDQ-39 irrespective of their classification can be

further explained in that the fluctuating patients may be grading their scores based on their impression or expectation of how their well being is impeded during the fluctuating state rather than how they really are during such a period and this may result in an inflation of the score reported. Also, it may be that the lack of association is simply a reflection of the difference between a distinct group (the fluctuating patients) and the group which comprises of both fluctuating and stable patients, and as we have seen the patients who fluctuate report worse HRQoL scores indicating poorer health. The analysis supports the suggestion that the PDQ-39 does not have the precision to detect change in the fluctuating group and should in future discriminate between the two possible groups of PD patients as otherwise this is the risk that important differences in such groups may be missed. The analysis indicates that the PIMS exhibits convergent construct validity for patients classified as stable as these patients are reporting their well being at the time of completion of the instrument as is the case for the PDQ-39 and not basing it on their impression of how they would be. The same cannot be conclusively said of the PIMS for patients who experience fluctuation of symptoms as none of the measures correlated with the conceptually related subscales of the established PDQ-39 and the reasons for this may indicate that the PIMS is detecting parameters of well being that are important to the fluctuating sub-group but these are not detected by the PDQ-39 hence the lack of association. Further investigation of these differences is required including an assessment of whether there was a difference between fluctuating patients and stable patients who completed the PDQ-39.

The PIMS physical and psychological domain scores were moderately correlated with the UPDRS score substantiating its validity in this sample of PD patients who were classified as stable. The ten items or factors that make up the PIMS cover four core areas which are particularly important in PD: mobility, self care, usual activities and anxiety and depression. They represent the aspects of life most affected by PD and its typical features. Other features of PD which are not covered by the PIMS include stigma, social support, communication and cognition which are addressed in the PDQ-39. However, the areas most relevant for HRQoL in PD, as found by Fitzpatrick et al (Fitzpatrick et al., 1997) are also those covered in the PIMS: difficulty with leisure activities (covered in "leisure"), difficulty looking after the home (covered in "Safety"), difficulties getting around in public (covered in "Travel"), worrying about the future (covered in "Self-negative"). This result suggests that the PIMS for patients

classified as stable (i.e. those who completed column 1 of the questionnaire) is able to measure the impact of PD on the HRQoL of patients in a comparable manner to the disease-specific questionnaire. The fluctuating patients on the other hand, exhibited negative relationships between all domains and the UPDRS. As these patients experience a greater degree of disability by the very nature of their fluctuating condition which is indicative of the medication failing to adequately control symptoms one should have expected strong correlations against the clinical measure of PD, the UPDRS. This was not the case. Similarly to the poor correlations reported between the PIMS and PDQ-39, this may again highlight how the patient's perception of quality life in the fluctuating step is not reflected in the UPDRS as the patient is focusing on the process of coping whereas the UPDRS is measuring physical signs of disease deterioration. Though these findings indicate an absence of convergent validity for the fluctuating group it does suggest that the PIMS is probably a useful instrument to utilise the in fluctuating population of PD as it is sufficiently sensitive to detect HRQoL changes which are not linked directly to a clinical measure such as the UPDRS. Exact agreement with other measures such as severity of disease, is not required since that would mean that quality of life scores were redundant (Fitzpatrick R et al., 1992).

The PDQ-39 exhibited strong correlations against the UPDRS and unlike the PIMS against the Hoehn and Yahr scale indicating divergent construct validity in PD. The low correlations between the PIMS and the UPDRS motor and Hoehn and Yahr scores and the higher correlations between PIMS scores and PDQ-39 scores provide empirical evidence suggesting that the domain content of the PIMS is different conceptually from that of the UPDRS and Hoehn and Yahr, two of the most commonly used outcome measures in PD.

The Kruskall-Wallis one-way ANOVA was used to assess whether or not the HRQoL instruments could discriminate between different levels of disease severity. This hypothesis is underscored by the assumption that there is direct relationship between clinical disease severity and HRQoL. The ability of the PIMS and PDQ-39 to discriminate between different states i.e. stable versus fluctuating and different degree of disease severity using the UPDRS and Hoehn and Yahr were examined. In the first assessment of discriminant validity patients were grouped as stable or else fluctuating patients. The fluctuating patients were further separated as being "on"

(best state) or "off" (worst state). It was assumed that patients classified as fluctuating i.e. experiencing motor fluctuations would have more severe disease as this was an indication of medication failing to control symptoms of PD and therefore by extension would be expected to have a negative impact on HRQoL. Overall, the PIMS was able to discriminate between stable and fluctuating patients, showing a reduced HRQoL in those patients who were fluctuating. The difference in scores between stable and fluctuating worse cohorts were greatest for all domains and showed to be highly discriminative ($p<0.001$) in comparison to differences between the stable and fluctuating best groups indicating good discriminative properties.
The PIMS for stable patients also exhibited discriminative validity as evidenced by its ability to define groups of PD patients with differing degrees of disease severity as measured by the Hoehn and Yahr and UPDRS. Though levels of significance were not observed with the Hoehn and Yahr a consistent pattern of elevated scores indicating worse health for physical, social and psychological scales were seen as severity increased. The same observation was noted when the PIMS was assessed on its ability to discriminate between different levels of motor impairment as measured by the UPDRS. Levels of significance were not observed with the UPDRS though a consistent pattern of elevated scores indicating worse health for physical, social and psychological scales were seen as severity increased. Though the PIMS demonstrated that it was able to discriminate between patients with stable and fluctuating disease it could not be categorically stated that it exhibited discriminative properties against the clinical measures of disease severity though there was a trend observed. An adequate analysis of the discriminative properties for the patients completing the columns 2a and 2b (fluctuating patients) was not possible due to the very small sample size.

The PDQ-39 was clearly able to discriminate between the various stages of the Hoehn and Yahr and also the UPDRS for mobility, ADL, stigma and the single index score. The results strongly support discriminant properties of the PDQ-39 whereas the PIMS appears to be less able to discriminate between the different levels of disease severity. Disease-specific HRQoL instruments are expected to have stronger discriminative properties than generic measures since the questionnaire is supposed to directly address PD related problems. With respect to the PIMS we notice, perhaps a loss of sensitivity which may indicate that some of the items are unrelated to PD e.g. financial security. These results do not mean that the PIMS

cannot be used as a discriminative measure in PD patients as when assessed in its entirety the instruments exhibits other forms of construct validity as discussed previously.

One of the limitations of the study which affected some of the analysis was the sample size. Forty seven subjects completed the questionnaire at visit 1, 35 subjects at visit 2 and 27 subjects at visit 3. The reason for the withdrawal of patients was due, primarily, to disease deterioration and the desire of the patient or their carer to not continue with the study. Other patients were simply lost to follow-up and not followed up due to resource limitations at the clinic. In contrast to more traditional end-points in clinical trials, it is often not estimated whether a trial has a sufficient sample size with regards to quality of life outcomes. As in any other sample size estimation, one has to specify the minimal clinically relevant difference in order to estimate sample size but the magnitude of this difference is often unknown. There is not yet sufficient evidence to define exactly what change or differences one should look for in clinical trials (Groenvold, 1999). Hence to assess the number of patients required to detect an arbitrary change of 20% in HRQoL score an estimation of the sample size was conducted. To detect such a change at the 90% power a total of 112 patients completing the PIMS would have been required which is considerable target to achieve and would not have been realistic or practical in this study.

Distilling the measure of HRQoL into a few key questions is a goal for clinical investigators. The desirability of a short, practical HRQoL scale in PD has been recognised by others (Peto et al., 1998), including the authors of the lengthier PDQ-39 and as such any short instruments that can exhibit the required psychometric properties of a HRQoL would be welcomed by clinicians and patients alike. Before accepting such an instrument into a research environment it is necessary to assess such its reliability and validity and this is what has been attempted here. The authors of the PIMS did not make any comparisons against other HRQoL instruments or assess construct validity using a disease-specific instrument or an established clinical measure of disease severity. This analysis reports such findings for the first time and reveals that the PDQ-39 exhibits good construct validity and reliability in PD whereas the PIMS exhibits validity and reliability but to a lesser extent. Nevertheless, the PIMS highlights the difference in perception of well-being between stable and fluctuating patients which is a useful contribution as it requires that in future

assessments the patient who experiences motor fluctuations is likely to report a very different range of HRQoL scores compared to stable patients. Further work on the structure and lay out of the PIMS is recommended as patients had difficulty completing the questionnaire as they in particular did not understand the differences between columns 1, 2a and 2b and this was reflected by a high number of missing items. It would be preferable to separate each of the columns into a separate page with detailed instructions to the patient explaining which columns they are to complete.

Summary

- The PQD-39 was found to be the disease-specific instrument of choice in PD studies. Taking into account practical and psychometric properties, it performed better that the PIMS in PD patients.

- The PDQ-39 was able to discriminate between different levels of disease severity for most domains whereas the PIMS showed a trend towards differentiating between different levels of disease severity.

- The PIMS was not well comprehended by the PD patients as is reflected by the high rate of missing data in what is a relatively short questionnaire. Some items such a "financial security" is deemed to be irrelevant in the UK population.

- The PIMS allows scoring of patients who are either stable or those that experience motor fluctuations. This is an important aspect of the PIMS as it highlights clearly that these two groups of PD patients experience differing levels of impairment and probably report different perceptions of well-being.

- Convergent construct validity was evident for the PDQ-39 and PIMS column 1 which is completed by patients who do not experience motor fluctuations. Validity was not found for the group who completed columns 2a and 2b (fluctuating patients).

- The PDQ-39 is recommended as an outcome measure in future longitudinal studies, although an assessment of responsiveness is required.

- It is concluded that the most appropriate method of assessing the HRQoL of PD patients appears to be to use the PDQ-39 although there were useful parameters identified by the PIMS in particular its ability to separate scores of stable and fluctuating patients and the shortness of the questionnaire which is helpful in both routine use in out-patient clinic and the clinical trial setting.

Chapter 6

The burden of dyskinesia and motor fluctuations in patients with Parkinson's disease: a longitudinal study

INTRODUCTION

The analysis of the PDQ-39 has so far confirmed that the PDQ-39 is reliable and valid instrument for the assessment of the disabilities on patients with PD. It has shown excellent psychometric properties including construct and discriminative validity. Comparison against the PIMS has shown that it was able to discriminate between different levels of disease severity and that it is the most appropriate method of assessing the HRQoL of PD patients. One of the key findings of the previous analysis was that the PIMS allows the differential assessment of HRQoL in "on" and "off" periods which differentiates this instrument from the PDQ-39. The assessment of the patients HRQoL in a fluctuating state, as was seen in the previous chapter, differs considerably from a PD patient in a stable state and these patients reported greater levels of disability in most of the domains reported. It is therefore useful to assess whether the PDQ-39, is able to assess and differentiate the impact of motor fluctuations in PD and the associated dyskinesia (abnormality in performing voluntary muscle movements) which is a reflection of the deterioration of the patients clinical status. The application of the PDQ-39 to clinical trials will contribute greatly to the continued validation and interpretation of the instrument.

Current PD treatment and complications

In the early stage of PD the response to levodopa is excellent, and the benefits persist even if an individual dose is missed. However, with disease progression, the benefit of levodopa wears off before the next dose and patients fluctuate between 'on' responses, in which they experience a good antiparkinson effect, and 'off' responses, in which levodopa does not adequately treat the parkinsonian features (Dodel et al., 2001). When 'on', the patient is mobile and symptoms such as akinesia (motionlessness), rigidity or rest tremors are mild or absent. The 'off' state results in the return of parkinsonism, often accompanied by other un-pleasant non-motor symptoms. Furthermore, in the advanced stage, 'on' periods can be associated with involuntary movements or dyskinesias. Thus, patients with PD may cycle between 'on' periods in which the antiparkinson effect is complicated by dyskinesias, and 'off' periods in which the parkinsonism is not controlled and patients can be akinetic and 'frozen'. Motor fluctuations may occur as sudden, extreme and occasionally unpredictable changes.

Fluctuations of parkinsonian symptoms can be either disease or therapy-related. They may take the form of short-duration (seconds to minutes), medium duration and diurnal (minutes to hours) and long duration (days). All types of response variations increase in frequency and severity with progression of the disease and duration of therapy. Dyskinesias are unwanted, excessive and abnormal involuntary movements that may occur at both the time of peak drug effect and at the end of a dose administration interval (Vidailhet et al., 1999). Dyskinesias may present as chorea, ballism, dystonia, myoclonus or tremor, or a combination thereof. Often, patients with levodopa-induced dyskinesias (LIDs) have motor response fluctuations in addition to their dyskinetic phases.

PD treatment complications and HRQoL

The diversity of motor and non-motor symptoms associated with PD and its management may interfere with several areas of HRQoL; for example, the motor symptoms can lead to impaired physical function, which may affect self-care activities, ambulation and mobility (Montastruc et al., 1999, Weiner, 1999). As a result such patients may depend more on others such as the carer or else utilise more recourses of the health-care system which is likely to lead to diminished sense of autonomy and self-image. These are facets of HRQoL which are likely to be affected by the onset of motor fluctuations and dyskinesia. An assessment of the impact of motor fluctuations and dyskinesia on HRQoL is therefore warranted as is the impact of such motor fluctuations on the carer. Care giving is not a stable phenomenon but rather it progresses as the complications of PD mount (Dodel et al., 2001). In the early stages of the disease there is a balance of the patient's need and the carer's autonomy, which steadily deteriorates to a state where the patient's demands dominate.

The results of a long-term study of ropinirole (a dopamine agonist) versus L-dopa showed that similar proportions of L-dopa treated patients and ropinirole treated patients were satisfactorily maintained for up to 5 years and that the risk of developing dyskinesias during this period was significantly reduced in the ropinirole treated patients (SB Clinical Report for Study 056, Manuscript in review with New England Medical Journal). However the impact of the reduction in dyskinesias specifically remains unclear. This aims of this study were therefore to determine the

impact of dyskinesia on quality of life and the impact it has on care-givers of PD patients.

Study Hypothesis

The hypothesis is that PD patients who suffer dyskinesia suffer both decreased levels of quality of life and results in an increased burden to their care-giver relative to PD patients who do not suffer dyskinesias. Also, though the validity of the PDQ-39 as a measure of HRQoL in PD has been considered previously (Ali and Salek, 2002, Ali and Salek, 2002) it has not been tested in such distinct groups of the PD populations such as dyskinetic and non-fluctuating patients. This analysis will address the ability of the instrument to highlight differences between such cohorts.

Study Aims and Objectives

- To determine the difference in the levels of quality of life between dyskinetic PD patients versus non-dyskinetic patients at baseline and 6 months using the PDQ-39.
- To determine the difference in carer burden for dyskinetic patients versus non-dyskinetic PD patients at baseline and 6 months.
- To assess the responsiveness of the PDQ-39 in patients classified as being dyskinetic and those classified as being non-dyskinetic.
- Determine whether there is a correlation between the Dyskinesia Subjective Rating Scale and the PDQ-39

METHODS

Study Design

This was a longitudinal cohort study of 6 months duration where 145 patients with PD were followed from November 2002 to July 2003 from out-patient neurology clinics at four hospitals. Two patient visits were involved and after the first visit (baseline), the patient was asked to return for the second visit 6 months later.

Study Instruments

At the first patient visit, the investigator collected preliminary medical and PD status information including UPDRS and dyskinesia rating for the purpose of stratification. The patients were given a disease specific HRQoL questionnaire; the PDQ-39

(Damiano et al., 1999, Jenkinson et al., 1997) and Dyskinesia Subjective Rating Scale (Lang and Fahn , 1999) to complete at each visit i.e. baseline and then the six month visit, prior to any other evaluation. Patients were instructed to read the brief directions on how to complete the instruments. After the patient's correct understanding had been confirmed, he/she were encouraged to complete every item without skipping any. Patients were encouraged to mark the response that is most applicable. The patient's carer was given a Carer Burden Interview questionnaire (Zarit et al., 1986) to complete at baseline and at the 6 month visit. The investigators completed the UPDRS (Fahn et al., 1987) and the severity score of the Hoehn and Yahr at each visit. For detailed information regarding the study instruments see chapter two and the relevant appendices.

Study Group Allocation

Three strata were used to classify patients by whether they experience:

1. no fluctuations
2. motor fluctuations, but NO dyskinesias
3. motor fluctuations, WITH dyskinesias

The study strata into which patients fell was determined by the existence of motor fluctuations (investigator discretion) and the patients score on the items 32-35 of the UPDRS which assess complications of therapy including dyskinesia.

1) No motor fluctuations (NDMF)
2) Motor fluctuations, but NO dyskinesias (MFND): scores of 0 on items 32-35
3) Motor fluctuations, WITH dyskinesias (MFD): score of =1 on item 32 (duration), regardless of scores on items 33-35.

Inclusion criteria

- Patients gave written informed consent
- Patients had a confirmed diagnosis of Parkinson's Disease (L-dopa responsive disease)
- Patients fell into one of three strata as described above
- Patients must have had stable disease for preceding 3-6 months
- Patients carer willing to complete Care giver burden questionnaire

Exclusion criteria

- Patients with atypical PD
- Patients unable to complete the PDQ-39 by themselves, due to disease severity or other reason
- Patients with severe co-morbidities/major medical illness

Data Processing and Analysis

Statistical tests were performed for descriptive purposes, in order to determine the relationship between patient quality of life and carer burden by each group. The main focus of the analysis was on the difference in levels of quality of life between the dyskinetic and non-dyskinetic groups. Statistical analysis was performed using SPSS software.

The PDQ-39 was scored according to the algorithm provided by the developers. Scores for each of the 8 scales were calculated by summing the responses to individual items within the scale and then linearly transforming to a 0 to 100 scale, where 0 represents no problem at all and 100 represents the maximum level of a problem. The hypothesis tested is that the dyskinetic group suffers lower levels of quality of life as indicated by the PDQ-39 than the fluctuating, but non-dyskinetic group. These differences were tested with data at baseline and 6 months.

Summary statistics for each visit were determined for each group (strata) in the study. Correlations of quality of life data with clinical variables were used to determine clinically meaningful differences in quality of life scores between groups. The Burden Interview was scored by summing the responses of the individual items. Higher scores indicate greater carer distress. Summary statistics for each visit were determined for each group (strata) in the study. Differences in carer burden were subsequently described. Also, group comparisons were conducted using Mann-Whitney test which performs a hypothesis test of the equality of two population medians and is appropriate to use as a nonparametric alternative to the two-sample t-test. In addition, summary statistics for the UPDRS, Dyskinesia Rating Scale and other clinical parameters have been tabulated.

RESULTS

Study Patient's Demographic Characteristics

Four PD investigational sites recruited a total of 145 patients satisfying the inclusion/exclusion criteria with the majority of patients recruited from one site (centre 1) who recruited 70% of the patients included in the analysis (Table 6.1).

Table 6.1 Demographic characteristics

Age - years	
Mean (S.D.)	**67.85 (12.13)**
Median (Range)	**70.05 (40 – 99.5)**
Sample size	
Centre 1 (%)	**101 (69.7)***
Centre 2 (%)	**39 (26.9)**
Centre 3 (%)	**3 (2.1)**
Centre 4 (%)	**2 (1.4)**
Gender	
Male (%)	**80.0 (55.2)**
Female (%)	**65.0 (44.8)**
Stratification	
NDMF (%)	**57 (39.6)**
MFND (%)	**23 (16.0)**
MFD (%)	**64 (44.4)**
Disease duration - years	
Mean (SD)	**9.0 (5.0)**
Median (Range)	**9.0 (2 – 27)**

***Only patients from this center were included in the analysis presented in this chapter**

The mean age of the cohort was 68 years and a median of 70 years with a range of 40 to 99.5 years reflecting a young and old population with a mean duration of disease of 9 years and balance of male (55.2%) and female patients (44.8%).

At baseline visit there were 57 patients with no dyskinesia or motor fluctuations (NDMF) compared to 23 patients with motor fluctuations and no dyskinesia (MFND) and 64 patients with motor fluctuations and dyskinesia (MFD). More than half of all patients recruited had a Hoehn and Yahr staging of II indicating mild disease severity

(Table 6.2). None of the patients were assessed as having severe disease (stage IV) but 13% of the group with motor fluctuations had Stage III disease compared to 3% with no motor fluctuations.

At the second visit (6 months), the same pattern to the baseline was found with more than half of the patients assessed as having stage II disease as measured by the Hoehn and Yahr.

Table 6.2: Hoehn & Yahr Staging based on stratification at baseline and 6 month visit:

H & Y		NDMF	MFND	MFD	Total
Baseline	**I**	7	3	4	*14*
	I.5	2	0	1	*3*
	II	26	8	33	*67*
	II.5	1	0	5	*6*
	III	1	3	5	*9*
	IV	0	0	0	*0*
	Total	*37*	*14*	*48*	*99*
Six months	**I**	4	2	4	*10*
	I.5	3	0	2	*5*
	II	28	9	29	*66*
	II.5	2	3	6	*11*
	III	0	1	4	*5*
	IV	0	0	1	*1*
	Total	*37*	*15*	*46*	*98*
		NDMF - no dyskinesia or motor fluctuations	**MFND - motor fluctuations and no dyskinesia**	**MFD - motor fluctuations and dyskinesia**	

The population characteristics were that very few patients recruited were considered to have severe disease and there appeared to be no significant change in the Hoehn and Yahr scores at the 6 month visit indicating a cohort with stable disease. This can be further investigated by assessing the mean Hoehn and Yahr scores at the two visits and comparing to see whether differences existed in the stratified cohorts (Table 6.3).

Hoehn and Yahr staging according to stratification

Table 6.3 shows there was no pattern in the Hoehn & Yahr (H&Y) staging which is used as a measure of symptom severity between the no dyskinesias (NDMF) and dyskinesias with motor fluctuations (MFD) cohorts.

Table 6.3: Comparison between three groups based on Hoehn & Yahr stage

	NDMF		MFND		MFD		
Hoehn & Yahr	Mean	Median	Mean	Median	Mean	Median	p-value*
Baseline	1.8	2.0	2.0	2.0	2.1	2.0	0.070
6-months	1.9	2.0	2.0	2.0	2.1	2.0	0.110

*** Kruskal Wallis Test**

There is no evidence to suggest that those patients with more severe symptoms (as determined by H&Y staging) were predominantly in the MFD strata. This was confirmed in further statistical analysis which showed that the magnitude of difference in H&Y scores/staging between the NDMF, MFND and MFD cohorts were not significant (Table 6.3).

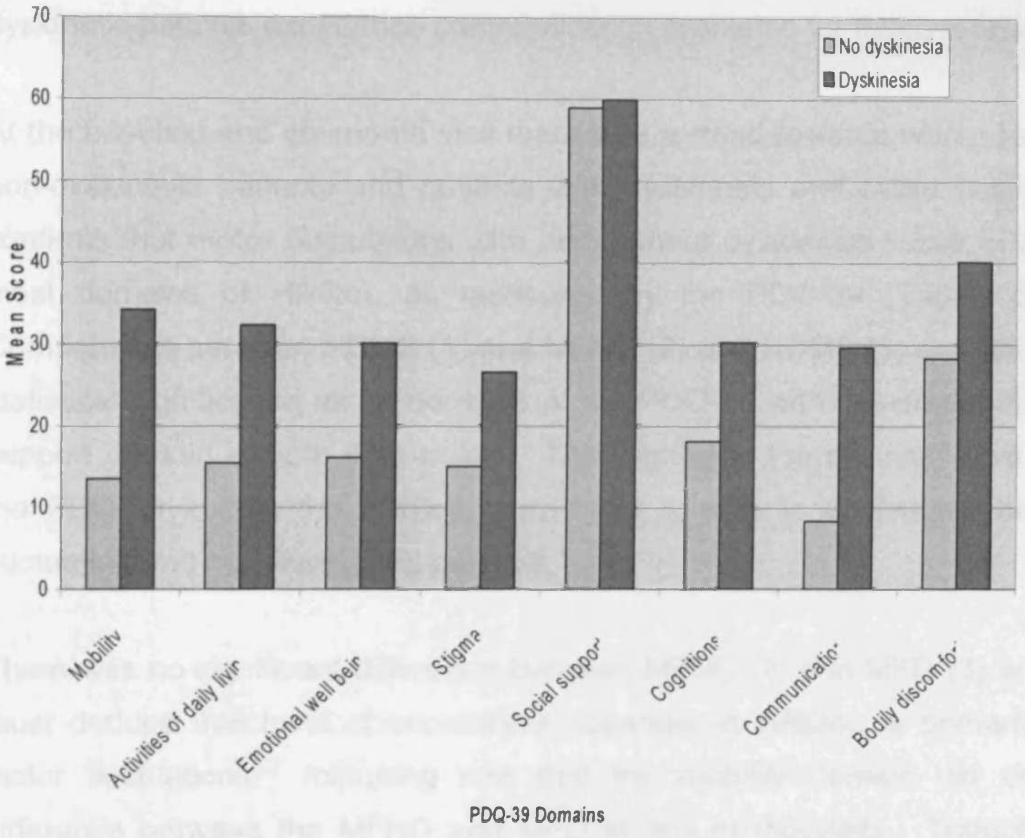
PDQ-39 Domain Scores in Dyskinesia and Non-Dyskinesia Patients

In this study population the domain of social support was most impaired at baseline and the six month visit for all groups of patients though the difference in mean scores between NDMF(1) and NDMF(2)/MFD(3) groups did not reach statistical significance (Figure 6.1). This indicates that PD patients, irrespective of their classification experience considerable impairment in this domain which is the opposite of what was found in the earlier comparison of the PDQ-39 and UKSIP which was administered by post to patients who were members of the UK Parkinson's Disease Society.

In this postal survey study which is discussed in chapter three, the social support dimension of the PDQ-39 had the lowest scores. This domain measures predominantly close ties and family support e.g. "*had problems with your close personal relationships*?"; "*lacked support in the ways you need from your family or close friends*?" Fitzpatrick et al 1997 and Schrag et al 2000 also found that social support scored the lowest compared to the other dimensions. This would suggest

that close relationships are less jeopardised by PD than are distant relationships but in this study quite the opposite seems to be the case.

Figure 6.1 Difference in levels of HRQoL between dyskinesia and non-dyskinesia PD patients at baseline



An explanation for this contradiction is that *Social support* is dependent on a variety of external variables at a particular point in time and can vary from the support of friends to the support of the health authority in the form of nursing/healthcare provision. These factors depend on circumstances that are independent of the patients' health status, namely the availability, affordability or access to these various forms of support. This particular cohort may indeed be highlighting concerns that patients have with access to healthcare provision.

Difference in HRQoL scores between the three stratified groups at baseline

The magnitude of difference in mean scores was greatest for the *communication* domain of the PDQ-39 at both visits between the non-fluctuating group and fluctuating group with dyskinesia (Tables 6.4 and 6.5). The results for

communication domain were statistically significant and the size of the significance was large ($p<0.001$) (Tables 6.4 and 6.5). Communication issues have been reported to affect social interactions (Fitzsimmons and Bunting, 1993, Jenkinson et al., 1995, Peto et al., 1995). Specifically, the mask-like facial symptom reduces the ability of individuals with PD to communicate non-verbal messages, and problems with speech and writing impair verbal and written messages. It is plausible that dyskinetic patients experience communication problems for these reasons.

At the baseline and six-month visit there was a trend towards worse health between non-dyskinesia patients and patients with dyskinesia and motor fluctuations. This confirms that motor fluctuations with and without dyskinesia result in a reduction in most domains of HRQoL as measured by the PDQ-39 (Tables 6.4 and 6.5). Comparisons between NDMS (1) and MFND (2) and NDMS (1) and MFD (3) showed statistical significance for all domains of the PDQ-39 with the exception of the social support domain at both time points. This highlights the discriminative properties of the PDQ-39 in that the HRQoL instrument is able to distinguish between motor fluctuating and non-fluctuating patients.

There was no significant difference between MFND (2) and MFD (3) and hence one must deduce that level of impairment observed in HRQoL is primarily due to the motor fluctuations. Intriguing was that the *mobility* domain did not show any difference between the MFND and MFD at any of the visits. Though the second largest difference in means scores were observed for this domain between the motor fluctuating and non-motor fluctuating (stable) group of patients and were statistically significant the same pattern was not seen between MFND and MFD. Dyskinesias are unwanted, excessive and abnormal involuntary movements which cause considerable distress to the patients and in this study group we observed that these patients experience worse HRQOL compared to stable patients their level of impairment are similar to patients with motor fluctuations and no dyskinesia. One would have expected mobility to have been greatly hampered in comparison with patients who experience such involuntary movements.

Table 6.4: Difference in HRQoL scores between the three stratified groups at baseline

Baseline PDQ-39 Category	Mean score (S.D.)						'p'-value†		
	1 (n=38)		2 (n=15)		3 (n=48)		(1) vs (2)	(1) vs (3)	(2) vs (3)
Mobility	13.5	(19.5)	36.8	(19.5)	34.2	(24.9)	0.0001	0.0001	0.533*
Activities of daily living	15.4	(14.6)	35.8	(16.4)	32.3	(17.6)	0.0001	0.0001	0.405*
Emotional well-being	16.1	(12.0)	36.7	(13.3)	30.1	(19.8)	0.0001	0.001	0.146*
Stigma	15.0	(19.6)	32.5	(22.4)	26.6	(19.1)	0.003	0.003	0.385*
Social support	58.9	(19.4)	57.2	(18.3)	59.9	(12.7)	0.258*	0.440*	0.610*
Cognitions	18.1	(15.7)	37.9	(11.9)	30.2	(16.9)	0.0001	0.002	0.061*
Communication	8.3	(11.8)	38.9	(15.0)	29.6	(18.5)	0.0001	0.0001	0.090*
Bodily discomfort	28.3	(23.4)	43.3	(17.6)	40.1	(25.1)	0.012	0.026	0.454*
	1= No dyskinesia or motor fluctuations (NDMF)		2= Motor fluctuations but no dyskinesia (MFND)		3= Motor fluctuations with dyskinesia (MFD)				

* = Not significant
†Mann-Whitney (p-values)

Table 6.5 Difference in HRQoL scores between the three stratified groups at 6 months

6 month PDQ-39 Category	Mean score (S.D.)						'p'-value†		
	1 (n=38)		2 (n=15)		3 (n=46)		(1) vs (2)	(1) vs (3)	(2) vs (3)
Mobility	15.9	(17.6)	31.3	(21.7)	32.7	(22.7)	0.011	0.001	0.947*
Activities of daily living	17.0	(14.1)	34.2	(19.7)	32.9	(19.3)	0.003	0.0001	0.805*
Emotional well- being	17.2	(14.5)	33.6	(19.1)	30.9	(17.5)	0.004	0.0001	0.536*
Stigma	14.2	(15.4)	33.8	(24.3)	23.0	(17.7)	0.005	0.014	0.148*
Social support	56.3	(20.9)	58.3	(24.3)	56.6	(18.7)	0.550*	0.908*	0.508*
Cognitions	19.9	(15.9)	35.8	(17.1)	30.6	(15.6)	0.004	0.002	0.241*
Communication	12.5	(14.5)	37.2	(17.8)	29.3	(20.0)	0.0001	0.0001	0.157*
Bodily discomfort	27.4	(20.6)	43.3	(15.8)	38.2	(21.1)	0.005	0.012	0.281*
	1= No dyskinesia or motor fluctuations (NDMF)		2= Motor fluctuations but no dyskinesia (MFND)		3= Motor fluctuations with dyskinesia (MFD)				

= Not significant
†Mann-Whitney (p-values)

Burden amongst carers of PD patients

The analysis of the impact on the carer of PD patients shows that those caring for dyskinesia patients experienced *mild* to *moderate* burden compared to non-dyskinesia carers who experience little or no burden (Table 6.6).

Table 6.6 Comparison of Carer Burden between dyskinesia and non-dyskinesia population

Burden Interview (% score)	1 n = 18		2 n = 8		3 n = 29		'p'-value†		
	Mean	S.D.	Mean	S.D.	Mean	S.D.	(1) vs (2)	(1) vs (3)	(2) vs (3)
Baseline total burden score (%)	17.6	(13.9)	22.7	(9.9)	(26.3)	(15.9)	0.165	0.082	0.605
6-month total burden score (%)	18.6	(8.5)	18.4	(12.0)	28.6	(18.5)	0.883	0.153	0.271

† **Mann-Whitney (p-values)** 1 = No dyskinesia or motor fluctuations (NDMF) 2 = Motor fluctuations but no dyskinesia (MFND) 3 = Motor fluctuations with dyskinesia (MFD)

Comparisons between groups NDMF and MFND, and NDMF and MFD were not statistically significant though the difference between MFD and NDMF just missed reaching statistical significance at baseline. The number of care giver burden interview questionnaires was lower at the second time of completion. To further investigate which dimensions of carers' interaction with the patient were most affected, the frequency that each item was affirmed by the carer was analysed. The percentage of carers who answered "sometimes", "quite frequently" or "nearly always" on the Burden Interview (BI) at baseline and 6-month follow-up are presented in Tables 6.7.

At the baseline and 6-month follow-up, more than sixty percent of the carers indicated that they at least sometimes "felt afraid about what the future holds for their relative" (burden Q7) and "felt that the relative was dependent upon them" (burden Q8). This was followed by the carer indicating in nearly half of the respondents that they "felt they should be doing more for the relative" (burden Q20). At both visits more than 70% of the carers who responded answered *Never* to the questions "do you feel uncomfortable about having friends over because of your relative?" (burden Q13) and "do you wish you could leave the care of your relative to someone else?" (burden Q18).

Table 6.7 Caregiver Burden Interview Questionnaire (CBIQ) affirmative responses

	Baseline & (6-month)							
CBIQ Items	Never (0)	Rarely (1)	Sometimes (2)	Quite frequently (3)	Nearly always (4)	Total	Baseline % (2,3, 4)	6-month % (2,3,4)
Burden Q1	8 (9)	13 (12)	9 (9)	3 (0)	0 (1)	33 (31)	36.4	32.3
Burden Q2	5 (10)	12 (11)	14 (8)	1 (0)	1 (1)	33 (30)	48.5	30.0
Burden Q3	7 (8)	10 (11)	12 (7)	3 (3)	1 (1)	33 (30)	48.5	36.7
Burden Q4	22 (21)	8 (6)	3 (3)	0 (1)	0 (0)	33 (31)	9.1	12.9
Burden Q5	14 (14)	15 (10)	3 (5)	1 (1)	0 (0)	33 (30)	12.1	20.0
Burden Q6	22 (15)	8 (11)	3 (4)	0 (0)	0 (0)	33 (30)	9.1	13.3
Burden Q7	0 (1)	5 (4)	18 (12)	10 (11)	0 (2)	33 (30)	84.8	83.3
Burden Q8	3 (3)	9 (5)	11 (13)	6 (4)	4 (5)	33 (30)	63.6	73.3
Burden Q9	16 (12)	12 (10)	4 (8)	0 (1)	1 (0)	33 (31)	15.2	29.0
Burden Q10	18 (17)	8 (10)	7 (3)	0 (1)	0 (0)	33 (31)	21.2	12.9
Burden Q11	20 (19)	7 (6)	5 (5)	0 (0)	1 (1)	33 (31)	18.2	19.4
Burden Q12	11 (11)	13 (13)	8 (4)	0 (2)	1 (1)	33 (31)	27.3	22.6
Burden Q13	25 (22)	6 (8)	2 (1)	0 (0)	0 (0)	33 (31)	6.1	3.2
Burden Q14	16 (11)	5 (10)	8 (5)	2 (1)	2 (4)	33 (31)	36.4	32.3
Burden Q15	21 (15)	4 (6)	6 (6)	2 (4)	0 (0)	33 (31)	24.2	32.3

Table 6.7 Caregiver Burden Interview Questionnaire (CBIQ) affirmative responses (continued)

	Baseline & (6-month)							
CBIQ Items	Never (0)	Rarely (1)	Sometimes (2)	Quite frequently (3)	Nearly always (4)	Total	Baseline % (2,3, 4)	6-month % (2,3,4)
Burden Q16	22 (18)	8 (8)	2 (4)	1 (1)	0 (0)	33 (31)	9.1	16.1
Burden Q17	17 (18)	9 (7)	6 (5)	0 (0)	1 (1)	33 (31)	21.2	19.4
Burden Q18	24 (24)	6 (4)	2 (3)	1 (0)	0 (0)	33 (31)	9.1	9.7
Burden Q19	10 (11)	16 (14)	6 (5)	1 (1)	0 (0)	33 (31)	21.2	19.4
Burden Q20	7 (8)	10 (10)	13 (9)	2 (2)	1 (2)	33 (31)	48.5	41.9
Burden Q21	8 (7)	12 (11)	11 (10)	1 (0)	1 (3)	33 (31)	39.4	41.9
Burden Q22	13 (14)	16 (12)	3 (4)	0 (0)	1 (1)	33 (31)	12.1	16.1

Table 6.8 Relationships between DSRS total and PDQ-39 domain scores at baseline and 6-month follow-up (n = 44)

PDQ-39 domains	Spearman correlation	
	Baseline	6-month
mobility	**0.338***	**0.339***
activities of daily living	**0.189**	**0.192**
emotional well being	**0.230**	**0.183**
stigma	**0.162**	**0.237**
social support	**0.104**	**0.112**
cognitions	**0.077**	**0.204**
communication	**0.169**	**0.157**
bodily discomfort	**0.259**	**0.274**

DSRS = dyskinesia subjective rating scale
*P<0.05

Relationships between the PDQ-39 and Dyskinesia Subjective Rating Scale

The results showed that patients who were dyskinetic experienced reduced HRQoL in all domains with the exception of social support. To further investigate the utility of the PDQ-39 in motor fluctuating patients Spearman's correlation coefficients were used to assess relationships between PDQ39 domains and the Dyskinesia Subjective Rating Scale (DSRS) – Table 6.8. The dyskinesia subjective rating scale did not correlate significantly with any of the PDQ-39 domains with the exception of mobility ($p<0.05$) at both baseline and 6 month follow-up. If one adjusts for multiple testing and applies the Bonferroni's adjustment with p value set at $p = 0.005$ significance level then none of the domains of the PDQ-39 would have reached statistical significance. At best, one can conclude that the mobility domain exhibits a modest correlation to DSRS total score. What is surprising is that the Activities of Daily Living domain of the PDQ-39 did not show moderate or high correlation with the

DSRS as this HRQoL domain contains items relating to writing, cutting food, dressing, and washing. These are domains of activities of daily living that are generally affected by dyskinesias. In the absence of such expected association it would suggest that the PDQ-39 is tapping into domains of interest that are not identified by the DSRS.

Table 6.9 Comparison of UPDRS section scores between dyskinesia and non-dyskinesia population at baseline

Baseline UPDRS sections	Mean score (S.D.)						'p'-value†		
	1 (n=38)		2 (n=15)		3 (n=48)		(1) vs (2)	(1) vs (3)	(2) vs (3)
Mentation, behaviour and mood	7.4	(7.7)	12.1	(8.3)	12.6	(11.9)	0.052	0.053	0.831
Activities of daily living	12.8	(7.6)	19.0	(12.0)	18.9	(10.3)	0.104	0.007	0.908
Motor examination	17.3	(7.1)	21.1	(9.1)	17.6	(8.3)	0.180	0.798	0.164
Complications of therapy	0.0	-	19.2	(6.3)	31.9	(15.2)	0.097	0.088	0.004
	1= No dyskinesia or motor fluctuations (NDMF)		2= Motor fluctuations but no dyskinesia (MFND)		3= Motor fluctuations with dyskinesia (MFD)				

†Mann-Whitney (p-values)

Table 6.10 Comparison of UPDRS section scores between dyskinesia and non-dyskinesia population at 6 months.

6-month UPDRS sections	Mean score (S.D.)						'p'-value†		
	1 (n=38)		2 (n=15)		3 (n=48)		(1) vs (2)	(1) vs (3)	(2) vs (3)
Mentation, behaviour and mood	6.7	(7.9)	9.6	(9.1)	12.1	(12.3)	0.306	0.037	0.677
Activities of daily living	14.3	(7.2)	20.8	(14.7)	19.6	(10.5)	0.147	0.016	0.946
Motor examination	18.0	(7.2)	18.8	(7.3)	18.0	(7.9)	0.550	0.708	0.755
Complications of therapy	26.1	-	20.7	(5.9)	30.4	(14.4)	0.268	0.679	0.007
	1= No dyskinesia or motor fluctuations (NDMF)		2= Motor fluctuations but no dyskinesia (MFND)		3= Motor fluctuations with dyskinesia (MFD)				

†Mann-Whitney (p-values)

Comparison of motor impairment as measured by the UPDRS between dyskinesia and non-dyskinesia patients

Table 6.9 and 6.10 above presents the differences in UPDRS section scores between fluctuating and non-fluctuating patients. There was a trend towards greater impairment in *activities of daily living (ADL)* and *mentation, behavior and mood* in the group MFD patients compared to NDMF group ($p<0.05$). Hence it is reasonable to propose that patients who present with motor fluctuations and dyskinesia are more likely to experience deterioration in their ADL compared to stable patients. The ADL section of the UPDRS covers speech, walking cutting food, dressing and handwriting which are likely to be affected by the involuntary movements characterised by dyskinesia. The PDQ-39 also identified statistically different mean scores for the ADL domain between these two groups of patients. *Mentation, behaviour and mood* section of the UPDRS assesses thought disorder, depression and motivation and the results showed a statistically significant difference between NDMF and MFD groups of patients indicating that those experiencing dyskinesias are more likely to experience deterioration in their mental state; a trend which was also detected in the *emotional well-being* domain of the PDQ-39. The non-motor symptoms associated with PD and its management represent impairments in emotional well-being and cognitive function (Menza et al., 1993, Peto et al., 1995). Individuals with PD may experience depression and anxiety and it has been shown that depression had a significant impact on the HRQoL of PD patients (Findley and Committee), 2002). These findings suggest that behaviour and mood has a greater impact in the dyskinetic PD patient compared to stable patients.

The mean score differences between MFND and MFD was statistically significant at both visits for *complications of therapy* with higher scores indicating greater severity for the MFD cohort. *Complications of therapy* section are a mixture of quantitative scores for dyskinesia and fluctuations in terms of duration and disability and qualitative scores for the presence or absence of specific types of dyskinesias and fluctuations. Qualitative scores are also provided for other complications, namely, sleep and blood pressure disturbances. This is compelling evidence that pharmacotherapy in the MFD group was not adequately controlling motor fluctuations and the associated dyskinesia. It shows that in this group these patients were not experiencing benefit in decreased dopaminergic complications and alternative treatment strategies must be considered.

UPDRS ability to discriminate between Hoehn & Yahr stages

An attempt was made to identify whether the UPDRS could discriminate between PD patients with different Hoehn & Yahr staging which is an indication of symptom severity. Plots of UPDRS total score against Hoehn and Yahr stage score are presented in Figure 6.2 – baseline and Figure 6.3 – six-month. This would have provided an arbitrary grading system to separate patients into those that are improving and those that are deteriorating at the follow-up visit and would provide a validated categorisation of the two patient groups for a sensitivity analysis of the PDQ-39. Currently there is no guidance on the categorisation of UPDRS scores that would allow the assessor to determine whether the patient has mild, moderate or severe signs and symptoms of PD.

Figures 6.2 and 6.3 show that there was no association between the two clinical instruments and that the UPDRS was unable to discriminate between the different Hoehn & Yahr stages. This may be explained because the Hoehn and Yahr staging presents a crude estimation of severity of the disease and was designed only for staging within the boundaries of the disease whereas the UPDRS scores the signs and symptoms of PD and hence the latter instrument is measuring many related multi-dimensional features of PD.

Though a validated grading system to separate patients into those that are improving and those that are deteriorating based on UPDRS scores would have facilitated the assessment of responsiveness of the PDQ-39, an alternative method is required in the absence of correlations between the Hoehn and Yahr and UPDRS.

Responsiveness (sensitivity) to change

It was necessary to assess whether the PDQ-39 was sensitive enough to detect changes in clinical symptoms or disease severity. Thus, responsiveness analyses were carried out using data from those patients who were deemed to have deteriorated as indicated in an increase in total UPDRS score for the following group of patients:

Figure 6.2: UPDRS total score against Hoehn and Yahr stage - baseline

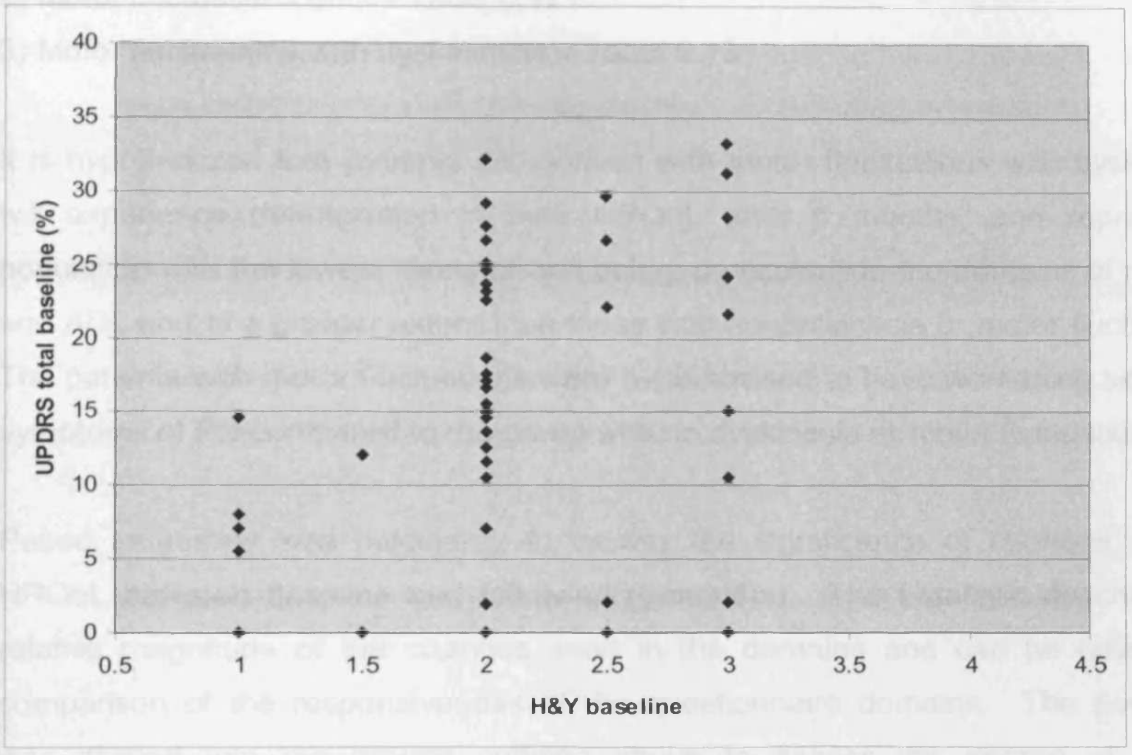
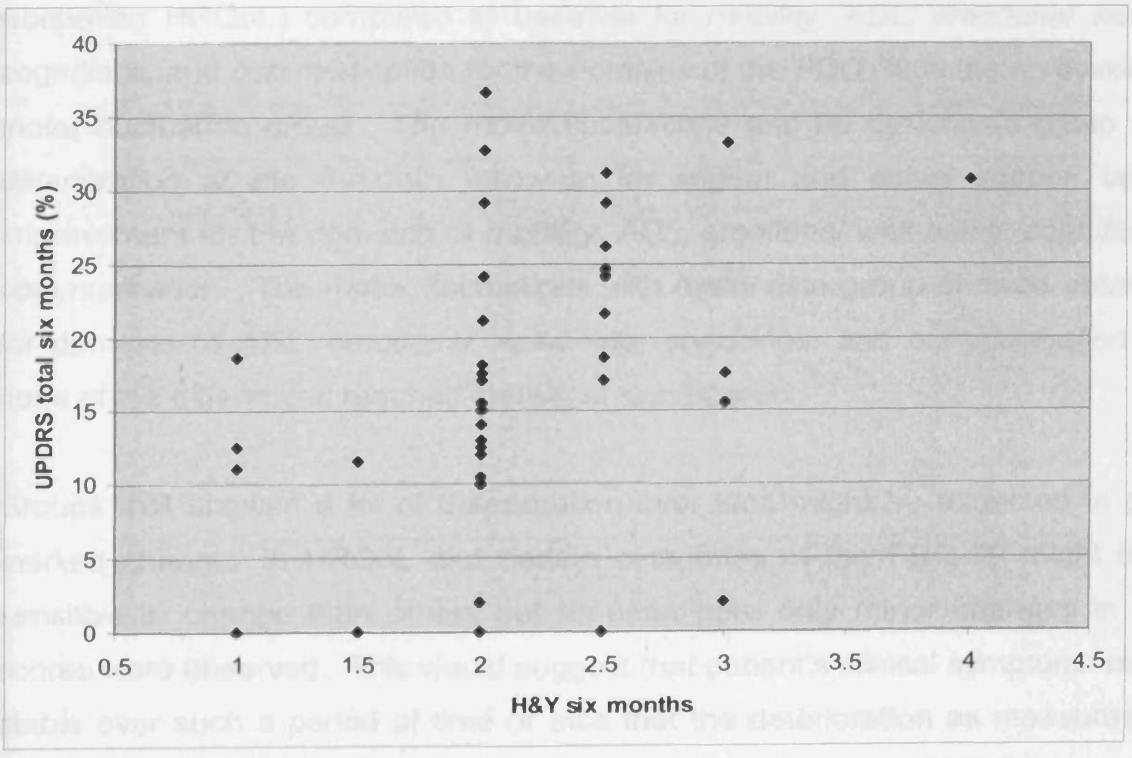


Figure 6.3: UPDRS total score against Hoehn and Yahr stage – six month



1) No dyskinesias or motor fluctuations – Table 6.11
2) Motor fluctuations only – Table 6.12
3) Motor fluctuations with dyskinesias – Table 6.13

It is hypothesized that patients categorised with motor fluctuations with dyskinesias will experience deterioration in their HRQoL after 6 months, and represent a population with the lowest levels of well being, particularly in the domains of mobility, and ADL and to a greater extent than those with no dyskinesia or motor fluctuations. The patients with motor fluctuations were hypothesised to have worsening signs and symptoms of PD compared to the group with no dyskinesia or motor fluctuations.

Paired t-statistics was calculated to assess the significance of changes seen in HRQoL between baseline and follow-up (6-months). The t-statistic describes the relative magnitude of the changes seen in the domains and can be taken as a comparison of the responsiveness of the questionnaire domains. The population was divided into the groups outlined above to assess the degree of change demonstrated during the 6-month period.

Mean scores at 6-month visit showed *deterioration* (higher PDQ-39 scores indicating worsening HRQoL) compared to baseline for *mobility, ADL, emotional well-being, cognitions,* and *communication* for the domains of the PDQ-39 in the no dyskinesia or motor fluctuation group. The motor fluctuations and no dyskinesia group showed *deterioration* at the 6-month follow-up for *stigma* and *social support* but slight improvement for the domains of *mobility, ADL, emotional well-being, cognitions*, and *communication*. The motor fluctuations with dyskinesia group showed *deterioration* for domains of *ADL*, *emotional well-being*, *cognitions*, and *communication* though none of the differences reached statistical significance.

Groups that showed a lot of deterioration over time might be expected to produce marked changes in HRQoL and certain categories of the PDQ-39 might be more sensitive to change than others but as seen here only minor changes in HRQoL scores were observed. This would suggest that patient's clinical symptoms remained stable over such a period of time or else that the deterioration as measures by the UPDRS did not reflect in deterioration in some of the domains of HRQoL.

The dyskinesia group and motor fluctuation groups exhibited higher domain scores compared to the no dyskinesias group indicating worse health but the dyskinesia and motor fluctuating domain scores were similar.

Table 6.11 Responsiveness of the PDQ-39 (no dyskinesias or other motor fluctuations)

PDQ-39 Category	Mean score		Paired differences mean	S.D.	'p'-value*
	Baseline	6-month			
Mobility	13.5	15.9	-2.4	12.1	0.237
Activities of daily living	15.7	17.0	-1.4	10.7	0.448
Emotional well-being	16.3	17.2	-0.9	8.8	0.537
Stigma	15.0	14.2	0.8	15.7	0.746
Social support	58.7	56.9	1.7	25.5	0.685
Cognitions	18.1	19.6	-1.6	9.8	0.347
Communication	8.3	12.5	-4.2	10.7	0.022**
Bodily discomfort	28.3	27.4	0.9	16.9	0.750
PDQ Total	19.9	21.3	-1.5	7.3	0.266

* paired t-test

**p<0.05

Table 6.12 Responsiveness of the PDQ-39 (motor fluctuations & no dyskinesias)

PDQ-39 Category	Mean score		Paired differences mean	S.D.	'p'-value*
	Baeline	6-month			
Mobility	36.8	30.2	6.6	10.8	0.040**
Activities of daily living	35.8	34.2	1.7	10.8	0.559
Emotional well being	36.7	33.6	3.1	11.8	0.334
Stigma	32.5	33.8	-1.3	19.9	0.812
Social support	57.2	58.3	-1.1	24.1	0.861
Cognitions	37.9	35.8	2.1	12.9	0.541
Communication	38.9	37.2	1.7	17.6	0.719
Bodily discomfort	43.3	43.3	0.0	15.7	1.000
PDQ Total	37.8	34.9	2.9	7.7	0.183

* paired t-test

**p<0.05

There were no statistically significant differences between the mean scores at baseline and follow-up for any of the patients in the 3 groups with the notable exception of *mobility* for the motor fluctuations and no dyskinesia group and *communication* for the no dyskinesia or other motor fluctuation group. The group that was considered to have the most severe signs and symptoms of PD, the motor fluctuations with dyskinesias cohort, all domain scores were found to be not responsive to change.

Table 6.13 Responsiveness of the PDQ-39 (motor fluctuations with dyskinesias)

PDQ-39 Category	Mean score		Paired differences mean	S.D.	'p'-value*
	Baseline	6-month			
Mobility	34.2	32.7	1.5	18.0	0.568
Activities of daily living	32.4	33.1	-0.6	14.5	0.768
Emotional well-being	30.4	30.8	-0.4	13.1	0.852
Stigma	26.6	23.0	3.5	16.8	0.154
Social support	59.9	56.6	3.3	18.2	0.216
Cognitions	30.2	30.3	-0.1	14.5	0.950
Communication	29.6	30.0	-0.4	16.3	0.882
Bodily discomfort	40.1	38.2	1.9	18.4	0.475
PDQ Total	33.9	33.3	0.6	10.2	0.706

paired t-test
**p<0.05

DISCUSSION

The morbidity of PD has improved dramatically over the past 25 years with the introduction of effective pharmacotherapy, most notably levodopa (synthetic dopamine). However, levodopa use is associated with long term complications such as motor fluctuations due to end of dose deterioration and in particular dyskinesia. At present, pharmacological approaches to these problems are limited (Clarke et al., 1995) though the use of dopamine agonists early in the course of the disease may prove valuable in the prevention of response fluctuations. Nevertheless, alternative methods must be sought to improve the quality of life of parkinsonian patients. Patients have described dyskinesia as feeling as if "My brain is sending the message

to the body, but the body wont listen"(Hickey, 1992) whereas motor fluctuations which result in the "on-off" phenomenon where the patient is ambulatory one minute and then they find they are frozen and unable to move the next has been described by patients as if someone has turned off the switch and their feet have sprouted roots (Fitzsimmons and Bunting, 1993). Clearly an investigation of the impact of motor fluctuations and dyskinesia on health-related quality of life is warranted but before research in this area can be pursued it is important to validate the use of the HRQoL instrument to be used in such a distinct cohort of the PD population. The PDQ-39 has been extensively assessed in the preceding chapters and by the authors of the instrument (Jenkinson et al., 1998, Jenkinson, 1997) but its application in such distinct cohorts (i. patients with no symptoms of motor fluctuation; ii. patients with motor fluctuations; and iii. patients with motor fluctuations and dyskinesias) has not been assessed and there is a need to assess whether it can measure differences between dyskinetic and non-dyskinetic PD patients and whether it is responsive to small but clinically important changes in such a cohort.

Carer burden is an area that has not been extensively addressed in the assessment of HRQoL but nevertheless remains an important consideration in PD which is predominantly a chronic disease affecting the elderly, a group who are reliant on close relationships. The burden of PD patients care falls mainly on family and relatives. The need for permanent care and help to the chronically disabled person leads to a decrease in the quality of life of carers (Glozman et al., 1998). The strain of care giving has been associated with adverse physiologic and psychological outcomes for both carers and PD sufferers and its impact and relationship with HRQoL in PD patients is also justified.

One of the key findings of this study is that patients with motor fluctuations and dyskinesia experience reduced HRQoL compared to patients with no symptoms of motor fluctuation or dyskinesia as measured by the PDQ-39. The domains which were most affected by dyskinetic patients compared to non-dyskinetic patients were *communication* and *mobility*. Though several studies have shown that health-related quality of life (HRQoL) is reduced when motor fluctuations occur (Calne et al., 1996, Chapuis et al., 2004, Schrag and Quinn, 2000); Dodel and colleagues (Dodel et al., 2001) stated that *"this may also be true of dyskinesias, but because of the limited number of studies a definite conclusion is not yet possible"* and hence it is hoped this

study will add to the understanding of HRQoL in dissimilar PD populations. This study substantiates the findings that all domains of the dyskinetic cohort, as assessed by the PDQ-39 with the exception of *social support*, are significantly different between the non-fluctuating group. The patients who are experiencing involuntary movements clearly experience deterioration in their quality of life which highlights the inadequacy of their treatment which is thought to contribute to the dyskinesia. The investigator may find that clinical assessment as measured for example by the Hoehn and Yahr shows that the patient's motor symptoms such as slowing of body movements and impairment of equilibrium on walking or rigidity are adequately controlled by medication but what he would not be aware of is the debilitating effects or perception of the dyskinetic patients, that the disease and/or medication is having on aspects of well-being which are relevant and important to the patient. It is for this reason that a HRQoL instrument should be completed in conjunction with the clinical assessment such that the investigator is able to identify issues perceived to be of importance to the patients, as such perceptions can relate to the patients perception of well-being. A HRQoL instrument can provide additional information that would allow the investigator to change treatment or management strategy for the patients. Once an alternative treatment strategy is implemented the investigator could use the HRQoL instrument to measure improvement in the domain of concern. The PDQ-39 highlights as mentioned that *communication* in dyskinetic patients is greatly affected and this is probably a manifestation of the multifactorial nature of the disease combined with the wearing off effect of medication and the deleterious side-effects of dopaminergic drugs such as levodopa with are known to cause dyskinesia. The patient who experiences dyskinesias will face problems related to *communication* as simple tasks, taken for granted by most including healthy elderly patients, are deemed to be problematic. Problems with hand writing and speech will be experienced due to the uncontrollable involuntary movements characterised by dyskinesia and problems with speech and writing impair verbal and written messages. The PDQ-39 *communication* domain includes the item "felt unable to communicate with people properly" and "felt ignored by people" which would mean that dyskinetic patients are unable to adequately communicate their motor and non-motor symptoms which are affecting their quality of life.

The PDQ-39 also identified *mobility* as the second most affected domain in dyskinetic and motor fluctuating patients. It is not surprising that the freezing of motor

functioning and abnormal muscular movement is impeding mobility of the patient more so than in the non-dyskinetic group. Such symptoms can interfere with several areas of HRQoL, for example, the hampering of mobility may affect self-care activities and mean that the patient is more dependent on others for assistance, which in turn may make the individual experience a diminished sense of autonomy. Such consequences reveal the importance of carers in the supervision of such patients.

Another intriguing finding from the analysis of the PDQ-39 was that the motor fluctuating and dyskinetic groups experienced considerably more impairment on the *emotional well-being* domain compared to the non-fluctuating group. At the baseline visit, the motor fluctuating patients with no dyskinesia (Group 2 in Table 6.4) reported a 128% increase in score in this domain compared to the no fluctuation group (Group 1 in Table 6.4) and 87% increase in the motor fluctuation with dyskinesia group (Group 3 in Table 6.4), compared again to the no fluctuation group. A similar pattern was observed at the 6-month visit showing the failure of treatment and management of the patient to address such a non-motor complication. The PDQ-39 clearly shows that emotional and psychosocial well-being is negatively affected. As the disease progresses, the response to levodopa typically decreases and various motor complications develop; these are difficult to treat and result in further declines in HRQoL which impact on the patient's well being resulting in depression and anxiety. It has been reported that the non-motor symptoms associated with PD and its management represent impairments in emotional well-being (de Boer et al., 1996, Stallibrass, 1997) but in this analysis we have seen that impairment in emotional well-being is more severe in motor fluctuating and dyskinetic patients compared to non-dyskinetic patients.

The PDQ-39 has hence shown that it can distinguish between patients with symptoms of motor fluctuation and those with no motor fluctuation as patients with motor symptoms have statistically significantly higher domain scores indicating reduced quality of life. This highlights the discriminative properties of the PDQ-39 in this study cohort.

Where this study has shown that a presence of motor fluctuations and dyskinesias were associated with poor quality of life, Schrag and co-workers (2000) in an earlier

study reported that a history of fluctuations of motor response to medication or presence of dyskinesias were not associated with poor quality of life. However, these results may be due to the relatively few patients with these complications which, if present, were usually not severe in this population based sample. Also, in another study by Schrag and Quinn, (2000) they state that quality of life was not significantly different between patients with and without motor fluctuations and dyskinesia whereas this analysis has shown a significant difference in HRQoL between patients with and without motor fluctuations and dyskinesia. Schrag used the summary score of the PDQ-39 in their assessment of HRQoL in dyskinesia patients identified from a sample of registered patients from the general practices in the London area in what was a cross sectional study. In this study by Schrag and Quinn (2000), most of the study information was reported retrospectively by the patients themselves or derived from the records of their GP's and thus it is possible that the frequency of motor complications or dyskinesia in the cross-sectional study was underestimated. The analysis in this study does show that there is no difference in HRQoL between patients with motor fluctuations and no dyskinesia and patients with motor fluctuations and dyskinesia, an item which has not been addressed so far in the literature.

The similarity of PDQ-39 domain scores between the patients with motor fluctuations and no dyskinesia and patients with motor fluctuations and dyskinesia suggests that though these two groups report worse HRQoL compared to the group with no motor symptoms of fluctuation, they both experience the same level of HRQoL impairment. It has been reported that the prevalence of both fluctuations and dyskinesias increase with increasing disease and treatment duration as well as advancing disease severity (Chapuis et al., 2004) and therefore the same variables are responsible for the onset of these debilitating symptoms of PD. It would appear that patients do not consider the onset of dyskinesias to be more severe than the onset of the freezing "on-off" phenomenon and vice versa and consider both conditions to be similarly debilitating on HRQoL. Dykinetic patients like motor fluctuations, affected the areas of *bodily discomfort*, *mobility* and *activities of daily living* the most. There was no statistical difference in mobility domain scores which was interesting in that one might have expected patients who are experiencing a "on-off" freezing phenomenon coupled with involuntary muscular movements to have reduced mobility in comparison to the group who were experiencing the "on-off" episodes without

dyskinesia. One would anticipate that the "on-off" and dyskinesias would have had a profound impact on the patient's ability to walk, get around the house, felt worried about falling over; items that make up the mobility domain of the PDQ-39.

It would be useful for future research to break down the types of dyskinesia reported as some forms of dyskinesia may be shorter in duration, less painful, etc, and hence impact less severely on domains of HRQoL compared to other forms of dyskinesia. This would have allowed assessing whether certain forms of dyskinesia had a greater impact on HRQoL compared to motor fluctuations and no dyskinesia group. No attempt was made to categorise the different forms of dyskinesia in this analysis.

It was decided to assess whether the UPDRS sub sections which include *Mentation, behaviour and mood*; *Activities of daily living*; *Motor examination* and *Complications of therapy* were able to distinguish between the different groups and whether there was a difference in scores between such groups. The UPDRS is used routinely by clinicians to assess the signs and symptoms of PD and is used extensively in clinical trials to assess improvement of PD symptoms. The *complications of therapy* section of the UPDRS which includes an evaluation of the presence and extent of dyskinesias and clinical fluctuations shows that the patients with motor fluctuations and dyskinesia (MFD) experienced more complications that those with motor fluctuations only (MFND). The results were statistically significant at both visits. This is interesting in that it shows that complications of therapy are more evident in the group exhibiting dyskinesia and therefore are experiencing more pain, disability and a longer duration of the symptoms, items which this section addresses. This was not picked up by the PDQ-39 where there was no difference in any of the domain scores between the two groups but nor would one expect a HRQoL instrument to identify that there are complications of therapy in the population. What is important here is that the UPDRS provides insight as to the difference between these two groups that the PDQ-39 could not. It would be strongly recommended that the complications of therapy section be used in conjunction with the PDQ-39 in future assessments of dyskinetic and fluctuating patients.

Greater statistically significant impairment was observed in *ADL* and *mentation, behaviour and mood* in the MFD patients compared to the NDMF group which reflects a similar finding from the PDQ-39 assessment. There was no difference in

motor scores between the NDMF, MFND and MFD groups which would suggest that motor functioning is similar irrespective of the effects of pharmacotherapy related symptoms. So far it can be elucidated that treatment strategies are effective in the control of motor function such as movement, tremor and rigidity but do not adequately alleviate the symptoms of motor fluctuations and dyskinesia. It would appear that to improve the quality of life of fluctuating and dyskinetic patients one would have to consider reduced dosages of levodopa, delayed onset of such medication or the introduction of alternative therapies which may in turn reduce motor control due to reduced efficacy as levodopa is considered to be the gold standard treatment in PD. Of the 101 patients for whom anti-parkinsonian medication was recorded all but seven were receiving levodopa. Of the seven patients not receiving levodopa, four were categorised as having no dyskinesia or other motor fluctuations, two were categorised as having motor fluctuation with dyskinesia and one had motor fluctuation and no dyskinesia. It was not feasible to assess whether there was a relationship between levodopa and non-levodopa use and the onset of dyskinesia due to the small number of patients not taking levodopa.

The impact of or burden felt by family members of PD patients is an area that has not been studies extensively and there is very little on the topic in the literature to date. Carer burden has been defined as "the physical, psychological or emotional, social and financial problems that can be experienced by family members caring for impaired older adults"(George and Gwyther, 1986). PD, in particular is an illness that places unrelenting and progressive demands for care on carers because, emotional, physical and self-care deficits restricts the ability of the impaired elder to perform activities of daily living. In this study higher scores indicating greater carer distress were observed for the MFD and MFND groups. The MFD groups showed a trend to greater carer distress compared to both MFND and NDMF groups though none of the comparisons achieved statistical significance. Norms for Burden Interview have not been completed but estimates of the degree of burden can be made from preliminary findings by Zarit (Zarit et al., 1986), these are: 0-20 indicates little or no burden; 21-40 indicates mild to moderate burden; 41-60 indicates moderate to severe burden. Based on these categories, the carers of the MFD group experienced mild to moderate burden, compared to little or no burden that was reported by the carers of the NDMF group. Hence the presence of dyskinesia has a negative impact on the carer more so that the carers of patients who are stable which ties in with a study

which found functional status in patients to be an important variable for carer burden (Caap-Ahlgren and Dehlin, 2002).

Carers reported most often that they "felt afraid about what the future holds for their relative", and "felt the relative was dependent upon them" and further a sense of helplessness was detected as nearly half of the carers "felt they should be doing more for their relative". There was a sense of responsibility on the part of the carer who in most cases was a spouse or family member in that more than 70% of the carer respondents answered *never* to wishing they could leave the care of the relative to someone else. Further examination of the Burden Interview showed that 88% of the carers who responded at baseline and 84% at 6-months *never* or *rarely* felt burdened in caring for their relative. It may be that the respondent is reluctant to truly provide a genuine response to this question for fear of upsetting the patient relative or for fear of admitting failure. As we have seen mild to moderate distress is experienced by the carer. Carer pessimism early in the carer role has been found to be a warning sign for poor current and future carer health(Lyons et al., 2004). High baseline pessimism signalled high levels of baseline depressive symptoms and poor physical health, as well as a faster decline in health over the 10-year study. Depressive status of the carer was not measured in this analysis but if the low levels of pessimism reported at baseline (assuming an indication of "not feeling burdened" relates to whether one is feeling pessimistic) one should not expect the carer role to be impacted by poor health. The lack of pessimism was observed six months later. Nevertheless, an assessment of carer burden provides a unique opportunity to detect and intervene with carers who show high levels of pessimism early in the care-giving trajectory. The value of maintaining such carers and their families is immense. It has been estimated that the NHS spent £99M in 1990 on bed costs alone for patients with PD (Clarke et al., 1995) and every effort should be made to understand the problems of the carer and how best to support them in the care of the PD patient.

It would have been useful to have measured the depressive states of the carer to examine the association between depression and the patients clinical status or level of HRQoL. Also, in future an assessment of the HRQoL of the carer using a generic measure would be required as it becomes apparent that their well-being is impeded by the mammoth task of being a carer of a chronically ill elderly patient over a

potentially long interval. This is surely likely to affect the patients psychological and physical well-being.

Although dyskinesia is a frequent and important problem in Parkinson's disease (PD), a reliable assessment measure has not been thoroughly developed and tested. Examination of the extent of agreement of quality of life scores with measures of severity of disease or in this case symptom severity as measured by the Dyskinesia Subjective Rating Scale (DSRS), provides information on the pattern of relations of the HRQoL instrument with other established measures. The DSRS was found to correlate significantly with the *mobility* domain of the PDQ-39 at both assessments and a trend to significance was noted for the *bodily discomfort* domain at both assessments. This illustrates that the DSRS is primarily concerned with physical aspects of the disease and that there is an association between dyskinesia score and mobility whereby the severity of the dyskinesia is likely to be detected by higher scores in the mobility domain of the PDQ-39. This is the first reported assessment of construct validity of the PDQ-39 using the DSRS. Exact agreement with the remaining domains of the PDQ-39 shows that that the quality of life instrument is not redundant in the assessment of dyskinesias as it is able to detect patients perceived impairment which is not associated with a measurement of dyskinesia severity.

Sensitivity (or responsiveness) to change in a patient's condition is an essential aspect of a quality of life instrument but unfortunately, most PD quality of life instruments have, thus far, not been measured for their responsiveness to change. Martinez-Martin (Martinez-Martin, 1998) states: "Responsiveness of the PDQ-39 is awaiting future studies. Fitzpatrick and colleagues (Fitzpatrick et al., 1997), the authors of the PDQ-39, reported that there was "some evidence of responsiveness" for the PDQ-39 but none of the changes in the PQ-39 were significantly related to changes in clinical scores. The reason for the relative lack of current information regarding the sensitivities of most Parkinson's disease quality of life instruments is the difficulty in designing and analysing appropriate sensitivity studies. Such studies must be longitudinal, require reasonably large samples and observe some induced or spontaneous change in the patients' conditions (Schulzer et al., 1999). This study investigated the best indicator of change for the purpose of measuring responsiveness of the PDQ-39.

The Hoehn and Yahr staging, a measure of disease severity, did not show a noticeable change between the 6-month visit and baseline in the study population, as statistical significance or a trend to significance was not achieved. Hence an alternative measure of change was required to measure the responsiveness of the PDQ-39 in the three stratified populations. An attempt was made to identify whether the UPDRS could discriminate between PD patients with different Hoehn and Yahr scores to provide an arbitrary grading system to separate patients into those that are improving and those that are deteriorating. Correlation between the UPDRS total score and the Hoehn and Yahr were not seen and may reflect the fact that the Hoehn and Yahr only provides a crude estimation of severity and may not be truly detecting the multidimensional nature of the disease progression over what might be considered a short duration. The majority of patients in this study were stage II on the Hoehn and Yahr scale and these patients have been reported in previous studies to have duration of stage II characteristics for 6±4 years (Hoehn, 1992). It would therefore seem that to measure spontaneous change in condition in PD using a clinical measure of severity, one should consider duration of up to 6 years in such longitudinal studies. This may not always be practical and other measures of patient change should be considered. Responsiveness was measured by assessing patients who were deemed to have deteriorated according to UPDRS score which was reflected by an increase in the total UPDRS score at 6-months.

The sensitivity to change of the PDQ-39 was examined by significance of change scores between assessments and level of agreement of changes in the PDQ-39 with changes over time in the clinical rating scale, the UPDRS. Changes in individual's scores between the two assessments were significantly worse for the PDQ-39 *mobility* domain for the MFND group and *communication* for the NDMF group. None of the changes in the PDQ-39 were significantly related to changes in UPDRS scores for the remaining domains in each of the stratified groups. This is the first report of the responsiveness of the PDQ-39 *mobility* domain in motor fluctuating (MFND) cohorts but responsiveness for the same domain in the MFD group was not observed as the PDQ-39 scores remained relatively stable over the two assessments. Fitzpatrick and colleagues, 1997 explained that the lack of responsiveness with clinical scales could be due to the fact that clinical scales focus more specifically on impairment and physical function, whereas the PDQ-39 focuses on patients' perception of health related quality of life. One must therefore surmise that the PDQ-

39 and other HRQoL instruments can only be sensitive to changes that matter to patients but not to those of primary focus to the physician. However, in an analysis of the responsiveness of the UKSIP which has been reported in a previous chapter, it was observed in the group showing deterioration according to the Hoehn and Yahr that the physical dimension score was significantly different between baseline and follow-up whilst there were trends towards significance on psychosocial dimension and overall score of the UKSIP. It therefore seems that measures of HRQoL can measure change in physical impairment which is the primary focus of the physician.

It is important to distinguish the different applications of quality of life measures because instruments that have proved useful when applied in one context may be less appropriate elsewhere. This study has shown the utility of the PDQ-39 for measuring HRQoL impairment in dyskinetic, fluctuating and stable patients and highlighted the problems experienced by carers.

Summary

- Patients with motor fluctuations and dyskinesia experience greater impairment in the domains of HRQoL as measured by the PDQ-39 compared to stable patients (who experience no symptoms of motor fluctuations).

- Communications and mobility are the domains most affected by motor fluctuations and dyskinesia and this study confirms that patients with both motor fluctuating and dyskinetic symptoms have poor HRQoL of similar magnitude to those with motor fluctuations only.

- Dyskinetic patients' pharmacotherapy may be adequately controlling the clinical symptoms of PD such as tremor and rigidity but they are resulting in poor quality of life due to the abnormal voluntary movements experienced.

- Dyskinetic patients experience higher scores for the domain of emotional well-being (indicating worse HRQoL) compared to patients with no motor fluctuations or dyskinesia.
- The PDQ-39 is able to distinguish between patients with motor fluctuations and dyskinesia and stable patients confirming its discriminative properties in the distinctive cohort of PD sufferers.

- There appears to be no significant differences in HRQoL between motor fluctuating patients with dyskinesia and patients with motor fluctuations but no dyskinesia.

- This study partially confirms the responsiveness of the PDQ-39 as level of significance was reached for the domain of mobility in the motor fluctuation group with no dyskinesia and communications for the group with no dyskinesia or motor fluctuations.

- The findings of this study show no association between the PDQ-39 and the Dyskinesia Subjective Rating Scale.

- Greater carer distress was observed for the motor fluctuation and dyskinesia groups compared to the carers of the stable cohort.

- Carers of patients with both motor fluctuations and dyskinesia showed a trend to greater carer distress compared to the motor fluctuating and stable cohorts.

- Carer's greatest concerns were for the future of their relative and more that half of the carers felt they should be doing more for their relative.

- Further longitudinal research in a larger population over a much longer period of time is indicated to build on the evidence provided by the analysis in this chapter and particularly to firmly establish the responsiveness of the PDQ-39 to change over time.

CHAPTER 7

General discussion

Trials of therapeutic regimens and disease management tended to focus on the prolongation of life and the prevention or management of such life-threatening conditions as heart disease or cancer. However, there is an increasing awareness of the impact of health and health care on the quality of human life. Therapeutic efforts are increasingly being directed to the improvement of quality, not quantity of life. Measurement of quality of life can only be undertaken if one can measure what was previously thought to be immeasurable, and assess in a reproducible and valid fashion those subjective states which cannot be converted into the position of a dial or laboratory reading. The growing importance of HRQoL assessment is illustrated by the Committee for Medicinal Products for Human Use (CHMP) who have published a position paper on the regulatory guidance for the use of HRQoL measures in the evaluation of medicinal products (reference *EMEA/CHMP/EWP/139391/2004*). The CHMP document states that "The second step (once efficacy has been established) may be to show effect on HRQoL. HRQoL measures might provide insight in the interpretation of the observed effect on the primary endpoint in terms of consequences for the daily life and social functioning". This ought to signify to clinical researchers and clinicians the intent of the European organisation that is responsible for the evaluation of medicinal products to recognise the inclusion of HRQoL measures in pharmacological intervention trial.

It is therefore intriguing why there is a paucity of published HRQoL data in the literature, specifically in the field of Parkinson's disease. One reason may be that the medical community have not fully understood the value or relevance of such instrument in clinical practice or the research setting. The concept of quality of life to many medical practitioners may remain controversial. In the clinical world, scepticism is exemplified by a comment by Wulff (1999) in the *Journal of the Royal Society of Medicine*:

"Scientists may use rating scales and visual analogue scales to measure pain, and they may even invent scoring systems quantifying types of handicaps; but when they talk about measuring quality of life they have gone too far"(Wulff, 1999)

The assessment of HRQoL can only be indoctrinated if there is a clear understanding and acceptance of the relevance of the measure or instrument. The application of the instrument must be validated and have utility in the environment it is to be

applied. The science of outcomes research has developed considerably and a framework exists to validate HRQoL instruments in a manner familiar to the medical sciences. Hence, it is possible to show, that though such instruments are measuring subjective constructs, they have the relevant psychometric properties of validity, reliability and sensitivity to detect clinical change. When such properties are examined it is the first step of obtaining the acceptance of the doubters that the instrument has utility. The second step is to pinpoint which particular measure has utility or preference in the field of investigation as the researcher/practitioner is bombarded with an explosion of options with respect to the choice of instrument that can be used.

In Parkinson's disease several disease-specific HRQoL instruments have become available in the past few years. Investigators who want to use such an instrument are faced with the choice between several scales, which differ in many respects. In the process of selecting the appropriate instrument, a comparison of the quality of these scales can be helpful. Streiner and Norman, 1995 states that "the most common error committed by the clinical researchers is to dismiss existing scales too lightly, and embark on the development of a new instrument with an unjustifiably optimistic and naïve expectation that they can do better". The proliferation of possible generic or disease-specific instruments can impede research. On one hand, the ample availability of measures has helped researchers to include HRQOL as an outcome in clinical trials, but on the other, the proliferation of such measures often turns out to be a barrier for decision makers having to evaluate HRQOL across disease stages and to compare the impact of alternative therapies (Sommers and Ramsey, 1999).

As mentioned in the General Introduction chapter, literature search of clinical trials in PD highlights a paucity of HRQoL measurement as an endpoint. Measurement of health related quality of life is an important outcome measure that needs to be considered in the treatment of patients with chronic diseases, where the aim of treatment is inevitably to give patients more control of their care, a greater say in what they consider to be of relevance with respect to their health and ultimately to improve their quality of life. Including standardised HRQoL assessments in therapeutic trials will broaden our understanding of treatment efficacy. The message to all those working in the field of HRQoL is that one of the main barriers to the acceptance of quality of life has been the lack of consensus surrounding the

instruments used. In the field of PD, the PDQ-39 was introduced at the time that this research began and since then four other disease-specific instruments have been introduced, with others in development. Marinus, 2002 reviewed the literature on the existing disease-specific measures and concluded that though there are some differences in the content of each instrument "in most other respects the scales differed only marginally". To date a review or comparison of generic quality of life instruments in PD has not been done. It is strongly argued that it would have been more pragmatic to extensively assess the PDQ-39 to highlight key deficiencies with the aim of improving upon it rather than developing a totally new instrument. For example, Marinus's, 2002 review of the literature of psychometric assessment of PD HRQoL instruments states that one of the newer instruments, the Parkinson's Quality of Life Questionnaire (PLQ), has insufficiently been assessed for construct validity and responsiveness. This obsession to develop new instruments leads to the proliferation of measures which further alienates the acceptance of HRQoL measurement by clinicians and arguably the central component, the patient.

With so many generic and disease-specific instruments on offer it can be confusing for clinicians to understand the concepts behind HRQoL measurement, but this is further confounded when there is little agreement about the best methodology to adopt in doing this. Clinician's and healthcare practitioners would readily accept the inclusion of such measures and gain from the understanding of the constructs that are deemed of importance to the patient if there was consensus on which instrument is the most appropriate. It is with this in mind that this project was undertaken. The overall aims of this project were firstly to evaluate the impact of PD on the HRQoL of patients, thus providing a useful body of descriptive evidence about how this disease affects the individual. Secondly, the project was undertaken to compare the measurement properties of different instruments that all measure HRQoL and yet are different in structure, design and content. It was expected that such comparative research would permit recommendations to be made about the most appropriate instrument for use in the measurement of HRQoL in clinical practice. It has been argued that HRQoL measurement in PD has merit and provides useful and complimentary data which will help to guide therapeutic decision making and ensure optimal allocation of resources to those treatments that have been shown to produce the best quality of life outcomes. Hence a HRQoL instrument was assessed in a trial setting of patients receiving standard PD therapy to assess the impact of drug related

side effects such as dyskinesia which are thought to impact on the patients' quality of life and compliance. Parkinson's disease was chosen as a suitable disease for study for a number of reasons: it is a chronic disease where the main aim of treatment, in the absence of a cure, is to alleviate symptoms, improve daily functioning and thus improve the patients quality of life; there is growing awareness of the impact of PD on HRQoL; clinical rating scores, such as the Unified Parkinson's Disease Rating Scale (UPDRS) and the Hoehn and Yahr Scale may provide useful information on disease severity or level of physical disability, but they are not correlated with the patients perspective of his or her overall well-being (Martinez-Martin, 1998).

At the time of initiating this research there was only one disease-specific instrument in Parkinson's disease that had been published; the PDQ-39 was introduced in 1997 and hence the continued evaluation of this instrument seemed like a good starting point. Validation of the PDQ-39 had been reported by the authors (Jenkinson, 1997, Jenkinson, 1995) but nevertheless required independent confirmation. Indeed, at the time there had been no comparisons of the PDQ-39 against other instruments by independent researchers and though the disease-specific instrument had been compared against the generic SF-36 by the authors of the PDQ-39 it had not been compared against any of the other generic instruments such as the Sickness Impact Profile. The PDQ-39 has been shown to be highly reliable in terms of internal consistency and test-retest reliability. The questionnaire has construct validity in that scale scores were found to be significantly associated with those scales of the SF-36 that measure related experience, and with assessment by the clinician (Peto et al., 1998). There was also a consistent trend for poorer PDQ-39 sores associated with more severe symptoms of PD.

The literature shows that there is little information on the use of generic instruments in PD. It was important to select another generic instrument for comparison as generic instrument take a broader assessment of quality of life dimensions though they do miss disease-specific effect. Generic instruments have an important role in the measurement of HRQoL as they allow comparisons with other disease states and trials. Perhaps more importantly, some may argue that HRQoL measurement is about measuring the broad aspects of quality of life and not just dimensions considered to be clinically of interest. After all we are interested in the holistic measurement of health which is not isolate to clinically defined interests and

therefore it is proposed that generic instruments have more value to quality of life measurement compared to disease-specific instruments.

Hence the first study compared the PDQ-39 against the UK version of the SIP, referred to as the UKSIP (United Kingdom Sickness Impact Profile). Also, though the SIP has been used to assess its utility in the measurement of HRQoL in patients with PD in the USA by Longstreth (Longstreth et al., 1992) its internal consistency and reliability had not been assessed. Neither has the construct validity of the instrument been reported in PD. The results from Longstreth showed that much of the dysfunction reported by PD patients attending neurology clinics were in the areas related to psychosocial difficulties rather than physical problems. The results revealed that the SIP was able to measure the impact of PD on HRQoL, showing significantly reduced HRQoL in the PD group, compared to the control group across most domains. The SIP physical category scores were significantly correlated with PD disability scores, indicating that the SIP can reflect the degree of PD related physical disability. Many of the psychosocial and independent category scores correlated weakly with the PD disability scales, giving evidence that the SIP assesses more functional domains than the Hoehn and Yahr. Hence the reduced scores in physical and psychosocial functioning illustrates the complex manifestation of chronic disease on the sufferer demonstrating that symptoms experienced by the patient have a variety of implications that can interfere with everyday life, beyond problems such as tremor, rigidity and bradykinesia.

However, the studies reported (above) were based on different methodological or study designs in patients recruited from disparate sources which necessitated the need to test the PDQ-39 and UKSIP head to head in patients, an approach that is very rarely presented in the literature. The advantage of head to head comparisons is that it allows confidence in the findings of the study as each instrument is used in every patient allowing direct comparisons to be made. The first study evaluated the PDQ-39 and the UKSIP in PD patients based in the community and who were approached via the UK Parkinson's Disease Society. The questionnaires were administered in a postal survey at two time points with a transition question form which was completed at the second time of questionnaire completion to determine whether the patient's condition had changed since the first time point.

Both instruments exhibited good internal consistency and reliability and the PDQ-39 was shown to have construct validity in that the overall UKSIP score was significantly associated with scale scores of the PDQ-39, in particular the *Mobility* and *Activities of Daily Living* domains. The PDQ-39 was also previously shown to have construct validity with the SF-36 as reported by the developers of the disease-specific instrument. Both instruments further confirmed their construct validity in that they correlated well with the global rating scale showing that patients perceptions of their health were detected by both the disease-specific and generic instruments. The PDQ-39 had high Chronbach a > 0.9 for the domain of *Mobility* which may be suggestive of redundancy of items or loss of content validity. There may be items within the *Mobility* domain which may be surplus to requirement. Nevertheless, all scales were similar to those reported previously. An intriguing finding was that there was a high number of patients who did not answer the question "lacked support in the ways you need support from your spouse or partner?" The PDQ-39 does not allow for the omission of answers. This item of the PDQ-39 may be considered irrelevant as patients feel they receive adequate support or more likely, the patient is uncomfortable answering such a question and would prefer to avoid it for fear of upsetting their partner. This highlights the importance of ensuring that items are phrased in a way that do not prevent a true response from being given and raises the question on how reliable such self-reported questionnaires may be. There were missing answers (16%) above the arbitrary limit set for seven items that make up the *Mobility* domain of the PDQ-39. This was surprising as *Mobility* is severely impaired in PD. Closer inspection of the items that make up this domain show that the questions were intended at activities concerned with activities such as house work, DIY, cooking , and walking half a mile. The cohort of patients in this analysis had a mean age of 70.5 years old and perhaps the missing responses reflect that an elderly population are not concerned with such "vigorous" activities. Such patients are likely to have a carer, home help support and very unlikely to make the type of journey which involves walking half a mile. It is therefore argued that these questions are inappropriate to a community based cohort of elderly patients and should be revised to reflect the type of activity such a population would normally embark on. Most of the patients recruited to this study were mostly home bound, not making regular visits to a hospital neurology clinic, had a carer or partner responsible for their care and were mostly under the clinical management of the General Practitioner or PD Nurse Specialist who would make home visits. Thus, this would indicate that such patients

would not be in a position to lead even a moderately active life. Questions that reflect and are more in tune with the age range of the patient would be more appropriate.

The UKSIP *Physical* and *Psychosocial* summary scores showed good internal consistency and reliability in PD which is the first time this has been reported. Much of the dysfunction reported by PD patients was related to non-physical difficulties rather than physical problems which are important to appreciate as clinical measures are primarily intended to assess physical impairment. The UKSIP would complement the clinical instrument used in the assessment of PD and would definitely be beneficial in the clinical trial setting to differentiate between two treatment options with similar efficacy. However, the *Work* section of the UKSIP was found to be redundant as very few patients completed this section either because respondents considered this section irrelevant as they were retired or not working. The UKSIP does ask respondents about their work status and a quarter of the patients reported that they were retired due to their health, information which is not captured by the PDQ-39. Also, the UKSIP revealed that more than half of the PD respondents reported that their sexual activity was reduced and this is a pertinent dimension of HRQoL which has been reported to have a negative impact on patients' general satisfaction from life. This is not an item which is included in the PDQ-39 instrument and should be considered in possible future revisions of the PDQ-39. The UKSIP was found to address another aspect of PD which is considered to be important in relation to HRQoL, that is sleep and fatigue, which is also not addressed adequately in the PDQ-39. Nearly half of the respondents completing the UKSIP reported that they "sleep less at night" or "sleep more during the day" and this may reflect the effects of some anti-parkinsonian medication which are thought to cause somnolence. These are examples of the advantage of a generic instrument in that they address broader aspects of HRQoL constructs and in this case sleep problems would not have been identified adequately by the PDQ-39. The UKSIP also identified that patients "sit for much of the day" and this aspect of fatigue is not addressed in the PDQ-39. Fatigue is known to be a risk factor for poor HRQoL in PD. The UKSIP identifies non-physical dimensions of health as having the greatest impairment in HRQoL in PD patients whereas the PDQ-39 identified physical dimensions as having the greatest impairment in HRQoL in PD patients. This is not surprising as the PDQ-39 focuses on the physical aspects of PD whereas the UKSIP

addresses in greater detail or depth the psychosocial or independent domains of HRQoL.

This study highlighted that the UKSIP is relevant to the measurement of quality of life in PD as it exhibited good internal consistency, test-retest reliability and construct validity. The PDQ-39 also, as previously reported in the literature, exhibits good psychometric properties. This analysis reveals important differences between the two instruments which suggest that the generic measure asks questions in the broader context and identifies problems of HRQoL which are not assessed by the PDQ-39. Head to head comparison of these instruments suggests that the UKSIP should be used along with the PDQ-39 to assess and monitor patients HRQoL, in particular non-physical dimensions of PD which are not adequately exposed by the disease-specific instrument. This does not mean that the PDQ-39 is not suitable for HRQoL assessment in PD as it does indeed highlight impairment in nearly all domains as disease severity increases, as reported by the developers of the PDQ-39. In spite of this the PDQ-39 is very much geared towards impairment of a physical nature and some of the questions or items included should be revised to take account of the age or lifestyle of an elderly, sedentary population suffering from a chronic illness. The HRQoL instrument should not, in my opinion, only reflect the disability or physical impairment that is assessed by the clinical measure of disease severity or disability but should also provide additional and perhaps unexpected information that adds to the value of the assessment. The PDQ-39 identified a range of problems experienced by the patient with PD which are also highlighted by the UKSIP, in the areas of home care, leisure and access to public areas.

The UKSIP would seem to fit the requirements for use in the assessment of HRQoL in PD in the current setting as it addresses areas of concern which are relevant to PD, identifies items not addressed adequately by the PDQ-39, has comparable psychometric properties when compared to the PDQ-39, and focuses more on the psychosocial domains of HRQoL which are likely to complement the standard clinical assessments.

The UKSIP was shown to have utility in assessing the impact of PD on HRQoL; however, it was necessary to compare the instrument against the SF-36 to determine which of the two most commonly used generic instruments would be suitable for the

measurement of HRQoL in PD and in the clinical (trial) setting. The relevance of using the SF-36 as the comparison is that it is a generic measure that has undergone a considerable amount of testing for reliability and validity. Summary and Mental Component Summary can be derived from the SF-36 and have been validated for use in a number of chronically ill populations. Also, the SF-36 has fewer items that the UKSIP and therefore it would be interesting to see which instrument captures the impact PD on HRQoL more effectively. The UKSIP and the SF-36 were administered to patient attending an out-patient neurology clinic and respondents were asked to complete these instruments at the time of their first scheduled visit and then again at the time of their next scheduled clinic appointment. This study assessed the measurement properties of the two generic instruments in out patient setting and specifically assessed convergent and discriminative validity, internal consistency and sensitivity to change. As these were generic instruments it was feasible to compare HRQoL scores against population norms. There has not been a direct comparison of these two instruments in the PD population and also the sensitivity to change has not been previously been reported. Clinical assessment was conducted by a neurologist using the Hoehn and Yahr scale as a measurement of disease severity. The purpose of this second study was to arrive at a consensus as to which of these instruments should be recommended as the generic instrument of choice for use in discriminative longitudinal PD studies.

The results showed that, as expected, both instruments had construct validity when compared to clinical indices and both instruments were able to discriminate between patients according to disease severity. Such a finding is likely to provide confidence to the clinician or clinical researcher that the generic instruments will be able to distinguish between populations whose disease is not responding to treatment and those whose symptoms are improving. Comparisons against population norms revealed significant differences in that both generic instruments were able to detect impairment in domains of relevance to PD patients. The greatest impairment was in physical dimensions of HRQoL when compared to data from the general population and hence there is evidence that though generic measures are concerned with broader aspects of HRQoL they can detect the physical impact of this chronic disease which provides assurance that such instruments are not necessarily skewed to detecting impairment of psychosocial domains at the expense of physical constructs of PD which are equally important. We have already noted that the UKSIP

is better at detecting non-physical impairment compared to the disease-specific instrument, the PDQ-39, but we now present data that shows that it can also detect and differentiate the impact of physical impairment on patients' HRQoL compared to a healthy population. Key differences between the two generic measures were based on the instruments item content which reveals that the UKSIP addresses more relevant questions for this population. The UKSIP detected impairment in cognitive symptoms, an item which is not addressed in the SF-36. When the authors of the PDQ-39 compared their instrument against the SF-36 they reported that the use of a disease-specific measure was justified on the basis that the PDQ-39 was able to determine that worsening PD was associated with greater disability in the areas of cognition and communication. In this research we report that the UKSIP was also able to detect impairment in cognitive function, which has a major impact on the patients well being and can lead to further deterioration of health and quality of life if not detected. Another major omission that was not detected in the comparison of the SF-36 against the PDQ-39 is that the SF-36 omits items referencing sleep disturbances which again is most relevant to PD patients. The body of evidence is showing a preference for the use of the UKSIP over the SF-36 in PD and this conclusion is further strengthened when addressing the types of questions asked in the SF-36 which would suggest it is not suitable for use in an elderly population. The SF-36 has items that refer to vigorous activities and time spent at work which raise the question of what bearing or relevance such items have on the elderly PD population. The key finding in this study was evidence supporting the responsiveness of the UKSIP over time particularly for the domains of *Recreation and Pastimes* and *Psychosocial* dimension in the group defined as deteriorating. The UKSIP also showed a trend towards significance for the other domains compared to the SF-36 which did not. The direction towards significance was also more notable for the UKSIP domains compared to the SF-36 and if the sample size had been larger we may have observed more statistically significant results. There was also evidence that the UKSIP is responsive to change in patients' clinical status over time in that the UKSIP *Physical* and *Psychosocial* dimensions exhibited a corresponding change in score with change in severity of symptoms measured by the Hoehn and Yahr. The PDQ-39 was not able to show sensitivity to change with the same clinical measure as reported by the developers of the PDQ-39. This is the first evidence of sensitivity of the UKSIP in PD and makes the instrument an important addition to outcome measures in clinical trials.

Overall, the findings from this research suggest that the UKSIP is the instrument of choice for use in PD and in fact it exhibits a number of key beneficial differences compared to the disease-specific PDQ-39. The PDQ-39 has shown to be valid and able to detect impairment of HRQoL that was not identified by the SF-36 therefore it would make sense to evaluate whether there might be a more relevant disease-specific instruments to compare against the PDQ-39 which is, if the literature is to be believed, becoming the instrument of choice in PD research.

The most recent disease-specific instrument in PD to be introduced was the PIMS which was developed to use with patients in a clinical setting. The items selected were identified from PD nurse specialists who were interviewed rather than directly from PD patients and neurologists. It was necessary to evaluate the PIMS in comparison to the PDQ-39 to ascertain whether this new disease-specific HRQoL instrument added to the knowledge already in existence and to determine which disease-specific instrument was ideally suited for the assessment of HRQoL in PD. The development of a new HRQoL instrument should not be taken lightly and it is necessary to subject such a new measure to extensive testing to assess what attributes it contains over existing instruments. It is believed that existing measures can be improved upon, made more applicable, and valid with further testing and one should avoid introducing a new measure unless it fulfils a need that is not currently met. Construct validity and specifically, convergent and discriminant validity of the PIMS had not been reported. One of the perceived advantages of the PIMS is that it is made up of 10 items which are broadly formulated and concern domains rather than specific situations. There can be a trade off regarding the length of an instrument between its measurement and practical properties(Ware and Sherbourne, 1992). The 10 items can be separated into four distinct factors; physical, social, psychological and economic. Also, the PIMS allows patients with fluctuating symptoms characterised by dyskinesia to score when they are in a best state and also to provide a score to assess the impact of their symptoms when they are in a worse state, defined as being "off" which is characterised by the wearing off of the medication.

Patients who were attending an out-patients PD clinic at a regional hospital were eligible for inclusion. Patients completed both the PDQ-39 and PIMS at baseline, along with the Beck Depression Inventory and the investigator, a neurologist or PD specialist nurse completed a clinical Hoehn & Yahr and UPDRS assessment. The

same assessments were completed at subsequent visits, usually three months and again six months later.

The PIMS was not well comprehended by the PD patients and this was characterised by the relatively high rate of missing data and a poor understanding of whether they were to complete the columns assigned for patients classified as fluctuating or else the column assigned for stable patients. This was in all likelihood related to the lay out of the questionnaire and a poor understanding of the instructions on how to completer the questionnaire. It would probably have been more beneficial to have obtained answers to each question by interview which would then negate the advantage of a shorter questionnaire. Also patients with fluctuating disease completed two columns, one for how they felt when the medication was satisfactorily controlling their symptoms ("on" state) and another column for when they were not adequately controlled ("off" state). There is then a propensity for a patient to score based on their perception of their condition during a particular state rather than a true assessment reflecting the actual state on the day of completing the questionnaire. One can therefore not be assured of a true response to each item as the answers provided are very much likely to be based on the perception of well being in a particular state, memory of what impact their condition really had on the individual at the point in time and how they really coped at such a time. It was found that patients identified as fluctuating answered much fewer questions on the PIMS compared to the stable patients who completed the PIMS and this is not what one should expect from such a short instrument.

Nevertheless, the PIMS does attempt to measure HRQoL of PD patients depending on their clinical state i.e. fluctuating or non-fluctuating and this is admirable as much has been reported on the impairment of well-being and poor HRQoL in PD patients but very little on the impact of fluctuating disease on HRQoL. The PDQ-39 on the other hand does not attempt to score HRQoL based on the clinical state of the PD patient and scores are obtained irrespective of the clinical state at the time of completing the questionnaire. The PDQ-39 could be further improved by adding an item to enquire about the patients exact clinical state (stable or fluctuating) at the time of completion so that the reviewer can judge whether HRQoL impairment is a consequence of the disease and its symptoms or else a consequence of the poor control of symptoms and ineffective medications. Such a mechanism would reveal

whether physical impairment was greater in the fluctuating population compared to the non-fluctuating group and treatment strategies could be adopted to compensate for the deterioration of health such as the introduction of a dopamine agonist which is thought to cause less symptom fluctuations compared to the standard levodopa treatment. The PIMS reveals useful information in this respect in that it highlights greater impairment in fluctuating patients.

Convergent validity between the PIMS and PDQ-39 was observed but the notable exception was an absence of correlation between the PIMS when completed by the fluctuating cohort and the PDQ-39 completed by the complete cohort. This key finding indicates that fluctuating patients report different perceptions of well being when assessed by the PIMS and that these perceptions are not in this scenario detected by the PDQ-39. This particular finding led to the next study which was the evaluation of the PDQ-39 in patients with fluctuating disease and is discussed later. Also, the PIMS exhibited discriminant validity in being able to differentiate between stable and fluctuating patients and showed a trend to being able to differentiate between different clinical states as measured by the Hoehn and Yahr and UPDRS. Once again this trend was for the stable cohort completing the PIMS and the association was greatly reduced for the fluctuating cohort completing the PIMS. However, the strength of correlations was much more significant for the PDQ-39 which was clearly able to discriminate between the various stages of the Hoehn and Yahr and UPDRS. Also the PDQ-39 was able to discriminate based on severity of depression, showing a consistent and significant pattern of elevated scores for domains which was not reproduced for the PIMS.

Overall the psychometric properties and performance of the PDQ-39 were better than the PIMS which was not easily comprehended by the respondents. The PDQ-39 was found to be an applicable measure of PD HRQoL and is the most appropriate disease-specific instrument to use in this group of patients.

The comparison of the two disease-specific instruments raised important questions: is there a difference in the HRQoL in PD patients between fluctuating disease and non-fluctuating disease; is the PDQ-39 sufficiently sensitive enough to assess the impact that these different states have on HRQoL; further, is the PDQ-39 able to discriminate between those who have fluctuating disease and no motor complications

and those with fluctuating disease and motor complications. An attribute of motor complications in this case would be the presence of dyskinesia. As the research to date has shown that the PDQ-39 has some of the attributes required of a HRQoL instrument such as reliability, and construct validity and it experienced much less missing item completion compared to the PIMS. Nevertheless the research had not evaluated the application of a PDQ-39 in a large trial setting to investigate these questions and responsiveness of the instrument in such a discrete PD population. One of the limitations of the previous studies reported had been the moderate sample sizes which make the assessment of responsiveness more difficult. Hence the final study in the thesis was undertaken to determine the difference in levels of HRQoL between dyskinetic PD patients versus non-dyskinetic patients and to assess the responsiveness of the PDQ-39 in such a cohort. Also, there was the opportunity to assess caregiver burden to determine the impact of dyskinesia in carer sample; this is an area of HRQoL that has not been extensively addressed before in PD.

A longitudinal study of six months duration was conducted recruiting one hundred and forty five patients from four hospital based neurology clinics. At baseline a UPDRS assessment and Hoehn and Yahr scale were completed by a neurologist. Patients completed the PDQ-39 and the Dyskinesia Subjective Rating Scale and the patients' carer completed the Care-giver Burden Interview questionnaire. These assessments and questionnaires were repeated at the follow-up visit six months later. All patients were receiving standard dopaminergic therapy mostly consisting of levodopa and/or a dopamine agonist. The patients were stratified into three groups which were those with no fluctuations; those with motor fluctuations but no dyskinesia; and those with motor fluctuations with dyskinesia.

This study revealed that dyskinetic patients experienced reduced HRQoL in most domains of the PDQ-39, and in particular for *communication* and *mobility* in comparison to the group who did not have dyskinesia or motor fluctuations. This is one of very limited studies to report that HRQoL is reduced when dyskinesia occurs. A closer look at the data revealed that patients with motor fluctuations and dyskinesia have much higher levels of impairment for the domain of *emotional well-being* compared to the group with no fluctuations. What is intriguing in this finding is that over the six months of treatment and follow-up no body would have been aware of the impact that medication was having without having implemented the HRQoL

instrument to the assessment. The PDQ-39 findings would suggest that the existing treatment approaches are not adequately alleviating the impact of the disease and that new PD treatments ought to be introduced or else treatments to alleviate the symptoms of depression or anxiety ought to be considered. The PDQ-39 was able to discriminate between fluctuating and non-fluctuating patients and hence has utility in this cohort of PD patients. It was not able to differentiate between fluctuating patients with dyskinesia and fluctuating patients without dyskinesia but this is probably because the cause of these two events are the same, i.e. medication, and patients do not consider one condition to have a greater impact on their well being compared to the other condition. The ongoing construct validity assessment of the PDQ-39 showed that the *mobility* and to a lesser extent the *bodily discomfort* domains correlated with the Dyskinesia Subjective Rating Scale but the PDQ-39 is also measuring the impact of the disease in the other domains which bear no relation to the severity of dyskinesia highlighting the merits of adding the PDQ-39 to the assessment of such populations.

The sensitivity to change of the PDQ-39 was also evaluated in the final chapter and revealed the first sign of responsiveness of the PDQ-39 *mobility* domain in motor fluctuating patients. In other words when there is a change in the clinical UPDRS score then this will be reflected by a change in the PDQ-39 *mobility* score. It would be realistic to state that there was some evidence of responsiveness and the changes reported were significantly related to changes in clinical scores, something the developers of the PDQ-39 were unable to demonstrate.

Carer burden in the PD population is often ignored which is surprising as family members caring for anyone with a chronic disease in an elderly population can expect to face, physical, psychological and social problems. The demand on such carers can be relentless. The results from this study show that greater carer distress is experience by the carers of the patients with motor fluctuations compared to the group with no motor fluctuations. Such a finding should challenge the medical communities approach to treating PD patients, particularly those with motor fluctuations. It is not sufficient to treat the symptoms only but also necessary to address the impact that medication or management of the disease has on HRQoL of the patient and also the carer of the patient. In essence to treat the totality of the

disease one must also recognise the impact the disease has on the person responsible for the daily care of the elderly patient.

Assessing the impact of PD related motor fluctuations and dyskinesia on HRQoL was certainly warranted as it has been inadequately investigated so far. However, before such investigations can be embarked upon it is necessary to validate the use of the HRQoL instrument to be used in such a population. This study has shown that the PDQ-39 is a suitable measure to use in the assessment of motor fluctuations in that it has exhibited construct and discriminative validity as well as partial responsiveness in the population under investigation. Additionally, the study has highlighted the level of impairment experienced by carers of patients with motor fluctuating PD.

One of the crucial findings of this project was that the UKSIP is an applicable measure of PD related HRQoL; it exhibited good construct validity as demonstrated by domain correlations with clinical measures such as the Hoehn and Yahr and discriminative validity in that it was able to differentiate based on disease severity. Also the UKSIP addressed the broad range of issues relevant to the PD population and exhibited responsiveness to change. The UKSIP identified issues such as sexual and sleep disturbances which are either not addressed or inadequately addressed in the PDQ-39. It is the most appropriate generic instrument to use in this group of patients and can be used to complement the PDQ-39 or when one would wish to limit the burden of questionnaires to complete, could arguably be considered as an appropriate alternative to the disease-specific PDQ-39. Though it is a lengthier questionnaire it was easy to comprehend (compared to the much shorter PIMS) and can quite easily be used in the clinical trial setting. Additionally, as it is a generic measure it can be used to compare HRQoL scores with other diseases.

This project also determined that the PDQ-39 has the required psychometric properties expected of a new instrument and that its performance against the newly developed PIMS was far superior. However, the comparison did reveal that the PDQ-39 would benefit with the addition of an item to determine whether the respondent was a stable or motor fluctuating patient to allow the clinician to make better sense of the domain scores. Nevertheless the PDQ-39 exhibited partial responsiveness in the motor fluctuating cohort and was sensitive enough to detect differences between fluctuating and non-fluctuating patients.

An additional finding of this project was the impact that the disease has on patient's carer and close family members. In particular, the impact of dyskinesia greatly impacts on the carer's distress and well-being compared to the carers of non-fluctuating patients. The revelation that drug related effects which are known to cause fluctuation of symptoms and dyskinesia can greatly impede the well-being of the carer who plays a central role in the care and management of the patient - this is crucial to address. Medications which reduce the incidence of fluctuations can be administered or greater care of the caregiver introduced to cope with these drug induced events. This is an exciting area of research with requires further investigation.

Furthermore, this project has consistently demonstrated that PD is a chronic disease that causes reduced HRQoL. Symptoms cause a variety of problems to the individual including reduced physical functioning and social limitations that can result in emotional problems such as depression, and anxiety. It is imperative that clinicians, nurses and trialists have a greater understanding of how PD affects patients' everyday lives and that there is more to patient management in the elderly than controlling disease symptoms. The goal of treatment should be multidisciplinary and aim to address the totality of the disease and this included the impact of the disease on health related quality of life. HRQoL should continue to play a key role in the assessment of patient well-being in the clinically dominated approach to treatment and care; HRQoL is without a doubt of paramount importance and through this research project we have highlighted the impact of PD on HRQoL and the caregiver, and more importantly arrived at recommendations on the suitability or applicability of a range of disease-specific and generic HRQoL instruments with the aim of making the choice of the clinician to be less convoluted.

If anything it is hoped that this research project will add to the body of knowledge in existence, make the choice of instrument more straight forward and abate the doubters on the benefits of HRQoL assessment. Progress can be claimed if we can see an end to comments such as by Wulff:

'Scientists may use rating scales and visual analogue scales to measure pain, and they may even invent scoring systems quantifying types of handicaps; but when they talk about measuring quality of life they have gone too far'(Wulff, 1999).

Limitations of the study

The study has limitations which may have resulted in selection bias:

- The study presented in chapter three, certain PDS branches, for reasons of confidentiality, preferred to send questionnaires to patients via the welfare officer. The welfare officer selected potential patients to participate based on the patient's ability to complete the questionnaire (could patient write, dementia, availability of a carer, and general condition of the patient). This was repeated at a number of other branches and means that the study could not ensure a true representation of the PD population. The absence of a carer usually precluded the patient's entry to the study as it was felt that a carer was necessary to help the subject complete the questionnaire where the subject had difficulty with writing, or some form of cognitive impairment.

- In the chapter three study some patients from various branches refused to participate because of other medical complications such as rheumatoid arthritis and it is feasible that the study was biased to patients presenting with milder morbidity than would be expected.

- It was not felt appropriate for patients recently diagnosed with PD to complete the HRQoL questionnaires, as it would highlight the types of problems these patients may experience in time. This was likely to ensure that patients in the study were familiar with the consequences of their disease and as such have modified their expectations and handling strategies to the condition.

- Some sites did not or refused to distribute "non-participation" questionnaire to collect information on non-responders. The reason for this varied from privacy to inability to send out these questionnaires.
- Some patients were unwilling to participate in the studies as they were reluctant to complete questionnaires because they were afraid to reveal too much private information even though confidentiality was assured.

- Some patients were lost to follow-up for a variety of reasons including death, relocation to a different healthcare region, or admission to a care or nursing home.

These patients were excluded from the analysis due to missing questionnaires and contributed to the small sample size.

- Recruitment of patients was hindered because some patients were keen to participate in treatment trials in the hope of gaining benefit from such participation. Also, the practicality of recruitment from single centre studies affected recruitment rates and multi-centre studies would in future be recommended.

- Clinical assessments were missing for some of the patients due to time constraints on the clinician and this contributed to the small sample size in the responsive to change assessments

- The numbers of patients with severe disease were limited in this study as such patients do not routinely attend out-patient clinics and are mostly found in care homes. Also, where patients with severe disease were available the clinician did not consider them suitable for participation in such a study.

- Some patients did not attend their follow-up visit as scheduled and therefore some follow-up assessments were not performed as outlined in the protocol.

FUTURE RECOMMENDATIONS

As the measurement of health-related quality of life is a relatively young science, there are still many issues that need to be resolved. Ideas and suggestions for future research directions as a result of this project are given below:

1. Further assessment in Parkinson's disease patients of the reproducibility and reliability of the UKSIP is required before there can be complete confidence in the results produced using this questionnaire in a longitudinal clinical study. In particular a longitudinal study in a larger sample size would be advocated as would a follow-up period of at least 12 months to ensure that there was likelihood of a change in disease severity.

2. The Work item of the UKSIP was found to be redundant because most patients had retired and therefore this section should be revised to better reflect the elderly PD population. The section should however retain the items asking whether the patient has retired due to their health.

3. In order to interpret the results more easily, it would be beneficial to determine the minimum change in score of a given UKSIP domain that is important to the PD patient. It would make sense to initially determine important changes for the Physical summary score or the domains which make up the Physical summary score, given that physical-related area of HRQoL are more liable to show therapy related changes in score in diseases such a PD which can be measured by established clinical instruments such as the UPDRS.

4. Further assessment in Parkinson's disease patients of the responsiveness and validity of the PDQ-39 is required once further modifications are implemented such as the addition of items which were either inadequately addressed or absent from the PDQ-39. Also in a longitudinal study a follow-up period of at least 12 months to ensure that there was likelihood of a change in disease severity is recommended. If spontaneous change in this period is not viable it would be necessary to assess the responsiveness of the PDQ-39 in a clinical trial setting of two treatment approaches.

5. The HRQoL of PD caregivers and PD patients should be considered using the UKSIP to ensure comparisons could be made. Whist we are now aware that caregivers experience considerable distress the impact of caring for a PD patient on their quality of life has not been fully assessed. Such a study should also include the measurement of depression of the caregiver.

This study found that the UKSIP is an applicable measure of Parkinson's disease related HRQoL and is the most appropriate generic instrument to use in this group of patients. It has both cross-sectional and longitudinal measurement properties that are comparable with disease-specific indices assessed and demonstrated good construct validity as demonstrated by domain correlations with clinical measures such as the Hoehn and Yahr. The UKSIP was also shown to assesses items of relevance to PD patients which were found not to be adequately addressed in the disease specific PDQ-39. Together with these results and an indication of responsiveness, the UKSIP will allow the broader effects of treatment on general quality of life to be determined. The PDQ-39 was also shown to be a valid and reliable disease specific instrument that was able to differentiate between fluctuating and stable PD patients. The PDQ-39 was able to differentiate the differences in these cohorts of patients and show that certain domains are more impacted than others in patients with fluctuating disease.

This project has gone some way in improving how we measure PD patient's quality of life so that the impact of treatment on patients can be more accurately assessed. Quality of life has a considerable impact on the ability to lead as 'normal' a life as possible and hence it is necessary to continue to improve the efficiency and accuracy of how we assess it, particularly in an age on unprecedented discoveries in the field of medicine which offer new treatment options to the patient. Through this research project, recommendations have been made about the most appropriate and optimal instruments to use in the assessment of HRQoL in Parkinson's disease.

REFERENCES

(2000). Pramipexole vs levodopa as initial treatment for Parkinson disease: A randomized controlled trial. Parkinson Study Group. Journal of the American Medical Association, 284, 1931-8.

Aarsland, D., Tandberg, E., Larsen, J. P. and Cummings, J. L. (1996). Frequency of dementia in Parkinson disease. Archives of Neurology, 53, 538-42.

Ali T and Salek S (2002). Comparison of the PDQ-39 with a Generic Measure of Health Status (UK Sickness Impact Profile) in Parkinson's Disease Subjects. Quality of Life Research, 11, 672.

Ali T and Salek S (2002). The Reliability of the PDQ-39 and a Generic Measure of Health Status (UK Sickness Impact Profile) in Parkinson's Disease. Quality of Life Research, 11, 671.

Anastasi A (1990) Psychological Testing, MacMillan Publishing Company, New York.

Anderson JS and Ferrans CE (1997). The quality of life of persons with chronic fatigue syndrome. Journal of Nervous and Mental Disease., 185, 359-367.

Andresen E M and Meyers A R (2000). Health-related quality of life outcomes measures. Archives of Physical Medicine and Rehabilitation.81.(12.Suppl 2):S30.-45.

Anthony-Bergstone, C. R., Zarit, S. H. and Gatz, M. (1988). Symptoms of psychological distress among caregivers of dementia patients. Psychology and Aging, 3, 245-8.

Arle J. E and Alterman R. L (1999). Surgical options in Parkinson's disease. Medical Clinics of North America, 83, 483-98, vii.

Baas H, Beiske A G, Ghika J, Jackson M, Oertel W H, Poewe W and Ransmayr G (1997). Catechol-O-methyltransferase inhibition with tolcapone reduces the "wearing

off" phenomenon and levodopa requirements in fluctuating parkinsonian patients. Journal of Neurology, Neurosurgery and Psychiatry, 63, 421-8.

Barone P, Bravi D, Bermejo-Pareja F and Marconi R (1999). Pergolide monotherapy in the treatment of early PD. Neurology, 53, 573-579.

Beck A T, Ward C H, Mendelson A, Mock J and Erbaugh J (1961). An Inventory for Measuring Depression. Archives of General Psychiatry, 4, 53-63.

Bergner M (1989). Quality of life, health status, and clinical research. Medical Care, 27, S148-56.

Bergner M, Bobbitt R A, Carter W B and Gilson B S (1981). The Sickness Impact Profile: development and final revision of a health status measure. Medical Care, 19, 787-805.

Berry R A and Murphy J F (1995). Well-being of caregivers of spouses with Parkinson's disease. Clinical Nursing Research, 4, 373-86.

Bland J M and Altman D G. (1997). Cronbach's alpha. British Medical Journal, 314, 572.

Block G, Liss C, Reines S, Irr J and Nibbelink D (1997). Comparison of immediate-release and controlled release carbidopa/levodopa in Parkinson's disease. A multicenter 5-year study. The CR First Study Group. European Neurology, 37, 23-7.

Bombardier, C., Ware, J., Russell, I. J., Larson, M., Chalmers, A. and Read, J. L. (1986). Auranofin therapy and quality of life in patients with rheumatoid arthritis. Results of a multicenter trial. The American Journal of Medicine, 81, 565-78.

Bowling A (1997) Measuring Health: A Review of Quality of Life Measurement Scales, Open University Press, Buckingham, England.

Brazier J.E, Harper R and Jone NM (1992). Validating the SF-36 health survey questionnaire: new outcome measure for primary care. British Medical Journal, 305.

Brook R H, Ware J E Jr, Rogers W H, Keeler E B, Davies A R, Donald C A, Goldberg G A, Lohr K N, Masthay P C and Newhouse J P (1983). Does free care improve adults' health? Results from a randomized controlled trial. New England Journal of Medicine, 309, 1426-34.

Brown B W and Hollander M (1977) Statistics: A Biomedical Introduction, Wiley, New York.

Bryman A and Cramer D (1990) Quantitive Data Analysis for Social Scientists, Routledge, London.

Caap-Ahlgren, M. and Dehlin, O. (2002). Factors of importance to the caregiver burden experienced by family caregivers of Parkinson's disease patients. Aging Clinical and Experimental Research, 14, 371-7.

Calman K C (1984). Quality of life in cancer patients--a hypothesis. Journal of Medical Ethics, 10, 124-7.

Calne S, Schulzer M, Mak E, Guyette C, Rohs G, Hatchard S, Murphy D, Hodder J and Gagnon S (1996). Validating a Quality of Life Rating Scale for Idiopathic Parkinsonism: Parkinson's Impact Scale (PIMS). Parkinsonism and Related Disorders, 2, 55-61.

Chapuis S, Ouchchane L, Metz O, Gerbaud L and Durif F (2005). Impact of the motor complications of Parkinson's disease on the quality of life. Movement Disorders. 20(2):224-30

Chrischilles E A, Rubenstein L M, Voelker M D, Wallace R B and Rodnitzky R L (1998). The health burdens of Parkinson's disease. Movement Disorders, 13, 406-413.

Clarke C E, Zobkiw R M and Gullaksen E (1995). Quality of life and care in Parkinson's disease. British Journal of Clinical Practice, 49, 288-293.

Cohen J (1977) Statistical power analysis for the behavioural sciences, Academic Press, New York.

Col N, Fanale J E and Kronholm P (1990). The role of medication noncompliance and adverse drug reactions in hospitalizations of the elderly. Archives of Internal Medicine, 150, 841-5.

Colcher A and Simuni T (1999). Clinical manifestations of Parkinson's disease. Medical Clinics of North America, 83, 327-47.

Coons S J and Kaplan R M (1992). Assessing health-related quality of life: application to drug therapy. Clinical Therapeutics. 14, 850-858.

Coons S J and Kaplan R M (1993). Quality of life assessment: understanding its use as an outcome measure. Hospital Formulary, 28, 486-90, 492, 497-8.

Crocker L M and Algina J (1986). Introduction to Classical and Modern Test Theory. In: Holt, Rinehart and Winston.

Cronbach L J (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297-334.

Cronbach L J (1955). Construct validity in psychological tests. Psycological Bulletin, 52, 281-302.

Cummings, J. L. (1992). Depression and Parkinson's disease: a review. American Journal of Psychiatry, 149, 443-54.

Damiano A M, Snyder C, Strausser B and Willian M K (1999). A review of health-related quality-of-life concepts and measures for Parkinson's disease. Quality of Life Research, 8, 235-243.

de Boer A G, Wijker W, Speelman J D and de Haes J C (1996). Quality of life in patients with Parkinson's disease: development of a questionnaire. Journal of Neurology, Neurosurgery and Psychiatry, 61, 70-74.

Dewey Richard B (2000) Parkinson's Disease and movement Disorders: Diagnosis and Treatment Guidelines for the Practicing Physician, Humana Press, Totowa, New Jersey.

Deyo R A, Inui T S, Leininger J and Overman S (1982). Physical and psychosocia function in rheumatoid arthritis. Clinical use of a self-administered health status instrument. Archives of Internal Medicine, 879-882.

Deyo R A and Patrick D L (1989). Barriers to the use of health status measures in clinical investigation, patient care, and policy research. Medical Care, 27, S254-68.

Dodel R C, Berger K and Oertel W H (2001). Health-related quality of life and healthcare utilisation in patients with Parkinson's disease: impact of motor fluctuations and dyskinesias. Pharmacoeconomics, 19, 1013-38.

Dodel R C, Singer M, Kohne-Volland R, Szucs T, Rathay B, Scholz E and Oertel W H (1998). The economic impact of Parkinson's disease. An estimation based on a 3-month prospective analysis. Pharmacoeconomics., 14, 299-312.

Dogu Okan, Sevim Serhan and Metin Ozmen (2002). Sertraline in Parkinson's Disease. Bulletin of Clinical Psychopharmacology, 12, 126-129.

Dorman, P., Slattery, J., Farrell, B., Dennis, M. and Sandercock, P. (1998). Qualitative comparison of the reliability of health status assessments with the EuroQol and SF-36 questionnaires after stroke. United Kingdom Collaborators in the International Stroke Trial. Stroke, 29, 63-68.

Dubois B and Pillon B (1997). Cognitive deficits in Parkinson's disease. Journal of Neurology, 244, 2-8.

Durif, F., Devaux, I., Pere, J. J., Delumeau, J. C. and Bourdeix, I. (2001). Efficacy and tolerability of entacapone as adjunctive therapy to levodopa in patients with Parkinson's disease and end-of-dose deterioration in daily medical practice: an open, multicenter study. European Neurology, 45, 111-8.

Ebrahim, S., Barer, D. and Nouri, F. (1986). Use of the Nottingham Health Profile with patients after a stroke. Journal of Epidemiology Community Health, 40, 166-9.

Factor S A, McClarney T and Sanchez-Ramos JR (1990). Sleep disorders and sleep effect in Parkinson's disease. Movement Disorders, 4, 280-285.

Faden, R. and Leplege, A. (1992). Assessing quality of life. Moral implications for clinical practice. Medical Care, 30, MS166-75.

Fahn S, Elton R L and And the UPDRS Development Committee (1987) In Recent Developments in Parkinson's Disease, Vol. 2 (Eds, Fahn S, Marsden C D, Calne D B and Goldstein M) Macmillan Healthcare Information, New Jersey, USA, pp. 153-163, 293-304.

Fairclough, D. L. (2002) Design and Analysis of Quality of Life Studies in Clinical Trials, Chapman and Hall/CRC Press, Boca Raton, FL.

Fairclough, D. L., Peterson, H. F., Cella, D. and Bonomi, P. (1998). Comparison of several model-based methods for analysing incomplete quality of life data in cancer clinical trials. Statistics in Medicine, 17, 781-96.

Fairclough, D. L., Peterson, H. F. and Chang, V. (1998). Why are missing quality of life data a problem in clinical trials of cancer therapy? Statistics in Medicine, 17, 667-77.

Felton B J and Revenson T A (1984). Coping with chronic illness: a study of illness controllability and the influence of coping strategies on psychological adjustment. Journal of Consulting and Clinical Psychology, 52.

Findley L, (on behalf of the Global Parkinson's Disease Steering Committee). (2002). Factors impacting on quality of life in Parkinson's disease: results from an international survey. Movement Disorders;17(1):60.-7.

Finlay A Y, Khan G K, Luscombe D K and Salek S S (1990). Validation of Sickness Impact Profile and Psoriasis Disability Index in psoriasis. British Journal of Dermatology, 123, 751-756.

Fitzpatrick R and Alonso J (1999) In Quality of Life in Parkinsons disease(Eds, Martinez Martin P and Koller WC) Masson SA, Barcelona, Spain, pp. 1-15.

Fitzpatrick R, Fletcher, A., Gore, S., Jones, D., Spiegelhalter, D. and Cox, D. (1992). Quality of life measures in health care. I: Applications and issues in assessment. British Medical Journal, 305, 1074-7.

Fitzpatrick R, Peto V, Jenkinson C, Greenhall R and Hyman N. (1997). Health Related Quality of Life in Parkinson's Disease: A Study of Outpatient Clinic Attenders. Movement Disorders, 12, 916-922.

Fitzpatrick R, Ziebland S, Jenkinson C, Mowat A. (1993). Transition Questions to Assess Outcomes in Rheumatoid Arthritis. British Journal of Rheumatology, 32:807-811.

Fitzsimmons, B. and Bunting, L. K. (1993). Parkinson's disease. Quality of life issues. Nursing Clinics of North America, 28, 807-818.

Fletcher, A., Gore, S., Jones, D., Fitzpatrick, R., Spiegelhalter, D. and Cox, D. (1992). Quality of life measures in health care. II: Design, analysis, and interpretation. British Medical Journal, 305, 1145-8.

Follick, M.J., Smith, T.W. and Ahern, D.K. (1985). The Sickness Impact Profile: a global measure of disability in chronic low back pain. Pain, 21, 67-76.

Folstein, M. F., Folstein, S. E. and McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research, 12, 189-98.

Friedman, J. and Friedman, H. (1993). Fatigue in Parkinson's disease. Neurology, 43, 2016-2018.

George L K and Gwyther, L. P. (1986). Caregiver well-being: a multidimensional examination of family caregivers of demented adults. Gerontologist, 26, 253-9.

Glozman J M, Bicheva K G and Fedorova N V (1998). Scale of Quality of Life Care-Givers (SQLC). Journal of Neurology, 245, S39-S41.

Gray A, McNamara I, Aziz T, Gregory R, Bain P, Wilson J and Scott R (2002). Quality of life outcomes following surgical treatment of Parkinson's disease. Movement Disorders. 7(1):68.-75.

Groenvold M (1999). Methodological issues in the assessment of health-related quality of life in palliative care trials. Acta anaesthesiologica Scandinavica, 43, 948-953.

Guyatt G, Walter S and Norman G (1987). Measuring change over time: assessing the usefulness of evaluative instruments. Journal of Chronic Diseases, 40, 171-8.

Guyatt G H, Feeny D H and Patrick D L (1993). Measuring health related quality of life. Annals of Internal Medicine, 118, 622-629.

Guyatt G H, Jaeschke R, Feeny D H and Patrick D L (1996) In Quality of Life and Pharmacoeconomics in Clinical Trials(Ed, B, S.) Lippincott-Raven, Philadelphia, pp. 41-48.

Guyatt G H, Kirshner B and Jaeschke R (1992). Measuring health status: What are necessary measurment properties? Journal of Clinical Epidemiology, 45, 1341-1345.

Guyatt G H, Veldhuyzen Van Zanten S J, Feeny D H and Patrick D L (1989). Measuring quality of life in clinical trials: a taxonomy and review. Canadian Medical Association Journal., 140, 1441-1448.

Happe S and Berger K (2001). The association of dopamine agonists with daytime sleepiness, sleep problems and quality of life in patients with Parkinson's disease--a prospective study. Journal of Neurol, 248, 1062-7.

Harrison JE, Preston S and Blunt SB (2000). Measuring symptom change in patients with Parkinson's disease. Age Ageing; 29(1):41.-5.

Hayes, V., Morris, J., Wolfe, C. and Morgan, M. (1995). The SF-36 health survey questionnaire: is it suitable for use with older adults? Age Ageing, 24, 120-125.

Heckman, J. and MaCurdy, T. (1981). New methods for estimating labor supply functions: A survey. Research in labor economics, 4, 65-102.

Hely, M. A., Morris, J. G., Traficante, R., Reid, W. G., O'Sullivan, D. J. and Williamson, P. M. (1999). The sydney multicentre study of Parkinson's disease: progression and mortality at 10 years. Journal of Neurology Neurosurgery and Psychiatry, 67, 300-7.

Herlofson K and Larsen J.P (2003). The influence of fatigue on health related quality of life in patients with Parkinson's disease. Acta anaesthesiologica Scandinavica, 107, 1-6.

Hickey J (1992) Nervous system degenerative diseases. In The Clinical Practice of Neurological and Neurosurgical Nursing, JB Lippincott, Philadelphia.

Hobart, J. C., Lamping, D. L. and Thompson, A. J. (1996). Evaluating neurological outcome measures: the bare essentials. Journal of Neurology Neurosurgery and Psychiatry, 60, 127-30.

Hobson, J. P. and Meara, R. J. (1997). Is the SF-36 health survey questionnaire suitable as a self-report measure of the health status of older adults with Parkinson's disease? Quality of Life Research, 6, 213-6.

Hoehn M M (1992). The natural history of Parkinson's disease in the pre-levodopa and post-levodopa eras. Neurologic Clinics, 10, 331-9.

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

Hoehn M M and Yahr M D (2001). Parkinsonism: onset, progression, and mortality. 1967. Neurology, 57, S11-26.

Hughes A J, Daniel S E, Kilford L and Lees A J (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. Journal of Neurology, Neurosurgery and Psychiatry, 55(3), 181-184.

Hunt S M and McKenna S P (1993). Measuring patients' views of their health. SF 36 misses the mark. British Medical Journal, 307, 125.

Hunt S M, McKenna S P, McEwen J, Williams J and Papp E (1981). The Nottingham Health Profile: subjective health status and medical consultations. Social science & medicine [A], 15, 221-9.

Hyland M E, Kenyon C A and Jacobs P A (1994). Sensitivity of quality of life domains and constructs to longitudinal change in a clinical trial comparing salmeterol with placebo in asthmatics. Quality of Life Research, 3, 121-6.

ICH (1998) ICH HARMONISED TRIPARTITE GUIDELINE STATISTICAL PRINCIPLES FOR CLINICAL TRIALS E9, http://www.ich.org, http://www.ich.org.

Jacobson, N. S. and Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. Journal of consulting and clinical psychology, 59, 12-9.

Jaeschke, R., Singer, J. and Guyatt, G. H. (1989). Measurement of health status. Ascertaining the minimal clinically important difference. Controlled Clinical Trials, 10, 407-15.

Jenkinson C, Fitzpatrick R and Peto V (1999). Health-related quality-of-life measurement in patients with Parkinson's disease. Pharmacoeconomics., 15, 157-165.

Jenkinson C, Fitzpatrick R, Peto V, Greenhall R and Hyman N (1997). The Parkinson's Disease Questionnaire (PDQ-39): development and validation of a Parkinson's disease summary index score. Age Ageing, 26, 353-357.

Jenkinson C, Peto V, Fitzpatrick R (1998). The Parkinson's disease questionnaire: user manual for the PDQ-39, PDQ-8 and PDQ-summary index. Oxford: Health Research Services Unit.

Jenkinson C, Peto V, Fitzpatrick R, Greenhall R and Hyman N (1995). Self-reported functioning and well-being in patients with Parkinson's disease: comparison of the short-form health survey (SF-36) and the Parkinson's Disease Questionnaire (PDQ-39). Age Ageing, 24, 505-509.

Jensen, M.P., Strom, S.E., Turner, J.A., and Romano, J.M. (1992). Validity of the Sickness Impact Profile Roland scale as a measure of dysfunction in chronic pain patients. Pain 50(2):157-62

Juniper E F, Guyatt G H, Willan A and Griffith L E (1994). Determining a minimal important change in a disease-specific Quality of Life Questionnaire. Journal of Clinical Epidemiology, 47, 81-7.

Juniper EF, Guyatt GH and Jaeschke R (1996) In Quality of life and pharmacoeconomics in clinical trials(Ed, B, S.) Lippincott-Raven, Philadelphia, pp. 49-56.

Kaplan R M and Anderson J P (1988). A general health policy model: update and applications. Health services research, 23, 203-35.

Kaplan R M, Anderson J P, Wu A W, Mathews W C, Kozin F and Orenstein D (1989). The Quality of Well-being Scale. Applications in AIDS, cystic fibrosis, and arthritis. Medical Care, 27, S27-43.

Kaplan R M, Bush J W and Berry C C (1976). Health status: types of validity and the index of well-being. Health services research, 11, 478-507.

Karlsen KH, Tandberg E, Arsland D and Larsen JP (2000). Health related quality of life in Parkinson's disease: a prospective longitudinal study. Journal of neurology, neurosurgery and psychiatry; 69.(5):584.-9.

Kirshner, B. and Guyatt, G. (1985). A methodological framework for assessing health indices. Journal of Chronic Diseases, 38, 27-36.

Koller W C, Hutton J T, Tolosa E and Capilldeo R (1999). Immediate-release and controlled-release carbidopa/levodopa in PD: a 5-year randomized multicenter study. Carbidopa/Levodopa Study Group. Neurology, 53, 1012-9.

Koller WC (1994) In Beyond the Decade of the Brain(Ed, Stern M) Wells Medical Ltd, Royal Turnbridge Wells, pp. 31-46.

Koplas, P. A., Gans, H. B., Wisely, M. P., Kuchibhatla, M., Cutson, T. M., Gold, D. T., Taylor, C. T. and Schenkman, M. (1999). Quality of life and Parkinson's disease. The Journals of gerontology. Series A, Biological sciences and medical sciences, 54, M197-M202.

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.-23.

Lang AE and Fahn S (1999). The Dyskinesia Subjective Rating Scale. Movement Disorders, 14, 79-80.

Larsen J P, Karlsen K and Tandberg E (2000). Clinical problems in non-fluctuating patients with Parkinson's disease: a community-based study. Movement Disorders.15.(5):826.-9.

Lees A J, Blackburn N A and Campbell V L (1988). The nighttime problems of Parkinson's disease. Clinical Neuropharmacology, 11, 512-9.

Liang M H, Fossel A H and Larson M G (1990). Comparisons of five health status instruments for orthopedic evaluation. Medical Care, 28, 632-42.

Lipsey MW (1990) Design Sensitivity: Statistical Power for Experimental Research Sage, Newbury Park, California, USA.

Longstreth W T, Nelson L, Linde M and D, M. (1992). Utility of the Sickness Impact Profile in Parkinson's disease. Journal of Geriatric Psychiatry and Neurology, 5, 142-8.

Lydick E and Epstein R S (1993). Interpretation of quality of life changes. Quality of Life Research, 2, 221-6.

Lydick E, Zimmerman S I, Yawn B, Love B, Kleerekoper M, Ross P, Martin A and Holmes R (1997). Development and validation of a discriminative quality of life questionnaire for osteoporosis (the OPTQoL). Journal of bone and mineral research, 12, 456-63.

Lyons K S, Stewart B J, Archbold P G, Carter J H and Perrin N A (2004). Pessimism and optimism as early warning signs for compromised health for caregivers of patients with Parkinson's disease. Nursing research, 53, 354-62.

Maier Hoehn, M. M. (1983). Parkinsonism treated with levodopa: progression and mortality. Journal of neural transmission. Supplementum, 19, 253-64.

Malek M (1997). (2) Measuring Quality of Life Impacts. The Pharmaceutical Journal, 258, 22-24.

Marinus J, Ramaker C, van Hilten J J and Stiggelbout A M (2002). Health related quality of life in Parkinson's disease: a systematic review of disease specific instruments. Journal of Neurology, Neurosurgery and Psychiatry, 72, 241-8.

Martilla RJ and Rinne UK (1976). Epidemiology of Parkinson's disease in Finland. Acta neurologica Scandinavica, 53, 81-102.

Martinez Martin P (1998). Quality of life in Parkinson's disease: validation study of the PDQ-39 Spanish version. Journal of Neurology, 245, S34-S38.

Martinez Martin P, Frades B, Jimenez Jimenez F J, Pondal M, Lopez Lozano J J, Vela L, Vazquez A, del Val Fernandez J, Divi M, Fabregat N, Tuldra A, Crespo Maraver M C and Molinero L M (1999). The PDQ-39 Spanish version: reliability and correlation with the short-form health survey (SF-36). Neurologia, 14, 159-63.

Martinez-Martin P (1998). An introduction to the concept of "quality of life in Parkinson's disease". Journal of Neurology, 245 Supplement 1, S2-6.

Martinez-Martin P (2001). Health-related quality of life in Parkinson's disease: outcomes of therapeutic interventions. Expert Review of Pharmacoeconomics & Outcomes Research, 1, 99-108.

Martinez-Martin P, Grandas F, Linazasoro G and Bravo J L (1999). Conversion to controlled-release levodopa/carbidopa treatment and quality of life as measured by the Nottingham Health Profile. The STAR Study Group. Neurologia, 14, 338-43.

McDowell I and Newall C. (1987) Measuring health: a guide to rating scales and questionnaires, Oxford University Press, New York.

McKenna S P (1993). Generic versus specific measures in clinical trials. British Journal of Medical Economics, 6C, 17-25.

Meltzer J W and Hochstim J R (1970). Reliability and validity of survey data on physical health. Public Health Rep, 85, 1075-86.

Menza M A, Robertson-Hoffman D E and Bonapace A S (1993). Parkinson's disease and anxiety: comorbidity with depression. Biological psychiatry, 34, 465-70.

Montastruc, J. L., Rascol, O. and Senard, J. M. (1999). Treatment of Parkinson's disease should begin with a dopamine agonist. Movement Disorders, 14, 725-30.

Moore O, Gurevich T, Korczyn A D, Anca M, Shabtai H and Giladi N (2002). Quality of sexual life in Parkinson's disease. Parkinsonism and Related Disorders, 8, 243-6.

Mutch W J, Dingwall-Fordyce I, Downie A W, Paterson J G and Roy S K (1986). Parkinson's disease in a Scottish city. British Medical Journal (Clin Res Ed), 292, 534-6.

Nunnally J C (1978) Psychometric theory, McGraw-Hill, New York.

O'Connor R (1993) Issues in the measurement of health related quality of life. Center for Health Program Evaluation. West Heidelberg, Australia.

Oertel WH and Quinn NP (1996) In Parkinsonism(Eds, Brandt T, Diener HC and Caplan LR) Academic Press, San Diego, pp. 715-72.

Osoba D and Zee B (1998). Completion rates in health-related quality-of-life assessment: approach of the National Cancer Institute of Canada Clinical Trials Group. Statistics in medicine, 17, 603-612.

Pahwa R, Lyons K, McGuire D, Silverstein P, Zwiebel F, Robischon M and Koller W C (1997). Comparison of standard carbidopa-levodopa and sustained-release

carbidopa-levodopa in Parkinson's disease: pharmacokinetic and quality-of-life measures. Movement Disorders, 12, 677-681.

Parkinson J. An essay on the shaking palsy. In. London; 1817

Patrick D L, Bush J W and Chen M M (1973). Toward an operational definition of health. Journal of health and social behavior, 14, 6-23.

Patrick, D. L. and Bergner, M. (1990). Measurement of health status in the 1990s. Annual review of public health, 11, 165-83.

Peto V, Jenkinson C and Fitzpatrick R (1998). PDQ-39: a review of the development, validation and application of a Parkinson's disease quality of life questionnaire and its associated measures. Journal of Neurology., 245 Suppl 1, S10-S14.

Peto V, Jenkinson C and Fitzpatrick R (2001). Determining minimally important differences for the PDQ-39 Parkinson's disease questionnaire. Age and Ageing, 30, 299-302.

Peto V, Jenkinson C, Fitzpatrick R and Greenhall R (1995). The development and validation of a short measure of functioning and well being for individuals with Parkinson's disease. Quality of Life Research, 4, 241-8.

Petrie A and Sabin C (2000) Medical Statistics at a Glance, Blackwell Publishing, Oxford.

Playfer J R (1999). Depression, cognition and quality of life in parkinsonian patients. Age and Ageing, 28, 333-334.

Price Martin J, Salek Sam S, Davies Brian H and et al. (1997). Evaluation of the Sickness Impact Profile and Short Form-36 as Measures of Quality of Life in Asthma Patients. Journal of Applied Therapeutics, 1, 247-256.

Rajput A H, Offord K P, Beard C M and Kurland L T (1984). Epidemiology of parkinsonism: incidence, classification, and mortality. Annals of neurology, 16, 278-82.

Rajput, A. H., Martin, W., Saint-Hilaire, M. H., Dorflinger, E. and Pedder, S. (1997). Tolcapone improves motor function in parkinsonian patients with the "wearing-off" phenomenon: a double-blind, placebo-controlled, multicenter trial. Neurology, 49, 1066-71.

Revicki D A (1989). Health-related quality of life in the evaluation of medical therapy for chronic illness. Journal of Family Practice, 29, 377-80.

Riazi J C, Lamping D L, Fitzpatrick R, Freeman J A, Jenkinson C and Peto V (2003). Using the SF-36 measure to compare the health impact of multiple sclerosis and Parkinson's disease with normal population health profiles. Journal of Neurology, NeuroSurgery and Psychiatry, 74, 710-714.

Rinne U K (1983). Problems associated with long-term levodopa treatment of Parkinson's disease. Acta neurologica Scandinavica. Supplementum, 95, 19-26.

Rubenstein L M, Voelker M D, Chrischilles E A, Glenn D C, Wallace R B and Rodnitzky R L (1998). The usefulness of the Functional Status Questionnaire and Medical Outcomes Study Short Form in Parkinson's disease research. Quality of Life Research, 7, 279-290.

Salek M S (1997). Measuring the Quality of Life of Patients with Arthritis: Validation and Clinical Applications of a Self Administered Health Status Instrument (UK-SIP). Journal of Applied Therapeutics, 1, 219-236.

Salek M S (1992) In: Pharmacy, University of Wales, Cardiff.

Salek M S (1993). Health-related quality of life measurement. A new challenge for pharmacy practitioners. Journal of pharmacy and pharmacology, 45 Suppl 1, 387-92.

Salek M S (1999). Compendium of quality of life instruments. John Wiley & Sons, Ltd.

Salek M S and Luscombe D K (1992). Health-Related Quality of Lise Assessment: A Review. Journal of Drug Development, 5, 137-153.

Sales, A. E., Plomondon, M. E., Magid, D. J., Spertus, J. A. and Rumsfeld, J. S. (2004). Assessing response bias from missing quality of life data: the Heckman method. Health and Quality of Life Outcomes, 2, 49.

SB Clinical Report for Study 056. A double-blind, L-dopa plus benserazide controlled, multicenture study of ropinirole at a flexible oral dose of 0.25 - 8mg three times daily for five years in the treatment of early parkinsonian patients not treated with L-dopa. In; Manuscript in review with New England Medical Journal.

Schipper H, Clinch JJ and Olweny CLM (1996) In Quality of life and pharmacoeconomics in clinical trials(Ed, B, S.) Lippincott-Raven, Philadelphia, USA, pp. 11-23.

Schrag A, Jahanshahi M and Quinn N (2000a). How does Parkinson's disease affect quality of life? A comparison with quality of life in the general population. Movement Disorders, 15, 1112-8.

Schrag A, Jahanshahi M and Quinn N (200b), What contributes to quality of life in patients with Parkinson's disease? Journal of Neurology, Neurosurgery and Psychiatry, 69, 308-12.

Schrag A, Jahanshahi M and Quinn N P (2001). What contributes to depression in Parkinson's disease? Psychological medicine, 31, 65-73.

Schrag A and Quinn N (2000). Dyskinesias and motor fluctuations in Parkinson's disease. A community-based study. Brain, 123 (Pt 11), 2297-305.

Schulz, G.M. (2002). The effects of speech therapy and pharmacological treatments on voice and speech in Parkinson's disease: a review of the literature. Current medicinal chemistry, 9, 1359-66.

Schulzer M, Calne S and Calne D B (1999) In Quality of Life in Parkinson's Disease(Eds, Pablo Martinez Martin and Koller, W. C.) Masson S.A., Barcelona, pp. 161 - 180.

Shih, W. (2002). Problems in dealing with missing data and informative censoring in clinical trials. Curr Control Trials Cardiovasc Med, 3, 4.

Shindler JS, Brown R and Welburn P (1995) In Quality of life assessment: key issues in the 1990's.(Eds, Walker SR and Rosser RM) Kluwer Academic Publishers, Boston, pp. 289-300.

Siegal S and Castellan Jr. N J (1988) Nonparametric statistics for the Behavioural Sciences, McGraw Hill Book Company, New York.

Sommers, S. D. and Ramsey, S. D. (1999). A review of quality-of-life evaluations in prostate cancer. Pharmacoeconomics, 16, 127-40.

Spilker B (1996) Quality of Life and Pharmacoeconomics in Clinical Trials, Lippincott - Raven, Philadelphia New Yok.

Stacy M (1999). Managing late complications of Parkinson's disease. Medical Clinics of North America, 83, 469-81, vii.

Stacy M (2002). Sleep disorders in Parkinson's disease: epidemiology and management. Drugs Aging, 19, 733-9.

Stallibrass, C. (1997). An evaluation of the Alexander Technique for the management of disability in Parkinson's disease--a preliminary study. Clinical Rehabilitation, 11, 8-12.

Staniszewska, S. (1998). Measuring quality of life in the evaluation of health care. Nursing Standard, 12, 36-9.

Stern M B (2000). The early treatment of Parkinson's disease: levodopa, dopamine agonists or both. Parkinsonism and Related Disorders 1;7.(1):27.-33.

Strawbridge, W. J. (1998). Quality of life: what is it and can it be measured? Growth Hormone and IGF Research, 8 Supplement A, 59-62.

Streiner D L and Norman G R (1995) Health measurement scales: A practical guide to their development and use. Oxford University Press Inc, New York.

Sutcliffe, R. L., Prior, R., Mawby, B. and McQuillan, W. J. (1985). Parkinson's disease in the district of the Northampton Health Authority, United Kingdom. A study of prevalence and disability. Acta Neurologica Scandinavica, 72, 363-79.

Tandberg, E., Larsen, J. P., Nessler, E. G., Riise, T. and Aarli, J. A. (1995). The epidemiology of Parkinson's disease in the county of Rogaland, Norway. Movement Disorders, 10, 541-9.

Temkin, N. R., Dikmen, S., Machamer, J. and McLean, A. (1989). General versus disease-specific measures. Further work on the Sickness Impact Profile for head injury. Medical Care, 27, S44-53.

Testa M A (2000). Methods and applications of quality-of-life measurement during antihypertensive therapy. Current Hypertension Reports, 2, 530-7.

Testa M A and Nackley J F (1994). Methods for quality-of-life studies. Annual Review of Public Health, 15, 535-59.

Testa M A and Simonson D C (1996). Assesment of quality-of-life outcomes. New England Journal of Medicine, 334, 835-40.

Thier S O (1992). Forces motivating the use of health status assessment measures in clinical settings and related clinical research. Medical Care, 30, MS15-22.

Thompson, D. R. and Yu, C. M. (2003). Quality of life in patients with coronary heart disease-I: Assessment tools. Health and Quality of Life Outcomes, 1, 42.

Timmerman, A. A., Anteunis, L. J. and Meesters, C. M. (2003). Response-shift bias and parent-reported quality of life in children with otitis media. Archives of Otolaryngology--Head and Neck Surgery, 129, 987-91.

Torrance, G. W. (1987). Utility approach to measuring health-related quality of life. Journal of Chronic Diseases, 40, 593-603.

Vickrey, B. G., Hays, R. D., Graber, J., Rausch, R., Engel, J., Jr. and Brook, R. H. (1992). A health-related quality of life instrument for patients evaluated for epilepsy surgery. Medical Care, 30, 299-319.

Vidailhet, M., Bonnet, A. M., Marconi, R., Durif, F. and Agid, Y. (1999). The phenomenology of L-dopa-induced dyskinesias in Parkinson's disease. Movement Disorders, 14 Suppl 1, 13-8.

Vingerhoets, G., Lannoo, E., van der Linden, C., Caemaert, J., Vandewalle, V., van den Abbeele, D. and Wolters, M. (1999). Changes in quality of life following unilateral pallidal stimulation in Parkinson's disease. Journal of Psychosomatic Research, 46, 247-55.

Visser, M. C., Fletcher, A. E., Parr, G., Simpson, A. and Bulpitt, C. J. (1994). A comparison of three quality of life instruments in subjects with angina pectoris: the Sickness Impact Profile, the Nottingham Health Profile, and the Quality of Well Being Scale. Journal of Clinical Epidemiology, 47, 157-63.

Wagner, E. H., LaCroix, A. Z., Grothaus, L. C. and Hecht, J. A. (1993). Responsiveness of health status measures to change among older adults. Journal of the American Geriatrics Society, 41, 241-8.

Ware J E and Sherbourne C D (1992). The MOS 36 Item Short Form Health Survey (SF-36). Medical Care, 30, 473-483.

Ware J E Jr (1987). Standards for validating health measures: definition and content. Journal of Chronic Diseases, 40, 473-80.

Ware JE, Kosinski M and Keller SD (1994) The Health Institute, Boston, MA.

Ware, J. E., Jr. and Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Medical Care, 30, 473-483.

Weiner, W. J. (1999). The initial treatment of Parkinson's disease should begin with levodopa. Movement Disorders, 14, 716-24.

Welsh M, Hung L and Waters C H (1997). Sexuality in women with Parkinson's disease. Movement Disorders, 12, 923-7.

Welsh M, McDermott M, Holloway R, Plumb S and Pfeiffer R (2003). Development and Testing of the Parkinson's Disease Quality of Life Scale: The PDQUALIF. Movement Disorders, 18, 637-645.

Welsh M D, Dorflinger E, Chernik D and Waters C (2000). Illness impact and adjustment to Parkinson's disease: before and after treatment with tolcapone. Movement Disorders, 15, 497-502.

Wenger, N. K., Mattson, M. E., Furberg, C. D. and Elinson, J. (1984). Assessment of quality of life in clinical trials of cardiovascular therapies. American Journal of Cardiology, 54, 908-13.

WHOQOL group (1993) In Quality of Life Research 2World Health Organisation, Geneva.

World Health Organisation (1991). World Health Statistics Annual. 35.

Wulff, H. (1999). The two cultures of medicine: objective facts versus subjectivity and values. Journal of the Royal Society of Medicine, 92, 549-52.

Zarit S H, Anthony C R and Boutselis M (1987). Interventions with care givers of dementia patients: comparison of two approaches. Psychology and Aging, 2, 225-32.

Zarit S H, Todd P A and Zarit J M (1986). Subjective burden of husbands and wives as caregivers: a longitudinal study. Gerontologist, 26, 260-6.

Zeller R A and Carmines E G (1980) Measurement in the Social Sciences, Cambridge University Press, Cambridge.

PUBLICATIONS

Abstracts

Ali T and Salek S (2002). Comparison of the PDQ-39 with a Generic Measure of Health Status (UK Sickness Impact Profile) in Parkinson's Disease Subjects. Quality of Life Research, 11, 672.

Ali T and Salek S (2002). The Reliability of the PDQ-39 and a Generic Measure of Health Status (UK Sickness Impact Profile) in Parkinson's Disease. Quality of Life Research, 11, 671.

Ali T, Lane R, and Salek S (2004). Convergent and Discriminative Validity of the UKSIP and SF-36 in Patients with Parkinson's Disease (PD). Quality of Life Research, 13, 1533.

Ali T and Salek S (2004). Evaluation of the Humanistic Burden of Dyskinesia in Patients with Parkinson's Disease. Quality of Life Research, 13, 1534.

APPENDICES

Appendix I:	Parkinson's Disease Questionnaire (PDQ-39)
Appendix II:	UK Sickness Impact Profile (UKSIP)
Appendix III:	Short Form-36 (SF-36)
Appendix IV:	SF-36 scales that make up summary scores
Appendix V:	Parkinson's Impact Scale (PIMS)
Appendix VI:	Hoehn and Yahr Scale (H&Y)
Appendix VII:	Unified Parkinson's disease Rating Scale (UPDRS)
Appendix VIII:	Dyskinesia Subjective Rating Scale
Appendix IX:	Burden Interview Questionnaire (CBIQ)
Appendix X:	Change in condition questionnaire form
Appendix XI:	Example of a Patient Information Sheet
Appendix XII:	Example of Informed Consent Form

APPENDIX I

DUE TO HAVING PARKINSON'S DISEASE, how often have you experienced the following, during the last month?

Due to having Parkinson's disease, how often during the last month have you	*Please tick* ***one box*** *for each question*				
	Never	**Occasionally**	**Sometimes**	**Often**	**Always** or cannot do at all
1. Had difficulty doing the leisure activities which you would like to do?	☐	☐	☐	☐	☐
2. Had difficulty looking after your home, e.g. DIY, housework, cooking?	☐	☐	☐	☐	☐
3. Had difficulty carrying bags of shopping?	☐	☐	☐	☐	☐
4. Had problems walking half a mile?	☐	☐	☐	☐	☐
5. Had problems walking 100 yards?	☐	☐	☐	☐	☐
6. Had problems getting around the house as easily as you would like?	☐	☐	☐	☐	☐
7. Had difficulty getting around in public?	☐	☐	☐	☐	☐
8. Needed someone else to accompany you when you went out?	☐	☐	☐	☐	☐
9. Felt frightened or worried about falling over in public?	☐	☐	☐	☐	☐

Please check that you have ticked one box for each question before going on to the next page

Due to having Parkinson's disease, how often <u>during the last month</u> have you ....

*Please tick **one box** for each question*

	Never	Occasionally	Sometimes	Often	Always
10. Been confined to the house more than you would like?	☐	☐	☐	☐	☐
11. Had difficulty washing yourself?	☐	☐	☐	☐	☐
12. Had difficulty dressing yourself?	☐	☐	☐	☐	☐
13. Had problems doing up buttons or shoe laces?	☐	☐	☐	☐	☐
14. Had problems writing clearly?	☐	☐	☐	☐	☐
15. Had difficulty cutting up your food?	☐	☐	☐	☐	☐
16. Had difficulty holding a drink without spilling it?	☐	☐	☐	☐	☐
17. Felt depressed?	☐	☐	☐	☐	☐
18. Felt isolated and lonely?	☐	☐	☐	☐	☐
19. Felt weepy or tearful?	☐	☐	☐	☐	☐

Please check that you have ticked <u>one box for each question</u> before going on to the next page

Due to having Parkinson's disease, how often during the last month have you ....

*Please tick **one box** for each question*

	Never	Occasionally	Sometimes	Often	Always
20. Felt angry or bitter?	☐	☐	☐	☐	☐
21. Felt anxious?	☐	☐	☐	☐	☐
22. Felt worried about your future?	☐	☐	☐	☐	☐
23. Felt you had to conceal your Parkinson's from people?	☐	☐	☐	☐	☐
24. Avoided situations which involve eating or drinking in public?	☐	☐	☐	☐	☐
25. Felt embarrassed in public due to having Parkinson's disease?	☐	☐	☐	☐	☐
26. Felt worried by other people's reaction to you?	☐	☐	☐	☐	☐
27. Had problems with your close personal relationships?	☐	☐	☐	☐	☐
28. Lacked support in the ways you need from your spouse or partner? *If you do not have a spouse or partner, please tick here* ☐	☐	☐	☐	☐	☐
29. Lacked support in the ways you need from your family or close friends?	☐	☐	☐	☐	☐

Please check that you have ticked <u>one box for each question</u> before going on to the next page

Due to having Parkinson's

disease, how often *Please tick* ***one box*** *for each question*

<u>during the last month</u>

have you	**Never**	**Occasionally**	**Sometimes**	**Often**	**Always**
30. Unexpectedly fallen asleep during the day?	☐	☐	☐	☐	☐
31. Had problems with your concentration, e.g. when reading or watching TV?	☐	☐	☐	☐	☐
32. Felt your memory was bad?	☐	☐	☐	☐	☐
33. Had distressing dreams or hallucinations?	☐	☐	☐	☐	☐
34. Had difficulty with your speech?	☐	☐	☐	☐	☐
35. Felt unable to communicate with people properly?	☐	☐	☐	☐	☐
36. Felt ignored by people?	☐	☐	☐	☐	☐
37. Had painful muscle cramps or spasms?	☐	☐	☐	☐	☐
38. Had aches and pains in your joints or body?	☐	☐	☐	☐	☐
39. Felt unpleasantly hot or cold?	☐	☐	☐	☐	☐

APPENDIX II

HEALTH STATUS QUESTIONNAIRE
U K S I P

PLEASE READ THE WHOLE INTRODUCTION BEFORE YOU COMPLETE THE QUESTIONNAIRE

INTRODUCTION

On the next few pages are statements which describe things people often do when they are not well.

As you read them, think of yourself today. When you read a statement that you are sure describes you TODAY AND IS RELATED TO YOUR HEALTH, please put a tick in the box.

FOR EXAMPLE

I am not driving my car.

If you have not been driving for some time because of your health, and are still not driving today, you should tick this statement.

On the other hand, if you never drive or are not driving today because your car is being repaired, the statement " I am not driving my car " is NOT related to your health and you should NOT tick it. If you are driving less, or are driving shorter distances, and feel that the statement only partially describes you, DO NOT TICK IT. In each of these cases you should leave the box to the right of the statement BLANK.

Read the introduction to each group of statements and then consider the statements in the order listed. While some of the statements may not apply to you, we ask that you please read ALL of them. Some of the statements will differ only in a few words, so please read each one carefully. While you may wish to go back to change a response, your first answer is usually the best. Please do not read ahead in the questionnaire.

Once you have started the questionnaire it is very important that you complete it within one day (24 hours).

If you find it hard to keep your mind on the statements, take a short break and then continue from where you left off.

Please do not discuss the statements with anyone, including family members, while doing the questionnaire.

Now turn to the questionnaire section and read each statement. Remember we are interested in recent or longstanding changes in your activities which are related to your health.

HOW WOULD YOU DESCRIBE YOUR PRESENT HEALTH?

Very Good ☐	Good ☐	Fair ☐	Poor ☐	Very Poor ☐

THANK YOU FOR YOUR TIME AND HELP

Patient Number ☐ ' Number ☐

Copyright M S Salek

PLEASE TICK ONLY THOSE STATEMENTS THAT YOU ARE SURE DESCRIBE YOU TODAY AND ARE RELATED TO YOUR HEALTH

A *These statements describe your sleep and rest today*

1. I spend much of the day lying down in order to rest ☐
2. I sit for much of the day ☐
3. I am sleeping or dozing most of the time - day and night ☐
4. I lie down more often during the day in order to rest ☐
5. I sit around half asleep ☐
6. I sleep less at night, for example, I wake up easily, I don't fall asleep for a long time, I keep waking up ☐
7. I sleep or doze more during the day ☐

B *The following statements describe your emotional behaviour*

1. I say how bad or useless I am, for example, that I am a burden on others ☐
2. I laugh or cry suddenly ☐
3. I often moan and groan in pain or discomfort ☐
4. I have attempted suicide ☐
5. I behave nervously or restlessly ☐
6. I keep rubbing or holding areas of my body that hurt or are uncomfortable ☐
7. I am irritable and impatient with myself, for example, I run myself down, I swear at myself, I blame myself for things that happen ☐
8. I talk about the future in a hopeless way ☐
9. I get sudden frights ☐

C *These following statements describe how you move about, bath, go to the toilet and dress yourself*

1. I make difficult movements with help, for example getting in or out of the bath or a car ☐
2. I do not get in or out of bed or chairs without the help of a person or mechanical aid ☐
3. I only stand for short periods of time ☐
4. I do not keep my balance ☐
5. I move my hands or fingers with some limitation or difficulty ☐
6. I only stand up with someone's help ☐
7. I kneel, stoop or bend down only by holding on to something ☐
8. I am in a restricted position all the time ☐
9. I am very clumsy ☐
10. I get in and out of bed or chairs by grasping something for support or by using a stick or walking frame ☐
11. I stay lying down most of the time ☐
12. I change position frequently ☐
13. I hold onto something to move myself around in bed ☐

PLEASE TICK ONLY THOSE STATEMENTS THAT YOU ARE SURE DESCRIBE YOU TODAY AND ARE RELATED TO YOUR HEALTH

14. I do not bathe myself completely, for example, I need help with bathing ☐
15. I do not bathe myself at all, but am bathed by someone else ☐
16. I use a bedpan with help ☐
17. I have trouble putting on my shoes, socks or stockings ☐
18. I do not have control of my bladder ☐
19. I do not fasten my clothing, for example, I require assistance with buttons, zips or shoelaces ☐
20. I spend most of the time partly dressed or in pyjamas ☐
21. I do not have control of my bowels ☐
22. I dress myself, but do so very slowly ☐
23. I only get dressed with someone's help ☐

D *These statements describe your daily work around the house*

1. I only do housework or work around the house for short periods of time or I rest often ☐
2. I am doing less of the daily household chores that I would usually do ☐
3. I am not doing any of the daily household chores that I would usually do ☐
4. I am not doing any of the maintenance or repair work that I would usually do in my house or garden ☐
5. I am not doing any of the shopping that I would usually do ☐
6. I am not doing any of the cleaning that I would usually do ☐
7. I have difficulty using my hands, for example, turning taps, using kitchen gadgets, sewing or doing repairs ☐
8. I am not doing any of the clothes washing that I would usually do ☐
9. I am not doing heavy work around the house ☐
10. I have given up taking care or personal or household business affair, for example, paying bills, banking or doing household accounts ☐

E *And these next statements describe how you get around the house and outside*

1. I am only getting about in one building ☐
2. I stay in one room ☐
3. I am staying in bed more ☐
4. I am staying in bed most of the time ☐
5. I am not using public transport now ☐
6. I stay at home most of the time ☐
7. I am only going out if there is a toilet nearby ☐
8. I am not going into town ☐
9. I only stay away from home for short periods ☐
10. I do not get about in the dark or in places that are not lit unless I have someone to help ☐

STILL THINKING ABOUT YOURSELF TODAY, PLEASE TICK ONLY THOSE STATEMENTS THAT YOU ARE SURE DESCRIBE YOU, AND YOU ARE SURE ARE RELATED TO YOUR HEALTH

F *These next statements describe your contact with your family and friends today*

1. I am going out less to visit people ☐
2. I am not going out to visit people at all ☐
3. I show less interest in other people's problems, for example, I don't listen when they tell me about their problems, I don't offer to help ☐
4. I am often irritable with those around me, for example, I snap at people or criticise easily ☐
5. I show less affection ☐
6. I take part in fewer social activities than I used to, for example, I go to fewer parties or social events ☐
7. I am cutting down the length of visits to friends ☐
8. I avoid having visitors ☐
9. My sexual activity is decreased ☐
10. I often express concern over what might be happening to my health ☐
11. I talk less with those around me ☐
12. I make many demands on other people, for example, I insist that they do. things for me, or tell them how to do things ☐
13. I stay alone much of the time ☐
14. I am disagreeable with my family, for example, I act spitefully of stubbornly ☐
15. I frequently get angry with my family, for example, I hit them, scream or throw things at them ☐
16. I isolate myself as much as I can from the rest of my family ☐
17. I am paying less attention to the children ☐
18. I refuse contact with my family, for example, I turn away from them ☐
19. I am not doing the things I usually do to take care of my children or family ☐
20. I am not joking with family members as I usually do ☐

G *The next set of statements describe walking and use of stairs*

1. I walk short distances or often stop for a rest ☐
2. I do not walk up or down hills ☐
3. I only use stairs with a physical aid, for example, a handrail, stick or crutches ☐
4. I only go up or down stairs with assistance from someone else ☐
5. I get about in a wheelchair ☐
6. I do not walk at all ☐
7. I walk by myself but with some difficulty, for example, I limp, wobble, stumble or I have a stiff leg ☐
8. I only walk with help from someone else ☐
9. I go up and down stairs more slowly, for example, one step at a time or I often have to stop ☐
10. I do not use stairs at all ☐

AGAIN, PLEASE TICK ONLY THOSE STATEMENTS THAT YOU ARE SURE DESCRIBE YOU, TODAY AND ARE RELATED TO YOUR HEALTH

11. I get about only by using a walking frame, crutches, stick, walls or hold onto furniture ☐
12. I walk more slowly ☐

H *These statements describe your feelings*

1. I am confused and start to do more than one thing at a time ☐
2. I have more minor accidents, for example, I drop things, I trip and fall or bump into things ☐
3. I react slowly to things that are said or done ☐
4. I do not finish things I start ☐
5. I have difficulty reasoning and solving problems, for example, making plans, making decisions, learning new things ☐
6. I sometimes get confused, for example, I do not know where I am, who is around, or what day it is ☐
7. I forget a lot, for example, things that happened recently, where I put things, or to keep appointments ☐
8. I do not keep my attention on any activity for long ☐
9. I make more mistake than usual ☐
10. I have difficulty doing things which involve thought and concentration ☐

I *These statements are about how you talk to other people and write*

1. I am having trouble writing or typing ☐
2. I communicate mostly by nodding my head, pointing, or using sign language, or other gestures ☐
3. My speech is understood only by a few people who know me well ☐
4. I often lose control of my voice when I talk, for example, my voice gets louder or softer or changes unexpectedly ☐
5. I don't write except to sign my name ☐
6. I carry on a conversation only when very close to other people or looking directly at them ☐
7. I speak with difficulty, for example, I get stuck for words, I stutter, I stammer, I slur my words ☐
8. I am understood with difficulty ☐
9. I do not speak clearly when I am under stress ☐

J THE NEXT GROUP OF STATEMENTS HAS TO DO WITH ANY WORK YOU USUALLY DO OTHER THAN LOOKING AFTER YOUR HOME - BY THIS WE MEAN ANYTHING THAT YOU REGARD AS WORK THAT YOU DO ON A REGULAR BASIS

Do you usually do work other than looking after your home ? ☐ YES ☐ NO

→ If you answered yes, go on to Section **K**

→ If you answered no, complete the following

Are you retired ? ☐ YES ☐ NO

If you are retired, was your retirement related to your health? ☐ YES ☐ NO

If you are not retired, but are NOT working, is this related to your health? ☐ YES ☐ NO

Now go on to Section **L**

K **IF YOU ARE NOT WORKING AND IT IS NOT BECAUSE OF YOUR HEALTH, PLEASE GO ONTO THE NEXT PAGE**

NOW CONSIDER THE WORK YOU DO AND TICK ONLY THOSE STATEMENTS THAT YOU ARE SURE DESCRIBE YOU TODAY AND ARE RELATED TO YOUR HEALTH. (IF TODAY IS A SATURDAY OR SUNDAY OR SOME OTHER DAY THAT YOU WOULD USUALLY HAVE OFF, PLEASE RESPOND AS IF TODAY WERE A WORKING DAY).

1. I am not working at all ---------- ☐
   (IF YOU TICKED THIS STATEMENT, GO ON TO THE NEXT PAGE)
2. I am doing part of my job at home ---------- ☐
3. I am not getting as much work done as usual ---------- ☐
4. I often get irritable with my workmates, for example, I snap at them and criticise them easily ---------- ☐
5. I am working shorter hours ---------- ☐
6. I am only doing light work ---------- ☐
7. I only work for short periods of time or often stop to rest ---------- ☐
8. I work at my usual job but with some changes, for example, I use different tools or special aids, or I swop jobs with someone else ---------- ☐
9. I do not do my jobs as carefully and accurately as usual ---------- ☐

PLEASE TICK ONLY THOSE STATEMENTS THAT YOU ARE SURE DESCRIBE YOU TODAY AND ARE RELATED TO YOUR HEALTH

L *The following statements describe the activities you usually do in your spare time, for relaxation, entertainment or just to pass the time*

1. I spend shorter periods of time on my hobbies and recreation ☐
2. I am going out to enjoy myself less often ☐
3. I am cutting down on some of my usual inactive pastimes, for example, I watch less TV, play cards less, or read less ☐
4. I am not doing any of my usual inactive pastimes, for example, I do not watch TV, play cards, or read ☐
5. I am doing more inactive pastimes in place of my other usual activities ☐
6. I am taking part in fewer community activities ☐
7. I am cutting down on some of my usual physical recreation or more active pastimes ☐
8. I am not doing any of my usual physical recreation or more active pastimes ☐

M *The following statements describe your eating and drinking habits*

1. I am eating much less than usual ☐
2. I feed myself but only with specially prepared food or special utensils ☐
3. I am eating special or different food, for example, soft food, bland diet, low-salt, low-fat, low-sugar ☐
4. I eat no food at all but I take liquids ☐
5. I just pick or nibble at my food ☐
6. I am drinking less fluids ☐
7. I feed myself with help from someone else ☐
8. I do not feed myself at all but have to be fed ☐
9. I am eating no food at all except by tubes or intravenous infusion ☐

NOW PLEASE LOOK THROUGH THE QUESTIONNAIRE AND MAKE SURE THAT YOU HAVE READ EVERY QUESTION

THANK YOU ONCE AGAIN FOR YOUR HELP

What date is it today? _____ 19 ____

Printed with permission of M S Salek, Medicine Research Unit, UWC, Cardiff

APPENDIX III

SF-36 HEALTH SURVEY

INSTRUCTIONS: This survey asks for your views about your health. This information will help keep track of how you feel and how well you are able to do your usual activities.

Answer every question by marking the answer as indicated. If you are unsure about how to answer a question, please give the best answer you can.

1. In general, would you say your health is:

(circle one)

Excellent 1

Very good 2

Good 3

Fair 4

Poor 5

2. Compared to one year ago, how would you rate your health in general now?

(circle one)

Much better now than one year ago 1

Somewhat better now than one year ago 2

About the same as one year ago 3

Somewhat worse now than one year ago 4

Much worse now than one year ago 5

Copyright © 1992 Medical Outcomes Trust. All rights reserved.(SF-36 Standard U.S. Version 1.0)

SF-36 HEALTH SURVEY

INSTRUCTIONS: This survey asks for your views about your health. This information will help keep track of how you feel and how well you are able to do your usual activities.

Answer every question by marking the answer as indicated. If you are unsure about how to answer a question, please give the best answer you can.

1. In general, would you say your health is:

(circle one)

Excellent 1

Very good 2

Good 3

Fair 4

Poor 5

2. Compared to one year ago, how would you rate your health in general now?

(circle one)

Much better now than one year ago 1

Somewhat better now than one year ago 2

About the same as one year ago 3

Somewhat worse now than one year ago 4

Much worse now than one year ago 5

Copyright © 1992 Medical Outcomes Trust. All rights reserved.(SF-36 Standard U.S. Version 1.0)

3. The following items are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

(circle one number on each line)

ACTIVITIES	Yes, Limited A Lot	Yes, Limited A Little	No, Not Limited At All
a. Vigorous activities, such as running, lifting heavy objects, participating in strenuous sports	1	2	3
b. Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf	1	2	3
c. Lifting or carrying groceries	1	2	3
d. Climbing several flights of stairs	1	2	3
e. Climbing one flight of stairs	1	2	3
f. Bending, kneeling, or stooping	1	2	3
g. Walking more than a mile	1	2	3
h. Walking several blocks	1	2	3
i. Walking one block	1	2	3
j. Bathing or dressing yourself	1	2	3

4. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

(circle one number on each line)

	YES	NO
a. Cut down on the **amount of time** you spent on work or other activities	1	2
b. Accomplished less than you would like	1	2
c. Were limited in the kind of work or other activities	1	2
d. Had difficulty performing the work or other activities (for example, it took extra effort)	1	2

Copyright ©1992 Medical Outcomes Trust.
All rights reserved.
(SF-36 Standard U.S. Version 1.0)

5. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

(circle one number on each line)

	YES	NO
a. Cut down the amount of time you spent on work or other activities	1	2
b. Accomplished less than you would like	1	2
c. Didn't do work or other activities as carefully as usual	1	2

6. During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or groups?

(circle one)

Not at all 1

Slightly 2

Moderately 3

Quite a bit 4

Extremely 5

7. How much bodily pain have you had during the past 4 weeks?

(circle one)

None 1

Very mild 2

Mild 3

Moderate 4

Severe 5

Very severe 6

Copyright ©1992 Medical Outcomes Trust.
All rights reserved.
(SF-36 Standard U.S. Version 1.0)

8. During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?

(circle one)

Not at all 1

A little bit 2

Moderately 3

Quite a bit 4

Extremely 5

9. These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks –

(circle one number on each line)

	All of the Time	Most of the Time	A Good Bit of the Time	Some of the Time	A Little of the Time	None of the Time
a. Did you feel full of pep?	1	2	3	4	5	6
b. Have you been a very nervous person?	1	2	3	4	5	6
c. Have you felt so down in the dumps that nothing could cheer you up?	1	2	3	4	5	6
d. Have you felt calm and peaceful?	1	2	3	4	5	6
e. Did you have a lot of energy?	1	2	3	4	5	6
f. Have you felt downhearted and blue?	1	2	3	4	5	6
g. Did you feel worn out?	1	2	3	4	5	6
h. Have you been a happy person?	1	2	3	4	5	6
i. Did you feel tired?	1	2	3	4	5	6

Copyright © 1992 Medical Outcomes Trust.
All rights reserved.
(SF-36 Standard U.S. Version 1.0)

10. During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting with friends, relatives, etc.)?

(circle one)

All of the time 1

Most of the time 2

Some of the time 3

A little of the time 4

None of the time 5

11. How TRUE or FALSE is each of the following statements for you?

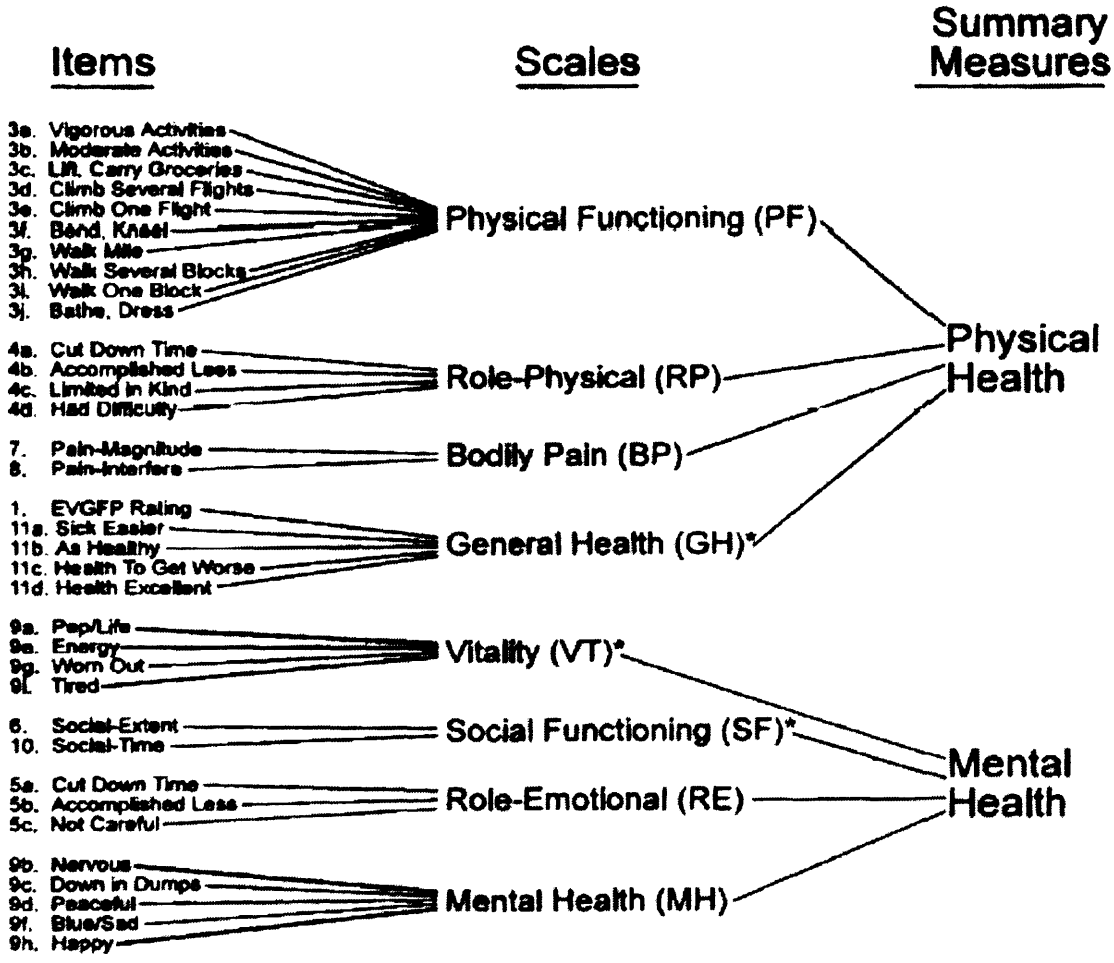
(circle one number on each line)

	Definitely True	Mostly True	Don't Know	Mostly False	Definitely False
a. I seem to get sick a little easier than other people	1	2	3	4	5
b. I am as healthy as anybody I know	1	2	3	4	5
c. I expect my health to get worse	1	2	3	4	5
d. My health is excellent	1	2	3	4	5

Copyright © 1992 Medical Outcomes Trust.
All rights reserved.
(SF-36 Standard U.S. Version 1.0)

APPENDIX IV

SF-36 Health Status Questionnaire
Scales that make up the Summary Measures (PH and MH)



* Significant correlation with other summary measure.

APPENDIX V

Please check that you have ticked <u>one box for each question</u> before going on to the next page

PIMS PARKINSON'S IMPACT SCALE PIMS

NAME: **DATE:**

YEAR SYMPTOMS BEGAN: **DATE OF BIRTH:**

- *Please indicate by a number (0 – 4) what impact Parkinsonism has had on your life.*
  0= no change 1= slight 2= moderate 3= moderately severe 4= severe
- *Use the definitions below to help you measure impact.*

Self: (Positive)	*Refers to how positive you feel about yourself (self-worth, happiness, optimism)*
Self: (Negative)	*Refers how negative you feel about yourself (level of stress, anxiety or depression)*
Family Relationships:	*Refers to your spouse, partner, children and relatives that you consider part of your immediate family*
Community Relationships:	*Refers to your neighbours, friends, people you work with and those who provide you with services (store clerk, doctor, pastor, etc.)*
Work:	*Refers to your job and/or the running of your home and your ability to support yourself and your family*
Travel:	*Refers to your ability to reach your destinations i.e.: work and/or social*
Leisure:	*Refers to your ability to continue enjoyable activities (hobbies, sports, volunteering*
Safety:	*Refers to your ability to do what you want without injuring yourself or others (driving, being outdoors, in the kitchen, in the bathroom, etc.)*
Financial Security;	*Refers to your ability to support yourself and your family and pay your medical costs*
Sexuality:	*Refers to your ability to maintain a satisfactory sexual relationship*

- *If your symptoms are stable complete column 1*
- *If your symptoms fluctuate complete columns 2a and 2b (best and worse)*

	Column 1	***Column 2a (Best)***	***Column 2b (Worst)***
1. Self-positive			
2. Self-negative			
3. Family Relationships			
4. Community Relationships			
5. Work			
6. Travel			
7. Leisure			
8. Safety			
9. Financial Security			
10. Sexuality			

Parkinson's Impact Scale (PMIS) Parkinsonism and Related Orders. 1996 Vol 2. No 2. pp 55-61

APPENDIX VI

Hoehn and Yahr Staging of Parkinson's Disease

Patient Initials: |__|__|__| **Date of Visit:** |__|__|__|__|__|__|
D D M M Y Y

Date of birth: |__|__|__|__|__|__|
D D M M Y Y **Sex: Male/Female** |__|

Assessor Name:

.....

...

Please indicate by ticking the appropriate box whether the patient is an early stage Parkinson patient or has advanced Parkinson's disease:

Early stage PD patient o

Advanced Parkinson's Disease o - **Please indicate which classification best describes the patient:**

Insufficient control of motor symptoms o
Motor fluctuations o
Stable o

Based on your clinical examination today, please indicate by (circling) **what you consider the patient's staging of Parkinson's disease to be:**

Stage 1 Stage 2 Stage 3 Stage 4 Stage 5

Reference

Stage One 1. Signs and symptoms on one side only 2. Symptoms mild 3. Symptoms inconvenient but not disabling 4. Usually presents with tremor of one limb 5. Friends have noticed changes in posture, locomotion and facial expression	**Stage Four** 1. Severe symptoms 2. Can still walk to a limited extent 3. Rigidity and bradykinesia 4 No longer able to live alone 5. Tremor may be less than earlier stages
Stage Two 1. Symptoms are bilateral 2. Minimal disability 3. Posture and gain affected	**Stage Five** 1. Cachectic stage 2. Invalidism complete 3. Cannot stand or walk 4. Requires constant nursing care
Stage Three 1. Significant slowing of body movements 2. Early impairment of equilibrium on walking or standing 3. Generalised dysfunction that is moderately severe	

APPENDIX VII

UNIFIED PARKINSON'S DISEASE RATING SCALE

MENTATION, BEHAVIOR AND MOOD

Score these items "as is," that is, regardless of exact etiology

		0	1	2	3	4
1.	Mentation	☐	☐	☐	☐	☐
2.	Thought disorder	☐	☐	☐	☐	☐
3.	Depression	☐	☐	☐	☐	☐
4.	Motivation/ initiative	☐	☐	☐	☐	☐

ACTIVITIES OF DAILY LIVING

Score this historical data when the patient is at their best and worst, i.e. when "ON" and when "OFF for those patients who have clinical fluctuations from levodopa therapy.

		"ON"					"OFF"				
		0	1	2	3	4	0	1	2	3	4
5.	Speech	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
6.	Salivation	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
7.	Swallowing	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
8.	Handwriting	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
9.	Cutting food	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
10.	Dressing	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
11.	Hygiene	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
12.	Turning in bed	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
13.	Falling	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
14.	Freezing	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
15.	Walking	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
16.	Tremor	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
17.	Sensory symptoms	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

UNIFIED PARKINSON'S DISEASE RATING SCALE (cont.)

Is the patient "ON" or "OFF"? ☐ "ON" ☐ "OFF"

MOTOR EXAMINATION

		0	1	2	3	4
18.	Speech	☐	☐	☐	☐	☐
19.	Facial expression	☐	☐	☐	☐	☐
20.	Tremor at rest					
	Face, lips, chin	☐	☐	☐	☐	☐

		RIGHT					LEFT				
		0	1	2	3	4	0	1	2	3	4
	Hands	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Feet	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
21.	Action tremor	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
22.	Rigidity – neck	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	UE	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	LE	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
23.	Finger taps	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
24.	Hand movement	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
25.	Hand (pronation – supination)	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
26.	Leg agility	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

		0	1	2	3	4
27.	Arise from chair	☐	☐	☐	☐	☐
28.	Posture	☐	☐	☐	☐	☐
29.	Gait	☐	☐	☐	☐	☐
30.	Postural stability	☐	☐	☐	☐	☐
31.	Body bradykinesia	☐	☐	☐	☐	☐

UNIFIED PARKINSON'S DISEASE RATING SCALE (cont.)

COMPLICATIONS OF THERAPY

This section should represent the status of the patient in the week prior to the examination

DYSKINESIAS		0	1	2	3	4
32.	Duration	☐	☐	☐	☐	☐
33.	Disability	☐	☐	☐	☐	☐
34.	Pain	☐	☐	☐	☐	☐
35.	Early morning dystonia	☐	☐	☐	☐	☐

CLINICAL FLUCTUATIONS		0	1	2	3	4
36.	"Offs" duration	☐	☐	☐	☐	☐
37.	"Offs" predictable	☐YES			☐NO	
38.	"Offs" unpredictable	☐YES			☐NO	
39.	"Offs" sudden	☐YES			☐NO	

OTHER COMPLICATIONS

40.	Anorexia, nausea, vomiting	☐YES	☐NO
41.	Sleep disturbances	☐YES	☐NO
42.	Symptomatic orthostasis	☐YES	☐NO

APPENDIX VIII

DYSKINESIA SUBJECTIVE RATING SCALE

1. Handwriting/ Drawing

(0) ☐	No dyskinesias are evident when you write which is accomplished as well as your Parkinson's disease allows at the best of times (this may be small or partly illegible).
(1) ☐	When dyskinesias are at their current maximum, writing is no different from the best performance without dyskinesias (writing may be small or illegible at the best of times but dyskinesias do not interfere further with writing).
(2) ☐	When dyskinesias are at their current maximum, they interfere with writing but allow you to complete the writing task successfully with only a slight increase in the time or effort required and no further increase in illegibility compared with the best writing performance without dyskinesias.
(3) ☐	When dyskinesias are at their current maximum, they cause substantial impairment in writing or drawing compared with the best performance without dyskinesias; result from the dyskinesias writing is more illegible than without dyskinesias and drawing shows more errors.
(4) ☐	When dyskinesias are at their current maximum, they make writing and drawing impossible. Because of dyskinesias, it is not possible to maintain contact between pen and paper long enough to write or draw (that is, it is not possible to produce even small, illegible script).

2. Cutting Food and Handling Utensils

(0) ☐	No dyskinesias are evident when you feed yourself which is accomplished as well as your Parkinson's disease allows at the best of times.
(1) ☐	When dyskinesias are at their current maximum, they do not interfere at all with cutting food and feeding yourself. The ability to feed yourself is no different from the best performance without dyskinesias (this includes slowness and clumsiness, possibly even requiring help at the best of times as a result of persistent problems with parkinsonism).
(2) ☐	When dyskinesias are at their current maximum, they definitely interfere with feeding but allow you to complete the necessary feeding task with only a slight increase in the time or effort required compared with the best performance without dyskinesias (any assistance required in cutting food or feeding relates mainly to the underlying parkinsonism and persists at the best of times without dyskinesias).
(3) ☐	When dyskinesias are at their current maximum, they cause substantial impairment in feeding compared with the best performance without dyskinesias; as a result of the dyskinesias, more help is required to cut food or feed yourself than at the best of times without dyskinesias.
(4) ☐	When dyskinesias are at their current maximum, they make feeding impossible and even with full assistance from someone else, feeding is exceedingly difficult or impossible because of the presence of dyskinesias.

3. Dressing

(0) ☐	No dyskinesias are evident when you get dressed (including outerwear to go outdoors) which are accomplished as well as your Parkinson's disease allows at the best of times.
(1) ☐	When dyskinesias are at their current maximum, they do not interfere at all with the performance of dressing tasks. These are completed no differently from their best performance without dyskinesias (this includes slowness and clumsiness, possibly even requiring help at the best of times as a result of persistent problems with parkinsonism).
(2) ☐	When dyskinesias are at their current maximum, they definitely interfere with the performance of dressing tasks but allow you to complete them with only a slight increase in time or effort compared with the best performance without dyskinesias (any assistance required in getting dressed, including doing up buttons or getting arms into sleeves, relates mainly to the underlying parkinsonism and persists at the best of times without dyskinesias).
(3) ☐	When dyskinesias are at their current maximum, they cause substantial impairment in the performance of dressing tasks compared with the best performance without dyskinesias; as a result of the dyskinesias, more help is required for dressing than at the best of times without dyskinesias.
(4) ☐	When dyskinesias are at their current maximum, they make the performance of dressing tasks impossible and even with full assistance from someone else, dressing is exceedingly difficult or impossible because of the presence of dyskinesias.

4. Hygiene

(0) ☐	No dyskinesias are evident when you you wash, shower, shave, brush your teeth, and so on, which are accomplished as well as your Parkinson's disease allows at the best of times.
(1) ☐	When dyskinesias are at their current maximum, they do not interfere at all with the performance of hygiene tasks. These are completed no differently from their best performance without dyskinesias (this includes slowness and clumsiness, possibly even requiring help at the best of times as a result of persistent problems with parkinsonism).
(2) ☐	When dyskinesias are at their current maximum, they definitely interfere with the performance of hygiene tasks but allow you to complete them with only a slight increase in time or effort compared with the best performance without dyskinesias (any assistance required in hygiene relates mainly to the underlying parkinsonism and persists at the best of times without dyskinesias).
(3) ☐	When dyskinesias are at their current maximum, they cause substantial impairment in the performance of hygiene tasks compared with the best performance without dyskinesias; as a result of the dyskinesias, more help is required for washing, grooming, and going to the bathroom than at the best of times without dyskinesias.
(4) ☐	When dyskinesias are at their current maximum, they make the performance of hygiene tasks impossible and even with full assistance from someone else, washing, grooming, and going to the bathroom are exceedingly difficult or impossible because of the presence of dyskinesias.

APPENDIX IX

BURDEN INTERVIEW

INSTRUCTIONS: The following is a list of statements, which reflect how people sometimes feel when taking care of another person. After each statement, indicate how often you feel that way; never, rarely, sometimes, quite frequently, or nearly always. There are no right or wrong answers.

1. Do you feel that your relative asks for more help than he/she needs?

   0. Never 1. Rarely 2. Sometimes 3. Quite Frequently 4. Nearly Always

2. Do you feel that because of the time you spend with your relative that you don't have enough time for yourself?

   0. Never 1. Rarely 2. Sometimes 3. Quite Frequently 4. Nearly Always

3. Do you feel stressed between caring for your relative and trying to meet other responsibilities for your family or work?

   0. Never 1. Rarely 2. Sometimes 3. Quite Frequently 4. Nearly Always

4. Do you feel embarrassed over your relative's behavior?

   0. Never 1. Rarely 2. Sometimes 3. Quite Frequently 4. Nearly Always

5. Do you feel angry when you are around your relative?

   0. Never 1. Rarely 2. Sometimes 3. Quite Frequently 4. Nearly Always

6. Do you feel that your relative currently affects your relationship with other family members or friends in a negative way?

   0. Never 1. Rarely 2. Sometimes 3. Quite Frequently 4. Nearly Always

7. Are you afraid what the future holds for your relative?

   0. Never 1. Rarely 2. Sometimes 3. Quite Frequently 4. Nearly Always

8. Do you feel your relative is dependent upon you?

   0. Never 1. Rarely 2. Sometimes 3. Quite Frequently 4. Nearly Always

9. Do you feel strained when you are around your relative?

0. Never 1. Rarely 2. Sometimes 3. Quite Frequently 4. Nearly Always

10. Do you feel your health has suffered because of your involvement with your relative?

0. Never 1. Rarely 2. Sometimes 3. Quite Frequently 4. Nearly Always

11. Do you feel that you don't have as much privacy as you would like, because of your relative?

0. Never 1. Rarely 2. Sometimes 3. Quite Frequently 4. Nearly Always

12. Do you feel that your social life has suffered because you are caring for your relative?

0. Never 1. Rarely 2. Sometimes 3. Quite Frequently 4. Nearly Always

13. Do you feel uncomfortable about having friends over, because of your relative?

0. Never 1. Rarely 2. Sometimes 3. Quite Frequently 4. Nearly Always

14. Do you feel that your relative seems to expect you to take care of him/her, as if you were the only one he/she could depend on?

0. Never 1. Rarely 2. Sometimes 3. Quite Frequently 4. Nearly Always

15. Do you feel that you don't have enough money to care for your relative, in addition to the rest of your expenses?

0. Never 1. Rarely 2. Sometimes 3. Quite Frequently 4. Nearly Always

16. Do you feel that you will be unable to take care of your relative much longer?

0. Never 1. Rarely 2. Sometimes 3. Quite Frequently 4. Nearly Always

17. Do you feel you have lost control of your life since your relative's illness?

0. Never 1. Rarely 2. Sometimes 3. Quite Frequently 4. Nearly Always

18. Do you wish you could just leave the care of your relative to someone else?

0. Never 1. Rarely 2. Sometimes 3. Quite Frequently 4. Nearly Always

19. Do you feel uncertain about what to do about your relative?

0. Never 1. Rarely 2. Sometimes 3. Quite Frequently 4. Nearly Always

20. Do you feel you should be doing more for your relative?

0. Never 1. Rarely 2. Sometimes 3. Quite Frequently 4. Nearly Always

21. Do you feel you could do a better job in caring for your relative?

0. Never 1. Rarely 2. Sometimes 3. Quite Frequently 4. Nearly Always

22. Overall, how burdened do you feel in caring for your relative?

0. Not at all 1. A little 2. Moderately 3. Quite a bit 4. Extremely

APPENDIX X

Change in your condition

Patient's Initials: ☐ **Age:** ☐

Gender: ☐ **Today's date:** ☐

(A) Overall, has there been any change in the effects of your Parkinson's disease on your everyday life since you completed the previous questionnaire?

Please tick one box.

Much better ☐

Better ☐

The same ☐

A little worse ☐

Much worse ☐

(B) *Since your last clinic appointment have you visited, been counselled by or treated by one of the following:*

Speech therapist:	Yes	☐	No	☐
Occupational therapist:	Yes	☐	No	☐
Parkinson's Disease Specialist Nurse:	Yes	☐	No	☐
Social Worker:	Yes	☐	No	☐
Other*:	Yes	☐	No	☐

*If you have ticked 'Yes' to Other, please write what other type of treatment you have had since the last clinic appointment:

_____

_____

_____

APPENDIX XI

Patient Information Sheet

Dear Sir or Madam,

Re: Invitation to participate in research on Health Related Quality of Life in Parkinson's Disease.

Please take a few minutes to read the following information on a research study assessing quality of life in Parkinson's patients.

What is the study for?
I am undertaking a research project at Cardiff University under the supervision of Professor David Luscombe, Dean of the Welsh School of Pharmacy. The purpose of my research is to assess the usefulness and relevance of **Quality of Life** measures for Parkinson's Disease in a clinic and clinical trial environment. The Quality of Life questionnaire is an important tool for assessing how well a patient is responding and functioning to their treatment and also to assess the effect of Parkinson's disease on your daily activities.

What is involved?
I would like to ask patients such as your-self, with a diagnosis of Parkinson's disease, to complete two different quality of life questionnaires; the Short Form Health Survey (SF-36) and the UK-Sickness Impact Profile (UK SIP). These questionnaires which will be posted to you (if you agree to participate), must be completed one to two days prior to your planned outpatient appointment at the neurology clinic at the Ashford Hospital. You will be expected to hand in the completed questionnaires to Dr. Russell Lane (consultant neurologist) at this visit.

An assessment by the consultant neurologist will take place as normal at this scheduled visit. After completion of the two questionnaires, which ought to take no longer than 15 to 20 minutes each, the doctor will assess your level of disability referring to a measurement called the "Hoehn and Yahr Index". The rest of your visit will be as normal.

In about three or six month's time when you are attending the neurology clinic for a scheduled follow-up, you will be asked to repeat the above questionnaires.

Confidentiality
Any information collected will be treated in the strictest of confidence and your name will not be collected on any document. Your name will not be identified in any report.

You may refuse to participate in this research at any time.

What are the benefits of participating?

It is very unlikely that you will obtain any personal benefit from participating in this research but I hope the study will help clinicians and clinical researchers better understand the progression of the disease with respect to quality of life and it's impact on the patient and the carer. Also it is hoped that this study will confirm which of the two questionnaires is most relevant to collecting quality of life data in the Parkinson's disease population.

Further Information
If you are still unsure about the purpose of the study or require further information please contact the Parkinson's Disease Nurse for your area, or else me at:

Contact details: Mr. Tauhid Ali
Research Assistant – Cardiff University
Tel. 0385 334116

Dr. Russell Lane, who is the Consultant Neurologist at the Ashford Hospital, Middlesex, will perform the clinical assessments.

Parkinson's Disease Nurse Specialist at Ashford and St Peters Hospital is Ms. Viv Nicholas (Tel. 01932 872000 ext. 2920)

What Next?
If you are interested in participating in this study you will be provided with an informed consent form to sign. Please do not hesitate to contact the people listed above should you require further information. When you have read and signed the consent form, please return it to Dr. Lane in the stamped addressed envelope, which will be provided. A few days later you will be sent the questionnaires to complete along with conformation of the date of your next outpatient appointment.

If you do not wish to participate you do not have to do anything further.

Thank you for your time and assistance.

Yours sincerely

Tauhid Ali
Research Assistant
Medicines Research Unit
University of Cardiff

APPENDIX XII

INFORMED CONSENT

QUALITY OF LIFE ASSESSMENT OF PATIENTS WITH PARKINSON'S DISEASE

Patient Name _____ (Block capitals)

1. I, undersigned voluntarily agree to take part in this study.

2. I confirm that a full explanation of the purpose and nature of the study has been given or sent to me and that I have read and understood the "patient information sheet".

3. I have been given the opportunity to ask questions about all aspect of the study and have understood the advice and information given as a result.

4. I also understand that failure to take part will in no way prejudice my further treatment.

5. I understand that any publication shall not reveal my name or any other personally identifying information.

Signature _____ Date _____

I confirm that I have explained the nature and purpose of the study to this volunteer and have honestly answered all questions they have raised.

Signature _____ Date _____

Witness _____ Date _____

