Measuring Outcome After Stroke

A thesis submitted to the University of Newcastle Upon Tyne for the degree of

Doctor of Medicine

1998

Paul Jacob Dorman

B Med Sci (Hons), MB BS, MRCP (UK)

TABLE OF CONTENTS

Acknowledgements

I would like to thank Martin Dennis and Peter Sandercock, my joint supervisors, for their continuing enthusiasm, encouragement and support. I am also particularly grateful to: Jim Slattery and Dave Signorini for statistical advice; David Perry for his programming skills; Barbara Farrell for her organisational expertise; Fiona Waddell for her collaboration; the many staff of the Neurosciences Trials Unit for their support; and, Paul Kind for many discussions on health related quality of life.

I would like acknowledge the cooperation of the many patients who agreed to participate in the studies described in this thesis. Without their help, none of this would have been possible.

Finally, I would like to thank Heather for helping me through the darkest days.

This work was supported by grants from the: Medical Research Council and Glaxo-Wellcome plc.

Abstract

Stroke is the second most common cause of death worldwide. However stroke is not invariably fatal and survivors may experience major physical, social and psychological problems. The United Kingdom government identified the improvement of the quality of life of stroke survivors as a key objective in the recent "Health of the Nation" consultative document. The concept of health related quality of life has developed over the past few years. Although there is no one universally agreed definition, there are several instruments which claim to measure at least some aspects of health related quality of life. These instruments have not been extensively tested in stroke patients, so the hypothesis which I shall test in this thesis is as follows: that a simple instrument can prove a feasible, valid, reliable, and clinically useful measure of health related quality of life in stroke survivors.

I selected a simple measure of health related quality of life (the EuroQol questionnaire) and evaluated its validity in a sample of patients with stroke. A small, but important proportion of patients were unable to complete EuroQol questionnaires either by themselves or by interview, so I investigated whether a proxy (e.g. a spouse or carer) could assess the patient's health status after stroke accurately and without bias. Previous studies comparing one or more different health status instruments did not involve strictly random allocation, so could not provide reliable information on the "best" measure of quality of life to use in stroke patients. I therefore performed a study in a sample of survivors of stroke which directly compared the EuroQol and SF-36 by using a strict random allocation of questionnaires. It was not possible to compare quantitatively the reliability and validity of the EuroQol and SF-36; however, a qualitative comparison suggested their reliability was similar and they appeared to be sampling broadly the same areas of health.

I finally investigated patients' perception of their own quality of life after stroke. The data suggested that many disabled stroke survivors might not view survival in a dependent state as badly as one might expect. This somewhat surprising finding will inform decisions about whether to accept the high risks associated with certain treatments (e.g. thrombolysis) in order to reduce the chances of survival in a dependent state. Assessments of health related quality of life may therefore provide a more comprehensive and relevant view of the patients' outcome than simple measures of disability or impairment.

1 Introduction: Measuring Health and Outcome after Disease

1.1 Historical aspects

1.1.1 Mortality

John Graunt, a 17th century tradesman turned statistician, is credited with taking the first systematic approach to measuring the health of the population. He used the crude data of the London Bills of Mortality to construct the first life table, a table of probabilities of surviving to a given age (Greenwood, 1970). His interest also extended to measuring the frequency and outcome of specific diseases. For instance, he reported that, in a series of years, out of more than a quarter of a million deaths, only 392 were assigned to syphilis, from which he inferred that syphilis had been significantly under-reported as a cause of death. The publication of the life table and the appreciation of its potential weaknesses established Graunt's reputation as a pioneer medical statistician. His interest in measuring the health of the population and the outcome of specific diseases was, however, restricted to mortality. This may reflect simply that the acute fatal infectious diseases were the major public health concern of the time. It is not surprising that non-fatal chronic diseases were given scant attention. A list of causes of death in London for the year 1798-1799 showed how infrequently stroke was diagnosed (Anonymous, 1800a; Anonymous, 1800b). Of the 18,134 deaths that year, 4,843 were due to consumption and 1,111 to smallpox; only 249 were classified as "apoplexy or suddenly" (the 12th commonest cause of death). Epidemiology became established as a useful discipline to assess the distribution, incidence and outcome (usually mortality) of infectious disease. This work saw practical application in the cholera epidemic of 1854, when John Snow, using data from a simple observational epidemiological study, was able to identify the

intervention that directly led to successful disease prevention (Lilienfeld and Lilienfeld, 1980): he took the handle off the public water pumps delivering contaminated drinking water.

1.1.2 Morbidity

Chronic diseases are, by definition, not immediately fatal, so mortality data cannot reliably assess the impact of chronic disease on the health of a population, yet few early studies aimed to examine the morbidity of the population. William Farr, a general practitioner and freelance medical journalist in the first half of the nineteenth century, was one of the first to estimate both the prevalence of morbidity and the proportion of patients permanently incapacitated by disease (Greenwood, 1970). Farr's data were scanty. As national data were not available, he had to base his analyses on reports from a few benefit societies and the returns relating to workers in the Royal Dockyards and employees of the East India Company. Starting with the assumption:

> "in manhood, for every death we may reckon two persons constantly sick"

He estimated that approximately 600,000 people were constantly sick in England and Wales at any one time (population 14 million) and that approximately 2% of labourers were kept constantly at home by illness. Farr suggested that this might reduce the productive power of the community by one-seventeenth part. Clearly, interpretation of Farr's estimates depends to a large extent on the basis by which he classified individuals as "sick".

Farr, like Graunt, appreciated the need for a uniform international classification of diseases:

" The advantages of a uniform statistical nomenclature, however imperfect, are so obvious, that it is surprising no attention has been paid to its enforcement in Bills of Mortality. Each disease has, in many instances, been denoted by three or four terms, and each term has been applied to as many different diseases..... The nomenclature is of as much importance in this department of inquiry as weights and measures in the physical sciences, and should be settled without delay."
(Registrar General of England and Wales, 1839)

An internationally accepted classification of the causes of death was first published in 1893 (World Health Organisation, 1967). It provided the first effective tool for measuring health within a population and comparing health between populations. When the major force of a disease was expressed as acute illness (of which acute infections provided the most notable example), the public health burden associated with it could be readily assessed by measurement of the occurrence of the disease, recovery from it and fatal outcome.

However, classification of the causes of death did not solve the problem of measuring the public health burden of chronic diseases:

"with the prominent place which the study of infectious and acute diseases has taken in the last few years, there has been a lamentable neglect ..... of the still more fundamental problems underlying the long list of constitutional and chronic diseases" (Anonymous 1902)

As life expectancy has increased in the second half of the twentieth century and, correspondingly, peoples' expectations of a disease-free life have risen, medical interest began to focus on chronic diseases such as stroke.

1.2 Stroke: a major burden on society

The burden of disease attributable to stroke is large. Stroke accounts annually for approximately 4.4 million deaths worldwide and is the second most common cause of death after ischaemic heart disease (Murray & Lopez, 1997). In Britain, stroke represents the third most common cause of death (after ischaemic heart disease and cancer) (The Secretary of State for Health, 1991). Approximately 105,000 British people suffer their first-ever stroke every year (Bamford *et al.* 1988). Of these, about 31% will be dead at one year after the stroke onset (Bamford *et al.* 1990).

However, death is only one aspect of the public health burden attributable to stroke. About one third of patients still alive one year after their stroke require help with activities of daily living (Bamford *et al.* 1990), and stroke is one of the major causes of severe disability in Britain (Martin et al. 1988). Furthermore, only about one-sixth of patients who survive the acute phase of their stroke recover completely (i.e. to their pre-stroke state) (International Stroke Trial Collaborative Group, 1997). Many stroke survivors who are apparently not disabled report problems in "higher level" activities such as domestic management, participation in social activities and in their psychological well-being (Lawrence & Christie, 1979; Duncan *et al.* 1997). In particular, various studies have reported that depressive illness affects between 23% and 63% of patients in the first year after a stroke (Burvill *et al.* 1995; Wade *et al.* 1987; Ebrahim *et al.* 1987), which is more than twice the proportion in the general elderly population (House, 1987) or in populations matched for physical disability (Folstein *et al.* 1977).

Stroke therefore places a large burden of physical, psychological and social difficulties on the families and the carers of patients with stroke. The disease places a similarly large financial burden on the providers of health care and society in

general. Isard and Forbes recently estimated that stroke accounts for approximately 4.3% of all Scottish NHS resources, and 5.5% of hospital resources (Isard & Forbes, 1992).

Stroke has been targeted as a public health priority in the United Kingdom in the recent "Health of The Nation" report. The Secretary of State for Health identified the following objective:

> "To reduce the occurrence of stroke and associated death and disability and to ensure the maximum quality of life for survivors" (The Secretary of State for Health, 1991)

In order to achieve this objective, appropriate tools to measure the impact of stroke in individuals and in the population as a whole are needed. A single instrument is unlikely to fulfil these distinct applications.

1.3 Assessing outcome after stroke

1.3.1 Early approaches: development of stroke specific measures

Dyken and White conducted the first randomised controlled trial in patients with acute stroke (Dyken & White, 1956). They wrote:

> "At first an elaborate system of physical grading of recovery was devised, but, as the experiment progressed, it became obvious that the most important point in evaluation was whether the patient lived or died"

Around the same time, Rankin appreciated that death was not the only bad outcome after stroke, since many patients survived, yet had severe disability. In his study of the factors which determined the prognosis of stroke patients, he reported (as an incidental finding) a system to grade the degree of functional recovery (Rankin, 1957):

> "Grade I. *No significant disability*: able to carry out all usual duties. 36 (19%) surviving patients.
>
> Grade II. *Slight disability*: unable to carry out some of previous activities but able to look after own affairs without assistance. 87 (45%) surviving patients.
>
> Grade III. *Moderate disability*: requiring some help but able to walk without assistance. 33 (17%) patients.
>
> Grade IV. *Moderately severe disability*: unable to walk without assistance and unable to attend to own bodily needs without assistance. 4 (2%) patients
>
> Grade V. *Severe disability*: bedridden, incontinent and requiring constant nursing care and attention. 32 (17%) patients"

About one third of the survivors in his study had moderate to severe disability. Rankin's system focused on the patient's ability to perform "usual duties", their level of mobility and ability to perform activities of daily living. Its content is therefore clearly relevant to patients with stroke. However, in common with other assessments of the time, it depended on a doctor (or other observer) making a supposedly objective assessment of the patient's outcome. The patient's subjective assessment of their own functioning made no contribution to the assessment. The paper did not specify how the observations were to be made; so the observer might, on the one hand, just grade the patients by quick reference to the medical records or, alternatively, make prolonged and detailed observation of the patient's abilities over several days. It is clear that administering the same scale with the same patient in

these two different ways could lead to very different gradings. Furthermore, Rankin did not describe how the scale was developed, how the different levels of outcome were arrived at and whether they were valid. Perhaps he considered the development of the scale an unimportant part of his study; in any event, although the Rankin scale was not developed with today's rigorous clinimetric methods, it is still widely used. This may be because it is so simple and relevant.

Barham Carter, like Dyken and Rankin, also recognised that randomised trials of treatment for stroke should measure more than just the effect of treatment on death (Barham Carter, 1961). He conducted a randomised trial of anticoagulant therapy in 76 patients with progressing stroke and, like Rankin, devised his own system for grading the patients' recovery:

> "The patients were followed for six months and then an assessment was made by clinical examination, dividing them into four groups: recovered, improved, not improved, and died.
>
> Recovered - This meant that the patient noticed slight or no disability and was able to resume normal life and to return to full work....
>
> Improved - This meant some useful movement was possible at the elbow and wrist and that the patient required only slight assistance to get about and live a useful life.....
>
> Not improved - This implied no return of arm movement below the shoulder and a poor recovery in the leg so that considerable assistance was needed for the patient to get about. Return to work was not possible, and this group included the bedridden patient."

Barham Carter's system shared Rankin's emphasis on mobility, return to work and normal activities. However, it differed in two important respects. Firstly, Barham

Carter also graded patients on the severity of their neurological deficit. Although this system was not adopted by other investigators, the idea of using aspects of the neurological examination to assess patients' outcome after stroke became a popular approach and within several years many alternative scales for grading the results of the neurological examination had been proposed (Table 1.1). Secondly, Barham Cater's system classified patients according to their change in health with treatment, whereas Rankin classified patients according to their health state at follow up. This difference highlights the ambiguity that exists over what constitutes a measure of "outcome" (Kilgour-Christie & Watt, 1993). Barham Carter's scale could be considered as a measure of outcome, and the Rankin scale a measure of health status. Alternatively, both could be viewed as a measure of outcome. I will follow the latter approach in this thesis.

All of the early methods to assess outcome after stroke were developed empirically, rather than on the basis of an underlying conceptual theory, by clinicians with an interest in stroke. Similar "specific" measures were developed at the same time for other chronic disorders (Kind, 1988). Subsequently, more general health measures have been developed which encompass a broader range of dimensions than any condition-specific system. Indeed, the measurement of health status has become a focus of interest for health researchers and policymakers and two distinct approaches to classify health have emerged. These are the International Classification of Impairments, Disabilities and Handicaps and the broader, overlapping, concept of **Quality of Life.**

1.4 International Classification of Impairments, Disabilities, and Handicaps

In response to the growing number of different classifications and methods of grading outcome after disease, the World Health Organisation developed the International Classification of Impairments, Disabilities and Handicaps (ICIDH) to clarify the taxonomy of the consequences of disease (World Health Organisation, 1980). Each of these three categories were developed to represent a discrete and independent plane of experience consequent upon disease.

1.4.1 Definition of terms

Impairments are defined as the loss or abnormality of psychological, physiological, or anatomical structure or function. They are associated with signs on clinical examination, and so represent the focus of the routine medical examination. The elements of impairment included in the neurological examination have been widely used in the construction of a variety of so called "stroke scales", (Table 1.1). These scales, derived by arbitrarily summing scores related to various physical impairments detected during neurological examination, were developed to provide a means of describing the severity of the patients' stroke in randomised trials of medical interventions. The Mathew Scale, for example, was first used to describe patients at baseline and at follow up in a trial of glycerol (Table 1.2) (Mathew *et al.* 1972).

Disability assesses function at the level of the person and is defined as:

> "any restriction or lack of ability to perform an activity in the manner or within the range considered normal for a human being."

It is therefore concerned with the compound or integrated activities expected of the person or of the body as a whole, such as are represented by tasks, skills, and behaviours.

Handicap is the highest level of measurement in this classification. It is defined as:

> "a disadvantage for a given individual (resulting from an impairment or a disability) that limits or prevents the fulfillment of a role that is normal (depending on age, sex and cultural factors) for that individual."

The ICIDH handicap section defines six basic "survival roles," which describe disadvantage in orientation (the ability to perceive and understand the immediate environment including sight, hearing, and cognition), physical independence (from human or mechanical assistance), mobility (the distance one can move from one's bed), occupation (employment, domestic work and recreation), social integration and economic self sufficiency (including the ability to earn an income and the possession of resources enabling problems to be overcome). Handicap may be classified according to the disadvantage associated with deficiencies in each dimension.

1.4.2 Application of the ICIDH to stroke

The theoretical model of ICIDH provided a useful means of describing the consequences of disease because it enabled a clear distinction to be made between the intrinsic experience of disease that causes impairments and disability, and the external factors (e.g. poverty, poor environment) that lead to handicap. Furthermore, the clarity of the definition makes it more likely that assessment of the effect of an intervention will be carried out at an appropriate level. However, there are advantages and disadvantages when measuring outcome at each of these levels.

Stroke scales became a popular means for assessing outcome after stroke. In the twenty year period between 1969 and 1990, over 20 different stroke scales were devised (van Gijn, 1992). Although these scales provided a relatively objective and sensitive means of assessing outcome, researchers recognised a number of shortcomings in the scales. The most serious difficulties were poor interpretability, little relevance to patients and limited generalisability (van Gijn, 1992; van Gijn & Warlow, 1992). In particular, many scales focus on impairments (i.e. the neurological signs), which are much less important to a patient's daily life than disability and handicap (i.e. what they can and what they actually do). In general, stroke scales, if they have any value at all, are probably more suitable for small scale trials being undertaken early in the development of an intervention. Such trials generally seek to examine whether the treatment has any effect at all on the disease process. Stroke scales are less suitable as a primary measure of outcome for large scale clinical trials which aim to determine the role of the intervention in routine clinical practice (van Gijn, 1992). Scales which assess impairment cannot provide estimates of the burden that stroke places on society and on health care providers; only scales that measure disability and handicap are relevant here.

Handicap is at the other end of this hierarchical classification of outcome. In contrast to impairments, assessments of handicap provide information which is more relevant to patients and is more generalisable. However, these assessments are limited by the subjective and relatively insensitive nature of this outcome (many factors other than the disease or the intervention determine the eventual handicap). Furthermore, until recently, the measurement of handicap has been less well defined and tested than that of impairment or disability (van Gijn, 1992; Harwood *et al.* 1994a). The Oxford Handicap Scale was derived from the Rankin scale (Bamford *et al.* 1989); however, although it is termed a handicap scale, it focuses on disability rather than handicap and only covers the physical independence dimension of handicap in the

ICIDH classification. Harwood and colleagues have recently developed a generic measure of handicap - the London Handicap Scale – based on the ICIDH dimensions of handicap (see above) (Harwood et al. 1994b). The London Handicap Scale comprises a classification questionnaire and a matrix of scale weights. The classification questionnaire has six questions, one for each dimension, each comprising a six point hierarchical scale of disadvantages in a self-completion format. The matrix of scale weights enables the severity of disadvantage in each dimension to be combined into an overall handicap score, ranging from 1 (no handicap) to 0 (maximum handicap). Initial studies suggest that it has acceptable measurement attributes for group applications after stroke (Harwood et al. 1994b).

Disability lies between impairment and handicap in the hierarchical classification proposed by the World Health Organisation and so represents a compromise between impairment and handicap in terms of relevance, sensitivity, objectivity and generalisability. It has become a widely accepted level at which to measure outcome after stroke. There are many different instruments which can measure disability. They generally focus on independence and the ability of patients to perform activities of daily living. Of these, the Barthel Index has become very widely used (Mahoney & Barthel, 1965; Wade & Langton Hewer, 1987; Wade, 1992). It includes ten activities of daily living and the maximum score is 100, Table 2. The scale is hierarchical, at least in patients with stroke, in that an ascending order of difficulty can be attributed to the activities listed (Wade & Langton Hewer, 1987). However, simple disability scales, such as the Barthel Index, have consistent ceiling effects. For example, a substantial proportion of patients after stroke report persistent problems in their daily life, yet score 100/100 on the Barthel Index which suggests they should have no significant disability (Wellwood *et al.* 1995; Duncan *et al.* 1997).

Holbrook and Skilbeck recognised the limitations of scales which focus on self care (such as the Barthel Index) and developed the Frenchay Activities Index to examine broader aspects of everyday living after stroke (Holbrook and Skilbeck, 1983). The Frenchay Activities Index measures lifestyle in terms of complex domestic, leisure and social functioning and can be administered, by interview or self-completed questionnaire, in a few minutes. It does not appear to have the ceiling effect noted with the Barthel Index and principal components analysis suggests that it shows two traits: instrumental disability and some aspects of handicap (Schuling et al. 1993).

Lindley and colleagues developed two simple questions to assess outcome after stroke in large scale epidemiological studies (Lindley et al. 1994). Their content includes aspects of disability ("do you require help from another person for everyday activities?") and handicap ("do you feel that you have made a complete recovery from your stroke?"). With this approach patients may be classified as "dependent", "independent, but not fully recovered", and "fully recovered" and so the ceiling effects associated with simple measures of disability may be avoided. The relationship between these questions and measures of health related quality of life is examined in Chapter Eight.

1.5 Quality of Life

1.5.1 Background

Life expectancy has increased substantially over the last 150 years. A century ago life expectancy at birth was only 44 years for males and 47 years for females; life expectancy is now 73 years for males and 78 years for females (The Secretary of State for Health, 1991). In spite of this, many major health problems remain

unresolved, and increased life expectancy has been associated with an increased prevalence of disability in the population (Gruenberg, 1977; Wilkins & Adams, 1983). This epidemiological transition has been associated with a change in attitudes; in that patients want to live better, and not merely longer.

Society's main response to these problems continues to be increased investment in health care, including preventative, caring and curing interventions, and the last fifty years has seen the emergence of many dramatic new health care technologies. However, rising health care expenditures have emphasised the need to look at the relative costs and benefits of different interventions, and increasingly health care providers demand that additional expenditure is justified by improved health outcomes.

Impairments and disabilities have been recognized as incomplete and inadequate measures of outcome in chronic disease (Harwood *et al*, 1994), and the growth of the consumer movement and criticism of the biomedical model has pointed to the need to bring patients' values and needs into the medical decision making process. These concerns created a demand for new indices of health which reflect patients' values and views.

1.5.2 Defining health status

Many definitions of health exist. Most are variants of the World Health Organisation definition:

> "Health is a state of complete physical, mental and social wellbeing and not merely the absence of disease or infirmity" (World Health Organisation, 1948)

Several groups have built on these constructs to develop concepts of health that can be operationalised. Bergner conceptualised health status as a multidimensional construct which contains those elements that are an integral part of the person but excludes those that exist and behave independently of that person (Bergner, 1985). She, therefore, has a notion of health status that "ends at the skin". Bergner identifies five interrelated dimensions of health status:

1. The genetic foundation: forms the basic structure on which all other aspects of health status must build.
2. The biochemical, physiological, or anatomic condition: includes disease, disability or handicap whether obvious or not.
3. The functional condition: includes performance of the usual activities of life, such as working, walking and thinking.
4. The mental condition: includes self-perception of mood and emotion.
5. The health potential of the individual: includes longevity, functional potential, and the prognosis of the disease or disability.

She proposed that four groups of other factors interact with health status: societal factors, health care system factors, social and familial factors and personal factors, see Figure 1.1.

Ware has proposed a similar model of health, see Figure 1.2 (Ware, 1984). This model shows disease in the centre, and personal functioning, psychological distress and wellbeing, general health perceptions, and social/role functioning in the surrounding boxes. This model indicates that health status moves from characteristics at the centre (intrinsic to an individual) to characteristics outside an individual, in a behavioural, social and cultural perspective that links the experience of disease and treatment at the level of the individual, the immediate social environment,

or the larger society (Patrick & Erickson, 1993). Ware's emphasis on general health perceptions is a major difference between his model and that suggested by Bergner.

1.5.3 What is Quality of Life?

The use of the term "quality of life" has grown exponentially over the last 30 years (Testa & Simonson, 1996). Spitzer reviewed the content of 33 papers that included the term in their title in the five year period 1970-1974. He found the specific subject areas included: the contribution of the physician to quality of life of people, the quality of life of hospital patients, the quality of life in schools and universities, the quality of life in old age, and the quality of life in poverty. In more recent years, it has also been evoked as a target outcome for treatment in a number of specific diseases, particularly cancer, renal failure, hypertension, coronary artery bypass surgery, arthritis, hearing impairment and specific mental illnesses (Spitzer, 1987). Thus, the concept of quality of life is subject to numerous interpretations, but clearly designates a prominent goal in a wide range of endeavours of patient care, health care systems and social programs. To the layperson, quality of life is of the highest importance and acts as the driving force behind all actions (Leplege & Hunt, 1997).

Quality of life has been conceptualised in a variety of ways. These meanings reflect the particular knowledge, experience, and values of each individual. Subjective definitions of quality of life include the following concepts: wellbeing, life satisfaction, morale and happiness. The judgement of how satisfied people are with their present state of affairs is based on a comparison with a standard that each individual sets for him or herself. Calman defined quality of life as the extent to which an individual's hopes and ambitions are matched by experience (Calman, 1984).

Alternatively, objective indicators may be used to assess quality of life. Social researchers have defined quality of life as the sum total of the individual's scores on characteristics that can be objectively determined, e.g. assets, housing, access to leisure facilities (Tate et al, 1996). Other less tangible aspects of human existence may also be considered, e.g. safety, respect, love and freedom. However, the health care system and its providers do not usually assume responsibility for these more global human concerns, even though they may adversely affect or be affected by disease and treatment. Therefore, a final more specific approach is to focus on those aspects of quality of life that are directly related to health, disease, treatment or policy (Brooks, 1991). Brooks argues that there is a risk in extending the definition of quality of life too far, and highlights the important potential distinction between health and non-health related aspects of quality of life:

> "If the focus is to be on health interventions, should we not be attempting to focus our analysis on health related entities. Are we in danger of becoming too holistic? One way to answer these questions would be to concentrate on what might be termed "health related quality of life",..... It would be necessary to select those domains and dimensions deemed of relevance, and to be explicit about defining such health related quality of life." (Brooks, 1991)

Torrance agrees that overall quality of life is an all-inclusive concept incorporating all factors that impact upon an individual's life; whereas health related quality of life includes only those factors that are part of an individual's health (Torrance, 1987). He suggests that physical functioning and emotional functioning, taken together, constitute health related quality of life. Social functioning, although important to an individual's overall quality of life, is considered "beyond the skin" and not an aspect of health related quality of life (Torrance, 1987). This view that, for health studies, investigators should focus on health related issues and preferentially use the term

"health related quality of life" has been widely supported by many workers (Lohr, 1992; Guyatt et al, 1993; Levine, 1995). However, the concept that quality of life can be dissected into its health and non-health-related components has been challenged:

> "This view fails to acknowledge the interconnectedness of health status with other aspects of existence such as changes in income, work status, personal relationships, coping strategies, responsibilities, self-image, and customary modes of being. This interconnectedness makes the project of specifically measuring health related quality of life improbable." (Hunt, 1995; Leplege & Hunt, 1997).

The lack of consensus regarding the conceptualisation of quality of life has created wide variation in its assessment. Gill and Feinstein attempted to clarify the meaning of the term "quality of life" by systematically reviewing 75 articles which included this term in their title (Gill & Feinstein, 1994). They were surprised to find that, despite the lack of a widely accepted operational definition of quality of life, very few articles included definitions of what they measured (Gill & Feinstein, 1994). Furthermore, no article distinguished quality of life, which takes account of how individuals might react to non-medical aspects of their lives (e.g. jobs, family, friends, and other life circumstances) from health related aspects of quality of life (Gill & Feinstein, 1994).

1.5.4 Is health related quality of life equivalent to health status?

Leplege and Hunt suggest that when the demand for quality of life measures arose, there were no such measures, and so the term "health related quality of life" was coined as a method of justifying the use of the existing health status measures under a new banner (Leplege & Hunt, 1997). This view is supported by the fact that many of the former health status instruments, such as the Sickness Impact Profile and the

Nottingham Health Profile, are now described as measures of health related quality of life. Alternatively, the absence of a MEDLINE term for health related quality of life, until recently, may have caused investigators to label their health related quality of life measures as measures of health status (personal communication, Paul Kind). However, most researchers now appear to use the terms "health status" and "health related quality of life" interchangeably (Spitzer, 1987; McDowell & Newell, 1996; Leplege & Hunt, 1997). This convention will be used from here on in this thesis. However, this approach has been criticised:

> "One of the striking differences between the notion of quality of life and that of health status is level of conceptualization. Quality of life as it is used in clinical research is a vague term without conceptual clarity. It is what investigators mean it to be. Conceptual frameworks for health status, on the other hand, have appeared in the literature, have been discussed and debated, and have provided the underpinnings of several measures." (Bergner, 1989)

1.5.5 Conceptual models for health related quality of life

At least two models of health related quality of life have recently been proposed. These models have striking similarities with the earlier conceptualisations of health status. Wilson and Cleary proposed that measures of health can be thought of as existing on a continuum of increasing biological, social, and psychological complexity (Wilson & Cleary, 1995). This relationship is illustrated schematically in Figure 1.3. As with Bergner's model of health status, the model begins with biological and physiological variables. The next layer are symptom reports, which are expressions of subjective experiences that summarise and integrate data from a variety of sources. Functional status is the next integrative level. In addition to symptoms, the individual's motivation, beliefs, and local environment determine functional status.

General health perceptions represent an integration of all of the previous health concepts. General health perceptions are recognised as amongst the best predictors of the use of general medical and mental health services (Hunt *et al.* 1981). These perceptions are predicted to influence the individual's overall quality of life. Psychological influences can be classified in a variety of different ways in this model, e.g. depression could be classified as a biological factor, a measure of symptom status, or a measure of psychological functioning.

This system provides a useful conceptual basis for linking traditional clinical variables with health related quality of life outcomes. It has much in common with another recent definition of health related quality of life:

> "Health related quality of life is the value assigned to duration of life as modified by impairments, functional states, perceptions and social opportunities that are influenced by disease, injury, treatment or policy" (Patrick and Erickson, 1993)

Patrick and Erickson suggest five distinct semi-hierarchical health related quality of life concepts: opportunity, health perceptions, functional states, impairments, and death or duration of life. However, their approach differs from that proposed by Wilson and Cleary in that it specifically draws attention to the relationship between the quantity and quality of life, and aims to provide a basis for the comparison of the costs of different health care interventions.

1.5.6 Current strategies for the measurement of quality of life

The method used to measure quality of life will depend on the target population and the purpose of the assessments. Assessments may be grouped into those that best measure quality of life at the level of an individual patient or at the level of a whole

population. Four distinct approaches can be identified: generic measures of health related quality of life, disease specific measures, patient generated indices, and utility measures.

1.5.6.1 Generic measures of health related quality of life (health status)

Generic instruments seek to assess health concepts that represent basic human values and are relevant to everyone's health status and well-being. These instruments are the most widely used approach for the measurement of quality of life. The measures are called generic because they are both universally valued and are not specific to any particular age, disease, or treatment group (Ware, 1992). However, the problem (or advantage) with generic multidimensional health rating scales is that the choice of dimensions and the values attached to them are imposed on the patient. The value judgments are those of a group of health professionals, or at best, represent the view of the general public, and may differ importantly from patients' own values. These assessments are therefore potentially less relevant to patients with a particular disease, and less sensitive to change, than instruments which are either disease- or individual- specific. However, from the public health perspective, generic measures of health related quality of life are of greater value than disease- or individual- specific measures. Generic measures can be used to compare health gains from different interventions in different groups of patients. This might then allow health resources to be allocated equitably between competing disease groups in a way that maximises population health gain (Cairns, 1996) (for example, many hip replacements versus a few heart transplants - which gives greater health gain?).

1.5.6.1.1 Development of generic measures of health related quality of life

Health measures may be developed from a specific conceptual theory or empirically. A conceptual theory for a measure may justify its content and relate it to a broader body of theory, showing how the results obtained can be interpreted in light of the theory. In the empirical approach, a smaller number of items may be selected by statistical techniques from a larger pool of items, to provide the best possible prediction of the outcome of interest. Although the empirical approach has a practical appeal, the interpretation of results subsequently obtained with the measure may prove more difficult. In practice, most health status measures have been developed using a combination of both approaches.

The Sickness Impact Profile was an early, influential measure. It was developed to provide an appropriate, valid, and sensitive measure of health status for use in assessing the outcome of health care services (Gilson *et al.* 1975). It was developed as a behaviourally based measure of sickness-related dysfunction. This measure was based on the assumption that the ultimately sought product of health services was the reduction in sickness, where sickness denotes the non-professional definition of illness based on lay observations. The content of the Sickness Impact Profile was developed empirically by asking patients, health care professionals, carers, and healthy individuals to provide statements of sickness related behaviour. In addition, the content of previous disease specific instruments was reviewed. This process yielded 312 unique statements that were grouped into 14 categories. This prototype was empirically refined to a final version with 136 statements in 12 categories. Patients' responses may be used to construct a score for each of the 12 categories, an aggregate score for each of the two dimensions (physical and psycho-social), or a global score for the questionnaire as a whole.

The Quality of Well-Being Scale was another early influential health measure. It summarised a patient's health status and symptoms in a single number that reflects a judgement on the social undesirability of the problem (Patrick *et al.* 1973). It consists of three ordinal scales on dimensions of daily activity: mobility, physical activity, and social activity. Scores on these scales can be linked with a separate classification of symptoms and problems. It has become widely used for health economic analyses.

The Sickness Impact Profile influenced the content and development of later health status measures, such as the Nottingham Health Profile. One major difference is that the Nottingham Health Profile asks about feelings and emotional states directly, rather than via changes in behaviour. This emphasis on perceived health status recognises its importance as a predictor of need for, and utilisation of, health services (Hunt *et al*, 1981).

The development of the Short Form 36 (SF-36), one of the newest health related quality of life measures, was strongly influenced by its underlying conceptual model (Ware & Sherbourne, 1992; Ware, 1992). The SF-36 resulted from a process of questionnaire development which began during the Health Insurance Experiment, conducted by the Rand Corporation between 1974 and 1982, to construct the best possible scales for measuring a broad array of functioning and wellbeing concepts for children and working age adults. The Health Insurance Experiment was followed by the Medical Outcomes Study that provided the opportunity for a large-scale test of the feasibility of self-administered patient questionnaires and generic health scales for adults with chronic conditions, including the elderly. The Medical Outcomes Survey was based on a two dimensional model of health: physical and mental. In this framework, social health is not considered to be conceptually equivalent to the dimensions of physical and mental health and is incorporated primarily as an indicator of these dimensions (Ware, 1992). The Medical Outcomes Survey framework

describes five categories of indicators of physical and mental health: 1) clinical status, 2) physical functioning and wellbeing, 3) mental functioning and wellbeing, 4) social/role functioning and wellbeing, and 5) general health perceptions. Functioning pertains to the ability to perform various daily activities and functions, whereas wellbeing refers to more subjective internal states not observable by others (such as symptoms and feelings).

The EuroQol instrument was developed, by a collaborative multinational European group, as a generic instrument for the description and valuation of health related quality of life (The EuroQol Group, 1990). It was intended that the EuroQol should complement other forms of outcome measures, and should facilitate the collection of a common data set for reference purposes. The instrument was designed as a self completed questionnaire for use in large scale surveys of the community. It has therefore balanced the desire to cover all potentially relevant topics against the need to be practical. The EuroQol questionnaire covers five dimensions of health: mobility, self-care, activities, pain and mood; and was designed to produce a single index value of health for any given health state. It also includes a visual analogue scale that allows respondents to report their valuation of their own overall health state. The dimensions were selected following a detailed examination of the descriptive content of existing health status measures including the Quality of Wellbeing Scale, the Sickness Impact Profile, the Nottingham Health Profile, and the Rosser Index (Rosser & Watts, 1972).

1.5.6.1.2 Empirical evidence for the determinants of generic measures of health related quality of life

Lay attitudes to health are important as they may affect the meaning and performance of health status assessments. Herzlich's study of French lay-people

demonstrated that they could distinguish between negative and positive health concepts (Herzlich, 1973). This is important because it lends support to the World Health Organisation definition of health, i.e. health is not merely the absence of disease or infirmity. This finding has been confirmed in several more recent studies (Blaxter, 1990; van Dalen et al, 1994). Blaxter examined attitudes to health in a large United Kingdom "Health and Lifestyles Study". She also found that good health was viewed as a positive concept with many degrees (Blaxter, 1990), and clear evidence for the multidimensionality of health. Her respondents reported the following concepts of health most frequently: 1) unable to answer, 2) health represents the absence of disease (negative concept), 3) health is being functionally able, 4) health is being psychologically fit, 5) health is leading a healthy life, and 6) health is being in good health for age. Blaxter also found that the concepts of health differed with age, sex, education and whether individuals were talking about health in general or their own health in particular.

The results from these studies provide empirical support for multidimensionality of health related quality of life, but also for the importance of positive aspects of health. However, generic health related quality of life measures have been criticised for containing items that may be irrelevant to people with the condition under study, and for omitting others that are pertinent (Bowling, 1995). Bowling examined which dimensions of life people perceive to be important in relation to quality of life and how different conditions impact on peoples' lives (Bowling, 1996). She found that respondents reported the following, most frequently, as being important to their lives: finances, standard of living, relationships, own health, health of relevant others, and social life and leisure. However, respondents identified a different set of issues as important for the effects of illness on life. For the effects of illness on life the most frequent responses were: ability to get out and about, ability to stand, ability to walk, ability to go out shopping; social life and leisure activities; work; symptoms;

instrumental activities of daily living; psychological problems; and other restrictions on activities. Responses differed somewhat with age and gender. For example, women were more likely than men to identify psychological problems as important. Responses differed even more substantially with illness experience. For example, respondents with mental health disorders were most likely to report "availability to work" or "ability to work" as the first most important effect; whereas respondents with digestive or endocrine disorders were most likely to report "dietary restrictions". Bowling concludes that this finding supports the wider use of disease specific measures of outcome (Bowling, 1996).

Williams and colleagues conducted a similar experiment to examine the content of existing generic measures of health related quality of life (Williams, 1995). They asked a random sample of the general public what the salient features of good or bad health were in themselves or in others. They found that simple measures (Rosser Index and EuroQoL) covered only about 26-36% of the items mentioned spontaneously by the respondents, and the more complex measures (Nottingham Health Profile, Quality of Wellbeing Scale, and the Sickness Impact Profile) provided between 50 and 60% coverage. However, the investigators considered approximately one third of the items mentioned by the public as not relevant to an index designed to appraise variations in health-related lifestyle. They also tested the importance of items in a prompted section of the interview. In these analyses they found there was little to choose between the various instruments.

1.5.6.2 Disease specific measures

Unlike a generic measure of health related quality of life, a disease-specific measure is designed specifically to examine the areas of concern for patients with that

particular disease. Including only the aspects of health related quality of life that are relevant to the disease should improve both the relevance and responsiveness of the measure (Guyatt *et al.* 1993). In some situations, disease specific measures are often no more responsive than generic measures because they are generally created by the addition of further items to existing generic measures of health related quality of life (Vickrey *et al.* 1992; Vickrey *et al.* 1995).

1.5.6.3 Patient generated index (individual quality of life)

Generic or disease specific measures of health related quality of life do not address goals or behaviours important to individual quality of life, such as attendance at religious services, playing darts, or personal finance. Moreover, apparently similar behaviours do not have the same importance for all individuals, and behaviours and events do not necessarily retain the same meaning for an individual over the course of an illness. Assessments of individual quality of life aim to quantify the person's level of functioning in those areas of life that he or she believes to be important and the relative importance of these areas (O'Boyle *et al.* 1992). Several groups have reported methods of assessing individual quality of life in patients with chronic problems (O'Boyle *et al.* 1992; Hickey *et al.* 1996; Paterson, 1996). However, it is unlikely that these assessments could be performed as an emergency, for example patients presenting with acute stroke.

1.5.6.4 Utility measures

These measures are derived from economic and decision theory and are designed to reflect the preferences of patients for differing outcomes in relation to death (Guyatt *et al.* 1993). Utility scores therefore aim to reflect a specific health state and its value to the patient. The utility is summarized as a single number along a continuum that

usually extends from death (0.0) to full health (1.0). Scores less than zero are possible and reflect health states worse than death. Utility scores are perhaps most relevant when they are derived directly from the individual patients who may be asked to rate the value of their own health state. Alternatively, patients' responses to certain multidimensional health profiles (such as the EuroQol) can also be converted to utilities using preferences derived from the general public (Dolan et al. 1995). Such utilities may be more useful from the public health perspective.

1.6 What makes a good measure of outcome?

1.6.1 Measurement must be related to purpose

Many investigators appear to select measures of outcome simply on the basis of "brand familiarity". Whilst this might enhance the interpretability, and potential acceptance, of any subsequent results, a variety of other issues should also be considered when outcome measures are selected. These include the aim of the study, its underlying ethical principle, the level of measurement (i.e. is outcome to be assessed at the level of the individual or the population?), the planned method of administration, and the attributes of the instrument in the population of interest (McDowell and Newell, 1996).

The underlying ethical principle should be considered when investigators are selecting measures to **value** health. Potential valuations include those of the patient, the community, and the health care professionals. Researchers and clinicians must consider the relative merits of the valuations derived from each of these groups, as they may well differ (Leplege and Hunt, 1997). The ethical standpoint of health care providers is that of utilitarianism, i.e. doing the greatest good for the greatest number.

This may create tensions with the ethical perspective of the individual clinician, which should be to make the care of the individual patient his first concern (General Medical Council, 1998).

The issue that the content of the measure must reflect the purpose of the study has already been discussed, see Section 1.4.2 and Section 1.5.6. The majority of the instruments discussed so far were developed primarily, for use in populations of patients, for the purposes of: monitoring trends in health, evaluating the effects of health and social policies, and to assist the allocation of resources in relation to need (Ebrahim, 1995). There is increasing interest in using the same instruments at the level of individual patients. At this level the purposes of measurement might include: the diagnosis of the nature and severity of disease, assessment of prognosis, evaluation of the effects of treatment and the study of aetiological factors (Ebrahim, 1995). Although some of these purposes appear similar, the same measure of outcome might not have the appropriate measurement attributes for both of these applications (Ebrahim, 1995).

1.6.2 Measurement Attributes

The essential attributes on which outcome measures are evaluated are their burden, validity, reliability, sensitivity to change and interpretability.

1.6.2.1 Respondent and Administrative Burden

The respondent burden is defined as the time, energy, and other demands placed on the person responding to the instrument (Scientific Advisory Committee, 1995). This has become particularly important as failure to respond (or not completing the data collection form) are the "Achilles heel" of quality of life assessments in clinical trials (Aaronson, 1992; Fallowfield, 1996). In stroke survivors, who often have cognitive,

speech and language deficits and may be unable to complete complex questionnaires, it is particularly important to minimise "respondent burden".

Administrative burden is defined as the demands (time, financial cost and difficulty) placed on those who administer the instrument (Scientific Advisory Committee, 1995). This is an important factor in determining the resources required for any study.

Large scale studies, involving many thousands of patients, require robust measures of health related quality of life with very low respondent and administrative burden. Such studies are being increasingly performed in patients with stroke. For example, two large randomised controlled trials, each involving about 20,000 patients, have recently been completed in patients with acute stroke (International Stroke Trial Collaborative Group, 1997; Chinese Acute Stroke Trial Collaborative Group, 1997). Similarly, public health researchers also need to assess the health related quality of life of large numbers of stroke patients to assess the quality of care given by health services (Working Group on Outcome Indicators for Stroke, 1997).

1.6.2.2 Validity

The validity of an instrument is defined as:

> "the degree to which the instrument measures what it purports to measure" (Scientific Advisory Committee, 1995)

Many different approaches and a myriad of different terms have been used to describe the components of validity (Streiner and Norman, 1989), but the most important ones are:

Face or content validity refers to the extent to which a measure samples the important areas of interest. This aspect of validity is generally assessed in a qualitative manner. This process may be formalised by the use of lay and expert panels to judge the clarity, comprehensiveness and redundancy of items and scales of an instrument.

Criterion validity is defined as the correlation of a scale with some other measure of the trait under study, ideally a "gold standard" which has been used and accepted in the field (Streiner and Norman, 1989). Criterion validity may be divided into two types: concurrent and predictive validity. With concurrent validity we correlate the new scale with a criterion measure. This is often impossible with quality of life assessments because of the absence of widely accepted criterion measures. Predictive validity refers to the extent to which patients' responses to a scale predict future important events, e.g. the predictive validity of a health related quality of life scale would refer to the extent to which patients' responses to the scale predicted their future health status.

Construct validity is concerned with the degree of agreement or disagreement between a health related quality of life scale and some predefined relationship. These predefined relationships are referred to as (hypothetical) constructs. Construct validity may be demonstrated by showing convergent relationships with related variables or poor correlation (discriminant validity) with dissimilar, unrelated variables.

Validation of an instrument may be considered as a process of hypothesis testing (Streiner and Norman, 1989). An alternative and more general definition is that validity describes the range of interpretations that can be appropriately placed on a measurement score (McDowell and Newell, 1996), i.e. what can we conclude about the person who produced particular scores on the test?

1.6.2.3 Reliability

Reliability is the extent to which a measure is free from random error in the population of interest (Guyatt *et al.* 1993; Scientific Advisory Committee, 1995; Testa & Simonson, 1996). The reliability of a measurement is defined as the proportion of observed variation in scores that reflects actual variation in the aspect of health being measured (McDowell and Newell, 1996). Reliability may be measured by assessing an instrument's internal consistency or its reproducibility. Internal consistency is concerned with the extent to which the component items of a scale are inter-related (Streiner and Norman, 1989). Items showing high levels of internal correlation are assumed to be measuring the same underlying concept. A measure's reproducibility is the degree to which it yields consistent scores over time among respondents who are assumed not to have changed (test-retest reproducibility), or the extent to which different observers may administer it to a particular patient and achieve similar results (inter-observer reproducibility).

Measures with poor reliability will be inefficient at distinguishing patients with different health states, because true differences in score may be obscured by random error. Instruments with good reproducibility can reliably distinguish changes in quality of life due to progression (or successful treatment) of the disease (true change) from changes due to random error in the measurement (change due to "noise"). In particular, instruments used to assess individuals longitudinally require better reproducibility than those which are used in cross-sectional studies.

1.6.2.4 Sensitivity to change

The sensitivity to change or responsiveness of an instrument refers to its ability to detect change, often defined as the minimum change considered to be important by the person being studied (or by their "significant others" or their carers). One may

think of it as the ratio of the "signal" (real change that has occurred over time) to the "noise" (the variability in the scores associated with the instrument's measurement error). Responsiveness is a function of both the validity and the reliability of the instrument.

Sensitivity to change can be measured in several ways. Effect size may be calculated as the mean difference in outcome between the two measurements divided by the variability (the standard deviation) of the baseline measurement (Kazis *et al.* 1989; Fitzpatrick *et al.* 1992b). However, this method of calculating effect size does not take account of the reproducibility of the measure and so instruments with poor test-retest reliability may yield spuriously high effect sizes with this approach (Ebrahim, 1995). Few measures of outcome which have been used in acute stroke trials to date have had their sensitivity to change explicitly measured. This is perhaps because, until recently, there has not been an effective treatment for stroke.

1.6.2.5 Interpretability

Interpretability is defined as the degree to which one can assign qualitative meaning to an instrument's quantitative scores (Scientific Advisory Committee, 1995). Instruments which place respondents into clinically sensible and distinct categories have immediate interpretability, but they are not necessarily valid, reliable or responsive. Numerical scores are less easy to interpret. It may be easier to interpret health related quality of life scores if the distribution of scores in different patient groups (and in the general population) is well characterised, or the change in scores associated with a particular type of life event, or the use of an effective treatment is known precisely.

1.6.2.6 Generalisability

Several health related quality of life instruments (such as the SF-36, the Nottingham Health Profile, the EuroQol and the Sickness Impact Profile) have become very widely used (McDowell and Newell, 1996). Their validity and other measurement attributes have been examined in large community samples and many different patient groups, and appear to be generally acceptable and qualitatively similar (Jenkinson *et al.* 1993; Brazier *et al.* 1992; Ware, 1992; Brooks & with the EuroQol group, 1996; McDowell and Newell, 1996). It is therefore unclear whether these instruments need to be re-examined in every new patient group or population of interest.

Although generic instruments aim to take a broad view of health, their content varies. Some instruments may be less relevant, and so perform less well, in certain patient or population groups. For example, Hayes and colleagues examined the suitability of the SF-36 questionnaire for the assessment of health status in older adults (Hayes *et al.* 1995). They found that two fifths of the sample were unable to complete the questionnaires by themselves and had to be assessed by interview. They also reported that missing responses were concentrated on a small number of questions whose emphasis was on work or vigorous activities. Perhaps elderly people viewed such activities as irrelevant to their own lives. However, Hayes et al did not assess whether these findings affected the instrument's other measurement attributes in the elderly.

The responsiveness of a measure may also be compromised by "ceiling effects" in which the patients with the maximal score may still have poor health related quality of life or "floor effects" in which patients with the worst score may deteriorate further (Guyatt *et al.* 1993). Floor and ceiling effects can only be identified if the actual

distribution of scores in the population of interest is known. This knowledge will also facilitate the interpretation of subsequent data generated with the instrument.

In summary, although examination of the properties of generic health status instruments in a new population of interest will almost always generate useful information, it is particularly important in the following settings. Firstly, if the new patient group differs importantly from other groups in which the instrument has already been evaluated. Secondly, if specific problems are anticipated, e.g. problems with feasibility, as in the elderly. Thirdly, if the instrument is under consideration for use in an important and potentially expensive study (because clinicians may not implement the results of a study if they are unsure about the measure of outcome).

1.7 Measuring health related quality of life after stroke

Conventional impairment and disability scales focus on residual physical problems after stroke and do not measure the difficulties with psychological and social functioning that may have a greater impact on the patients' life. Furthermore, recovery in the physical domain of health may not lead to parallel improvement in psychological or social functioning. Psychological and social outcomes are very relevant to the patients' quality of life (Goodare & Smith, 1995). They need to be enquired of specifically since doctors tend to focus on the patient's physical problems (Rothwell *et al.* 1997).

Measuring quality of life may also be particularly relevant if the treatment being evaluated does not have a major impact on death or disability, yet improves outcome in more subtle ways that matter to the patient. Similarly, if the treatment has a mixture of beneficial and undesirable effects, measurement of quality of life may help clarify the overall balance of risk and benefit (Friedman et al. 1984; Croog *et al.*

1986). Furthermore, some measures of health related quality of life generate an overall estimate of health status or utility (see below), which is useful in assessing the economic impact of a treatment (e.g. cost-benefit analyses), and will probably become a standard requirement of the regulatory agencies for the licensing and approval of new treatments (Freemantle *et al.* 1995).

Although it is now routine to measure health related quality of life in some areas of medical and public health research, little work has been done on the measurement of health related quality of life after stroke (de Haan et al. 1993a), and the selection of the most appropriate instrument is not clear.

1.8 Summary of Chapter One

1. Stroke is a major public health problem. The disease places a great burden on patients, their families and on society at large. The burden is not just in human suffering, but also a financial one.

2. Policymakers, health professionals, researchers and patients need detailed information on the burden of stroke as well as information on which treatments can reduce the burden. To measure outcome after stroke, and how it is influenced by treatment, requires suitable tools. Tools for the assessment of outcome at the level of the population, or groups of patients, will not necessarily be suitable for assessing outcome in individual patients.

3. Health related quality of life, which focuses on the patient's subjective view of their own health, is now commonly measured. The United Kingdom government health policy aims to improve the quality of life of stroke survivors.

4. Measures of health related quality of life must be valid, reliable, sensitive to change, and interpretable in the population of interest. Their administration must not place an excessive burden on either the subjects or the researchers.

5. Little is known about health related quality of life after stroke. A systematic review of this area is needed to summarise the results of the completed studies, highlight any methodological problems, inform the design of appropriate future studies and help to set research priorities.

Table 1.1: Stroke scales (adapted from van Gijn, 1992)

Author	Year of publication
Barham Carter	1961
Meyer et al	1965
Tuthill et al	1969
Mathew et al	1972
Fugl-Meyer et al	1975
Norris	1976
Kaste et al	1976
Larsson et al	1976
Fawer et al	1978
Mulley et al	1978
Admani	1978
Woollard et al	1978
Demeurisse et al	1980
Norris and Hachinski	1982
Feigenson	1982
Hamrin and Wohlin	1982
Orgogozo et al	1983
Scandinavian Stroke Study Group	1985
EC/IC Bypass Study Group	1985
Cote et al	1986
Orgogozo and Dartigues	1986
Adams et al	1987
Olesen et al	1988
Reding	1990
Loewen and Anderson	1990

Table 1.2: The Mathew Scale (Mathew *et al.* 1972)

Factor	Score
Mentation	
Level of consciousness	
Fully conscious	8
Lethargic but mentally intact	6
Obtunded	4
Stuporous	2
Comatose	0
Orientation	
Oriented x 3	6
Oriented x 2	4
Oriented x 1	2
Disoriented	0
Speech	
Reitan test	0-23
Cranial nerves	
Homonymous hemianopia	
Intact	3
Mild	2
Moderate	1
Severe	0
Conjugate deviation of eyes	
Intact	3
Mild	2
Moderate	1
Severe	0
Facial weakness	
Intact	3
Mild	2
Moderate	1
Severe	0
Motor power	
Normal strength	5
Contracts against resistance	4
Elevates against gravity	3
Gravity eliminated	2
Flicker	1
No movements	0
Performance or disability status scale	
Normal	28
Mild impairment	21
Moderate impairment	14
Severe impairment	7
Death	0
Reflexes	
Normal	3
Asymmetrical or pathological reflexes	2
Clonus	1
No reflexes elicited	0
Sensation	
Normal	3
Mild sensory abnormality	2
Severe sensory abnormality	1
No response to pain	0

Table 1.3: The Barthel Index (adapted from Manoney and Barthel, 1965)

What the patient ACTUALLY DOES ?			
1) FEEDING	Independent Needs some help Needs to be fed	= 10 = 5 = 0	
2) BATHING	Able to wash all over Needs help	= 5 = 0	
3) GROOMING	Totally independent Dependent in some way	= 5 = 0	
4) DRESSING	Totally independent Needs help with some items Unable to do any without help	= 10 = 5 = 0	
5) BOWELS	No accidents Occasional accident / help with enema Incontinent	= 10 = 5 = 0	
6) BLADDER	No accidents Occasional accident / use of device Incontinent	= 10 = 5 = 0	
7) TOILET	Independent Minor assistance Unable to use	= 10 = 5 = 0	
8) TRANSFER	Totally independent Minimal help needed Sit unaided, major help for transfer Unable	= 15 = 10 = 5 = 0	
9) AMBULATION	Independent for 50 metres Walk 50 metres with help Independent in wheelchair for 50 metres Immobile	= 15 = 10 = 5 = 0	
10) STAIRS	Independent Needs physical / verbal support Unable	= 10 = 5 = 0	
		TOTAL:	

Figure 1.1: The dimensions of health status and the factors that affect them (Bergner, 1985)

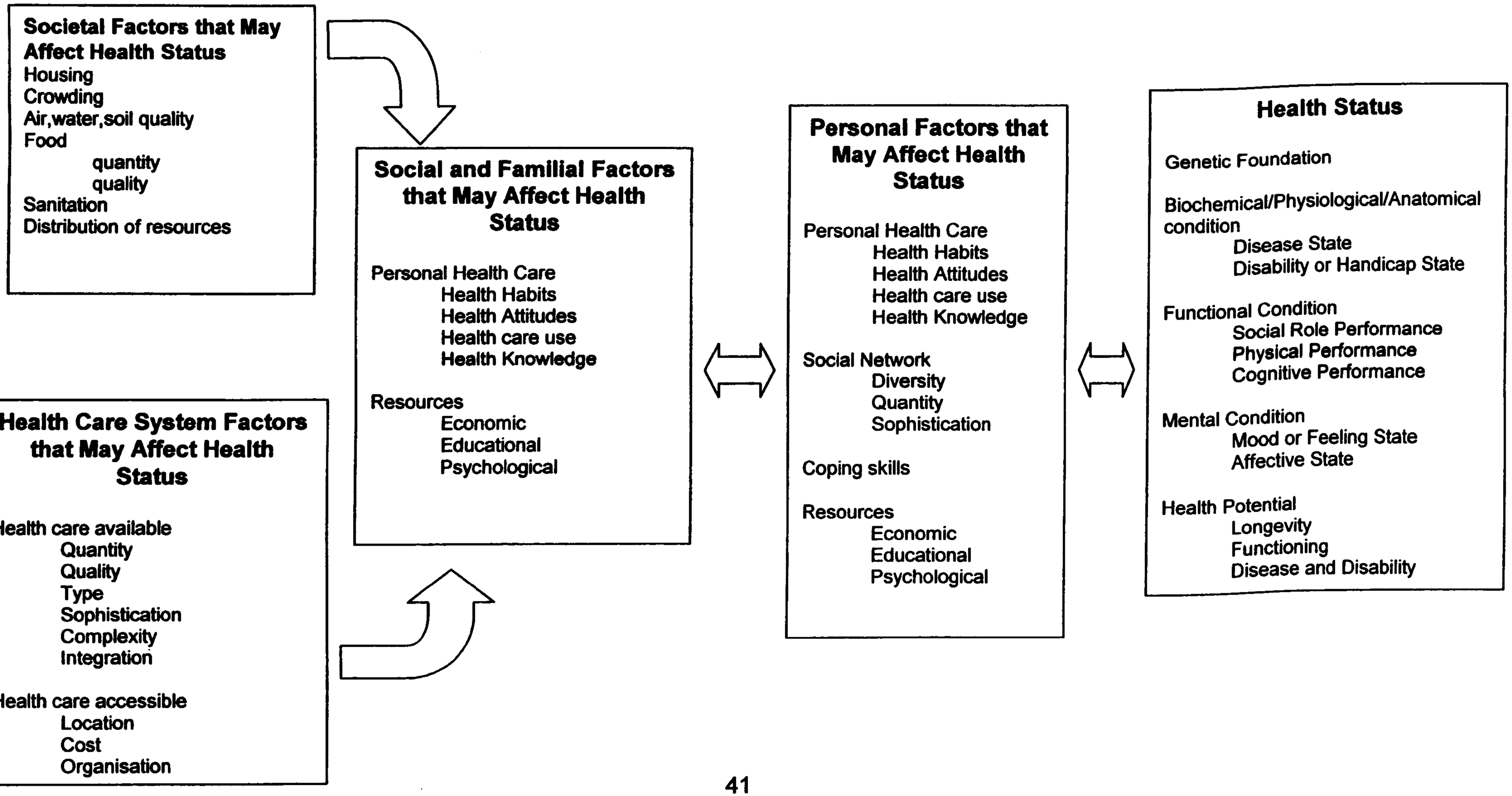


Figure 1.2: Framework for discussing health status and quality of life (Ware, 1984)

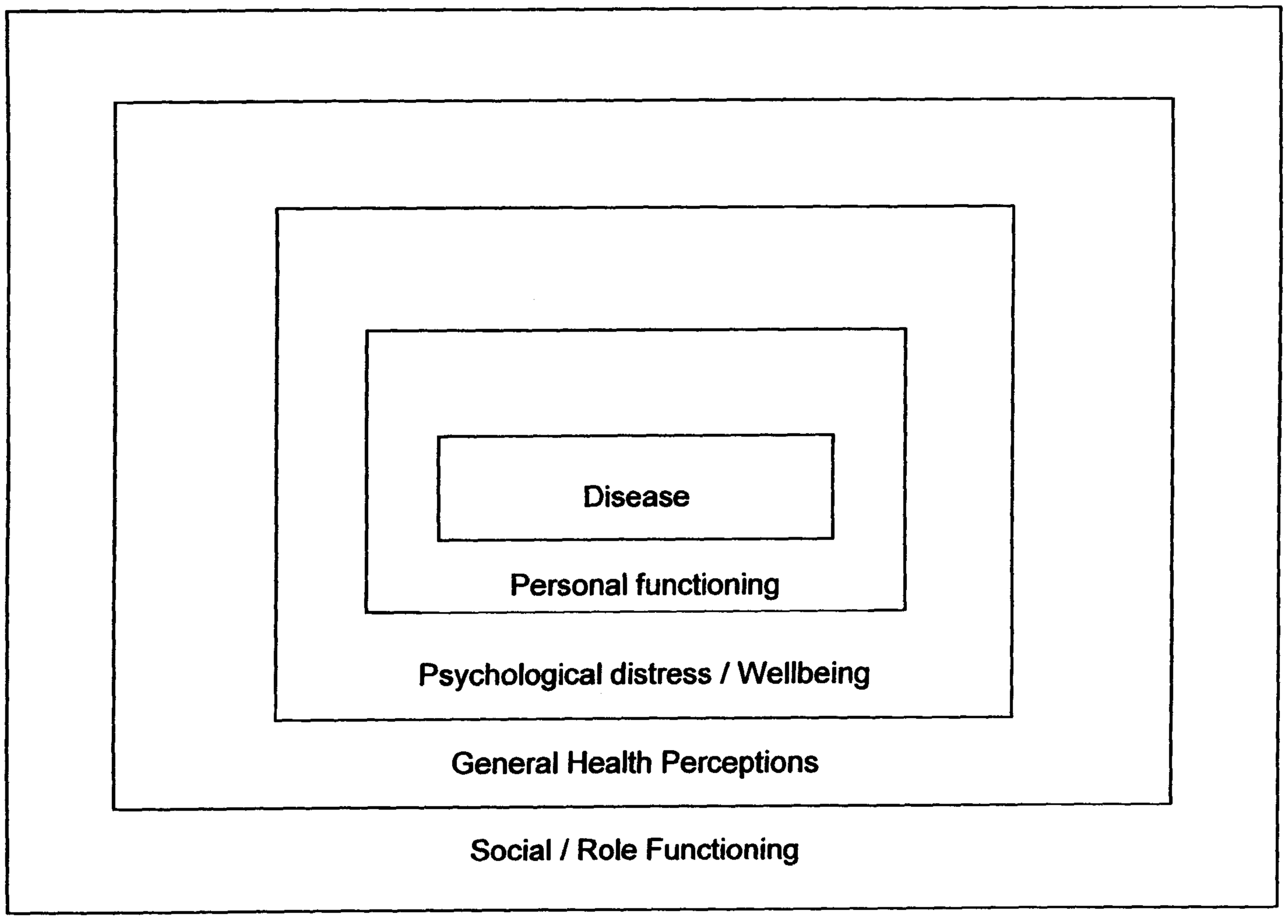
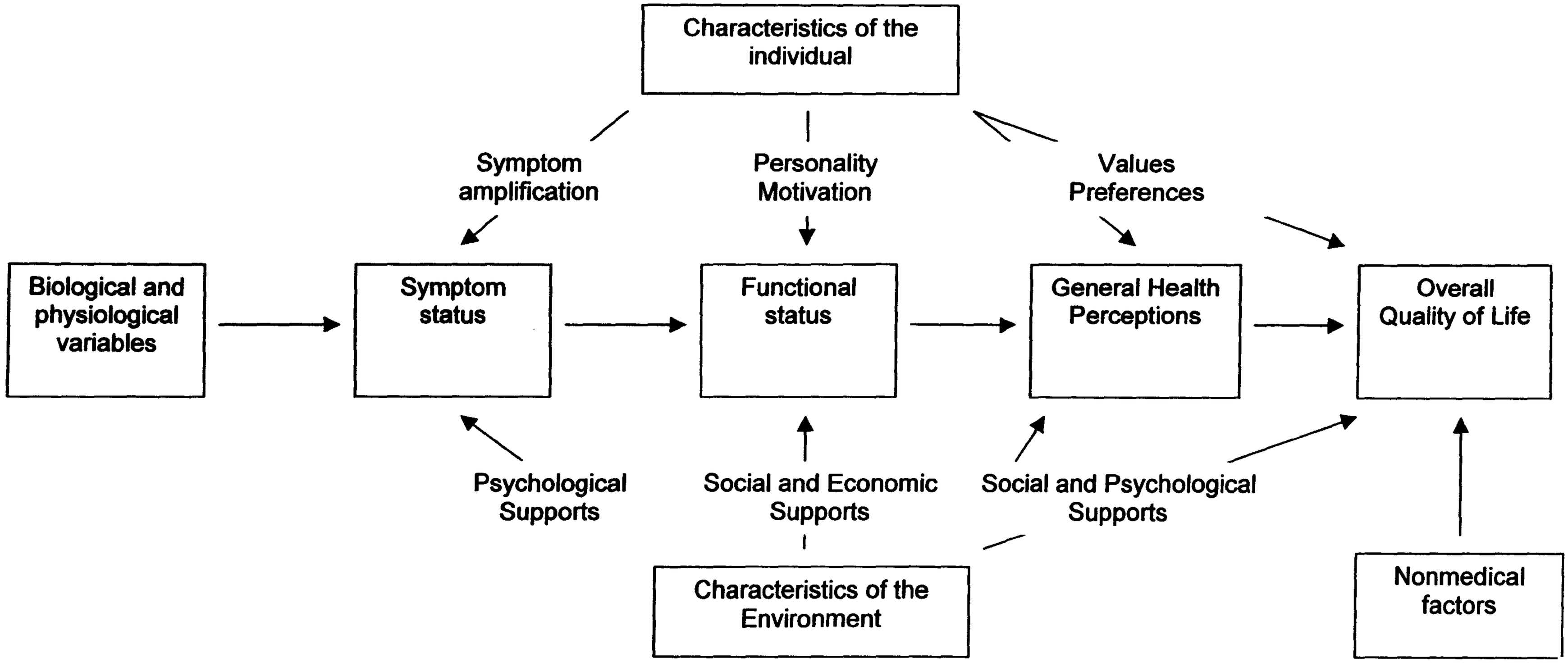


Figure 1.3: Relationships among measures of patient outcome in a health related quality of life conceptual model (Wilson and Cleary, 1995)



2 The measurement of health related quality of life after stroke: a systematic review of existing studies

2.1 Introduction

In 1994, the Neurosciences Trials Unit at the Western General Hospital, Edinburgh, began to plan a randomised controlled trial of a neuroprotective agent (BW619C89) in patients with acute stroke. The primary outcome measure was to be the proportion of patients "dead or dependent" in activities of daily living six months after the stroke. This measure is robust but relatively insensitive. Neuroprotective agents act on the cerebral cortex and we predicted that they might well have effects on cortical function that would not be detected by an instrument to measure activities of daily living. In view of this we wished to explore the possibility of using a measurement of health related quality of life as a more sensitive and comprehensive measure of treatment effect. In particular, neuroprotective therapy might have long term adverse neuropsychiatric effects, which a health related quality of life instrument could detect (Dorman & Sandercock, 1996). Secondly, neuroprotective therapy might improve cerebral cortical function and so improve cognitive function; assessments which only measured physical disability could not detect improvements in cognitive function (Dorman *et al.* 1996). Thirdly, clinicians and purchasers of health care would be more likely to adopt and pay for neuroprotective treatment if it was not merely clinically effective, but also cost-effective. Furthermore, economic analyses are more useful, and informative, if they can assess the impact of treatment on health related quality of life as well as on survival and disability.

In 1994 relatively little work had been published on measurement of health related quality of life after stroke (de Haan *et al.* 1993a; Adkins, 1993; Tate *et al.* 1996). In particular, we were unsure whether it was feasible to measure health related quality of life in a large scale stroke study. The selection of the "best" health related quality of life instrument for the planned randomised controlled trial was therefore critical. Firstly, we wished to be sure that the measurement properties of the relevant instruments was appropriate for the assessment of health related quality of life after stroke. Secondly, there seemed to be relatively little easily available literature, and so an extensive and thorough search (to detect what little there was) was particularly important. I therefore decided to systematically review the existing literature on the measurement of health related quality of life after stroke. A systematic review, rather than a traditional (non-systematic) one, would have several advantages for us. Firstly, the use of explicit criteria for the inclusion of studies and an extensive search strategy to identify all the potentially eligible studies, should reduce the bias that may occur when only certain subsets of data are reviewed. Secondly, it should allow other investigators to replicate ones review (Mulrow, 1987; Counsell *et al.* 1995). Thirdly, the diversity of studies reviewed allows the reasons for any consistencies and inconsistencies in the results to be explored (Mulrow, 1987).

The aim of the review was to: identify all relevant studies which examined the **measurement** of health related quality of life after stroke, assess their methodological quality; and, on the basis of these results, select the "best" measure for use in a future large randomised controlled trial.

2.2 Methods of review

2.2.1 Inclusion criteria

I primarily sought to review all studies which assessed the measurement properties of measures of health related quality of life after stroke. I also sought to review articles that had assessed health related quality of life after stroke, because these studies might provide indirect evidence on the measurement of health related quality of life after stroke.

There is no universally agreed definition of health related quality of life and the concept has evolved with time (see Chapter One). I therefore took a broad approach and aimed to identify: studies which used accepted quantitative measures of health related quality of life or health status (McDowell & Newell, 1996); articles which used a single questionnaire to measure multidimensional outcomes (in at least the physical, social and psychological dimensions of health); and, articles which included instruments to provide an assessment of overall (health related) quality of life. I excluded studies which used a battery of accepted unidimensional instruments to assess health related quality of life after stroke since this approach is likely to place an excessive burden on patients, carers and administrators, and so would not be practical in a large randomised controlled trial. Some studies used a qualitative approach to assess health related quality of life after stroke. These were also excluded since this approach was unlikely to be applicable in our proposed trial or other large epidemiological studies.

Although my search strategy included articles published to the end of 1997, I did not consider studies published after August 1997 because some of these articles included work from this thesis (see Appendix 10).

2.2.2 Search strategy

I used several strategies to identify relevant studies. Only studies published in full text, in English, were eligible. The primary strategy was a computerised bibliographic searching of the MEDLINE database from 1966 to 1997. The final search was performed in June 1998 using the *Ovid for Windows* software. I used a modified version of the Cochrane Stroke Group's search strategy for stroke related articles (Counsell, 1998) and combined this with a search for articles related to outcome (Table 2.1). The latter search was devised by examining the Medline subject headings and the indexing of studies known to be relevant. This search strategy was then validated by comparing its sensitivity against a gold standard of relevant articles produced by searching five volumes of the journal *Stroke* by hand (Counsell, 1998). I used the same strategy to search the EMBASE database from 1980 to 1997. I reviewed the titles and abstracts of all articles identified by these searches. I obtained copies of, and read in full, articles that appeared to meet the inclusion criteria for this study.

I also used several other methods to identify relevant articles. I hand-searched the journals *Cerebrovascular Diseases* and *European Journal of Neurology* (which are not indexed in MEDLINE). I reviewed the bibliography of an existing systematic review of the stroke outcome literature (Warbuton and Long, 1994) and the bibliographies of other relevant studies and reviews (Anderson, 1992; De Haan *et al*, 1993a; Adkins, 1993; Tate *et al*, 1996); and reviewed my personal collection of studies relating to the measurement of health related quality of life after stroke. I also had many informal discussions with other researchers. I did not, however, formally survey the opinions of experts regarding the measurement of health related quality of life after stroke. I sought the expert opinion of Carl Counsell regarding the development of the search strategy for this systematic review.

2.2.3 Data extraction and synthesis

I assessed the quality of evaluation of health related quality of life instruments after stroke by extracting the following data, as suggested by previous authors (Scientific Advisory Committee, 1995; McDowell & Newell, 1996):

i) Aim of study

Was the purpose of the study clearly stated?

ii) Description of measure and conceptual basis

Was the instrument adequately described, if not well known? Was the conceptual model explained or was an appropriate reference cited?

iii) Generalisability

Were the characteristics of sample (e.g. socioeconomic status, comorbid conditions and clinical details) described?

iv) Testing conditions

How and where was the instrument administered?

v) Practicality

Was information provided on the time required to complete (administer) the instrument? Was the feasibility of the instrument and the frequency of missing data or non-response reported? Can the instrument be administered in alternative forms? Is the instrument available in different languages?

vi) Size of study

Was it clearly documented?

vii) Validity

Were the different aspects of validity considered? See Chapter One.

viii) Reliability

Was test-retest reliability examined as well as internal consistency? See Chapter One.

ix) Sensitivity to change

Was this assessed longuitudinally?

x) Interpretability

Are reference standards available?

2.3 Results

2.3.1 Search strategy

I screened 1531 titles and abstracts retrieved by the Medline search, and read 38 of them in full. Of these, 20 fulfilled the inclusion criteria. The EMBASE search retrieved a total of 1420 titles and abstracts. However, the majority of the relevant articles had already been identified by the Medline search. I read a further seven articles in full and included two of them in the review. Two further studies were identified from expert comment and reading reference lists of published articles. In total, I identified 24 articles that fulfilled the inclusion criteria for the systematic review.

Hand-searching five volumes of the journal *Stroke* (years 1993 to the end of 1997) identified 14 articles which satisfied the inclusion criteria for the current review, Table 2.2. The final Medline search identified all but one of these articles (Table 2.2). Therefore, this electronic search strategy had an excellent sensitivity for identifying relevant articles in the indexed journals (sensitivity = 93%). The EMBASE search had equivalent sensitivity, Table 2.2. The other handsearching did not identify any additional relevant articles.

2.3.2 Measurement attributes of quality of life measures after stroke

I identified 24 articles which fulfilled the inclusion criteria. These referred to the results from 20 independent studies and involved at least 2330 patients, Table 2.3. Three studies which used the Sickness Impact Profile generated multiple publications (Nydevik et al, 1991 and 1993; Schuling et al, 1993(a) and 1993(b); De Haan et al, 1993 and 1995 and Sneeuw et al 1997). The most commonly used instruments were the Sickness Impact Profile, the Nottingham Health Profile and SF-36. They accounted for 18 of the 24 articles. Two studies did not examine multiple dimensions of health and exclusively assessed overall health related quality of life (Ahlsio et al, 1984; Kwa et al, 1996).

The instruments and the conceptual basis for measurement was well described in the studies which used the visual analogue scales, Sickness Impact Profile, Nottingham Health Profile, and SF-36. The questionnaires created by Niemi, Viitanen and Astrom were poorly described and explained, Table 2.3. A substantial minority of the articles (11 of the 24) provided inadequate descriptions of the baseline characteristics of their sample, Table 2.3.

The primary aim for the majority of the studies was to assess quality of life after stroke and determine which factors influenced it. Only six articles primarily assessed the measurement properties of health related quality of life instruments in stroke survivors. This subgroup of studies included only 682 patients. Of these studies, one aimed to assess validity, two aimed to examine the validity of assessments by proxies, one examined both feasibility and test-retest reliability, one solely examined test-retest reliability, and one study assessed sensitivity to change.

However, many of the remaining studies produced useful indirect information on the measurement of health related quality of life after stroke, Table 2.5.

2.3.2.1 Feasibility and practicality of health related quality of life assessments after stroke

In nearly all the studies, some patients were unable to have their health related quality of life assessed. In general, the more severely affected stroke patients could not be assessed because of communication, cognitive or other problems which limited their ability to participate in interviews or to complete complex questionnaires. Overall, about one-quarter of all surviving patients were unable to complete health related quality of life assessments, Table 2.4.

Only one study sought formally to measure the feasibility of health related quality of life assessments (Visser *et al.* 1995). The study included only 16 patients, and all had made an excellent recovery, so the sample was not representative. Only three studies assessed the time required to complete the relevant questionnaires, Table 2.4. Anderson attempted to use the Nottingham Health Profile in The Greenwich Stroke Study:

> "in the pilot study it became clear that the number of statements and their random ordering posed problems for patients. Ultimately only the statements relating to emotional distress and social isolation were presented to patients." (Anderson R., 1992).

The setting of the study, method of questionnaire administration and frequency of missing data was poorly reported, Table 2.4.

2.3.2.2 Validity

Anderson and colleagues in Perth were the only investigators to primarily examine the validity of health related quality of life assessments. They administered the SF-36 by interview to 90 patients one year after the index stroke. They found evidence for construct validity. The mean scores for patients dependent in self care and with mental ill-health were significantly different from patients without these problems (Anderson *et al.* 1996). However, they found less evidence for concurrent validity; scores on the social functioning domain of the SF-36 were only poorly correlated with those measured with the Adelaide Activities Profile (Anderson *et al.* 1996). Ahlsio and coworkers examined the validity of quality of life assessments with a simple visual analogue scale (Ahlsio *et al.* 1984). They used interviews to establish the concurrent validity of this method as well as demonstrating construct validity in other ways.

Many of the other articles provided indirect evidence for the validity of assessments with the Sickness Impact Profile, Nottingham Health Profile and SF-36, Table 2.5. For instance, de Haan and colleagues examined the relationship between patients' responses to the Sickness Impact Profile and measures of disability and impairment. They found evidence for discriminant validity: the correlation with scales of impairment decreased from the Barthel (mean r^2=47.5%) to the Rankin (mean r^2=36.5%) to the Sickness Impact Profile (mean r^2=33%) (de Haan *et al.* 1993b).

Segal and Schall assessed the validity of carers' assessments of the patients' functioning with the SF-36 (Segal & Schall, 1994). They found that agreement between the proxy and direct patient assessment was only "poor to moderate" for all the domains except physical functioning. However, they did not investigate whether there was a systematic bias between the response from the proxy and the patient.

Sneeuw and colleagues also assessed the validity of proxy assessments of the patients' functioning with the Sickness Impact Profile (Sneeuw et al, 1997). They found better agreement for most of the domains of the Sickness Impact Profile, but they also found some evidence that proxies systematically rated patients as having more problems than patients did themselves.

2.3.2.3 Reliability

Two studies examined the test-retest reliability of health related quality of life assessments after stroke: Visser and colleagues examined the test-retest reliability of the NHP and SIP in patients with myocardial infarction or stroke (Visser *et al.* 1995). Unfortunately, they reported the reliability of the assessments for both groups of patients combined. The value of their study is also limited because they used an inappropriate statistical measure (association rather than agreement) to examine reliability (McDowell and Newell, 1996). Gompertz and his colleagues examined the test-retest reliability of the Nottingham Health Profile in a small group of 21 patients (Gompertz *et al.* 1993). They selected these individuals from patients who responded to the original questionnaire without prompting. They found the reproducibility of the Nottingham Health Profile was inadequate for monitoring individual patients.

2.3.2.4 Sensitivity to change

Schuling and colleagues attempted to investigate the sensitivity to change of the Sickness Impact Profile in patients' with stroke. They found little difference in patient scores between 8 and 26 weeks and concluded that the instrument had inadequate sensitivity to change. However, this conclusion may not be justified because they made no efforts to identify patients who had actually changed and overall there was no significant difference in the patients' level of disability.

2.3.3 Pattern of health related quality of life in cohorts of stroke patients

Many of the articles examined the impact of stroke on patients' quality of life. The results were surprisingly consistent: irrespective of the measure used, the majority of the large studies with controls found that patients reported problems in all of the domains, as well as their overall quality of life, Table 2.5. Similarly, the predictors of poor overall quality of life were remarkably consistent: poor physical functioning and low mood appeared to be independent predictors of poor overall quality of life whenever it was assessed. The Greenwich Stroke Study provided insights into the consequences of stroke for both the patients and also their supporters by the use of a combination of qualitative and quantitative assessments (Anderson R., 1992).

2.4 Discussion

This study identified a small number of published studies that had formally investigated the measurement of health related quality of life after stroke. Most of the available studies were published in the last five years (1992-1997), and used established generic measures of health related quality of life, particularly the Sickness Impact Profile, Nottingham Health Profile and SF-36. By contrast, several of the earlier studies created instruments specifically for use in their studies; unfortunately these were generally poorly described and perhaps for this reason had not used subsequently.

2.4.1 Choice of instrument for proposed randomised controlled trial

The Sickness Impact Profile measures health status by assessing the way sickness changes daily activities and behaviour (Gilson *et al*, 1975). The instrument is relatively long and consists of 136 statements in 12 categories. I found indirect evidence for its concurrent, construct and discriminant validity after stroke. Its concentration on behaviour has several advantages over recording feelings in survivors of stroke. Behaviours are observable and so more accessible to external validation (McDowell & Newell, 1996). Sneeuw and colleagues support this since they found moderate agreement between proxy and patient assessments of patients' functioning (Sneeuw *et al*, 1997). They also examined the internal consistency of each of the subscales. This was relatively poor and ranged from 0.36 to 0.85 ("Eating" to "Bodycare and Movement" respectively). As yet, there have been no good studies on its test-retest reliability. Visser and colleagues examined its feasibility. However, they selected a small group of non-disabled stroke survivors who were all able to self complete the questionnaire with a mean time of 24 minutes (Visser et al, 1995). This contrasts with the other studies, in which it was administered by interview and a substantial proportion of patients were unable to provide appropriate responses. Its length and these concerns regarding its feasibility make it unlikely to be suitable for use in large scale epidemiological studies.

The Nottingham Health Profile was designed to give a brief indication of perceived physical, social and emotional health problems (Hunt *et al*, 1981). It differs from the Sickness Impact Profile in that it asks directly about feelings and emotional states. The Nottingham Health Profile has two parts: Part 1 contains 38 items grouped into six domains, whereas Part 2 provides a brief indicator of handicap, and contains seven items that record the effect of health problems on occupation, jobs around the

house, personal relationships, social life, sex life, hobbies and holidays. Most of the studies simply used Part 1 (Part 2 was used by Visser and colleagues and Harwood and his colleagues (Visser et al, 1995; Harwood et al, 1994)). Ebrahim and colleagues provided indirect evidence for its construct and concurrent validity when used after stroke (Ebrahim et al, 1986). They also found that the reproducibility of the NHP, although moderate, was inadequate to support its use in individual patients (Gompertz et al, 1993). The Nottingham Health Profile may also be too complex for use in a large epidemiological study - it was found to be excessively complex in the pilot study of the Greenwich Stroke Study (Anderson, 1992). The Nottingham Health Profile has been copyrighted and should only be used with the written permission (and possibly the payment of royalties to) of its authors.

The SF-36 was designed as a generic indicator of health status (Ware & Donald-Sherbourne, 1992). The items were drawn from a larger 245 item Medical Outcomes Study questionnaire. It assesses outcome in 8 domains and has become one of the most widely used health questionnaires. The New England Health Institute estimated that by 1992 a million forms were being administered each year. Although it appears to have reasonable validity when administered by interview after stroke, its validity when completed by proxies appears questionable (Segal & Schall, 1994). Furthermore, its feasibility, reliability and sensitivity to change remain unclear.

The choice of health related quality of life questionnaire for the future proposed study remains unclear. No single instrument has been adequately evaluated in patients with stroke; and so, none of the instruments currently identified, can be recommended for use after stroke on what is known about their measurement attributes in this particular patient group. I did not identify any studies that attempted to compare directly the measurement attributes of the available instruments. The

instruments could also be distinguished on their conceptual basis and content; but, as yet, this has not been done.

2.4.2 Implications for future research

Measurement of health related quality of life after stroke is likely to be unreliable if the instrument places too great a burden on the patient. If a health related quality of life instrument is easy to administer (or self-complete), it is likely to be more widely adopted in stroke research studies. In the majority of the studies reviewed the questionnaires were administered by interview, but even with this help patients could often not complete the assessment. Simpler instruments are more likely to be completed – so improving the power of the study, reducing the risk of bias and also reducing the resources required to conduct the study. The benefits and disadvantages of simpler instruments should be examined in future studies. Future studies should also compare directly the feasibility and other measurement properties of the available instruments.

2.4.3 Limitations of review

My resources were limited, so I aimed to be systematic but focussed. My searching was limited in several respects. I only searched Medline and EMBASE, rather than any of the other electronic databases. I also excluded non-English language studies and those published in abstract form only, since I did not have access to translation facilities and because abstracts do not contain enough information to allow a reliable assessment of the study quality. I did not attempt to identify unpublished studies because searching for this type of material is very time-consuming and – unlike searches for unpublished randomised controlled trials – the results would be of uncertain value. I also did not attempt a comprehensive survey of experts in this

area, though I did have extensive discussions with Dr Paul Kind, University of York. It is also possible that biases in interpretation of the inclusion criteria and extraction of the data have affected my results and that I have missed some relevant studies, but my search strategy was more comprehensive than that used in previous reviews of this area. Furthermore, I was able to demonstrate the high sensitivity of my Medline search strategy by comparing its results with those obtained by manually searching the journal *Stroke*. Therefore, despite the above limitations, I feel the results of this review are valid.

2.5 Summary of Chapter Two

1. A systematic search of the literature identified 24 eligible studies, but the majority were small, used complex instruments and had methodological weaknesses.

2. Several different generic instruments (Nottingham Health Profile, Sickness Impact Profile and the SF-36) were used to assess health related quality of life after stroke. However, their validity and other measurement properties were not adequately evaluated. There were no studies directly comparing one instrument with another in a group of stroke patients.

3. A significant number of patients were unable to complete or participate in assessments of their health related quality of life. This may reflect the complex nature of the instruments used so far. The feasibility and accuracy of health related quality of life assessments by a surrogate for the patient has not been fully explored.

4. Short and simple instruments may reduce the burden on respondents and thereby yield better response frequencies. Future research should directly compare the feasibility, validity and reliability of current health related quality of life instruments with a shorter and less complex instrument.

Table 2.1 Medline/Embase search to identify studies on the measurement of health related quality of life after stroke (modified from Counsell, 1998)

Indexing Terms		Notes
1	cerebrovascular disorders/	
2	cerebral artery diseases/	
3	cerebral embolism.ti,ab,sh. and thrombosis/	
4	carotid artery thrombosis/	
5	wallenberg's syndrome/	
6	cerebral hemorrhage/	
7	cerebral infarction/	
8	cerebral ischemia/	
9	stroke$.tw.	
10	cerebrovascular$.tw.	
11	(cerebral or cerebellar or brainstem or vertebrobasilar).tw.	
12	(infarct$ or ischaemi$ or thrombo$ or emboli$).tw.	
13	11 and 12	
14	(cerebral or intracerebral or intracranial or parenchymal or brain or intraventricular or brainstem or cerebellar or infratentorial or supratentorial).tw.	
15	(haemorrhage or hemorrhage or haematoma or hematoma).tw.	
16	14 and 15	
17	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 13 or 16	The final search for articles related to stroke
18	health status/	
19	health status.tw.	
20	health status.tw.	
21	quality of life/	
22	quality of life.tw.	
23	outcome.ti,ab,sh. and "Process Assessment (Health Care)"/	
24	disability evaluation/	
25	stroke assessment/	
26	questionnaires/	
27	depressive disorder/	
28	"Outcome Assessment (Health Care)"/	
29	18 or 19 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28	A broad search for any article to do with QoL or outcome
30	17 and 29	
31	limit 30 to human	
32	limit 31 to journal article	
33	limit 32 to english language	

"/" indicates Medline index term
".tw." is a search for title or abstract words. "$" is a truncation term.
The "or" operator means an article only has to include one of the terms identified.
The "and" operator means an article must be found in both sets.

Table 2.2 Articles identified by handsearching *Stroke* 1993-1997

Year	Month	Title	First author	Retrieved by Medline	Retrieved by EMBASE
1993	Aug	The Frenchay Activities Index...	Schuling	N	Y
1993	Aug	A comparison of 5 stroke scales with ..	De Haan	Y	Y
1994	July	Prognosis of young adults with ischemic stroke	Kappelle	Y	Y
1994	Dec	Determining functional/health status...	Segal	Y	Y
1995	March	Impact of stroke type and lesion location on QoL..	De Haan	Y	Y
1995	Nov	Clinical meaning of Rankin handicap grades...	De Haan	Y	N
1996	Sep	QoL after stroke...	King	Y	Y
1996	Oct	Validation of SF-36...	Anderson	Y	Y
1997	March	Health status of individuals with mild stroke..	Duncan	Y	Y
1997	Aug	Assessing QoL after stroke: proxy	Sneeuw	Y	Y
1997	Oct	Use of healh utilities index with stroke...	Mathias	Y	Y
1997	Oct	Is the EuroQol a valid...	Dorman	Y	Y
1997	Oct	Proxy EuroQol	Dorman	Y	Y
1997	Nov	A stroke adapted version of the SIP..	Van Straten	Y	Y

Table 2.3: Systematic review of studies of quality of life after stroke: aims

Instrument	Study	Year	Eligibility	Sample	Size	Conceptual basis	Description of measure	Aims of study
Questionnaire	Niemi et al	1988	Stroke unit, <65 yrs	Y	46	N	Y	Investigate QoL after stroke
Questionnaire	Viitanen et al	1988	Stroke unit	Y	62	N	N	Investigate effect of stroke on patients satisfaction
Questionnaire	Astrom et al	1992	Stroke unit, Proxies	Y	50	N	N	Identify factors associated with poor life satisfaction after stroke
Visual analogue scales (VAS)	Ahlsio et al	1984	Stroke unit	Y	59	Y	Y	Investigate influences on QoL after stroke Assess validity of VAS
	Kwa et al	1996	Hospital series	Y	129	Y	Y	Assess impact of cognitive impairment on QoL
Sickness Impact Profile (SIP)	Nydevik et al	1991	Hospital series	N	57	Y	Y	Assess QoL after stroke and relationship with disability
	Nydevik et al	1993	Follow up earlier series	N	36	Y	Y	
	Schuling et al	1993	Community	N	80	Y	Y	Impact of stroke on SIP, Is SIP sensitive to change, Validity of FAI
	Schuling et al	1993	Community	N	77	Y	Y	
	De Haan et al	1993	Hospital series	N	81	Y	Y	Assess relationship between ICIDH and SIP, Impact of stroke type on QoL
	De Haan et al	1995	Hospital series or proxies	Y	441	Y	Y	
	Sneeuw et al	1997	Hospital series and proxies	Y	437	Y	Y	Assess value proxy SIP
	Yoon, H.	1997	Hospital/ >65 yrs	N	119	Y	Y	Investigate factors which impact on QoL

Table 2.3 ctd

Instrument	Study	Year	Eligibility	Sample	Size	Conceptual basis	Description of measure	Aims of study
Nottingham Health Profile (NHP)	Ebrahim et al	1986	Not cognitive impaired*	N	198	Y	Y	Describe social and psychological problems after stroke, and association between NHP and disability, use of services, length stay, and GHQ
	Johansson et al	1992	Hospital series/first strokes	Y	224	Y	Y	Evaluate medical and social outcome after stroke
	Anderson	1992	Community	Y	NS	Y	Y	Examine experience of stroke and patterns of coping
	Gompertz et al	1993	Hospital	N	21	Y	Y	Examine reproducibility of outcome measures including NHP
	Harwood et al	1994	Hospital	N	89	Y	Y	Examine new handicap scale
	Visser et al	1995	Hospital	N	16	Y	Y	Examine feasibility and reproducibility of outcome measures (including NHP & SIP) after myocardial infarction or stroke
SF-36	Segal et al	1994	Post-rehab and proxies	Y	38	Y	Y	Assess proxy agreement for FAI and SF-36
	Kapelle et al	1994	Hospital / <45yrs	Y	212	Y	Y	Assess prognosis of young adults with ischaemic stroke
	Anderson et al	1996	Hospital	N	90	Y	Y	Assess concurrent and construct validity of SF-36
	Duncan et al	1997	Hospital	Y	304	Y	Y	Assess health status for individuals with mild stroke
Quality of life index	King	1996	Hospital	Y	86	Y	Y	Examine global and domain specific QoL in stroke survivors Identify variables which predict QoL

Notes:

"Sample" relates to how well baseline sociodemographic and clinical data have been described
"Conceptual basis" relates to whether the conceptual basis of measurement has been explained or referred to
"Description of measure" relates to whether the instrument has been cited or described well enough for independent investigators to replicate the study
N=not well described
Y=well described
QoL = Quality of life
ICIDH = International Classification of Impairments, Disability and Handicap
FAI = Frenchay Activities Index
GHQ = General Health Questionnaire

Table 2.4: The feasibility and practicality of health related quality of life assessments after stroke

Instrument	Study	Year	Setting	Method of Administration	N study	N willing	N alive	Missing data	Time to complete (minutes)	Alternative forms / languages
Questionnaire	Niemi et al	1988	NS	NS	46	52	59	NS	NS	No
Questionnaire?	Viitanen et al	1988	NS	Interview	62	88	96	NS	NS	No
Questionnaire	Astrom et al	1992	NS	Interview?	50	57	61	NS	NS	No
Visual analogue scale	Ahlsio et al	1984	NS	Self completed	59	76	78	NS	NS	No
	Kwa et al	1996	Home	Self completed	97	129	171	NS	NS	No
Sickness Impact Profile	Nydevik et al	1991	Home/hospital	Interview	57)	90	91	NS	60*	Yes
	Nydevik et al	1993	Home	Interview	36	41	43	NS	NS	Yes
	Schuling et al	1993	Home/hospital	Interview	80	NS	121	NS	NS	Yes
	Schuling et al	1993	Home/hospital	Interview	89	NS	NS	Yes	NS	Yes
	De Haan et al	1993	NS	NS	81	NS	87	NS	NS	Yes
	De Haan et al	1995	NS	Interview	441	485	502	NS	NS	Yes
	Visser et al	1995	NS	Self completed	16	16	16	NS	24	Yes
	Sneeuw et al	1997	NS	NS	437	485	502	NS	NS	Yes
	Yoon, H.	1997	Hospital	Interview	119	NS	NS	NS	NS	Yes
Nottingham Health Profile	Ebrahim et al	1986	NS	NS	198	222	235	NS	NS	Yes
	Johansson et al	1992	NS	NS	224	NS	260	NS	NS	Yes
	Anderson	1992	Home	Interview	NS	NS	NS	NS	NS	Yes
	Gompertz et al	1993	Home	Self completed	21	NS	233	NS	NS	Yes
	Harwood et al	1994	Home	Self completed	89	94	170	NS	NS	Yes
	Visser et al	1995	NS	Self completed	16	16	16	NS	9	Yes

Table 2.4 ctd

Instrument	Study	Year	Setting	Method of Administration	N study	N willing	N alive	Missing data	Time to complete (minutes)	Alternative forms / languages
SF-36	Segal et al	1994	Hospital	Interview	38	NS	107	NS	NS	Yes
	Kappelle	1994	Home	Self completed	212	NS	235	NS	NS	Yes
	Anderson et al	1996	Home	Interview	90	103	124	Yes	8	Yes
	Duncan et al	1997	Home	Telephone	304	NS	NS	NS	NS	Yes
Quality of life index	King	1996	Home	Interview	86	NS	NS	NS	NS	No

Notes:

NS = not stated or reported
Gompertz et al, 1993: 21 patients were selected from larger number of prompt respondents
Schuling et al, 1993: "SIP puts a heavy burden on patients as well as investigators"
Anderson, 1992: "In the pilot study it became clear that the number of statements and their random ordering (of NHP) posed problems for patients. Ultimately only the statements relating to emotional distress and social isolation were presented to patients."

Nydevik et al, 1991: time included an activities of daily living assessment (Katz)

Table 2.5: Systematic review of studies of measurement of quality of life after stroke: summary of results

Instrument	Study	Year	Evidence for validity Direct	Evidence for validity Indirect	Reliability	Sensitivity to change	Pattern of QoL problems	Predictors of poor QoL
Questionnaire	Niemi et al	1988	-	-	-	-	All domains	Low mood; problems with mobility, ADL & memory
Questionnaire	Viitanen et al	1988					Global, self care, leisure, social, and sexuality	
Questionnaire	Astrom et al	1992					-	Major depression, disability, impaired social network
Visual analogue scales	Ahlsio et al	1984	Concurrent	-	-	-	-	Depression, anxiety, poor physical functioning
	Kwa et al	1996	-	-	-	-	-	Disability, infarct volume and aphasia
Sickness Impact Profile (SIP)	Nydevik et al	1991	-	Concurrent	-	-	All domains except eating and work	Problems with ADL, ↑age
	Nydevik et al	1993	-	Construct	-	-		Problems with ADL
	Schuling et al	1993	-	Construct	-	No	Worse than controls in all domains of SIP	-
	Schuling et al	1993	-	Discriminant	-	-		-
	De Haan et al	1993	-	Discriminant	-	-	-	-
	De Haan et al	1995	-	Construct	-	-	All domains SIP	Age, comorbidity, stroke severity and type
	Sneeuw et al	1997	Proxy validity	-	Internal consistency	-	-	-
	Yoon, H.	1997	-	-	-	-	Worse overall score	Physical functioning
	Visser et al (see below)	1995	-	-	Test-restest for SIP (and NHP)		Emotional behaviour and household management	

Table 2.5 ctd

Instrument	Study	Year	Evidence for validity Direct	Indirect	Reliability	Sensitivity to change	Pattern of QoL problems	Predictors of poor QoL
Nottingham Health Profile (NHP)	Ebrahim et al	1986	-	Construct, concurrent and content	-	-	All domains NHP	Disability
	Johansson et al	1992	-	-	-	-	No controls, similar patter to Ebrahim	Disability
	Anderson	1992	-	-	-	-	NHP only used in pilot	-
	Gompertz et al	1993	-	-	Test-retest	-	-	-
	Harwood et al	1994	-	concurrent, and divergent	-	-	All domains	-
	Visser et al	1995	-	-	Test-restes for NHP (and SIP)	-	Energy	-
SF-36	Segal et al	1994	Proxy validity	-	-	-	-	-
	Kapelle et al	1994	-	-	-	-	-	-
	Anderson et al	1996	Construct and concurrent	-	Internal consistency	-	Physical functioning, general health and vitality	-
	Duncan et al	1997	-	construct	-	-	Worse in all domains	-
Quality of life index	King	1996	-		Internal consistency	-	-	Low mood, poor social support, disability

Notes to Table 2.5

QoL = quality of life

ADL = activities of daily living

↑ = increased

3 Is the EuroQol a valid measure of health related quality of life after stroke?

3.1 Introduction

There are several valid instruments for the measurement of single aspects of psychological, social or physical outcome after stroke (Johnson *et al.* 1995; Wade *et al.* 1985; Lindley *et al.* 1994; Wellwood *et al.* 1995). However, the use of a series of such assessments may subject patients to an unacceptable burden and so reduce the overall frequency and completeness of response.

The EuroQol is a generic instrument for the measurement of health related quality of life (The EuroQol Group, 1990), (see Appendix 2). It measures aspects of quality of life that are highly relevant to stroke patients. It is short and simple enough that many stroke patients (despite cognitive, motor and sensory deficits) may be able to complete the form without help. It provides a simple descriptive profile of health in five dimensions (mobility, self care, social, pain and psychological), each with three levels. The patient's health state can therefore be classified into one of 243 (3^5) theoretically possible health states, each of which has been assigned a utility (i.e. value to the patient) (Dolan et al. 1995). These utilities were assigned by a group of stroke-free individuals and so probably require further validation, but they might allow the EuroQol instrument to be used for the economic evaluation of health care interventions and also the relative cost-utility of treatments for stroke compared with interventions for other diseases (e.g. cardiac transplantation). The EuroQol also includes a visual analogue scale on which patients rate their own health between 0 and 100, so providing an overall numeric estimate of their health related quality of life.

Although the EuroQol is a valid assessment of health related quality of life in the general population (The EuroQol Group, 1990; Brazier *et al.* 1993; Brooks & with the EuroQol group, 1996), its validity has not been adequately assessed after stroke. I therefore assessed certain aspects of its validity by comparing it with a variety of widely used and previously validated instruments in a group of prospectively studied stroke patients.

3.2 Methods

3.2.1 Validity

Validity is the degree to which an instrument measures what it is intended to measure, (Chapter One). There are at least four aspects to validity (Hobart *et al.* 1996). Face validity involves a subjective assessment of whether or not the instrument measures what it is intended to measure. Content validity is a subjective assessment of how well the domain of interest is sampled. I did not examine these aspects of validity quantitatively in the current study. Neither did I assess criterion validity, which involves comparison of the results of the instrument under study against those of a gold-standard, because there is no gold standard for the measurement of health related quality of life. I therefore focused on the concurrent (convergent) validity and compared outcome in each domain of the EuroQol with an assessment of function in the same domain using a standard instrument (see below). I examined the discriminant validity of the EuroQol in two ways: by assessing the relationships between the individual domains of the EuroQol and also by examining whether it could distinguish groups of patients with different types and severities of stroke (and therefore likely to have different health related quality of life outcomes).

3.2.2 Patients

All patients with acute stroke (first or recurrent) who attend our hospital, as either inpatients or outpatients, were prospectively identified and assessed by experienced stroke physicians as part of a prospective stroke register. These assessments included examination of the patients' clinical status as well as an estimate of the patients' pre-stroke level of functioning using the Oxford Handicap Scale (Bamford et al. 1989). Details of eligible patients were included in the stroke register, (Appendix 1 for registration form). I selected patients from the register who had survived at least 3 months after their stroke and who lived within an approximate 10 mile radius of the hospital (determined by scrutiny of their post codes). The NHS central registry Office of Population Censuses and Surveys notified me of the death of any patients enrolled in the register. I excluded 12 patients whose vital status could not be confirmed by the Office of Population Censuses and Surveys on the day the sampling frame was assembled. For the current study, I aimed to examine the validity of the EuroQol in consecutive surviving patients from two distinct time periods. As a pilot study, I examined the validity of the EuroQol in a group of patients who might be considered longer term survivors. Of 98 consecutive patients registered between 1 October 1990 and 18 May 1991, 36 were still alive at the start of the present study. Twenty eight of this group (78% of survivors) were contactable and willing to participate. Three hundred and forty five patients were assessed and registered at our hospital between 31 May 1993 and the 20 April 1995. Of these, 193 were alive at the start of the current study. One hundred and twenty four of these patients (64% of surviving patients) were contactable and willing to be interviewed. Both groups of patients will be considered together for the purpose of this study.

3.2.3 Assessments

All patients were visited by a research nurse (FW) in their place of residence. The nurse administered the modified simple questions (SQ) (see Chapter Eight), EuroQol, Frenchay Activities Index (FAI) (Wade *et al.* 1985), Hospital Anxiety and Depression Scale (HADS) (Zigmond & Snaith, 1983) and VAS pain scale as questionnaires for self-completion when possible. When patients could not complete these questionnaires by themselves, they were administered by interview. The modified SQ and EuroQol were always administered first, to limit interaction with the subsequent questionnaires. The nurse assessed the Barthel Index (BI) and the Office of Population Censuses and Surveys (OPCS) disability scores by direct questioning at the end of the interview (Mahoney & Barthel, 1965; Wellwood *et al.* 1995). The OPCS disability instrument includes a communication subscale which I used to identify patients who had significant problems with communication.

3.2.4 Analysis

I initially assessed the concurrent validity by calculating the median score and interquartile range for the appropriate unidimensional instrument for each level in the corresponding domain of the EuroQol. I used the Kruskal Wallis one way analysis of variance to assess the significance of the differences between these distributions. I also assessed the convergence between each domain of the EuroQol and the relevant standard instrument with the Spearman rank method. For the assessment of discriminant validity I determined the outcomes of the EuroQol for patients with different stroke syndromes (defined at baseline by clinical examination) (Bamford *et al.* 1991). I tested the discriminant validity further by examining responses to the EuroQol for patients with differing stroke severities. I based my assessment of stroke severity on their predicted prognosis at baseline. This predicted prognosis was calculated using a validated prognostic model designed to predict each patients'

probability of being alive and independent at six months (Counsell *et al.* 1996). The variables in this model were: age, pre-existing disability, marital status, verbal component of the Glasgow coma score, the ability to lift both arms against gravity and the ability to walk without help from another person (Counsell *et al.* 1996). I also estimated the correlation between the domains of the EuroQol using the Spearman rank method with a two-tailed test of significance to determine the degree by which the different domains of the EuroQol discriminate between the different constructs of health related quality of life.

I determined the independent explanatory factors of overall health related quality of life (based on the visual analogue scale) by multiple linear regression, using a method of forward selection of variables. I examined whether the assumptions of linearity and homogeneity of variance held for the data by plotting the distributions of the residuals. All analyses were performed using "Access 2.0" (Microsoft Corporation) and the statistical software package "SPSS for Windows" (Release 6.1).

3.3 Results

152 patients participated in the study. Of these, 92 patients (61%) were able to complete the questionnaires without help; the remaining 60 (39%) could only be assessed by interview. Using the OPCS communication subscale (a score of over five implies "very difficult for strangers to understand or worse") the interviewer rated six of the interviewed patients as having significant difficulties in communication. I excluded the data on these six patients from the analyses.

The patients were assessed at a median interval of 72 weeks after the onset of the index stroke (interquartile range: 43 to 104 weeks). Thirty-six percent of patients had not been admitted to hospital for treatment during the acute phase of their stroke. Patients are normally only admitted to our hospital during the acute phase if in-patient nursing or rehabilitation is required. The characteristics of the patients recorded at the time of registration following their index stroke are shown in Table 3.1 and their functional status at the time of assessment is shown in Table 3.2. About one third of patients reported dependency in activities of daily living, one third were independent but had persisting problems and the remaining third were independent and with no stroke related problems. Patients who could not complete the EuroQol questionnaire themselves had significantly worse functional ability than those who could complete it ($p < 0.0001$).

The median score (and interquartile range) with the standard instruments are shown for groups defined by their responses to the mobility, self care, social, pain and psychological functioning domains of the EuroQol (Table 3.3). The median scores with the standard instruments were ordered appropriately (increasing dysfunction reported with the EuroQol was associated with lower scores on the standard instruments) and differed significantly from each other ($p \leq 0.0002$ in all domains).

I also examined convergence between each domain of the EuroQol and the standard instrument with the Spearman rank method (Table 3.4). Bivariate correlations were moderately good for all domains except psychological functioning with the EuroQol and the depression subscale of the HADS. However, the patient estimates of overall health related quality of life correlated most closely with this subscale of the HADS (Table 3.4).

The pattern of outcomes reported by the questionnaire component of the EuroQol for each of the major stroke syndromes (defined by clinical examination at baseline) are shown in Figure 3.1. For all domains, the worst outcomes were observed in patients with the most extensive cortical strokes (total anterior circulation strokes, TACS). The best outcomes were observed in patients with posterior circulation strokes (POCS). A similar pattern was observed for the numeric estimates of overall health related quality of life. I also assessed discriminant validity in patients ordered by baseline stroke severity (lower tertile, the predicted probability of good outcome was 0.006 to 0.339; middle tertile, the predicted probability of good outcome was 0.339 to 0.530; upper tertile, the predicted probability of good outcome was 0.530 to 0.908). With the exception of the psychological functioning domain, better predicted prognosis was associated with better reported health status at follow up, Figure 3.2. Patients with the highest predicted probability of good outcome at baseline had significantly higher reported overall health related quality of life (mean score 73/100 for patients in top tertile versus 62/100 for patients in the middle tertile and 62/100 for patients in the lower tertile, $p<0.05$).

The bivariate correlations between each of the individual domains of the EuroQol instrument are shown in Table 3.5. Mobility correlated best with social functioning

(Spearman rank correlation coefficient 0.56) and least well with the psychological outcome (Spearman rank correlation coefficient 0.28).

The significant independent explanatory factors of the patient estimates of overall health related quality of life are shown in Table 3.6. An examination of the residuals confirmed that the assumptions of linearity and homogeneity had been met.

3.4 Discussion

I have investigated some aspects of the validity of the EuroQol in stroke patients by comparing responses to the EuroQol with those to validated instruments for the assessment of mobility, self care, social functioning, pain and psychological functioning. Concurrent validity was good; patients who reported problems on the EuroQol also reported dysfunction with the relevant standard instrument for that domain. The EuroQol was valid both in patients who could complete questionnaires themselves and among more severely affected patients who could only be assessed by interview.

3.4.1 Construct validity: does health related quality of life differ according to stroke type and severity?

The discriminant validity of the EuroQol was demonstrated by outcome profiles which distinguished between the major stroke syndromes classified by the Oxford Community Stroke Project method (Bamford *et al.* 1991). This classification has several properties: it describes groups of patients with a characteristic natural history and prognosis (Bamford *et al.* 1991; Anderson *et al.* 1994), it has been shown to be moderately reproducible for patients seen in the acute phase of stroke (Lindley *et al.* 1993), and it classifies patients according to the underlying mechanism and pathology

(Lindgren *et al.* 1994; Wardlaw *et al.* 1996). de Haan and coworkers have reported that lesion location and stroke severity influence health related quality of life measured six months after stroke with the Sickness Impact Profile (de Haan *et al.* 1995); in particular, patients with infratentorial strokes had better health status outcomes than patients with supratentorial strokes. Like de Haan *et al*, I found that health related quality of life was worst in all domains for patients with total anterior circulation strokes and best in patients with posterior circulation strokes. Outcomes were similar in patients with partial anterior circulation strokes and with lacunar strokes, which is consistent with the epidemiological data which suggested that partial anterior circulation strokes and lacunar strokes have a very similar prognosis (Bamford *et al.* 1991). The different stroke syndromes studied were defined by the extent rather than the severity (i.e. prognosis) of the neurological deficit, so I also looked to see if health status was different for patients with differing degrees of stroke severity. Patients with less severe strokes (i.e. good predicted prognosis) had better outcomes as assessed by the EuroQol. The convergent relationships between the mobility, self care and social functioning domains of the EuroQol further supported its construct validity, as functioning in all these domains was closely related to the patients' physical functioning. I found good discriminant validity since mobility, self care and social functioning correlated much better with each other than with the domains assessing pain and psychological functioning.

3.4.2 Validity of overall estimates of health related quality of life

It is difficult to assess the validity of the numeric estimates of overall health related quality of life as this domain is difficult to define and is highly subjective. I could not assess its concurrent convergent validity as I could not identify other validated instruments which claimed to measure a similar outcome. However, the visual

analogue scale for estimating overall health related quality of life does at least appear to have discriminant validity since I found differences in the mean estimates of overall health related quality of life across the different stroke syndromes and severities. These follow a similar trend to the health outcome profiles; the lowest reported mean health related quality of life was observed in patients suffering TACS and highest mean estimate of health related quality of life was reported by patients who had suffered POCS. These observed trends were not statistically significant. Linear multiple regression modeling revealed that the statistically significant independent explanatory factors of good overall health related quality of life were the absence of depression on the HADS depression subscale, good social functioning on the Frenchay Activities Index and the absence of pain on the VAS pain scale. These three variables explained approximately 38% of the variability in the estimates of overall health related quality of life. The relationship between psychological, social functioning and pain and the estimates of overall health related quality of life in our study, provides strong support for the validity of the measurement of overall health related quality of life by the EuroQol, as psychological outcome has been reported to be as important as physical disablement in determining quality of life after stroke (see Chapter 2). Kwa and colleagues, using a similar visual analogue scale to investigate quality of life after stroke (Kwa *et al.* 1996), reported that dependency (measured with the Rankin score), infarct volume and aphasia were significant independent predictors of quality of life. However, only 22% of the total variation in the quality of life scores was explained by their model. In my model, physical functioning was not a significant independent explanatory factor of overall health related quality of life after stroke. This was probably not because physical functioning was unimportant to the patients, but merely a reflection of the close relationship between a patient's physical and social functioning (see Table 3.5).

3.4.3 Face and content validity

The face and content validity of a measure refers to its clarity, clinical credibility and completeness (McDowell & Newell, 1996). Although, I selected the EuroQol for this study because of its simplicity, clarity and content, I did not assess these subjective parameters quantitatively. However, our qualitative experience provided some support for its face and content validity - we found few patients asked for clarification of the questions (personal communication, Fiona Waddell). Further, albeit indirect, support for its face and content validity is provided by my study of the feasibility of the EuroQol after stroke (see Chapter Five). Future studies should examine which dimensions of life are perceived to be important by stroke patients, their families, and their health care personnel.

3.4.4 Appropriateness of study population

Although many of the patients were studied more than one year after the stroke, the study cohort included a good mix of "dependent", "independent but not fully recovered" and "fully recovered" patients. Very few patients (<10%) were unable to complete the EuroQol, either by themselves or by interview. This study population was ideal, as the measurement of "quality of life" outcomes is only meaningful in patients who can either complete assessments without help or communicate their views to an interviewer; the validity of proxy completed EuroQol questionnaires is unclear and is examined in detail in Chapter Four. Furthermore, it may not be appropriate to test the validity of the EuroQol until at least one year after the index stroke as health related quality of life may not be stable before then (Astrom *et al.* 1992).

3.4.5 Usefulness of the EuroQol in different study designs

These data support the EuroQol as a useful measure of health status after stroke. Its simplicity is a definite advantage, as many stroke survivors find more complex instruments difficult to complete without help. Moreover, simple categorical data of the type generated by the EuroQol can convey a surprisingly large amount of information. Simple measures of this type are particularly well suited for use in large randomised controlled trials, audits and screening projects (Lindley *et al.* 1994; Mahoney *et al.* 1994

The EuroQol appears to be a reasonably valid measure which can be administered as either a questionnaire for self completion in patients with mild to moderate stroke or by interview in patients with significant neurological problems. If further studies confirm its reliability and feasibility in survivors of acute stroke, it could be usefully applied in a variety of ways ranging from routine clinical screening of patients for psychosocial problems after stroke, to the measurement of outcome in large randomised controlled trials and audit studies.

3.5 Summary of Chapter Three

1. The EuroQol is short and simple and many stroke patients can complete it without help. It measures several key aspects of health that are highly relevant to patients with stroke: mobility, self care, social functioning, pain, psychological functioning and overall health related quality of life. This gives it excellent face and content validity.

2. The concurrent validity of the EuroQol questionnaire was good. Patients who reported problems with the EuroQol also reported problems with standard outcome instruments.

3. The discriminant validity of the EuroQol was confirmed. The scores could distinguish between patients with different stroke syndromes and different stroke severities. Its construct validity was confirmed by the convergent relationships between the mobility, self care and social functioning domains. Divergent relationships between these three domains and the domains which assessed pain and psychological functioning provided further support for its construct validity.

4. The numeric estimates of overall health related quality of life also appeared to be valid in that they were best explained by the patient's psychological functioning, social functioning and the level of pain.

5. The EuroQol was valid amongst patients who could complete questionnaires themselves and also among more severely affected patients who could only be assessed by interview.

Table 3.1: Characteristics of 146 patients without severe communication difficulties at the time of inclusion in the stroke register

	n	(%)
Admitted for acute stroke	90	(62)
Level of function before index stroke		
OHS ≤ 2 [a]	133	(91)
OHS ≥ 3 [b]	13	(9)
Living alone prior to stroke	37	(25)
Employed prior to stroke	24	(16)
Stroke syndrome		
Total anterior circulation stroke syndrome	11	(8)
Partial anterior circulation stroke syndrome	59	(40)
Lacunar stroke syndrome	40	(27)
Posterior circulation stroke syndrome	25	(17)
Uncertain	11	(8)
Stroke severity at baseline		
Reduced conscious level	13	(9)
Unable to walk without help	46	(32)
Age		
<50	8	(5)
50 to 70	67	(46)
>70	71	(49)

[a] Oxford Handicap Score ≤ 2 implies that the patient was independent in activities of daily living

[b] Oxford Handicap Score ≥ 3 implies that the patient was dependent in activities of daily living

TABLE 3.2: Functional ability of patients assessed by responses to the "simple questions" in patients unable to complete them without help (interview-completed) and in patients able to complete the questionnaires at the time of the interview (self-completed)

Functional ability	Interview-completed* (n=54) n (%)	Self-completed* (n=92) n (%)	All patients (n=146) n (%)
Independent and no problems	2 (4)	39 (42)	41 (28)
Independent, and persisting problems	19 (35)	30 (33)	49 (34)
Dependent	31 (57)	20 (22)	51 (35)
Unknown	2 (4)	3 (3)	5 (3)

* Patients who required interview had significantly worse functional ability (chi squared = 30.6, df = 2, p< 0.0001)

TABLE 3.3: The concurrent validity of the EuroQol in survivors of acute stroke

Domain of the EuroQol	Number with this response	Median score on unidimensional instrument	Interquartile range	Chi-square (2p)*
"Mobility"		OPCS locomotion subscale		
No problems	52	0	(0 - 0)	
Some problems	89	3	(0 - 7)	54.5
Confined to bed	4	11.5	(11.5 -11.5)	(<0.0001)
"Self Care"		Barthel Index		
No problems	93	20	(20 - 20)	
Some problems	37	18	(17 - 20)	65.0
Unable	13	9	(6 - 12)	(<0.0001)
"Social"		Frenchay Activities Index		
No problems	61	31	(25 - 36)	
Some problems	60	18	(10 - 26)	52.1
Unable	23	7	(3.5 - 12.5)	(<0.0001)
"Pain"		Visual analogue pain scale		
No pain	68	0	(0 - 0)	
Moderate pain	61	20	(0 - 45)	71.0
Extreme pain	11	60	(50 - 70)	(<0.0001)
"Mood"		HADS Anxiety Score		
Not anxious or depressed	78	3	(1 - 5)	
Moderately anxious or depressed	49	8	(6 - 11)	41.3
Extremely anxious or depressed	7	12	(8 - 13)	(<0.0001)
"Mood"		HADS Depression Score		
Not anxious or depressed	78	3	(2 - 6)	
Moderately anxious or depressed	49	7	(3 - 8)	16.9
Extremely anxious or depressed	7	7	(4 - 9)	(0.0002)

*Kruskal Wallis 1-way ANOVA used to assess the significance of the differences between the distributions.

TABLE 3.4: Convergent validity: correlation between the domains of the EuroQol and the standard unidimensional instruments.

Domain of the EuroQol	Unidimensional instrument	Correlation coefficients between domains of EuroQol and relevant standard instrument		Correlation coefficients between standard instrument & overall HRQoL with EuroQol	
		r	(n)	r	(n)
Mobility	OPCS locomotion subscale	0.61	(145)	-0.42	(142)
Self care	Barthel Index	-0.64	(144)	0.26[a]	(141)
Social functioning	Frenchay activities index	-0.60	(145)	0.49	(142)
Pain	VAS pain scale	0.71	(143)	-0.41	(141)
Psychological functioning	HADS anxiety subscale	0.56	(135)	-0.25[b]	(135)
Psychological functioning	HADS depression subscale	0.35	(137)	-0.54	(137)

r = Spearman rank correlation coefficient
[a] p=0.002
[b] p=0.003
For all other correlations p<0.0001

Table 3.5: Convergent-discriminant validity: correlations between each domain of the EuroQol health status instrument with the other five.

Independent variable	*Self care*		*Social functioning*		*Pain*		*Mood*		*Overall HRQoL*	
	r	(n)	r	(n)	r	(n)	r	(n)	r	(n)
Mobility	0.46	(145)	0.56	(145)	0.43	(144)	0.28[a]	(143)	-0.43	(142)
Self care	*		0.61	(145)	0.34	(141)	0.24[b]	(143)	-0.28[a]	(142)
Social	*		*		0.35	(144)	0.35	(144)	-0.49	(142)
Pain	*		*		*		0.31	(143)	-0.40	(141)
Mood	*		*		*		*		-0.34	(140)

r = Spearman rank correlation coefficient

n = number included in analysis (number not constant because of missing data)

[a] p=0.001

[b] p=0.005

For all other correlations shown, p<0.0001

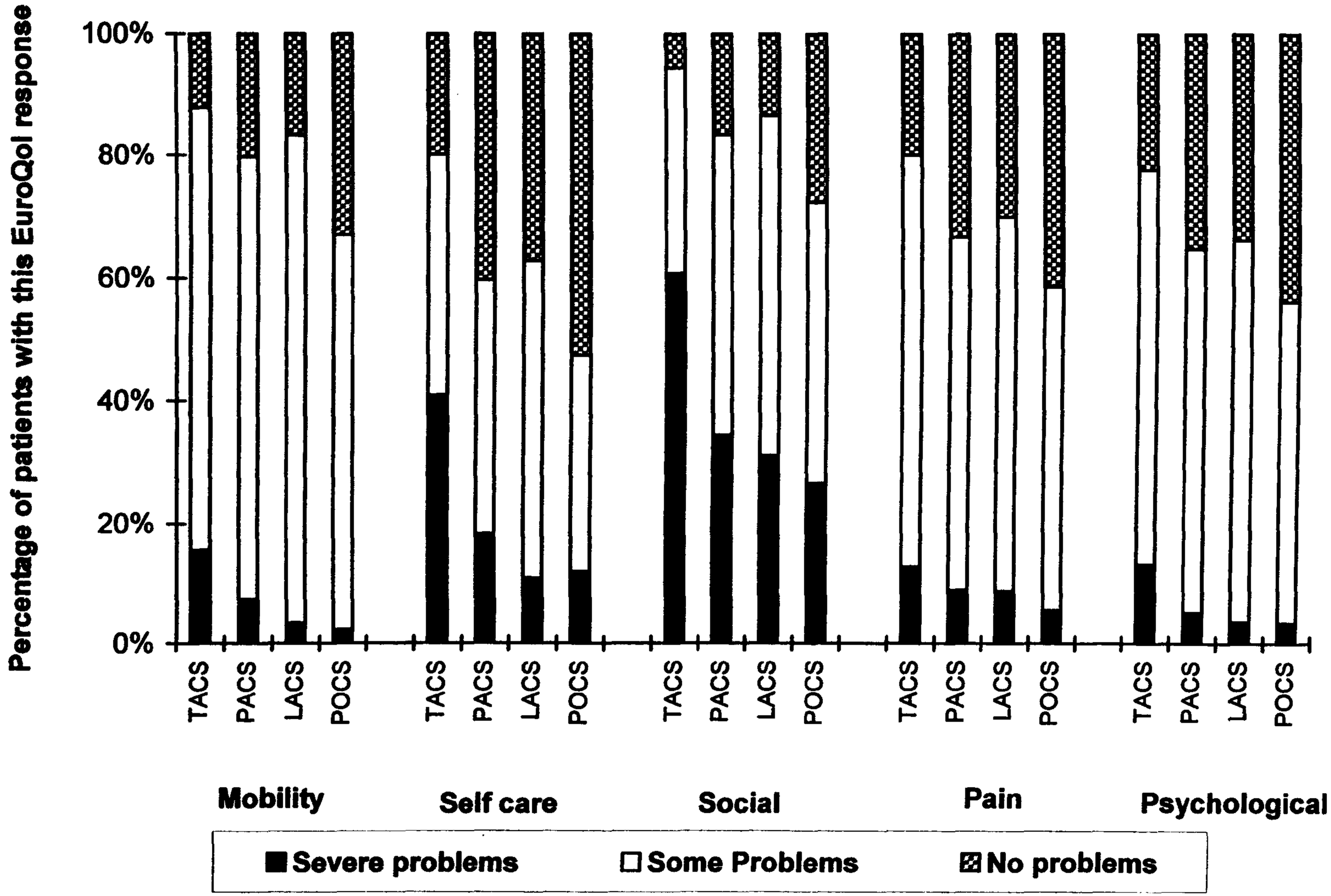
Table 3.6: Multiple linear regression explanatory factors of numeric estimates of overall health related quality of life.

Variable (possible range of scores)†	B	SE B	p value
Significant independent explanatory factors of HRQoL*			
HADS depression subscale (21 to 0)	-2.41	0.48	0.0001
Frenchay Activities Index (0 to 46)	0.38	0.22	0.008
Visual analogue pain scale (100 to 0)	-0.17	0.07	0.01
Non-significant explanatory factors of HRQoL			
Barthel Index (0 to 20)			
HADS anxiety subscale (21 to 0)			
OPCS locomotion subscale (11.5 to 0)			
Dependency question (1 or 0)			
Recovery question (1 or 0)			

* Adjusted $R^2 = 0.38$. This implies that these three significant independent variables explain 38% of the total variation in the estimates of overall HRQoL.

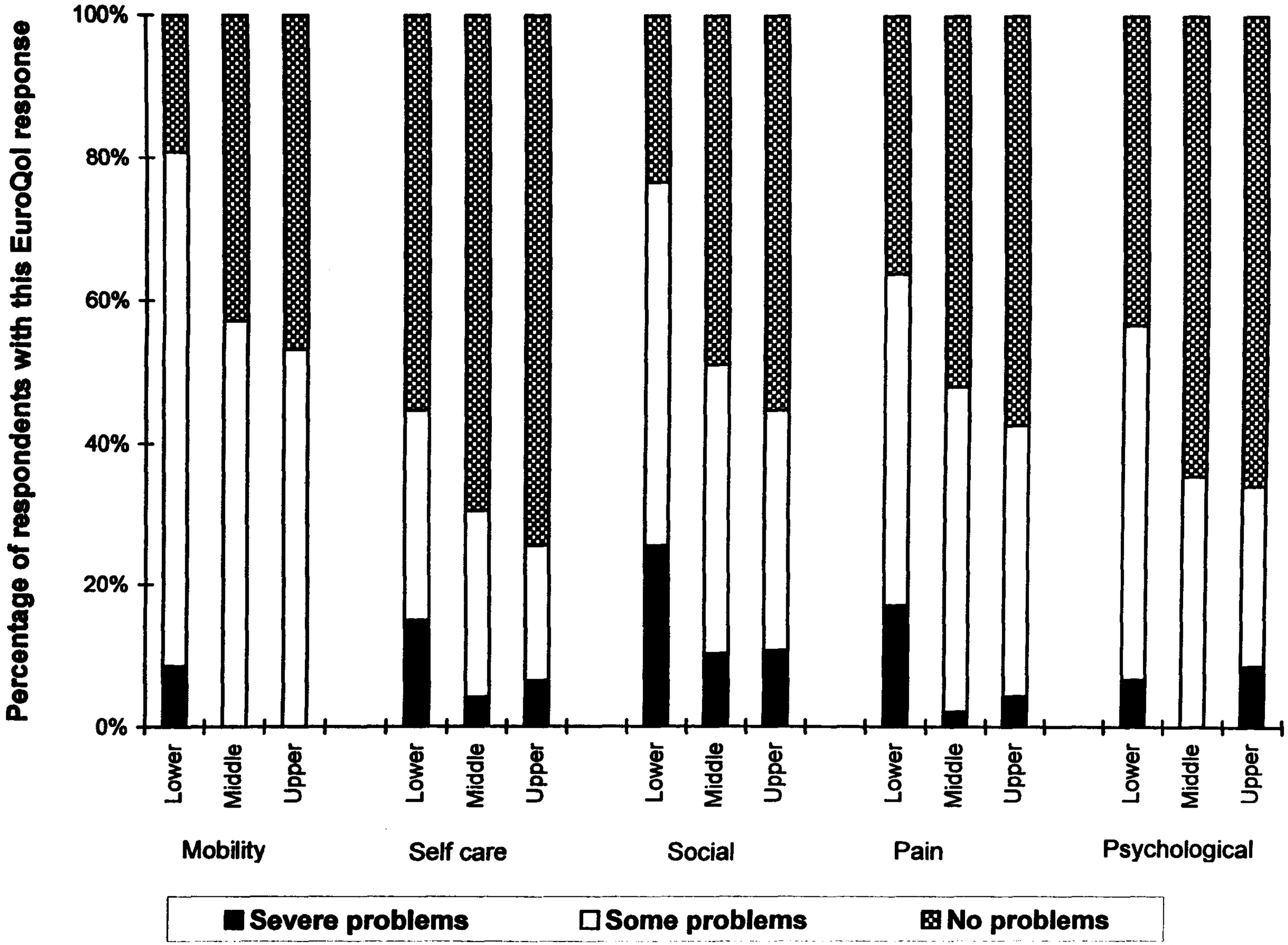
† worst score to best

Figure 3.1: Responses to the EuroQol questionnaire according to stroke syndrome at initial assessment



Legend: TACS = total anterior circulation stroke; PACS = partial anterior circulation stroke; LACS = lacunar stroke; POCS = posterior circulation stroke

Figure 3.2: Responses to the EuroQol questionnaire ordered by tertiles of baseline stroke severity (predicted probability of being alive and independent at 6 months)



Legend: predicted probability of good outcome: lower tertile 0.006 to 0.339, middle tertile 0.340 to 0.530, upper tertile 0.531 to 0.908

4 Are proxy assessments of health status after stroke with the EuroQol questionnaire feasible, accurate and unbiased?

4.1 Introduction

It is often difficult to measure the health related quality of life of patients with stroke because physical and cognitive (e.g. dysphasia, confusion, attention, or visuo-spatial neglect) problems limit their ability to complete complex questionnaires or participate in interviews (Kwa *et al.* 1996). Asking someone else, such as the carer, may be the only way to assess quality of life for a patient who is unable to complete the questionnaire themselves (this is often referred to as a proxy measure). Proxy measures of the SF-36 were disappointingly inaccurate (Segal & Schall, 1994). However, rating of a patients' functioning on the EuroQol by a proxy could prove to be more accurate, since much of the information sought is concrete and observable. I therefore examined whether a proxy could assess a stroke patient's health related quality of life with the EuroQol accurately and without bias.

4.2 Methods

4.2.1 Patients & Assessments

I described, in Chapter 3, a study of the validity of the EuroQol questionnaire in a series of 152 patients from our prospective registry of inpatients and outpatients with first (or recurrent) stroke. I also used this series of patients to examine the feasibility and validity of proxy assessments of health related quality of life after stroke. Patients could select the person they considered to be the most appropriate proxy

for them; we asked all patients to choose someone who knew them well (e.g. close relative, friend or carer) and who could be available at the time of the interview. The study nurse asked these proxies to complete a EuroQol questionnaire (preferably in a separate room) on behalf of the patient. If a proxy was not available at the time of the interview, I contacted them by post.

4.2.2 Analysis

I calculated the level of agreement for the categorical data items of the EuroQol between the assessments by the patient and their proxy of the patients' health status. I did not use a correlation coefficient (e.g. Spearman rank) to assess agreement because it only measures association and would be constant under deviations of scale or bias (Brennan & Silman, 1992). I therefore used the kappa statistic, which measures the amount of agreement beyond that which could be expected by chance (Brennan & Silman, 1992; Altman, 1993). I calculated the variance of the kappa statistic using Altman's method (Altman, 1993). As the scale items had three levels of response I used all three levels for the estimation of kappa. I assessed agreement separately for patients who were able to complete the EuroQol questionnaire themselves and those who could not complete the EuroQol themselves and consequently had to be interviewed. Unfortunately, there are no absolute definitions for the interpretation of any given kappa statistic. I planned to base my interpretation of the kappa statistic on the following widely cited guidelines: <0.2 implies poor agreement, 0.21 to 0.40 implies fair agreement, 0.41 to 0.60 implies moderate agreement, 0.61 to 0.80 implies good agreement, 0.81 to 1.00 implies very good agreement (Landis & Koch, 1977; Altman, 1993; Brennan & Silman, 1992).

The analysis of the continuous data from the visual analogue scale on overall health related quality of life was more complex. Differences between the patient and their proxy in their estimates of the patients' overall health related quality of life might be due to observer error, systematic differences (i.e. bias), or random effects (i.e. the play of chance). To display the raw data, I planned to plot a simple scatterplot and calculate a linear correlation coefficient. However, this plot gives little information on systematic differences, so I also performed a Bland and Altman analysis which plots the difference between the two estimates against the mean of the two estimates (Bland & Altman, 1986). The EuroQol is bounded at 0 and 100, which limits the value of a Bland and Altman plot, so I also used a factorial analysis of variance (SPSS for Windows, Release 6.1) to calculate the intraclass correlation coefficient, an appropriate measure of agreement for continuous data (Morton & Dobson, 1989; McDowell and Newell, 1996).

4.3 Results

152 patients participated in the study. Of these, 92 (61%) were able to complete the questionnaires independently and the remaining 60 patients (39%) could only be assessed by interview, see Chapter 3. The interviewer rated six of these 60 patients as having significant difficulties in communication by the OPCS communication subscale (all scored > 5 which implies they are very difficult for strangers to understand or worse). I excluded the data on these six patients from the analyses as it was almost exclusively derived from the carer, not the patient, and so was not informative for the current analyses (which required that the patient should be able to provide information directly and equally well by interview or self-completed questionnaire). A proxy was available and completed a form for 130 patients (86%): 94 of 130 forms (72%) were completed at the time of the home visit and 36 were returned later by post (of these, 16 were completed within one day of the patient

assessment and all but one were completed within 7 days of the patient assessment).

Agreement between the proxies' and the patients' estimate of health related quality of life is shown in Table 4.1. Agreement was better for patients who were able to self-complete the EuroQol than for patients who required the EuroQol to be administered by interview. For both groups, agreement was best for the self care domain and worst for the domain assessing psychological outcome. For the more severely affected patients (assuming that the reasons for being unable to self-complete are generally stroke related), agreement was only fair for the pain and social functioning domains and no better than chance alone for the psychological functioning domain (kappa=0.05, 95% CI: zero to 0.43).

Plotting the differences between the patient and proxy estimate of overall health related quality of life against the mean score (Bland and Altman plot) showed an expected distribution for a score bounded at zero and 100 (Figure 4.1). For all patients combined, the mean of the differences between the patient and proxy estimates of overall health related quality of life was 2 (95% confidence interval for a pair of differences = -38 to 42, Table 4.2); this indicates that the proxy estimates of overall health related quality of life were not significantly different to the patients. A factorial analysis of variance also suggested that there was no statistically significant variance between patient and proxy numeric estimates of overall health status. For all patients combined, the intraclass correlation coefficient (a measure of the agreement between the patient and proxy estimates of overall health status) was moderate with an intraclass correlation coefficient of 0.49 ($p < 0.0001$). Agreement for the estimates of overall health related quality of life was better for the subgroup of patients who were able to complete the EuroQol themselves. The intraclass

correlation coefficients for those able to complete was 0.53 compared with 0.32 for patients unable to complete by themselves.

Using the categorical data, a higher proportion of patients reported "no problems" in each of the five domains than did their proxies (Table 4.3). In these categorical domains of the EuroQol, the proxy estimated the level of functioning to be the same as that reported by the patient for 466 of the potential 640 outcomes. For 100 outcomes, the proxy estimated the functioning to be worse than that estimated by the patient. By contrast, there were only 74 outcomes for which the proxy estimate of the patients' functioning was better than that reported by the patient (test for symmetry, $p<0.05$).

4.4 Discussion

Many stroke survivors are unable to complete questionnaires measuring health status by themselves. The use of a proxy to assess a patient's health related quality of life should help increase the proportion of patients in trials and surveys of stroke therapy who have complete data. This should improve the quality and generalisability of the data. In the current study, I could obtain proxy assessments for 86% of the patients which suggested that proxy assessment of health related quality of life after stroke was generally feasible. My analyses suggested that the patient and their proxy agreed reasonably well in their assessment of some aspects of the patients' health related quality of life after stroke, particularly for mobility and self care. Agreement was less good for social functioning, pain and the overall estimate of health related quality of life and even worse for psychological functioning. The degree of agreement between proxy and patient varied and was better among less severely affected patients who completed the initial EuroQol themselves. However, proxy assessments would be of value if they could also be used in more

severely affected patients, who were unable to complete questionnaires themselves. The degree of agreement among patients who were more severely affected and had to have the EuroQol by interview may therefore provide a more realistic guide to the value of proxy assessments.

4.4.1 Are proxy assessments less accurate among more severely affected patients?

In this study, the agreement was apparently less among more severely affected patients. This loss of agreement could have been due to observer error by the proxy or a systematic difference due to the different mode of questionnaire administration. The latter notion is supported by a recent report which suggests that patients give a more optimistic picture of their health status when assessed by interview than by self-completed questionnaire (Weinberger *et al.* 1996). Furthermore, random errors may be important, as the sample size was quite small (especially for the subgroup analysis in Table 1) so the 95% confidence interval around each estimate of agreement is wide and does not exclude the possibility of substantially better agreement. There are a number of other possible sources for less than perfect agreement. When a patient and their proxy appear to disagree about the patients' health status after a stroke, the following factors may contribute to the disagreement: the domain under study, systematic differences in perceived health (i.e. bias), relationship of the proxy to the patient, random error and the choice of statistic to measure agreement. The poor agreement for social functioning, pain, psychological functioning and overall health related quality of life probably reflected the subjective nature of these domains.

4.4.2 Whose assessment is most valid?

The proxy tended to report the patients' problems as more severe than did the patient. This suggests that proxy assessments of health related quality of life do indeed differ systematically from self-assessments. Were the patients more optimistic about their health status than their proxies, did the patients adjust to or fail to perceive their own deficits, or were the proxy responders being pessimistic? The patients' view is likely to be more valid, as health related quality of life instruments primarily aim to assess the patients' subjective perception of their own health. However, I cannot be certain, since there is no accepted "gold standard" for the measurement of HRQoL.

4.4.3 Choice of proxy

I allowed the patient to decide who could act as their proxy (rather than stipulate that they must choose a spouse or a close family member). It is possible that some of the proxies were selected simply because they were available and so might not have known the patient well enough to complete the assessment accurately. If allowing the patient to choose the proxy does introduce some extra measurement error, the error might not be reduced by insisting that a family member is used as the proxy: regrettably not all blood relations are sufficiently familiar to assess their relatives' health related quality of life reliably! Furthermore, many patients do not have any family members living nearby and so a relatively imprecise estimate by a close friend may well be better than a very imprecise estimate from a distant family member and is likely to be better than no estimate at all.

4.4.4 Are these results generalisable to more severely affected patients?

In my study, patients had to be able to complete the EuroQol either by themselves or by interview. I could not have assessed whether the proxy responses were valid for the patients who were unable to complete the EuroQol. Although I observed worse agreement for the patients who required the EuroQol to be administered by interview, I cannot necessarily infer that the agreement would have been even worse for even more severely affected patients (who have greater difficulties with communication) because the observed differences in agreement may have been due to the method of questionnaire administration (Weinberger *et al.* 1996). However, it seems likely that the use of proxies for patients who have difficulties with communication will have greater bias and measurement error because their relatives, friends and carers will almost certainly have less insight into their perceived HRQoL.

4.4.5 Other factors influencing proxy agreement

The distribution of the random error is likely to be strongly influenced by the reproducibility of the EuroQol. In other words, some domains may be more prone to measurement error than others. It is possible that the more subjective domains have the worst reproducibility. A number of methodological factors may have caused me to underestimate the true level of agreement between patients and their proxies. Firstly, as the EuroQol assesses the patients' health related quality of life on the day of completion, any delay in getting assessments from proxies who were not available at the time of the interview might have reduced the true level of agreement (as some of the patients could have changed). This effect is unlikely to be important as the majority of assessments (72%) were performed at the time of the home visit and nearly all of the remaining assessments were completed within 7 days of the home

visit. An unweighted kappa statistic may also underestimate the true level of agreement, because it ignores the ordering of the three levels of the EuroQol. Furthermore, the interval differences between each of the three levels of the EuroQol ("no problems", "some problems" and "severe problems") are unlikely to be equal. The difference between "no problems" and "some problems" may be bigger than that between "some problems" and "severe problems". Weighting the kappa statistic to get round these problems is not necessarily the solution, since any weights will inevitably be arbitrary (Brennan & Silman, 1992; Svensson, 1993). Finally, the dependence of the kappa statistic on the prevalence of the underlying attribute being measured complicates its interpretation (Brennan & Silman, 1992). Alternatively, it is also possible that I have overestimated the true level of agreement because I cannot be sure that some of the questionnaires returned by post were not completed with some input from the patient.

4.4.6 Use of proxies in randomised controlled trials

In a randomised trial or survey which measures HRQoL, allowing a proxy to respond on behalf of the patient has potential disadvantages: it may increase random error and so reduce the statistical power of the study to detect the treatment effect, particularly for the domain of psychological functioning (Woods, 1995) and it may also introduce bias. In an observational study, such bias might make the overall outcome appear worse than if the patient had responded. In a randomised trial, if the treatment were effective, this might reduce the number of patients with poor outcome who can only be assessed by proxy in the treatment group (but not in the control group) and so exaggerate the treatment effect. This type of bias would, however, not be expected to affect the direction of the treatment effect or its statistical significance (personal communication, Jim Slattery). Furthermore, the above bias is not unique to the EuroQol as proxy assessments of more objective

outcomes, e.g. disability, are affected by a similar bias (Lindley *et al.* 1994). In general, such "second order" biases are unlikely to be important. However, the use of proxy responses is also likely to ensure a higher overall response rate which might reduce the risk of random error and bias.

In summary, a proxy assessment appears feasible in a wide variety of patients. The proxy assessed the domains of mobility and self care accurately and without major bias, although there was a slight tendency for them to take a generally somewhat more pessimistic view of the patients' overall HRQoL. Therefore, for at least these domains, it seems reasonable to use proxy responses for the EuroQol in stroke patients who cannot complete questionnaires by themselves (especially if face-to-face interviews are not practicable). Proxy assessments of social functioning, pain, and overall health related quality of life were associated with more error and must be interpreted more cautiously. Proxy assessments of psychological functioning were the least reliable, particularly in patients who required the EuroQol to be administered by interview, in whom they were no more accurate than chance alone. These findings are consistent with other evaluations of ratings by proxies (Guyatt *et al.* 1993; Rothman *et al.* 1991). In general, allowing the use of proxy response where necessary may be preferable to forbidding them in randomised controlled trials and many types of observational studies. However, where the focus of an observational study is an aspect of health related quality of life other than physical functioning, the use of proxy responses may not be a good idea.

4.5 Summary of Chapter Four

1. Proxy assessments of health related quality of life in a wide range of stroke survivors were found to be feasible. We were able to obtain proxy assessments easily in approximately 86% of our patients.

2. The proxy assessed the domains of mobility and self care accurately. Proxy agreement was less good for social functioning, pain and the overall estimate of health related quality of life, and even worse for psychological functioning. Therefore, proxy respondents are capable of assessing disability, dependency and some impairments. They are unlikely to be able to assess quality of life.

3. The degree of agreement between proxy and patient varied and was better among less severely affected patients who completed the initial EuroQol themselves. This may be because the relatives, friends and carers of more severely affected patients (who have greater problems with communication) may have less insight into the patients' perceived health related quality of life. Alternatively, it may have been due to the different mode of questionnaire administration (interview rather than self completion).

4. The proxy tended to report the patients' problems as more severe than did the patient. The patients' view is likely to be more valid, as health related quality of life instruments primarily aim to assess the patients' unique perception of their own health.

Table 4.1: Agreement between categorical data items on EuroQol questionnaires completed by the patient and by a proxy[a] for the same patient

Health outcome domains (categorical data)	n[b]	agreement (%)	kappa (95 % confidence interval)
Agreement for patients who self completed initial EuroQol			
Mobility	80	78	0.57 (0.39 - 0.74)
Self care	78	83	0.62 (0.43 - 0.81)
Social functioning	79	76	0.57 (0.41 - 0.74)
Pain	78	74	0.54 (0.36 - 0.71)
Depression and/or anxiety	76	67	0.38 (0.18 - 0.58)
Agreement for patients who completed initial EuroQol by interview			
Mobility	50	84	0.48 (0.16 - 0.81)
Self care	50	76	0.62 (0.44 - 0.81)
Social functioning	48	67	0.37 (0.12 - 0.62)
Pain	50	60	0.30 (0.06 - 0.54)
Depression and/or anxiety	50	54	0.05 (0.00 - 0.43)
Agreement for all patients combined			
Mobility	130	80	0.60 (0.46 - 0.74)
Self care	128	80	0.64 (0.51 - 0.77)
Social functioning	127	72	0.56 (0.44 - 0.69)
Pain	128	69	0.45 (0.30 - 0.59)
Depression and/or anxiety	126	62	0.30 (0.14 - 0.45)

[a] this could be a carer, relative or close friend

[b] number with available data

Table 4.2: Agreement between patient and proxy assessments of overall health related quality of life

	n*	Mean overall HRQoL† (SD)		Mean of paired differences (SD)	Intraclass correlation coefficient
		Patients	Proxies		
Self-completed	76	69.8 (18.0)	65.4 (19.0)	4.4 (17.2)	0.53
Interview	46	55.0 (22.3)	56.2 (16.8)	-1.2 (23.8)	0.32
All	122	64.2 (20.9)	62.0 (18.7)	2.2 (20.0)	0.49

* number of patients for whom there was an assessment of overall health related quality of life for both patient and proxy

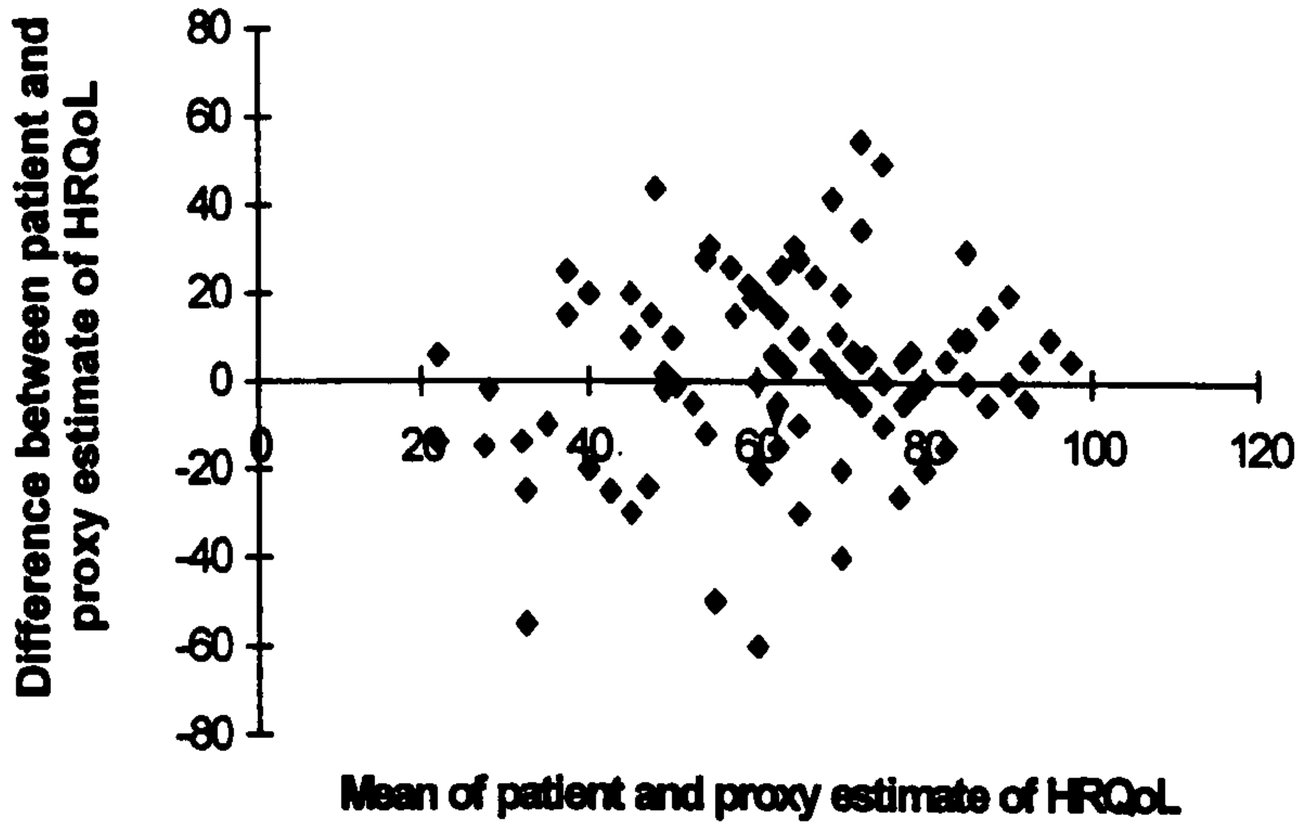
† analysis based only on patients (or proxies) for whom there was a corresponding proxy (or patient) assessment of health related quality of life. None of the differences were statistically significant.

Table 4.3: Proportion of patients with "no problems" in each of the five domains as assessed by the patients themselves and by their proxy

Health outcome domains	Number of patients who reported: "no problems"		Number of patients for whom the proxy assessed: "no problems"	
	n	(%)	n	(%)
Mobility (N=130)	44	(34)	43	(33)
Self care (N=128)	80	(63)	73	(57)
Social functioning (N=127)	51	(40)	45	(35)
Pain (N=128)	60	(47)	51	(40)
Depression \ anxiety (N=126)	70	(56)	59	(47)

N = number of patients for whom there was an assessment in the domain of patients health related quality of life by both the patient and their proxy

Figure 4.1: Difference between patient and carer estimate of overall health related quality of life plotted against mean overall health related quality of life (Bland and Altman plot).



A negative difference implies that the patient scored their overall health related quality of life as less than that estimated by the carer. Mean of differences = 2; 95% confidence interval for a pair of differences = – 38 to 42.

An overall score of 100 implies best imaginable health related quality of life and zero implies worst imaginable HRQoL.

5 A randomised parallel group comparison of the feasibility of the EuroQol and SF-36 after stroke

5.1 Introduction

The EuroQol and SF-36 are widely used generic instruments for the measurement of health status which have been validated after stroke (Anderson *et al.* 1996, Chapter 3). Both instruments were designed to be delivered by post and then be completed by the patient, although a variety of other methods of administration have also been employed. The SF-36 is very widely used (over a million SF-36 questionnaires were completed in 1992) (The MOS Trust, 1992), but it is also longer and more complex, and so may not be feasible in patients with stroke. As yet, however, these instruments have only been compared indirectly (i.e. not concurrently) or compared in studies with a non-random crossover design (i.e. both instruments were administered sequentially in the same order to all patients) in a variety of different patient groups (Brazier *et al.* 1993; Hollingworth *et al.* 1995; Brazier *et al.* 1996; Essink-Bot *et al.* 1997). These studies could be prone to biases whose effects might be as large as, if not larger than, any true differences between instruments. In particular, the indirect comparisons are particularly vulnerable to biases arising from differences between the study populations, whereas the non-random crossover studies are vulnerable to biases arising from ordering effects (i.e. the response to an instrument may differ if it is administered independently). These biases would be avoided in a direct randomised comparison as the process of randomisation aims to ensure that all study groups are similar. Furthermore, none of these studies have examined the properties of either instrument in patients with stroke, and so their relevance is debatable.

When investigators choose a health status instrument for a particular study they must take account of the intended application, the planned method of administration and the attributes of the instrument in the population being examined. The ease of completion is important (Scientific Advisory Committee, 1995), as it is likely to determine both the frequency of response and the number of items of missing data. If many recipients do not respond, or return only incomplete questionnaires this may introduce a number of biases, make interpretation difficult and reduce the generalisability of the results. Furthermore, if poor outcome is more common among non-respondents a non-linear interaction might occur between response frequency and the statistical power of the study to detect moderate differences between the two treatment groups, i.e. small falls in response frequency may lead to large reductions in statistical power. Moreover, if the patients have cognitive, motor, speech or language deficits (such as those present in stroke survivors) they may not be able to complete complex questionnaires.

Despite these considerations, response frequency was not mentioned in several recent papers on the selection of health related quality of life instruments (Guyatt *et al.* 1993; Coste *et al.* 1995; Testa & Simonson, 1996). As far as I am aware, there have been no randomised comparisons of the response to differing health related quality of life instruments. I postulated that the brevity and simplicity of the EuroQol questionnaire (five separate questions and a visual analogue scale) would achieve a significantly better response in stroke survivors than the SF-36 (34 separate questions). I therefore performed a randomised controlled trial to compare the response to postal versions of the EuroQol and SF-36 instruments.

5.2 Methods

5.2.1 Selection of patients

All United Kingdom (UK) patients who had been entered in the International Stroke Trial (IST) between 2 March 1993 and 31 May 1995 were considered for the current study. The IST is a multicentre international randomised controlled trial of antithrombotic therapy in patients presenting within 48 hours of onset of ischaemic stroke (International Stroke Trial Collaborative Group, 1997). I included UK patients randomised in the IST who were not known to be dead at the time of the survey. The NHS central registry Office of Population Censuses and Surveys notified me of deaths of patients in the study. I excluded patients whose vital status could not be confirmed by the Office of Population Censuses and Surveys. I also excluded patients if there were insufficient details to allow postal contact with either the patient or their general practitioner, or if they had previously contacted the Trial Coordinating Centre and asked not to be followed-up.

5.2.2 Randomisation

I randomised eligible patients using an allocation code generated by an adaptive randomisation algorithm (minimization)(White & Freedman, 1978), to postal follow-up with either the EuroQol or the SF-36 instrument. The algorithm aimed to balance the two groups for age, sex, stroke syndrome (using the Oxford Community Stroke Project Classification) (Bamford *et al.* 1991) and the time from stroke onset to follow-up in weeks.

5.2.3 Instruments

EuroQol provides a simple descriptive profile of health status in five domains and an overall numeric estimate of health related quality of life (The EuroQol Group, 1990). The SF-36 was developed in the USA from a larger battery of health status instruments employed in the Medical Outcomes Study and covers eight domains of health related quality of life (Ware & Donald Sherbourne, 1992; Brazier *et al.* 1992; Jenkinson *et al.* 1993). I incorporated both instruments, unaltered in format or content, into questionnaire booklets which included additional questions recording the patients address, type of residence, functional outcome after stroke (modified simple questions, see Chapter Eight) and whether or not the patient completed the form independently, (see Appendix 2 & 3). The questionnaire booklets were identical in all respects other than the nature of the health related quality of life instrument. I posted questionnaire booklets containing the appropriate instrument to all eligible patients with a personalised letter and a reply-paid envelope. The letter explained the purpose of the study and asked subjects to respond without help if possible, and if not, to pass the questionnaire on to a close relative or care-giver willing to respond on the patients behalf. A reminder letter and further questionnaire was sent to any patient who had not responded within two weeks. Thereafter, no further attempts were made to contact non-respondents. I marked individual questionnaire booklets with labels which included details of the patient's name, address, trial identifying number and questionnaire allocation. I generated all letters and labels directly from the database using a computerised mail-merge programme.

5.2.4 Outcome assessment

The primary measures of outcome for each instrument were: the frequency of response after both the first mailing and a reminder, and the number of forms with "no domains of missing data". For the SF-36, the domain was defined as missing when there were insufficient data to calculate an overall score for that domain after

interpolation of missing values (Medical Outcomes Trust, 1994). As the EuroQol was much shorter, any missing data resulted in a "domain of missing data". I planned exploratory analyses to examine the relationships between stroke syndrome and the frequency of response.

5.2.5 Power Calculations & Statistical analysis

The study was powered (power = 0.95 = (1-β), α=0.05) to detect an absolute difference in overall response frequency of 5%, i.e. of 50 forms per 1,000 between the two groups, assuming an overall mean response frequency of 75%. Odds ratios and confidence intervals were calculated using EpiInfo Version 5 (Center for Disease Control, Atlanta, Georgia).

5.3 Results

Of the 4,016 patients randomised by the UK centres in the International Stroke Trial between March 1993 and May 1995, I excluded 1,763. The reasons for exclusion were: 1,154 patients had died prior to the start of the study, 247 patients could not be easily traced by the Office of Population Censuses and Surveys to confirm survival, 238 patients had insufficient general practitioner or patient contact details and 124 patients were involved in other ongoing studies of health related quality of life. The remaining 2,253 patients were randomised; 1,125 to receive a EuroQol questionnaire and 1,128 a SF-36 questionnaire. The study groups were well matched for age, sex and distribution of baseline stroke syndromes, (Table 5.1). The median time interval between stroke onset and form completion was 56 weeks in both groups (range: 17 to 125 weeks in both groups). No errors in form allocation or dispatch were detected for either the EuroQol or SF-36 questionnaires.

Response was significantly more frequent in patients allocated the EuroQol instrument, 80% after one reminder versus 75% for the SF-36, (equivalent to a 35% increase in the odds of response with the EuroQol compared to the SF-36, 95% confidence interval: 10 to 66% odds increase, p=0.003),Table 5.2. A significant proportion of all respondents responded only after a reminder (17% of all respondents for the EuroQol versus 20% for the SF-36, p>0.05). For both instruments about half of all completed forms were completed by the patients (51% for the EuroQol and 50% for SF-36), and the remainder were filled in by carers. Allocation to the EuroQol increased the odds of response with complete data by 64% (95% confidence interval: 38 to 95% odds increase, p<0.0001) and also significantly reduced the overall level of missing data.

Missing data for the SF-36 was concentrated in the physical role limitations and emotional role limitations domains, (Table 5.3). Missing data for the EuroQol was most frequent in the domain which assessed overall health related quality of life (Table 5.4).

Respondents to the EuroQol questionnaire reported dependency in activities of everyday living as assessed by the modified "simple dependency question" significantly more frequently than patients allocated to follow-up with the SF-36 (40% increase in odds, 95% confidence interval 18% - 166% increase in odds), (Table 5.5).

Patients were less likely to complete and return the forms themselves if they had had a more extensive neurological deficit (total anterior circulation infarct). Conversely, patients with less extensive strokes, not causing deficits of cognitive function

(lacunar and posterior circulation strokes) were most likely to complete and return the questionnaires without help, (Table 5.6).

5.4 Discussion

To my knowledge, this is the first study which has directly compared the relative merits of two commonly used health status measures in a formal randomised manner. I found a significantly higher overall response frequency in stroke patients allocated to follow-up with the EuroQol questionnaire, and among responders to the EuroQol the proportion of missing data was smaller than for the SF-36. The observed difference in overall response frequency, although only modest in **absolute** terms (approximately 50 additional forms returned per thousand mailed) is of definite practical significance and would translate to a shortfall of approximately 1000 completed forms in a large survey with 20 000 subjects studied by SF-36 rather than the EuroQol.

The amount of missing data among responders was significantly different between the two instruments. Allocation to the EuroQol increased the odds of "response with the questionnaire having no domains of missing data" by 64% (95% CI 38% to 95% increase). The instrument with the best "response" and "response with no missing data" frequency is preferable. A higher response rate increases the efficiency of any study and reduces the resources used to chase-up non-respondents and seek items of missing data. The more complete the response the less the risk of bias by the empirical allocation of arbitrary values (e.g. worst possible outcomes in a sensitivity analysis) to missing items of data. This is particularly an issue where the intervention might affect response frequency and data quality; for instance, an

effective treatment for stroke might reduce the frequency of non-response in the treatment group and so bias any estimate of the size and direction of the treatment effect.

In this study, less than half of all patients completed the questionnaires themselves. This suggests that the problems experienced with completion of either instrument were independent of its characteristics, and are more likely to be a consequence of the frequent cognitive and motor deficits found after stroke. Nevertheless, the significantly better overall response frequency and data quality observed in patients randomised to receive the EuroQol suggests that it is the more practicable of the two for stroke. This is probably because the EuroQol is shorter, simpler and asks questions that are more relevant to survivors of stroke. Previous studies of the SF-36 in an elderly population have demonstrated that missing responses tended to be concentrated on items regarded as inappropriate to patients e.g. questions with an emphasis on work or vigorous activity (Hayes *et al.* 1995). A similar pattern of missing data was observed in the current study.

5.4.1 Proxy completed responses

Relatively few patients completed the EuroQol or the SF-36 themselves, so it is legitimate to ask whether or not the carer completed questionnaires gave a valid assessment of patients' health status. Segal examined agreement between survivors of acute stroke and their proxies for responses to the SF-36. Agreement was poor for all domains except physical functioning (Segal & Schall, 1994). This is likely to reflect the relatively subjective nature of many of the items in the SF-36, and raises concerns about the use of proxy assessments of health related quality of life with the SF-36. I have discussed the validity of proxy assessments of health related quality of life with the EuroQol in detail in Chapter 4.

5.4.2 Differences in response by patients in different categories of stroke type

In this study, patients with large cortical strokes (TACI) were the group that were least likely to complete the questionnaire themselves. About one fifth of this group completed the forms themselves, which is half the average of 50% self-completed. This probably reflects the particularly severe nature of the motor, cognitive and visual problems associated with this particular clinical syndrome. The low response frequency confirms the difficulty of measuring quality of life in a group of patients who are likely to have the lowest health related quality of life. This study illustrates how response frequency may differ in different types of patients, and highlights how non-randomised comparisons of different instruments might be confounded by differences of case mix in subgroups of the study population. Thus, it is always essential to consider whether or not it is appropriate to generalise the performance of an instrument in a specific type of patient to a mixed population of patients.

5.4.3 Response frequency may have important effects on assessment of outcome

Although feasibility and the acceptability of an instrument are important, there are many other attributes which guide the choice of instrument for a particular study (Guyatt *et al.* 1993; Scientific Advisory Committee, 1995). Other attributes include conceptual model, reliability, validity, responsiveness, interpretability and availability of the instrument in different forms and languages. A major problem with current statistical approaches to testing an instrument's measurement attributes, e.g. test-retest reliability or sensitivity to change, is that these assessments are based solely on data from respondents. An instrument's calculated measurement attributes may be affected by its response frequency. For instance, a complex and detailed instrument might demonstrate excellent sensitivity to change (in patients who

responded), but poor overall frequency of response. However, this may not be the consequence of the responsive nature of the instrument, but simply selection by the instrument of patients who had actually changed (and so were more motivated to respond). The converse is also a practical problem, as poor frequency of response might occur in those patients with poor health related quality of life. This was certainly noted among patients with the severest strokes in the current study. An instrument's responsiveness is therefore likely to be particularly dependent upon it being acceptable and feasible for use by such groups. In this study, respondents allocated to the EuroQol instrument reported significantly worse functional outcomes. As the treatment groups were well balanced for baseline prognostic risk, it suggests that the EuroQol instrument was more acceptable to dependent patients and their care-givers. This higher level of response from patients with poor outcomes is likely to improve the power of the EuroQol questionnaire for detecting differences in health related quality of life between the treatment and control groups at follow-up. Thus, paradoxically, a simple brief instrument with a high response frequency (e.g. the EuroQol) may have more power to detect differences at follow-up than a more detailed and complex measure with a lower frequency of response (e.g. the SF-36). Therefore, any critical assessment of an instrument's measurement characteristics must include consideration of the overall level of response.

5.4.4 Maximising follow up

Large randomised controlled trials, audit studies and surveys need to use practicable and cost effective means to obtain rapid, standardized, and blinded follow-up in large numbers of patients. Postal administration of questionnaires could meet these needs, but the validity of follow-up by post would be compromised if the response frequency were low. Postal follow-up alone may not be acceptable in randomised controlled trials, as analysis and interpretation of results based on data from only

80% of the study population - the maximal level of response observed in the current study - could be misleading. One solution might be to use other modes of administration (e.g. interviewer or telephone) to obtain data from patients who did not respond to a postal questionnaire. However, certain parts of these questionnaires (e..g. the EuroQol visaul analogue scale) may be more difficult to administer by phone. The validity of these latter alternatives requires further study.

5.4.5 Conclusions

In conclusion, I found higher response frequency and superior data quality in stroke survivors followed up by the EuroQol questionnaire compared with the SF-36. I felt that the response with the EuroQol was better and more complete because the questionnaire was simpler and shorter. The principle that simple questionnaires with high response frequencies may be preferable is relevant to researchers who use health related quality of life instruments (and researchers who develop new instruments). Whether these issues are relevant when the instrument is administered by alternative means, e.g. face-to-face interview, remains to be established.

5.5 Summary of Chapter Five

1. Non-randomised comparisons of response frequencies of different health related quality of life measures are unreliable because they may be confounded by differences in patient populations and effects from the co-administration of other questionnaires.

2. I studied response frequency to the postal administration of the EuroQol and SF-36 after stroke using a direct randomised parallel group design.

3. I found a significantly higher response frequency and significantly less missing data in patients allocated to follow up with the EuroQol. The use of the EuroQol rather than the SF-36 could therefore reduce bias and improve the generalisability of the study.

4. A significantly higher proportion of patients with poor outcome responded to the EuroQol questionnaire than the SF-36. Thus, the EuroQol may have more power to detect differences in health related quality of life between groups of stroke patients with marked impairment than a more detailed and complex measure with a lower frequency of response.

5. Patients with large cortical strokes were least likely to respond; only about one fifth of this group completed the forms themselves. Patients with less extensive strokes, not causing deficits of cognitive function (lacunar and posterior circulation strokes) were most likely to complete and return the questionnaires without help. Thus non-randomised comparisons of different instruments might be confounded by differences of case mix in subgroups of the study population.

TABLE 5.1: Characteristics at the time of entry to the International Stroke Trial in the 2,253 patients randomised between the EuroQol and SF-36

	Questionnaire allocated			
	SF-36		EuroQol	
	n=1128		n=1125	
	n	(%)	n	(%)
Male sex	600	(53)	598	(53)
Age				
< 50	45	(4)	56	(5)
50-60	129	(11)	119	(11)
60-70	259	(23)	274	(24)
70-80	421	(37)	416	(37)
>80	274	(24)	260	(23)
Stroke syndrome				
TACS	244	(22)	247	(22)
PACS	474	(42)	476	(42)
LACS	292	(26)	285	(25)
POCS	118	(11)	117	(10)

TACS = total anterior circulation stroke syndrome

PACS = partial anterior circulation stroke syndrome

LACS = lacunar stroke syndrome

POCS = posterior circulation stroke syndrome

TABLE 5.2: Comparison of response frequency and completeness of data for the EuroQol and SF 36

Measure of performance	Questionnaire allocated SF-36 n=1128 n (%)	EuroQol n=1125 n (%)	Absolute difference	Odds ratio of response (95% CI)	p
Response					
Response to first mailing	679 (60%)	747 (66%)	6%	1.31 (1.10 to 1.56) [a]	0.002
Response after two mailings	849 (75%)	905 (80%)	5%	1.35 (1.10 to 1.66) [a]	0.003
Complete Data					
No missing data	616 (55%)[c]	747 (66%)	11%	1.64 (1.38 to 1.95) [b]	<0.0001

[a] Odds of response, comparing Euroqol with the SF-36 (odds > 1 indicate EuroQol better).

[b] Odds of response with no missing data , comparing Euroqol with the SF-36 (odds > 1 indicate EuroQol better).

[c] Questionnaires with no missing data (after interpolation of missing values where possible) for the SF-36.

[d] Questionnaires with no missing data (for the EuroQol any missing data resulted in a missing domain).

TABLE 5.3: Frequency of missing data in each domain of the SF-36

SF-36 n=849 Questionnaires with ≥ 1 missing domain*		
Domains	n	(%)
Physical functioning	75	(9)
Bodily pain	26	(3)
Physical role limitations	137	(16)
General Health	99	(12)
Vitality	54	(6)
Social functioning	16	(2)
Emotional role limitations	162	(19)
Mental health	52	(6)

* Domains with insufficient data to calculate score (after interpolation of missing values where possible) for the SF-36.

Table 5.4: Frequency of missing data in each domain of the EuroQol questionnaire

	EuroQol n=905 Questionnaires with ≥ 1 missing domain*	
Domain	n	(%)
Mobility	35	(4)
Self care	30	(3)
Usual activities	25	(3)
Pain	34	(4)
Anxiety / depression	49	(5)
Overall numeric estimate of HRQoL	75	(8)

*For the EuroQol any missing data resulted in a missing domain.

HRQoL = health related quality of life

TABLE 5.5: Comparison of functional outcome reported by respondents

	Follow up instrument allocation				Absolute difference	Odds ratio (95 % CI)	p
	SF 36 n=1,128		EuroQol n=1,125				
	n	(%)	n	(%)			
Overall number returned	849	(75%)	905	(80%)			
Number of respondents reporting dependency in ADL	564	(50%)	657	(58%)	8%	1.40 (1.18 - 1.66)	0.00006
Number of questionnaires completed without help by respondents dependent in ADL	188	(17%)	243	(22%)	5%	1.38 (1.11 - 1.71)	0.003

TABLE 5.6: Influence of stroke syndrome at baseline on whether the patient or the carer completed the questionnaire

	Completed by patient		Completed by carer		Total	
	n	(%)	n	(%)	n	(%)
TACS (n=491)	97	(20)	251	(51)	348	(71)
PACS (n=950)	382	(40)	348	(37)	730	(77)
LACS (n=577)	281	(49)	186	(32)	467	(81)
POCS (n=235)	113	(48)	80	(34)	193	(82)
All Syndromes n=2253	873	(39)‡	865♦	(38)	1738	(77)*

* chi squared for heterogeneity amongst stroke syndromes 17.41, DF=3 p=0.0006

‡ chi squared for heterogeneity amongst patient completed data 107.9, DF=3 p < 0.0001

♦ 16 forms returned with missing data regarding who completed the form

TACS, PACS, LACS, POCS abbreviations as in Table 1

6 Is the EuroQol reliable after stroke?

6.1 Introduction

Reliability is the extent to which a measure is free from random error in the population of interest (Scientific Advisory Committee, 1995; Guyatt *et al.* 1993; Testa & Simonson, 1996), and refers to its internal consistency as well as its reproducibility, (Chapter One). A measure's reproducibility is the degree to which it yields consistent scores over time among respondents who are assumed not to have changed (test-retest reproducibility) or the extent to which different observers may administer it to a particular patient and achieve similar results (inter-observer reproducibility). Measures with poor reliability will be less efficient at distinguishing patients with different health states because true differences in score may be obscured by random error.

The Nottingham Health Profile is the only health related quality of life which has had its reproducibility specifically examined after stroke (Gompertz *et al.* 1993). Gompertz and colleagues found it had inadequate reproducibility for monitoring individual patients after stroke, but adequate reliability for assessing groups. The EuroQol and SF-36 also appear to have inadequate reliability for individual-patient applications (Brazier *et al.* 1992; van Agt *et al.* 1994; McHorney & Tarlov, 1995; Brooks & with the EuroQol group, 1996; Andresen *et al.* 1996). However, their reliability in stroke patients has not been assessed (see Chapter Two). I therefore aimed to assess the test-retest reliability of both instruments in a group of patients after stroke.

6.2 Methods

6.2.1 Patients and allocation to the EuroQol or SF-36

In the previous chapter, I examined response rates to postal versions of the EuroQol and SF-36; I randomly allocated patients to receive either the EuroQol or the SF-36. I have already described the methods used to identify patients and the format of the instruments in Section 5.2. I then randomly sampled one third of the patients who had responded within approximately three weeks to the first questionnaire, for repeat testing with the same health related quality of life instrument (test-retest reliability). I posted the second questionnaire booklet containing the appropriate instrument to all eligible patients with a personalised letter and a reply-paid envelope. The letter explained the purpose of the repeat questionnaire and asked the subjects to respond if possible without the help of another person, and if not, to give the questionnaire to a close relative or care-giver who was willing to respond on the patient's behalf. I sent a reminder letter and further identical questionnaire to any patient who had not responded within 14 days. I made no further attempts to contact non-respondents thereafter as this was unlikely to yield significant further improvements in response. I marked individual questionnaire booklets with labels that included details of the patient's name, address, trial identifying number and questionnaire allocation. I generated all letters and labels directly from the randomisation code using a computerised mail-merge programme.

6.2.2 Statistical analysis

Reliability is a generic term used to indicate both the internal consistency of a scale and its reproducibility (Deyo *et al.* 1991). I assessed the internal consistency, the extent to which items within a dimension are correlated with each other, among the items comprising each of the domains of the SF-36 using Cronbach's alpha

coefficient (SPSS for Windows, Release 6.1). I calculated alpha coefficients for each of the eight SF-36 domains using responses to the initial questionnaires. Accepted minimal standards for alpha coefficients are 0.7 for group comparisons and coefficients greater than 0.9 for comparisons between individual patients or the same patient over time (Scientific Advisory Committee, 1995).

I examined test-retest reliability in several ways. The primary method of analysis was by calculating agreement statistics. I only performed these analyses for patients who had complete data on test and retest for any particular domain. For the categorical domains of the EuroQol, I used an unweighted kappa statistic to calculate agreement beyond that which might be expected by chance (Morton & Dobson, 1989). I used the intraclass correlation coefficient (ICC) to examine agreement for the continuous data generated by the eight domains of the SF-36 and the visual analogue scale (and utilities) of the EuroQol (Morton & Dobson, 1989). For these data, I also calculated the arithmetic mean and standard deviation of the differences between the test and retest administration to provide additional information on the reproducibility of these assessments.

To aid the clinical interpretation of the findings, I also aimed to determine the frequency of "potentially important differences" between test and retest for both instruments. For the five categorical domains of the EuroQol, I considered that any change in score was potentially important, since each of the three levels are all explicitly defined. Since "important clinical change" is harder to define for the SF-36, I examined the frequency of differences of varying size.

6.3 Results

Of the 4,016 patients randomised by the United Kingdom centres in the International Stroke Trial between 2 March 93 and 31 May 1995, 2,253 (56%) patients were known to be alive and at a known address at the start of the present study. Of these, 1,125 were randomised to receive a EuroQol questionnaire and 1,128 to a SF-36 questionnaire, (Figure 6.1). Patients received the intial questionnaires after a mean period of 64 weeks (SD 30) from their stroke. The response frequency was significantly greater in patients allocated to the EuroQol (80% allocated EuroQol versus 75% SF-36 responded after one reminder, p=0.003). Of these respondents, 271 were selected at random to receive a further EuroQol questionnaire and 253 were randomised to follow-up with an additional SF-36 (fewer patients received a repeat SF-36 because a smaller proportion responded to the intial questionnaire). Both groups had had similar characteristics at the time they entered the International Stroke Trial, (Table 6.1).

Two hundred and thirty four (86%) of the patients allocated to a second EuroQol responded and, of these, 122 (52%) completed it without help. Of the 111 repeat EuroQol questionnaires completed with the help of another person (data on who completed it was missing for one patient), 94 were completed with the help of the same individual that helped with completion of the first. Of the 122 patients who managed to complete the EuroQol without help, 54 were independent in activities of daily living. Only seven of the 112 patients (6%) who required help with the questionnaire were independent in activities of daily living.

A similar proportion (83%) of patients allocated to a second SF-36 responded and, of these, 106 (51%) patients completed it without help (58 of these 106 patients were independent in activities of daily living). Of the 101 remaining forms completed with

the help of another person (data on who completed it was missing for two patients), 79 were completed with the help of the same individual that helped with completion of the first. Of these 101 patients, 16 (16%) were independent in activities of daily living. The mean period between completion of the initial questionnaire and posting of the repeat questionnaire was 21 (SD 7) days for the SF-36 and 21 (SD 9) days for the EuroQol. There were no significant differences in time from the stroke which lead to entry into the IST for patients who did or did not require help with form completion for either instrument.

Reproducibility ranged from moderate to good for the five descriptive domains of the EuroQol (kappa statistics range from 0.63 to 0.80), (Table 6.2). For these domains, reproducibility was best for mobility (kappa = 0.80) and worst for psychological functioning (kappa = 0.63). There were no significant differences in score between test and retest for any of these five domains (Table 6.2). Test-retest reliability was consistently better for questionnaires completed by the patients than for those completed with the help of proxies. Although the overall assessments of HRQoL with the EuroQol and the EuroQol utilities had excellent reproducibility (judged by the intraclass correlation coefficients), the standard deviation of the differences between test and retest for these domains were significant (Table 6.2).

Table 6.4 shows the internal consistency for the SF-36. Cronbach alpha reliability coefficients were 0.8 or greater for all the domains, suggesting very good or excellent internal consistency. I have reported test-retest reliability separately for the forms completed by the patients, for the forms completed on behalf of patients by a proxy and for all forms combined (Table 6.5). The mean of the difference between test and retest ranged from -1.8 to 3.1 for the different domains. The standard deviation of the differences between test and retest were substantial for all domains, but particularly for the two role functioning domains. The intraclass correlation coefficients were generally acceptable or good for all the domains, except mental health (intraclass correlation coefficient = 0.28). For all eight domains, reproducibility was better when the patient assessed HRQoL than when a proxy did.

Tables 6.3 and 6.6 report the frequency of potentially important disagreements between test and retest for both instruments. For the self care, activities, pain and psychological functioning domains of the EuroQol, more than 15% of respondents report a "potentially important difference" in health between test and retest (Table 6.3). A substantial proportion of patients reported differences of 20 points, or greater, between test and retest for all the domains of the SF-36 (Table 6.6).

6.4 Discussion

I found that the test-retest reliability of the EuroQol and SF-36 was generally moderate when assessed after stroke. For both instruments, I observed the worst reproducibility in the domains which examined psychological functioning. Mental health measured with the SF-36 had particularly poor reliability (intraclass correlation coefficient=0.28). This may be explained by the subjective nature of this domain, or

there may be a mathematical explanation. The intraclass correlation coefficient compares the variance between the patients with the total variance (Deyo *et al.* 1991); since all the patients had relatively similar outcomes for this domain the intraclass correlation coefficient would therefore be expected to be small in this particular sample.

6.4.1 Interpretation of agreement statistics

The difficulty with grading the results of agreement statistics has been raised in Chapter Four. For example, there is no consensus about the "meaning of kappa=0.5" (Brennan & Silman, 1992; McDowell and Newell, 1996). Furthermore, authors are inconsistent in reporting the clinical significance of any given kappa value or intraclass correlation coefficient (Deyo *et al.* 1991). I therefore examined the mean and standard deviation of the differences and the frequency of potentially important disagreement for both the SF-36 and EuroQol to try to clarify the practical implications of our findings. I did not find substantial mean differences between test and retest for any of the domains of the SF-36, or for the assessment of overall HRQoL using the EuroQol "thermometer". However, although the agreement statistics suggested that the reliability was generally moderate or good, I found the standard deviations of the differences were large for most domains (approximately ± 20) and even larger for the physical and emotional role functioning domains. This degree of variability means that neither instrument would be suitable for serial studies within the same stroke patient or for making serial comparisons between individual patients after stroke. "Potentially important disagreement" was also frequent. However, interpretation of the frequency of disagreements is complicated by the different number of levels for each of the domains of the SF-36. My findings do indicate that either instrument would function adequately to compare groups of patients, such as in a parallel group randomised controlled trial.

6.4.2 Explanations for less than perfect reproducibility

There were a number of potential sources for poor test-retest reproducibility in the current study. These included the nature of the domain under study, whether the patients completed the questionnaires themselves, change in the patients' health state between test and retest and measurement error.

6.4.2.1 Stroke severity

I consistently observed better reliability when the patients completed the questionnaire by themselves than when a proxy completed it on the patient's behalf. However, these instruments were designed to assess a patient's uniquely personal view of their own health state and so, by definition, were not designed for use by a proxy (Testa & Simonson, 1996).

I might have underestimated the reproducibility of the assessments in patients who required help to complete the questionnaires, because about a fifth of patients sought help from a different person for the first and second questionnaires. It might also be that HRQoL is less stable for more severely affected patients who are unable to complete the questionnaires themselves (usually because of physical and cognitive deficits after the stroke) (Kwa *et al.* 1996). In this situation, rating of the patient's health status by individuals other than the patient (e.g. a family member, friend or carer) may be the only means of assessing the patient's HRQoL. Although these proxy assessments were not as reproducible, and may not be as valid (Segall & Schall, 1994; Chapter Four), as those performed by the patients themselves, they appeared to be at least reasonably reliable in the current study.

6.4.2.2 Change in patient's health between test and retest

Poor test-retest reproducibility may be due in part to change in the patients' health state between the initial test and the subsequent retest. For both instruments, I observed worst reproducibility in the domains which assessed psychological functioning. This is not surprising, as this is arguably the most subjective domain and therefore likely to be subject to the greatest day-to-day variation.

I assessed reproducibility over an interval of several weeks, when the patients' neurological status was likely to be stable. I considered this period to be long enough to minimise memory effects, but short enough that a real change in the patients' health was unlikely. Some investigators suggest that patients who report a change in health state during the study period should be excluded from comparisons of test-retest reliability to identify the "noise" associated with the instrument (Scientific Advisory Committee, 1995; Ruta *et al.* 1994). I did not do this, as this does not give an indication of the true "noise" in the population of interest and this is the variability above which a measure must be responsive to detect change in a treatment group.

6.4.2.3 Content or wording of instrument

Measurement error associated with the instrument can result from either a lack of intelligibility or ambiguity in its wording. It may also occur if patients find the content lacks relevance to their situation. Elderly people may not regard some of the questions of the SF-36 about work or vigorous activities (domains of physical and emotional role functioning) as relevant to them (Hayes *et al.* 1995). In my study (in which the mean age of the patients was 70 years), these domains had relatively poor test-retest reliability.

6.4.3 SF-36 in this study compared with others

The internal consistency of the SF-36 has been consistently estimated as very good or excellent (McDowell and Newell, 1996). The estimates for the internal consistency of the SF-36 by postal questionnaire in this study, are therefore consistent with these estimates, as well as those reported by Anderson and coworkers in their study of the validity of the SF-36 when administered by interview after stroke (Anderson *et al.* 1996). The estimates from this study for the reproducibility of the SF-36 in stroke patients are also consistent with those obtained in other patient groups which suggest it is only adequate for group applications (McHorney & Tarlov, 1995; Weinberger *et al.* 1996; Ruta *et al.* 1994; Andresen *et al.* 1996). Weinberger and colleagues reported that the mode of administration (face-to-face interview, self-completed questionnaire or telephone interview) did not appear to affect the reproducibility of the SF-36 (Weinberger *et al.* 1996). The SF-36 is, therefore, likely to be reasonably reliable in stroke patients regardless of the mode of questionnaire administration.

6.4.4 Comparison of the reliability of the EuroQol and SF-36

I was only able to compare the test-retest reliability of the EuroQol and SF-36 indirectly in a qualitative manner as no one statistical technique could be used to assess the agreement for both categorical and continuous data. Within this limit, both instruments appeared to have similar reliability. I could have reclassified the outcome data with the SF-36 into several new categories to make a direct comparison with the EuroQol possible. However, this kind of arbitrary approach would be hard to validate. I therefore reported the frequency of what I considered might be "potentially important differences" for both instruments. This would at least allow a broad qualitative comparison of the reliability of the two instruments. There is no consensus over what a clinically meaningful change for either instrument might be, so even this approach has limited value. As the mobility, self care, social functioning,

pain and psychological domains of the EuroQol have just three distinct levels (for example mobility: I have no problems in walking about (1), I have some problems in walking about (2), I am confined to bed (3)) we considered any change for these domains as potentially important. The definition of a potentially important change with the SF-36 is more controversial. Some investigators consider differences of five points in any of its domains as potentially important (Ruta *et al.* 1994). However, this difference is not directly comparable with a change of one level for the EuroQol. I therefore reported the frequency of disagreement for four empirically chosen differences in score (5,10,20 and 40 points). These results need to be interpreted in the context of the number of items contributing to the scoring of each domain (e.g. there are only four levels for physical role functioning, see Table 7.2). Overall, they support the conclusion that unless investigators are seeking to identify very large differences (e.g. >40 points with the SF-36), neither instrument is likely to be effective at reliably identifying change over time in HRQoL within an individual patient after a stroke.

I was only able to compare the reliability of the EuroQol and SF-36 indirectly. The groups who received the initial EuroQol and SF-36 were similar, but there were inevitably some differences between the groups who were sent repeat questionnaires, since some selection bias had taken place at this stage. An alternative approach would have been to give all patients both instruments twice (test-retest). I felt, however, that this would place an unacceptable burden on patients and so might have adversely affected the response rates. The comparison may also have been biased because the EuroQol asks patients to report their health state on that particular day, whereas the SF-36 asks patients about their health over the previous four weeks. I was therefore surprised that the qualitative estimates of reliability of the SF-36 and EuroQol were so similar. This suggests that either day-to-day fluctuation

in a patient's health state was small or that the patients did not pay much attention to the exact wording of the questionnaires.

In summary, both the EuroQol and SF-36 have acceptable, and qualitatively similar, test-retest reliability when administered after stroke and completed by patients or their proxies. Either instrument might function effectively as a discriminatory measure for assessing HRQoL outcomes in groups of patients, as in a large parallel group randomised controlled trial or an audit study. Sample size calculations for observational studies and randomised trials must take the reliability of both instruments into account. Doing so will generally increase the sample size, but should reduce the risk of a "false negative" or type II statistical error. These data do not support the use of either instrument for serial assessments in individual patients, unless very large differences over time are expected.

6.5 Summary of Chapter Six

1. Reproducibility ranged from moderate to good for the five descriptive domains of the EuroQol (kappa statistics ranged from 0.63 to 0.80). The overall assessments of health related quality of life and the EuroQol utilities had excellent reproducibility.

2. All the domains of the SF-36 had very good or excellent internal consistency.

3. Test-retest reliability for the SF-36, assessed by intraclass correlation coefficients, was moderate or good for all the domains except mental health (ICC=0.28). The mean of the absolute difference between test and retest was less than 4 for all domains. However, the standard deviations of the differences were large for all domains.

4. For both instruments, the frequency of potentially important differences between test and retest was substantial. Neither instrument is therefore likely to be reliable at identifying change within an individual patient after stroke. However, either instrument might function effectively as a discriminatory measure for assessing health related quality of life outcomes in groups of patients after stroke.

5. Both the EuroQol and SF-36 have qualitatively similar reproducibility after stroke.

Table 6.1: Characteristics at time of randomisation in the International Stroke Trial in the 524 patients participating in the reliability studies.

	Questionnaire allocation	
	SF-36 (n=253)[a]	EuroQol (n=271)[b]
	n (%)	n (%)
Male	140 (55)	147 (54)
Age		
<50	12 (5)	19 (7)
50 to 60	32 (12)	26 (10)
60 to 70	63 (25)	67 (25)
70 to 80	99 (39)	95 (35)
>80	47 (19)	64 (23)
Stroke syndrome		
TACS	52 (21)	51 (19)
PACS	102 (40)	117 (43)
LACS	69 (27)	74 (27)
POCS	30 (12)	29 (11)

[a] Randomly sampled from 849 patients who responded to an initial questionnaire

[b] Randomly sampled from 905 patients who responded to an initial questionnaire

No statistically significant differences between the groups were observed

Table 6.2: The test-retest reliability of the EuroQol after stroke (n=234)

Functional domains	Median difference between test and retest* (interquartile range)	Kappa statistic (95% confidence interval; n) Completed by patient	Completed by proxy	All
Mobility	0 (0 to 0)	0.85 (0.72 to 0.94; 116)	0.31 (0.00 to 0.66; 103)	0.80 (0.71 to 0.90; 219)
Self care	0 (0 to 0)	0.74 (0.62 to 0.86; 112)	0.63 (0.50 to 0.77; 108)	0.73 (0.65 to 0.81; 220)
Usual activities	0 (0 to 0)	0.66 (0.54 to 0.78; 116)	0.61 (0.47 to 0.76; 105)	0.68 (0.59 to 0.76; 221)
Pain	0 (0 to 0)	0.71 (0.59 to 0.83; 115)	0.61 (0.45 to 0.78; 106)	0.68 (0.58 to 0.77; 221)
Psychological	0 (0 to 0)	0.73 (0.61 to 0.86; 110)	0.49 (0.32 to 0.65; 103)	0.63 (0.53 to 0.73; 213)
	Mean difference between test and retest* (sd of difference)	Intraclass correlation coefficient (n) Completed by patient	Completed by proxy	All
Overall HRQoL	0.2 (11.8)	0.86 (108)	0.74 (95)	0.86 (203)
Utility	-0.003 (0.19)	0.83 (97)	0.81 (96)	0.86 (193)

* All patients combined
n = number of patients with complete data for test and retest in that particular domain

Table 6.3: Test-retest reliability of the EuroQol: the frequency of "potentially important differences" between test and retest *

Domain	% with potentially important differences (95% CI)
Mobility	6 (4 to 11)
Self care	17 (12 to 22)
Activities	20 (15 to 26)
Pain	16 (11 to 21)
Mood	20 (14 to 25)

* defined as score for that domain changing by at least one level between test and retest

Table 6.4: Internal consistency of the SF-36 among the 849 initial respondents

Functional domains	Internal consistency
	Cronbach α
Physical functioning	0.95
Physical role functioning	0.94
Social functioning	0.80
Bodily pain	0.87
Mental health	0.86
Emotional role functioning	0.96
General health	0.83
Vitality	0.81

Table 6.5: The test-retest reliability of the SF-36 among 209 respondents to a repeat questionnaire

Functional domains	Test-retest reliability			
	mean difference between test and retest† (sd of difference)	intraclass correlation coefficient completed by patient (n)	completed by proxy (n)	both (n)
Physical functioning	3.1 (22)	0.80 (90)	0.59 (81)	0.74 (172)
Physical role functioning	-1.8 (29)	0.77 (89)	0.45 (61)	0.67 (151)
Social functioning	0.9 (21)	0.79 (105)	0.76 (94)	0.80 (200)
Bodily pain	1.2 (21)	0.81 (104)	0.65 (92)	0.75 (197)
Mental health	1.4 (15)	0.30 (94)	0.24 (80)	0.28 (175)
Emotional role functioning	1.2 (40)	0.60 (85)	0.50 (56)	0.57 (142)
General health	1.3 (15)	0.81 (86)	0.71 (78)	0.79 (165)
Vitality	0.2 (18)	0.77 (94)	0.55 (81)	0.70 (176)

† negative mean difference indicates that mean score on retest was greater than on initial testing

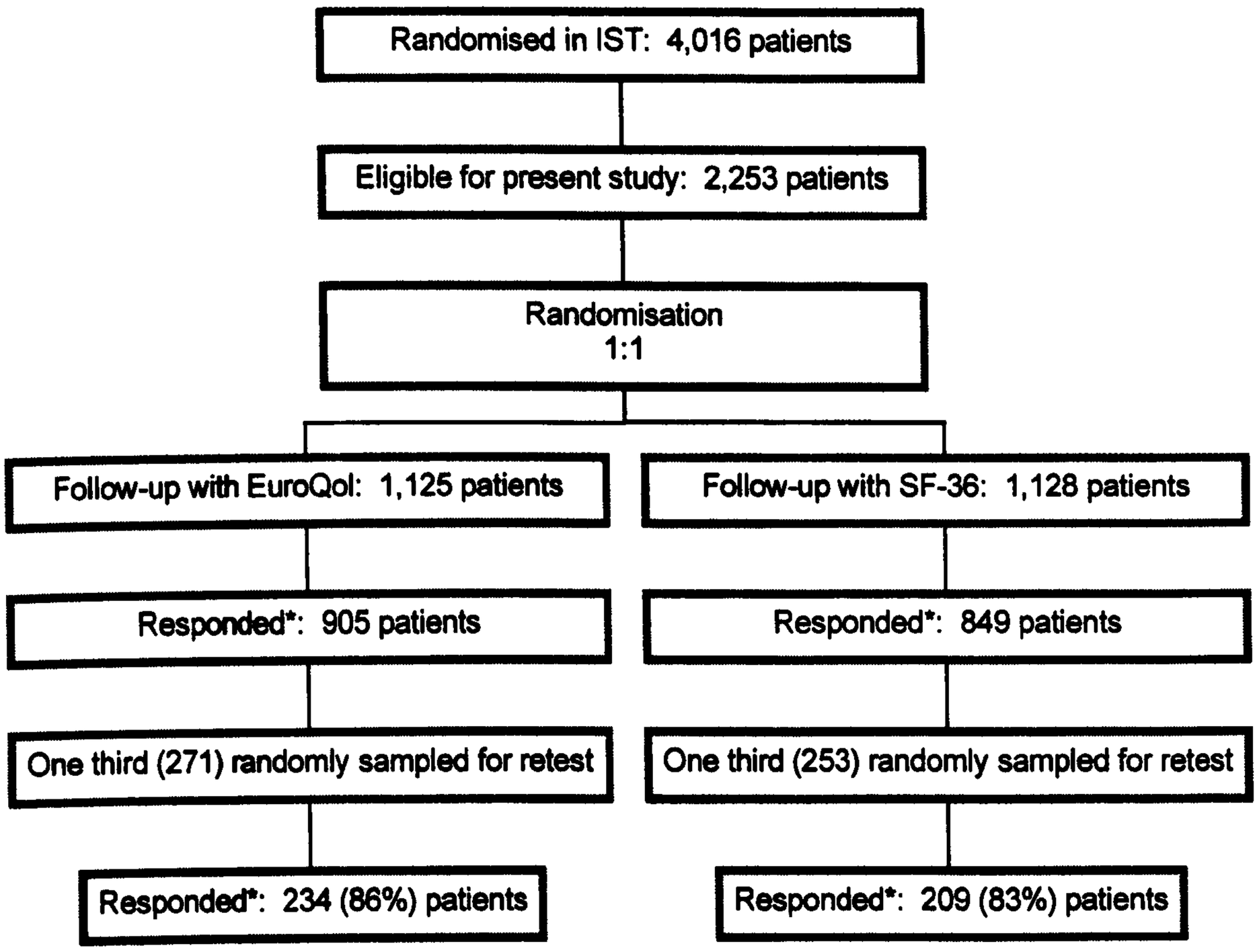
sd = standard deviation

n= number of patients with complete data for both test and retest in each domain. Note that for one patient it was not clear if the form was completed by the patient or with the help of a proxy.

Table 6.6: Test-retest reliability of the SF-36 in 209 patients: the frequency of changes of different sizes in HRQoL between two measurements several weeks apart

Domain	% with > 5 point difference (95% CI)	% with > 10 point difference (95% CI)	% with > 20 point difference (95% CI)	% with > 40 point difference (95% CI)
Physical functioning	39 (32 - 46)	25 (19 - 32)	12 (7 - 17)	5 (2 - 9)
Physical role functioning	29 (21 - 36)	29 (21 - 36)	29 (21 - 36)	13 (7 - 18)
Social functioning	65 (58 - 71)	65 (58 - 71)	29 (23 - 35)	5 (2 - 8)
Bodily pain	62 (56 - 69)	44 (37 - 51)	24 (18 - 30)	8 (5 - 13)
Mental Health	60 (53 - 67)	37 (29 - 44)	13 (8 - 18)	2 (1 - 6)
Emotional role functioning	32 (24 - 39)	32 (24 - 39)	32 (24 - 39)	21 (14 - 28)
General Health	56 (49 - 64)	35 (28 - 42)	11 (6 - 16)	2 (1 - 6)
Vitality	57 (50 - 65)	42 (35 - 49)	19 (14 - 25)	2 (1 - 6)

Figure 6.1: Flow of patients through study



* to two mailings

7 How do scores on the EuroQol relate to scores on the SF-36 in the same patient?

7.1 Introduction

The EuroQol and SF-36 are widely used to measure several aspects of health related quality of life (Anderson *et al.* 1993; McDowell and Newell, 1996). The EuroQol assesses outcome in six broad areas (mobility, self care, activities, pain, psychological functioning, and self reported overall health related quality of life) and also provides a utility score for overall health related quality of life (The EuroQol Group, 1990). The SF-36 assesses eight domains: physical functioning, physical role functioning, social functioning, bodily pain, mental health, psychological role functioning, vitality and general health (Ware & Donald Sherbourne, 1992; Medical Outcomes Trust, 1994). At first sight, many of the domains of the SF-36 are similar to those of the EuroQol: for example, the mobility question in the EuroQol appears to relate closely to the physical functioning questions on the SF-36. However, the relationship between the domains of each of these instruments has not been well-defined (Brazier *et al.* 1993; Brazier *et al.* 1996). Moreover, the degree to which a change in health related quality of life is reflected in changes in scores of the relevant domains of the EuroQol and SF-36 is unknown.

A clearer understanding of the relationship between these instruments might help improve the interpretation of a change of score with either instrument. Furthermore, as both instruments aim to measure health related quality of life there should be a strong correlation between responses on the two instruments. A poor correlation might suggest poor validity of one or both of the measures. I therefore administered

both the EuroQol and SF-36 to a group of patients after stroke to compare their responses to these instruments.

7.2 Methods

7.2.1 Patients and allocation to the EuroQol or SF-36

In the previous two chapters, I compared the frequency of response to the postal administration of the EuroQol and SF-36, and the reproducibility of these assessments. I described the methods used to identify patients and the format of the instruments in Chapter Five. I subsequently described the method used to select a subsample of respondents to the initial questionnaires for a comparison of their reproducibility in Chapter Six. At the same time, I randomly selected one third of patients who had responded within approximately three weeks to the first EuroQol for repeat testing with the SF-36, and two thirds of the patients who had responded within approximately three weeks to the first SF-36 for repeat testing with the EuroQol - see Figure 7.1. These patients were completely separate from those included in the study of the reproducibility described in the preceding chapter. The planned (and actual) flow of patients is laid out in Figure 7.1.

I posted the second questionnaire booklet containing the appropriate instrument to all eligible patients with a personalised letter and a reply-paid envelope. The letter explained the purpose of the repeat questionnaire and asked the subjects to respond if possible without the help of another person, and if not, to give the questionnaire to a close relative or care-giver who was willing to respond on the patient's behalf. I sent a reminder letter and further identical questionnaire to any patient who had not responded within 14 days. I made no further attempts to contact non-respondents thereafter. I marked individual questionnaire booklets with labels that included details

of the patient's name, address, trial identifying number and questionnaire allocation. I generated all letters and labels directly from the randomisation code using a computerised mail-merge programme.

7.2.2 Statistical analysis

The ability of a questionnaire to discriminate between different levels of health is an important aspect of validity (Streiner and Norman, 1989; Guyatt *et al.* 1993). This is determined in part by whether a measure can define a full range of potential health states, and whether it is sensitive over this range. Patients who are at the lowest score on a measure will have no scope to show any further decline of health ("floor" effects) (Streiner and Norman, 1989; Guyatt *et al.* 1993). Similarly, if the majority of patients score near the top of the measure, it will have little scope to show improvements in health ("ceiling" effects) (Streiner and Norman, 1989; Guyatt *et al.* 1993). I therefore initially assessed the distribution of scores, and levels of missing data, for both instruments.

"Ordering" effects are a potentially important source of bias in an unbalanced cross-over study. For instance, completing the EuroQol questionnaire first might affect the patients' subsequent response to the SF-36. I therefore used a simple factorial one way analysis of variance to investigate whether ordering effects occured after the administration of either the EuroQol or SF-36. I restricted these analyses to the comparable domains of both instruments, to avoid the problems which may arise from multiple testing.

The construct validity of both instruments was assessed further by testing the relationship between the EuroQol and SF-36 domains. Thus, the relationship between comparable domains on the EuroQol and SF-36 (such as physical

functioning on the SF-36 and mobility on the EuroQol) should be higher than between less comparable domains (such as physical functioning on the SF-36 and psychological functioning on the EuroQol). In contrast, the domains which examine more general aspects of health (such as overall health related quality of life on the EuroQol) should be moderately correlated with all the other domains. I examined these relationships in two separate ways. I initially calculated patients' mean score, for each domain of the SF-36, for patients categorised according to their response to the corresponding EuroQol domain. These analyses were performed to facilitate the interpretation of patients' scores with the SF-36. I subsequently calculated correlation coefficients between the domains of the EuroQol and each of the domains of the SF-36. All analyses were performed using "Access 2.0" (Microsoft Corporation) and the statistical software package "SPSS for Windows" (Release 6.1).

7.3 Results

Of the 905 respondents to the initial EuroQol, 272 (one-third of respondents) were selected at random to receive a subsequent SF-36. A separate 505 patients (two-thirds of respondents) were selected at random from the respondents to the initial SF-36 to receive a EuroQol questionnaire (Figure 7.1). Four hundred and fifty eight (91%) of those allocated to the EuroQol questionnaire responded. A slightly lower proportion (85%) of the patients allocated to the SF-36 questionnaire responded (Figure 7.1).

I performed a simple one way factorial analysis of variance to assess the effect of the questionnaire ordering (i.e. "EuroQol then SF-36" or "SF-36 then EuroQol") on the relationship between EuroQol and SF-36 scores for comparable domains. The "ordering term" was not a significant determinant of the relationship between the

EuroQol and SF-36 scores in the eight analyses performed (Table 7.1). I therefore combined all the data for the remaining analyses.

The distribution of scores for the SF-36 are described in Table 7.2. The proportion of responses with missing data ranged from 2% to 16% (social functioning and psychological role functioning respectively). A substantial proportion of respondents scored the minimum possible score (zero out of a possible 100, i.e. the floor of the scale - the worst possible outcome) for the domains of physical role functioning and emotional role functioning. About one-quarter of patients scored the maximum score for the bodily pain and psychological functioning domains of the SF-36.

The distribution of patients' responses to the categorical domains of the EuroQol are described in Table 7.3. The proportion of missing data (approximately 3%) was very similar for each of these five domains (Table 7.3). Although each of the domains had only three potential levels of response, the data were not particularly skewed (Table 7.3). Examination of the distribution of overall estimates of health related quality of life with the EuroQol visual analogue scale or the EuroQol utility scores does not suggest problems with ceiling or floor effects (Figures 7.2 and 7.3).

The relationships between patients' responses to the EuroQol and SF-36 questionnaires are presented in Tables 7.4 and 7.5. Table 7.4 presents the mean scores for the relevant SF-36 domains for patients categorised according to their response to the comparable EuroQol domain. For almost all of the domains, the mean scores were ordered appropriately and were significantly different between the groups. Indeed, physical functioning, social functioning and pain measured with the SF-36 were particularly closely related to the corresponding domains on the EuroQol. However, there was no difference in the mean scores for the physical role functioning domain between patients reporting "some" or "severe" problems with the EuroQol. Futhermore, there was only a weak relationship between the mental health domain (SF-36) and psychological functioning domain (EuroQol) (Table 7.4)

Table 7.5 reports the correlation between each of the domains of the EuroQol and those of the SF-36. The physical functioning domain on the SF-36 correlated most closely with the mobility, self care and activities domain of the EuroQol; it correlated less closely with the pain and psychological domains of the EuroQol. Social functioning on the SF-36 was moderately correlated with all the domains of the EuroQol. Bodily pain was most closely correlated with the pain domain of the EuroQol. In contrast, mental health correlated only poorly with psychological functioning measured with the EuroQol. The vitality and general health domains of the SF-36 correlated particularly strongly with the overall health related quality of life domain of the EuroQol, but also moderately with the other domains of the EuroQol.

7.4 Discussion

7.4.1 Relationship between the EuroQol and SF-36

I observed a close relationship between the domains which assessed physical functioning, social functioning, bodily pain and overall health related quality of life. Therefore, although the EuroQol and SF-36 questionnaires differ substantially in their background, structure, content, length, and also the time period to which they refer, these results suggest that they are generally sampling similar areas of health. This finding supports the notion that there are several key dimensions which comprise health related quality of life, as well as providing further support for the construct validity of the assessments of these domains with either instrument.

The correlation between patients' responses to the mental health domain of the SF-36 and the psychological functioning domain of the EuroQol was not very impressive. There are several possible explanations for this. Firstly, it is possible that these domains, although superficially similar, are measuring different constructs. This is supported by the fact that the EuroQol item focuses on anxiety and depression, whereas the SF-36 mental health scale includes positive emotions as well (e.g. feeling calm and peaceful). The psychological role functioning domain of the SF-36, which emphasises anxiety and depression, correlated much better with the EuroQol psychological functioning domain than did the mental health domain of the SF-36 (Spearman rank correlation coefficient 0.43 versus 0.21). However, an alternative explanation is that one, or both, of these domains has poor measurement properties in patients with stroke. There are several pointers to this. Firstly, the assessments of mental health with the SF-36 were clustered around the middle of the scale (mean score 61, standard deviation 12) and so did not appear to take full advantage of the potential breadth of the scale. Secondly, the reproducibility of the mental health assessments with the SF-36 was also particularly poor (intraclass correlation

coefficient = 0.28) (Chapter Six). Finally, approximately half of all these questionnaires were completed with the help of proxies and the validity of these proxy assessments are particularly questionable for the domain of psychological functioning (Segal & Schall, 1994)(Chapter Four).

It has been difficult to establish the validity of the numerical assessments of overall health related quality of life with the EuroQol because this domain is difficult to define and is highly subjective (Chapter Three). However, the general health domain of the SF-36 appears to examine a similar construct (Ware, 1992). It aims to assess an individual's general health perceptions and satisfaction and, as with the EuroQol, these general health perceptions appear to provide an approach by which different components of health such as disease, functioning, symptoms and feelings can be integrated (Ware, 1992). The strong correlation (Pearson correlation coefficient = 0.66) between patients' responses to these domains supports the view that both these domains are measuring the same underlying trait. The validity of these assessments is further supported by the moderate correlation of the assessments of overall health related quality of life with the other domains of the SF-36.

7.4.2 Interpretability

Studies of interventions must show that the observed changes in patients that are due to the intervention are important and substantial enough to warrant further consideration in medical practice and policy planning (Testa & Simonson, 1996). One approach to the definition of clinical meaningfulness is the use of anchor-based interpretations (Lydick & Epstein, 1993). These definitions represent instances where the changes in quality of life measures were compared, or anchored, to other clinical changes or results. The descriptive nature of the categorical levels of the EuroQol questionnaire could be considered as potential anchors e.g. in the current study, a

change of 40 points in the physical functioning scale of the SF-36 appeared to be equivalent to the difference between "no problems" and "some problems" in the categorical mobility domain of the EuroQol. However, several factors limit the usefulness of this approach. Firstly, clinicians may be unsure about the significance of the above change in the EuroQol, i.e. what is the meaning of "some" problems. Secondly, the amount of change judged significant may differ with the population and the treatment under study. Thirdly, most scales are not linear, i.e. not an interval scale; therefore, a change of 10 units at the high end of the scale may not be the same as a similar-sized change at the low end of the scale.

The problem with defining clinical significance with health related quality of life measures reflects the newness of these measures and our lack of experience with them (Lydick & Epstein, 1993). Therefore, presenting these correlations should improve familiarity with these measures and help clinicians develop an intuitive feeling about the relevance of any change.

7.4.3 Distribution of scores

The large number of patients scoring the minimum score (worst outcome) in the physical and emotional role functioning domains of the SF-36 suggests that floor effects may be present in these domains. The observation that the mean scores for the physical role functioning domain did not distinguish between patients classified as having "moderate" or "severe" problems by the mobility or self care domains of the EuroQol confirms this suspicion. The role functioning domains may therefore not measure the consequences of more severe disabilities and this might reduce responsiveness in these domains. Other investigators have reported similar findings in studies of the SF-36 in the elderly (Brazier *et al.* 1996), and groups of patients with other diagnoses (Anderson *et al.* 1993).

7.4.4 Methodological Issues

The lower frequency of response and lower levels of data completeness in patients followed up with the SF-36 compared with the EuroQol are consistent with the result of the direct randomised comparison of their feasibility after stroke (Chapter Five). I used an interpolation procedure to reduce the proportion of missing data (missing items were substituted with the mean response to other items) (Medical Outcomes Trust, 1994), and so these results underestimate the underlying level of missing data for the SF-36. Brazier and colleagues have expressed concerns over the validity of these interpolation procedures (Brazier *et al.* 1996). They suggest that where patients omit items because they do not appear relevant to them, this may indicate that the respondent is in fact unable to perform that particular activity or function and so the average response to the other items could be misleading if interpolation is used for missing values (Brazier *et al.* 1996).

The cross-over design employed in this study seems to have been valid as there did not appear to be any significant carryover or other ordering effects. Furthermore, the study of test-retest reliability demonstrated that the patients did not change significantly in any of the domains of health related quality of life between test and retest (Chapter Six). These findings justified the combined analysis of all the data irrespective of the order of questionnaire administration.

7.4.5 Conclusions

In summary, despite fundamental differences in their background, design and format the domains of the EuroQol and SF-36 measured broadly similar aspects of health related quality of life. The weak relationship between the assessment of mental health with the SF-36 and psychological functioning with the EuroQol may reflect a difference in content or more fundamental problems with the validity or reliability of

the items in either of these domains. Unfortunately, it is difficult to resolve which of these explanations applies as no reference instruments were administered concurrently. This study has provided the first empirical qualitative evidence by which data on the SF-36 after stroke may be interpreted.

7.5 Summary of Chapter Seven

1. The EuroQol and SF-36 both aim to measure health related quality of life. However, the relationship between comparable domains of each instrument has not been defined.

2. The domains for both instruments which assesssed physical functioning, social functioning, bodily pain and overall health related quality of life correlated closely. This also provides further support for the construct validity of these domains in both instruments.

3. The mental health domain of the SF-36 correlated only poorly with the psychological functioning domain of the EuroQol. This is likely to represent either differences in content for both domains or measurement error.

4. A significant number of patients scored the minimum score in the physical and emotional role functioning domains of the SF-36. This suggested that these domains may not measure the consequences of more severe disabilities and this might reduce the responsiveness in these domains, i.e. a "floor" effect.

Table 7.1: Significance of "ordering term" in simple factorial ANOVA performed to investigate possible ordering effects

Domains of SF-36	Domains of the EuroQol Questionnaire				
	Mobility	Self care	Activities	Pain	Psychological
Physical functioning	NS	NS	-	-	-
Physical role functioning	NS	NS	-	-	-
Social functioning	-	-	NS	-	-
Bodily pain	-	-	-	NS	-
Mental health	-	-	-	-	NS
Psychological role functioning	-	-	-	-	NS
Vitality	-	-	-	-	-
General Health	-	-	-	-	-

- indicates significance of ordering term was not examined for these domains
NS indicates the ordering term was not significant

Table 7.2: Number of patients with maximum, minimum and missing scores for each domain of the SF-36 (n=688)

	Range of potential levels*	N	Mean score in domain (SD)	No. with maximum score in domain n (%)	No. with minimum score in domain n (%)	No. with missing data for domain n (%)
Physical functioning	20	635	30 (31)	15 (2)	177 (26)	53 (8)
Physical role functioning	4	593	20 (36)	83 (12)	416 (61)	95 (14)
Social functioning	9	678	49 (33)	112 (16)	83 (12)	10 (2)
Bodily pain	10	669	57 (31)	164 (24)	22 (3)	19 (3)
Mental health	22	648	61 (12)	3 (<1)	0 (0)	40 (6)
Psychological role functioning	3	581	37 (45)	178 (26)	327 (48)	107 (16)
Vitality	20	647	38 (22)	3 (<1)	35 (5)	41 (6)
General Health	19	601	46 (24)	3 (<1)	14 (2)	87 (13)

*Number of potential levels for raw scores before transformation onto zero to 100 scale

N= number of patients with no missing data for the domain

Table 7.3: Number of patients in each response category for each domain of the EuroQol

Domain of EuroQol	Range of potential levels	N	Number of patients with:			
			"No problems" n (%)	"Some problems" n (%)	"Severe problems" n (%)	"Missing" n (%)
Mobility	3	666	147 (21)	479 (70)	40 (6)	22 (3)
Self care	3	669	267 (39)	282 (41)	120 (17)	19 (3)
Activities	3	674	111 (16)	317 (46)	246 (36)	14 (2)
Pain	3	669	226 (33)	397 (58)	46 (7)	19 (3)
Psychological functioning	3	664	257 (37)	369 (54)	38 (6)	24 (4)

N= number of patients with no missing data for the domain

Table 7.4: The relationship between the EuroQol and SF-36 after stroke

SF-36 Domain	EuroQol Domain	Mean score for SF-36 domain for patients classified according to response to EuroQol domain (SD)			p
		EuroQol response:			
		"No problems"	"Some problems"	"Severe problems"	
Physical functioning	Mobility	63 (27)	23 (26)	3 (15)	<0.0001
Physical functioning	Self care	52 (28)	20 (26)	5 (18)	<0.0001
Physical role functioning	Mobility	46 (44)	13 (30)	11 (30)	<0.0001
Physical role functioning	Self care	35 (43)	10 (25)	11 (30)	<0.0001
Social functioning	Activities	83 (23)	52 (28)	29 (29)	<0.0001
Bodily pain	Pain	84 (22)	46 (25)	15 (16)	<0.0001
Mental health	Psychological functioning	64 (10)	61 (13)	53 (13)	<0.0001
Psychological role functioning	Psychological functioning	61 (46)	23 (40)	8 (23)	<0.0001
Vitality	Psychological functioning	50 (23)	33 (19)	19 (20)	<0.0001
General Health	Psychological functioning	59 (23)	39 (20)	23 (18)	<0.0001

p = significance for between the groups in one way ANOVA to compare means

Table 7.5: Correlation between EuroQol and SF-36

Domains of SF-36	Domains of the EuroQol Questionnaire						
	Mobility	Self care	Activities	Pain	Psychological	Overall HRQoL	Utility
	r*	r*	r*	r*	r*	r†	r†
Physical functioning	0.57	0.65	0.63	0.39	0.34	0.51	0.64
Physical role functioning	0.38	0.33	0.39	0.36	0.29	0.40	0.39
Social functioning	0.47	0.53	0.54	0.43	0.41	0.56	0.63
Bodily pain	0.40	0.35	0.33	0.66	0.37	0.49	0.53
Mental health	0.10†	0.10†	0.06‡	0.12†	0.21	0.08‡	0.17
Psychological role functioning	0.27	0.24	0.32	0.29	0.43	0.33	0.37
Vitality	0.40	0.43	0.45	0.39	0.41	0.60	0.54
General Health	0.39	0.39	0.43	0.44	0.44	0.66	0.54

*Spearman rank correlation coefficients (two tailed). All were negative.
†Pearson correlation coefficient (two tailed).
‡$p>0.05$; †p is between 0.002 and 0.01; all other correlations were highly significant ($p<0.0001$).
All correlations based on data from 531 patients or more.

Figure 7.1: Flow of patients through cross validation study

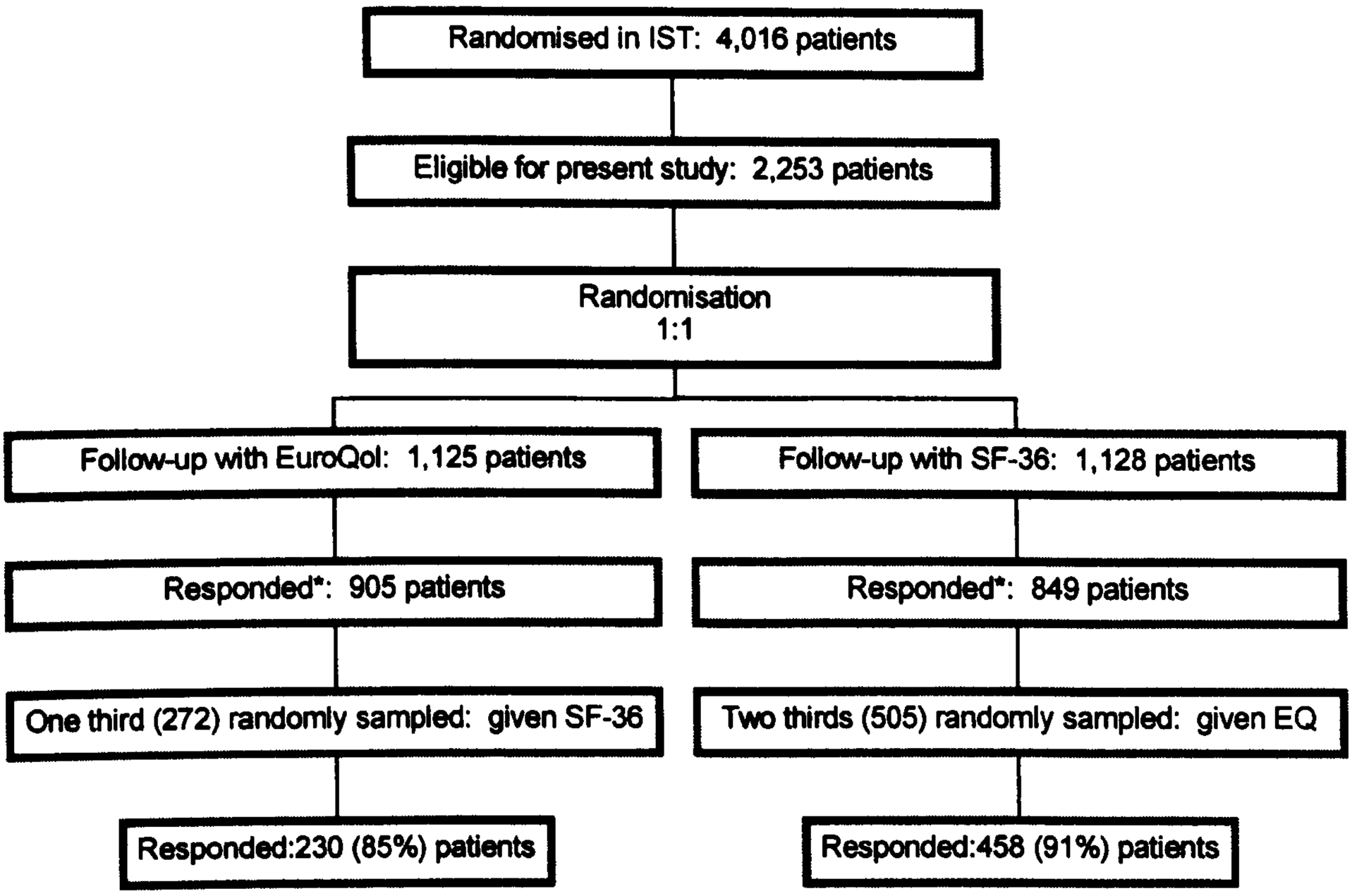


Figure 7.2: Distribution of estimates of overall health related quality of life with the EuroQol visual analogue scale (n=636)

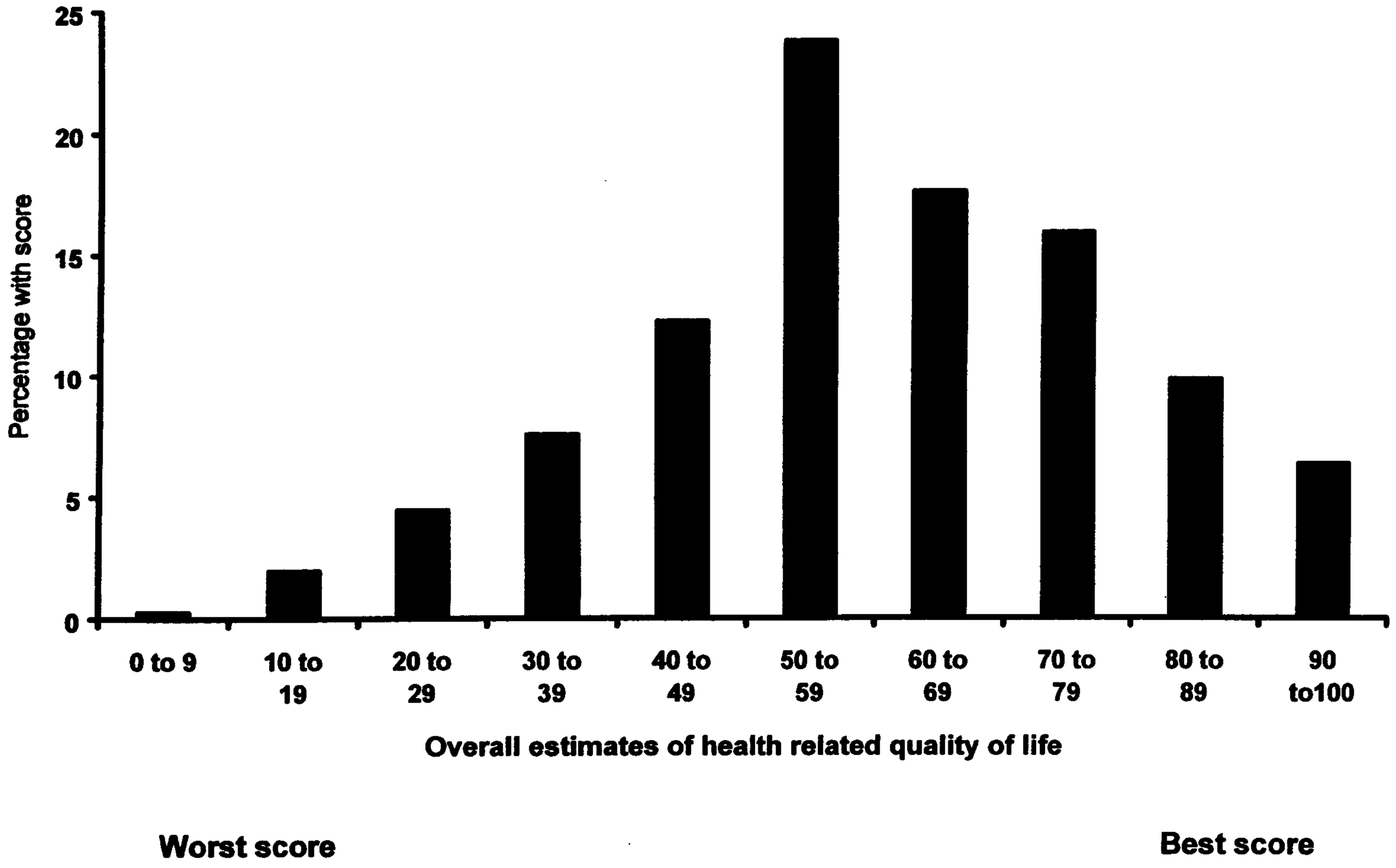
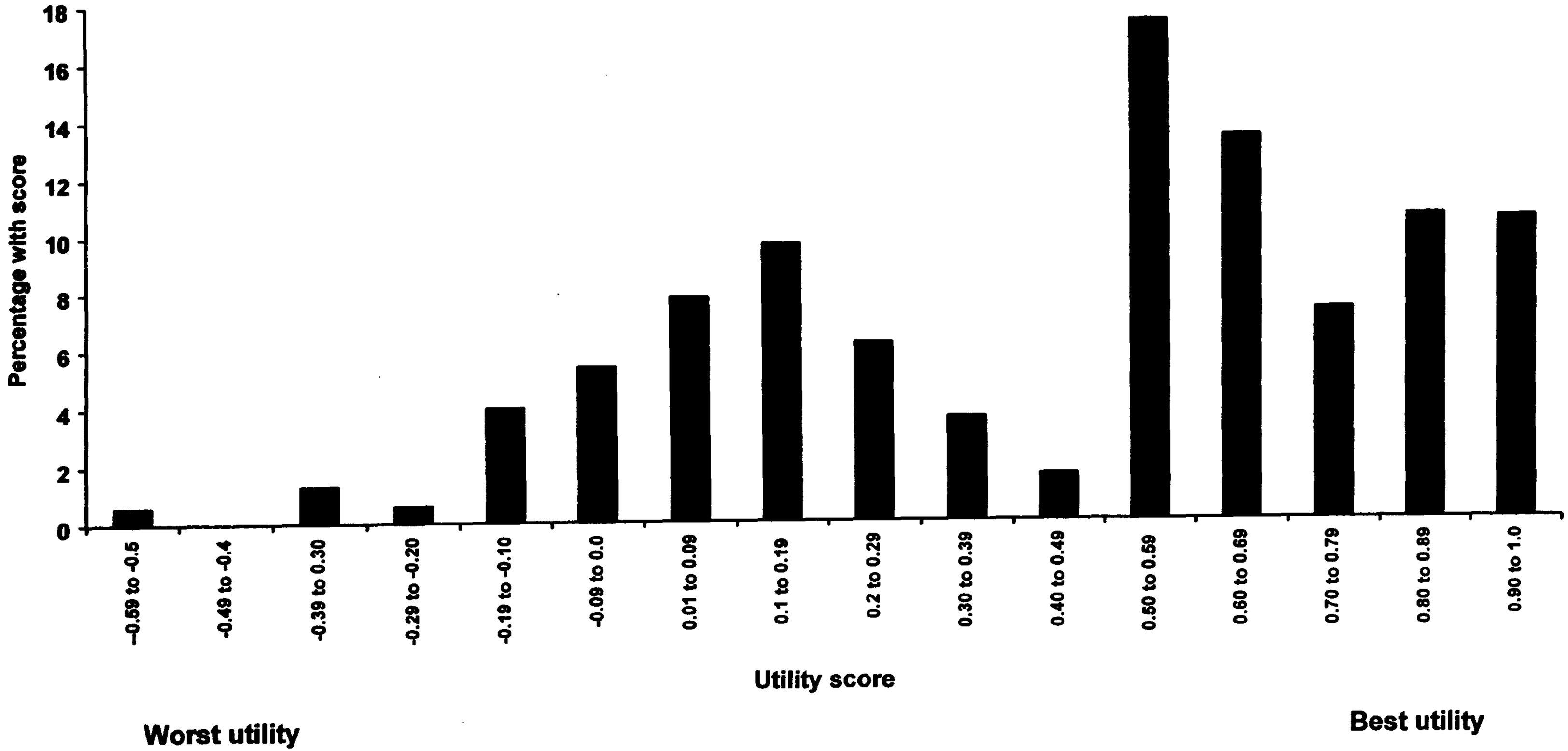


Figure 7.3: Distribution of estimates of overall utility scores with the EuroQol (n=620)



8 Are the modified "simple questions" a valid and reliable measure of health related quality of life?

8.1 Introduction

Lindley and colleagues developed two "simple questions", one to assess dependency and one to assess recovery after stroke (see Table 8.1 for original wording). Their aim was to determine whether a minimalist measurement tool could be a valid and inexpensive means to assess outcome in large randomised controlled trials and epidemiological studies (Lindley *et al.* 1994). These questions have been used in two recent randomised controlled trials in patients with stroke (Kay *et al.* 1995; International Stroke Trial Collaborative Group, 1997). The data were analysed in different ways in each of these studies: in the trial of low molecular weight heparin, the simple dependency question was used to provide an estimate of the proportion of patients disabled at follow up (Kay *et al.* 1995). In the IST, both questions were used **together** to classify patients into one of three hierarchical levels of outcome (dependent, independent but not recovered, independent and fully recovered) (International Stroke Trial Collaborative Group, 1997). Although this hierarchical classification has intuitive appeal, its relationship with other classifications of outcome (e.g. health related quality of life or the WHO Classification of Impairments, Disabilities and Handicaps) is not clear.

Lindley's initial assessment of the measurement attributes of the questions suggested they had good validity and acceptable reliability (Lindley *et al.* 1994). However, a more recent study highlighted some ambiguities in the wording (Dennis *et al.* 1997a). Some patients who answered yes to the dependency question meant they needed

help from a person, in addition to their normal helper, and not that they needed help from any person at all (Dennis *et al.* 1997a). Similarly, other patients answered yes to the recovery question when they meant they had stopped recovering, and not that they had returned to their pre-stroke state (Dennis *et al.* 1997a). These ambiguities reduced the questions' validity and reliability and so reduced their power to detect true differences in outcome.

Dennis and colleagues proposed a modification to both questions to try and improve their clarity, validity and reliability (Dennis *et al.* 1997a). They removed the clause referring to "another person" from the dependency question to give "Do you need help **from anybody** with everyday activities?" and reworded the recovery question to give "Has the stroke left you with any problems?". I examined the validity of these modified questions in the current study. I examined whether the modified dependency question is a valid measure of dependency and whether the combined use of the questions to classify patients into one of three outcome groups (dependent, independent with problems and independent with no persisting problems) is a valid measure of overall health related quality of life. I also assessed the test-retest reliability of this classification and the agreement between patients and their proxies for responses to the modified simple questions, because patients are often unable to complete questionnaires or participate in interviews after stroke.

8.2 Methods

8.2.1 Selection of patients

8.2.1.1 Lothian Stroke Register Series

I studied the validity, and the proxy validity, of the modified simple questions in a series of 152 patients from our prospective registry of inpatients and outpatients with

first (or recurrent) stroke. The patients, and their proxies, were asked the questions as part of a longer interview to examine the validity of the EuroQol questionnaire after stroke (Chapters Three and Four). I have described how we identified the patients and their proxies and how we administered the instruments in Sections 3.2 and 4.2. Briefly, all patients were visited by a research nurse at home. The nurse administered the modified simple questions, the EuroQol (The EuroQol Group, 1990), the Frenchay Activities Index (Wade *et al.* 1985), the Hospital Anxiety and Depression Scale (HADS) (Zigmond & Snaith, 1983) and a VAS pain scale as questionnaires to be completed by the patient as far as possible. When patients could not complete the questionnaires by themselves, the nurse administered the questionnaires by interview. The nurse always administered the modified simple questions first, to limit interaction with the subsequent questions. The nurse assessed the Barthel Index and the OPCS disability scores (Mahoney & Barthel, 1965; Wellwood *et al.* 1995) by direct questioning at the end of the interview. We asked patients to ensure that a friend or relative (a proxy) who knew them well was available at the time of the interview. The nurse asked each proxy to independently complete a questionnaire booklet including the modified simple questions on behalf of the patient.

8.2.1.2 International Stroke Trial Series

I also studied the relationship between responses to the modified simple questions and the assessments of health related quality of life, measured with the EuroQol and SF-36, in the cohort of patients included in the randomised comparison of these instruments (Section 5.2.1). I have described the selection of patients, allocation of patients to the questionnaires and the methods of questionnaire administration in detail in Section 5.2. The modified simple questions were re-administered with the health related quality of life questionnaires for both the study of the reliability and relationship between the different quality of life assessments (Chapters Six and

Seven), see Figure 8.1. I used data from both these populations to determine the test-retest reliability of the individual questions and that of the overall classification of outcome using the modified simple questions.

8.2.2 Statistical analysis

I assessed the concurrent validity of the modified simple questions by calculating the sensitivity, specificity and accuracy with which the responses to each question predicted whether the patient scored as "good" or "bad" (i.e. score above or below the cut-offs for the appropriate standard instrument which separated "good" from "bad" outcome). The cut-offs were selected either if they had face validity or if they had been used in previous studies for each subscale (Lindley *et al.* 1994; O'Rourke, 1996; Dennis *et al.* 1997a; Dennis *et al.* 1997b).

The modified simple questions may also be used to classify surviving patients into one of three hierarchical outcome groups: patients who respond with "yes" to the dependency question are classified as dependent, patients who respond with "no" to the dependency question and "yes" to the recovery question may be classified as independent and patients who respond with "no" to both questions may be considered independent and fully recovered, Figure 8.2. I investigated the validity of this classification by plotting histograms which showed the distribution of responses to our standard measures for patients who were dependent, independent and fully recovered. I further assessed the concurrent validity of this classification by calculating the median score and interquartile range for each standard instrument for each of these three outcome groups. I used the Kruskal-Wallis one way analysis of variance to compare the distribution of scores for each group.

I examined the agreement between patients' responses to the modified simple questions on test and retest using the kappa statistic (Brennan & Silman, 1992). I also examined the agreement between patients and their proxies for their classification of the patients' outcome using the kappa statistic (Brennan & Silman, 1992).

All analyses were performed using "Access 2.0" (Microsoft Corporation) and the statistical software package "SPSS for Windows" (Release 6.1).

8.3 Results

8.3.1.1 Lothian Stroke Register Series

The patients were assessed at a median interval of 72 weeks after the onset of their index stroke (interquartile range: 43 to 104 weeks). Of the 152 patients who participated in this study, 92 were able to complete the questionnaires themselves; the remaining 60 patients could only be assessed by interview. Their characteristics at the time of registration following their index stroke have been reported in detail in Chapter Three - (Table 3.1).

147 (97%) patients completed both modified simple questions (three patients did not complete the modified dependency question and another two patients did not complete the modified recovery question). 54 (37%) patients replied "yes" to the dependency question and were classified as dependent, Table 8.2; eight of these 54 dependent patients replied "no" to the question about problems from their stroke (but in fact five of these eight had been rated as having an Oxford Handicap Score of at least two or more before their index stroke) (Bamford et al. 1989; Rankin, 1957). The remaining 93 (63%) patients were independent in everyday activities; of these, 52

reported being left with problems after their stroke, whereas the other 41 appeared to have made an excellent recovery in that they denied any problems.

8.3.1.1.1 Validity

The sensitivity, specificity and accuracy of the modified simple questions for the assessment of outcome in the domains of mobility, self care, social functioning, pain and psychological functioning are reported in Table 8.3. The modified dependency question had excellent sensitivity, specificity and accuracy for the assessment of activities of daily living as defined by the Barthel Index. Not surprisingly, it was less accurate at predicting psychological dysfunction as defined by the HADS. The modified recovery question proved a highly sensitive question for the detection of problems in all of the domains assessed, but it lacked specificity.

The median score and interquartile range for the standard instruments are shown for groups defined by their responses to both questions (Table 8.4). The median scores with the standard instruments were ordered appropriately and were statistically distinct. Figure 8.3 shows the distribution of responses to the Frenchay Activities Index for patients who were classified as dependent, independent or fully recovered according to the modified simple questions. Figure 8.4 shows the distribution of responses to the EuroQol questionnaire for patients classified by their responses to the modified simple questions. The best outcome in all domains (including overall health related quality of life with the "thermometer" - Table 8.5) was reported by the fully recovered group of patients. The mean utility scores with the EuroQol differed significantly among groups of these patients classified by their responses to the simple questions, Table 8.5.

8.3.1.1.2 Validity of the proxy assessments

The study nurse, using the OPCS disability measure, rated six patients as having significant difficulties in communication. These six patients did not therefore provide data for the assessment of agreement between the patients and their proxies. We obtained proxy assessments of the patients' outcome with the modified simple questions for 121 of the 147 patients. Agreement between the patient and their proxy for the outcome classification with the modified simple questions was good (agreement = 87/121 = 72%; kappa = 0.57, 95% confidence interval 0.45 to 0.69). There was a trend towards proxies assessing the patients' functioning as worse than that assessed by the patients themselves (21 versus 13, sign test $p > 0.05$).

8.3.1.2 International Stroke Trial Series

8.3.1.2.1 Relationship between the simple questions and the EuroQol

Figure 8.5 shows the distribution of responses to the EuroQol questionnaire for the patients classified by their responses to the modified simple questions. As before, each of the groups have distinct responses for all of the domains of the EuroQol and the pattern of responses was very similar to that observed in the other (Lothian Stroke Register) cohort (see Figure 8.4). For this larger cohort of patients, the mean estimates of overall health related quality of life were also ordered appropriately and were statistically distinct for each of the three levels (Table 8.5). The EuroQol utilities showed a similar pattern (Table 8.5).

8.3.1.2.2 Relationship between the simple questions and the SF-36

Although the responses to the modified simple questions appeared to be closely related to the patients' health related quality of life as defined by the EuroQol, I then

confirmed the relationship by relating the simple questions with the SF-36 (Table 8.6). For all but two of the domains (physical role functioning and mental health), the median scores for each domain were ordered appropriately and differed from each other. In the physical role functioning domain, the dependent and independent patients both scored zero out of a possible 100. This provides further evidence for a floor effect in this domain, see Section 7.3. Similarly, the median score in the mental health domain of the SF-36 was the same for patients who were classified as independent or recovered.

8.3.1.2.3 Test-retest reliability of the modified simple questions

In the study of the reliability of the EuroQol and SF-36 questionnaires, 443 patients returned repeat assessments of outcome - see Figure 6.1. These assessments included responses to the modified simple questions. A further 688 patients also returned repeat assessments of outcome which included the modified simple questions as part of the comparison of the EuroQol and SF-36 questionnaires - Chapter Seven, see Figure 7.1. Thus, 1,131 patients potentially completed the simple questions at test and retest, see Figure 8.1. Test-retest reliability was very good for both modified questions and the classification of overall outcome with the modified simple questions (Table 8.7, 8.8, 8.9).

8.4 Discussion

8.4.1 Modified dependency question

I found that the modified dependency question was a valid measure of dependency in activities of daily living after stroke. It had excellent sensitivity, specificity and

accuracy for identifying dependency after stroke. Although the current study was not primarily designed to directly compare the validity of the modified dependency question with that of the original, it had a greater sensitivity than the original at identifying dependency after stroke (85% versus 61%); however, its specificity was slightly worse than that of the original question (85% versus 96%) (Lindley *et al.* 1994). A multivariate test of the null hypothesis that the modified question has equivalent sensitivity and specificity to the original was marginally significant (Chi squared 5.93, df=2, p=0.052) and therefore suggests that the modification may have altered the question's measurement properties. The modified dependency question was also highly accurate at identifying poor mobility and social functioning after stroke. It was, not surprisingly, less accurate at identifying patients with psychological dysfunction. These indirect comparisons suggest that the modified version of the question has, at least, equivalent concurrent validity to the original as well as improved face validity (Lindley *et al.* 1994; Dennis *et al.* 1997b). I found the modified dependency question also had excellent test-retest reliability (kappa = 0.81); this compares very favourably with the inter-rater reliability of the original question (kappa = 0.51)(Dennis *et al.* 1997a). This difference may reflect the improved wording of the new question. Alternatively it may just reflect the differences between inter-rater and test-retest reliability or the method of questionnaire administration. Dennis and colleagues employed two raters who administered the simple questions on different occasions by face-to-face interview (Dennis *et al.* 1997a), whereas we administered the questions by post.

8.4.2 Modified recovery question ("problems" question)

We completely reworded the recovery question because patients found the original wording ambiguous and our assessments of its validity also revealed uncertainty about which aspect of outcome it addressed (Dennis *et al.* 1997a; Dennis *et al.*

1997b). The reworded question aims to detect those patients who are left without significant problems resulting from the stroke, rather than to identify patients who recover to the point of having no problems at all, whatever their cause. This alternative emphasis, although relatively subjective and non-specific, may be more relevant to both patients and their families. The reworded question had excellent sensitivity and moderate specificity for the detection of problems in all the domains examined (mobility, self care, social functioning and psychological functioning). These measurement properties differ clearly from those of the original recovery question (Dennis *et al.* 1997b) and support the validity of the reworded question. The reworded question also had substantially better reliability than the original (kappa = 0.78 versus kappa = 0.61) (Dennis *et al.* 1997a). However, it is not clear whether this is because of the improved wording or differences in the study design - see above. It is also not clear from the current data how patients with problems after their stroke relate to those who reported that they had not made a full recovery with the original question.

8.4.3 Combined use of modified questions to assess health related quality of life

The modified simple questions may be used to classify surviving patients into one of three hierarchical outcome groups. Approximately half of the independent patients were left with significant problems as a consequence of the stroke. This illustrates the **ceiling effect** with simple measures of disability that focus exclusively on activities of daily living, i.e. if the dependency question was the only measure of outcome, then other significant limitations (e.g. problems with household maintenance, social and psychological functioning) would not be captured. Additional assessments of health are therefore required in independent patients who might otherwise be considered as having achieved a "good outcome" after their stroke. The

statistically distinct profile of scores for this group of patients with the standard instruments support the validity of this hierarchical classification.

The validity of this classification is further supported by its close relationship with patients' responses to the assessments of health related quality of life. The same pattern of responses were observed, in both cohorts of patients, for assessments with the EuroQol and in the IST cohort for assessments with the SF-36. Therefore, the combined use of these modified simple questions, to classify patients as dependent, independent and fully recovered, appears to offer a simpler approach to the assessment of global health related quality of life, than with even the EuroQol. These assessments also had very good test-retest reliability which was of a similar order of magnitude to that observed for the assessments of overall health related quality of life with the EuroQol or general health with the SF-36 (Tables 6.2 and 6.5 respectively). Moreover, agreement between patients and their proxies for their responses to the modified simple questions was also moderately good (kappa = 0.57). There would be several potential advantages associated with the use of these simple questions for the assessment of health related quality of life after stroke. Firstly, because of their brevity and simplicity they are likely to place less of a burden on patients, and so may be more feasible than more complex measures. The advantages of short and simple measures, with a high frequency of response, are discussed in full in Chapter Five. Secondly, analysis of data based on a single classification of outcome by the simple questions would be simpler to both perform and interpret, than analysis of multidimensional data from a health related quality of life instrument, see Section 10.3.1. Finally, treatment effects described in terms of the proportion of patients reporting a change in response to the simple questions would have more immediate interpretability than a change in numerical score or utility.

However, there are also disadvantages to replacing detailed assessments of health related quality of life with the cruder information provided by these questions. Firstly, the modified simple questions could only be used to provide a broad and global assessment of health related quality of life. Therefore, the opportunity of obtaining information about outcome in specific domains, for instance psychological functioning, would be lost. This would reduce the opportunity to explain how a treatment improves overall health related quality of life, e.g. by improving psychological functioning or by improving mobility. Secondly, the modified simple questions were developed as a disease-specific instrument. They do not, therefore, provide outcome information which can be used to compare directly the effects of different treatments in groups of patients with different diagnoses. However, such comparisons could still be performed, albeit indirectly, by using the data from Table 8.5 to map patients' responses to the simple questions into health related quality of life utilities with the EuroQol.

8.4.4 Would qualitative research methods help understand the meaning of patients' responses to the simple questions?

The above studies were based on quantitative research principles. Quantitative research may be censured for lacking contextual detail and deriving trivial insights. Qualitative research belongs to the contrasting epistemological stance. It necessitates the study of relatively small numbers of non-randomly selected patients and employs deliberately unstructured, in-depth, face-to-face interview and observational techniques (Henwood & Nicolson, 1995). Findings are logged and analysed in a semi-structured way and grounded hypotheses emerge and are refined as the study progresses. Qualitative studies aim to unlock the "black box" of human thought and behaviour, and might be valuable in understanding the findings produced by the cross-tabulation of the modified simple questions. However, qualitative

methodology may be criticised as biased, lacking in power and generalisability (Carey, 1993). Furthermore, the validity of mixing of qualitative and quantitative techniques in one study is controversial (Carey, 1993).

8.4.5 Conclusions

Patients completed these modified simple questions either by interview or questionnaire in the presence of the study nurse. Although the study nurse did not routinely ask patients about their interpretation of either question, no significant ambiguities were noted. These modifications therefore appear to have improved both the face and content validity as well as the reliability of these simple questions. The current study has provided strong support for the concurrent validity of the modified questions when used either individually or together to classify patients into one of three hierarchical groups which are relevant to the patients' overall health related quality of life. Proxies were able to provide a moderately accurate and unbiased assessment of the patients' outcome after stroke with these modified simple questions. Therefore, the "modified" versions of the simple questions appear to be valid, and although they have not been formally compared, they seem preferable to the original.

8.5 Summary of Chapter Eight

1. The responses of patients to both simple questions may be used to classify patients as dependent, independent but not fully recovered, and independent and fully recovered. The relationship between this classification and assessments of health related quality of life had not previously been tested.

2. A previous study of the validity of the original simple questions had revealed ambiguities in their wording. For the present study, we examined the clarity, validity and reliability of a modified version of these simple questions.

3. The modified dependency question had excellent sensitivity, specificity and accuracy for identifying dependency after stroke. Indirect comparison with the original wording suggest that it has, at least, equivalent concurrent validity to the original as well as improved face validity.

4. The reworded recovery question aimed to detect patients left with significant problems as a consequence of their stroke. It had excellent sensitivity and moderate specificity for the detection of problems in a broad range of domains (mobility, self care, social functioning and psychological functioning).

5. The combined use of the modified dependency and recovery questions, to classify patients as dependent, independent and fully recovered, provided a valid and very simple overall measure of health related quality of life after stroke. This classification had very good test-retest reliability and was also moderately accurate when assessed by proxies on behalf of the patients.

Table 8.1: The simple questions

Original wording:

Dependency question "In the last two weeks, did you require help from another person for everyday activities?"
Recovery question "Do you feel that you have made a complete recovery from your stroke?"

Modified wording:

Modified dependency question "Do you need help **from anybody** with everyday activities?"
Modified recovery question "Has the stroke left you with any problems?"

Table 8.2: 2x2 table illustrating the responses to the modified dependency and recovery questions (n=147)

	"Has the stroke left you with any problems" [b]		
	Yes	No	Totals
*"Do you need help **from anybody** with everyday activities?"* [a]			
Yes (dependent)	46	8	54
No (independent)	52	41	93
Totals	98	49	147

[a] modified dependency question

[b] modified recovery question

Table 8.3: The sensitivity, specificity and accuracy of the modified simple questions in predicting patients' responses to more complex measures

	"Do you need help from anybody with everyday activities?"			"Has the stroke left you with any problems"		
	Sensitivity (%)	Specificity (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
OPCS locomotion (>0)*	72	95	84	84	49	65
Barthel Index (<20)	85	85	85	87	42	56
(<18)	91	79	82	85	38	49
Frenchay Activities Index (>34)	100	42	49	65	72	71
(<22)	67	90	79	77	42	59
(<17)	77	88	84	81	41	56
Visual analogue pain scale (>0)*	52	73	64	70	36	50
HADS depression subscale (<7)*	60	71	67	88	42	56
HADS anxiety subscale (<7)*	50	70	61	83	46	62

*low scores indicate better outcome

Table 8.4: Concurrent validity of the modified simple questions (n=147)

Standard instrument	Median score on standard instruments (interquartile range) for groups defined by response to modified simple questions			p
	"Dependent" (n=54)	"Independent" (n=52)	"Recovered" (n=41)	
OPCS locomotion score	7 (3 - 9.5)	0 (0 - 3)	0 (0 - 0)	<0.0001
Barthel Index	17 (12 - 20)	20 (20 - 20)	20 (20 - 20)	<0.0001
Frenchay Activities Index	9 (5 - 15)	26 (21 - 31)	33 (26 - 36)	<0.0001
Visual analogue pain scale	15 (0 - 50)	0 (0 - 20)	0 (0 - 10)	0.001
HADS depression subscale	6 (4 - 8)	4 (2 - 7)	3 (2 - 4)	<0.0001
HADS anxiety subscale	7 (2 - 9)	4 (2 - 10)	3 (1 - 5)	0.0075

p = significance of differences between levels using the Kruskal Wallis test (df=2)
"Dependent" = needing help for activities of daily living (ADL)
"Independent" = not needing help for ADL, but persisting problems after stroke
"Recovered" = not needing help for ADL and no persisting problems after stroke

Table 8.5: Relationship between patients' responses to the modified simple questions and overall health related quality of life assessed by the EuroQol

	Mean score on EuroQol for groups defined by their responses to modified simple questions (95% CI of the mean)			
	"Dependent"	"Independent"	"Recovered"	p
LSR Series (n=147)				
Overall HRQoL	58 (52 to 64)	65 (58 to 71)	77 (74 to 80)	NS
EuroQol utility	0.38 (0.29 to 0.47)	0.74 (0.69 to 0.79)	0.88 (0.80 to 0.96)	<0.01
IST Series (n=867)				
Overall HRQoL	48 (47 to 50)	67 (64 to 70)	77 (74 to 80)	<0.0001
EuroQol utility	0.31 (0.29 to 0.34)	0.71 (0.68 to 0.74)	0.88 (0.84 to 0.92)	<0.0001

Table 8.6: Relationship between patients' responses to the modified simple questions and health related quality of life assessed by the SF-36 (n=849)*

Domain of SF-36	Median score on domains of SF-36 for groups defined by their responses to modified simple questions (interquartile range)			
	"Dependent" (n=564)	"Independent" (n=137)	"Recovered" (n=123)	p†
Physical functioning	5 (0 to 25)	50 (30 to 65)	68 (50 to 85)	<0.0001
Physical role function	0 (0 to 0)	0 (0 to 25)	75 (0 to 100)	<0.0001
Social functioning	25 (13 to 50)	63 (50 to 75)	88 (63 to 100)	<0.0001
Mental health	60 (52 to 68)	64 (56 to 68)	64 (60 to 68)	0.02
Emotional role function	0 (0 to 25)	25 (0 to 100)	100 (25 to 100)	<0.0001
Bodily pain	41 (22 to 62)	62 (41 to 100)	100 (62 to 100)	<0.0001
Vitality	30 (15 to 45)	40 (30 to 55)	60 (52 to 75)	<0.0001
General Health	35 (25 to 52)	52 (35 to 72)	72 (55 to 87)	<0.0001

* simple question data missing in 25 patients
† between groups by Kruskal-Wallis one way analysis of variance

Table 8.7: Test-retest reliability of modified dependency question (n=1,033)

		First assessment	
		Dependent	Independent
Second assessment	Dependent	679	46
	Independent	37	271

Agreeement = 92%
Kappa = 0.81 (0.77 - 0.85)

Table 8.8: Test-retest reliability of modified recovery question (n=1,083)

		First assessment	
		Problems	No problems
Second assessment	Problems	679	46
	No problems	37	271

Agreeement = 94%
Kappa = 0.78 (0.73 - 0.83)

Table 8.9: Test-retest reliability of the overall classification of outcome derived using both modified simple questions (n=1,020)

		First assessment		
		Dependent	Independent	Recovered
Second assessment	Dependent	679	38	8
	Independent	27	98	21
	Recovered	9	11	129

Agreeement = 89%
Kappa = 0.76 (0.71 - 0.80)

Figure 8.1: Flow of patients through study

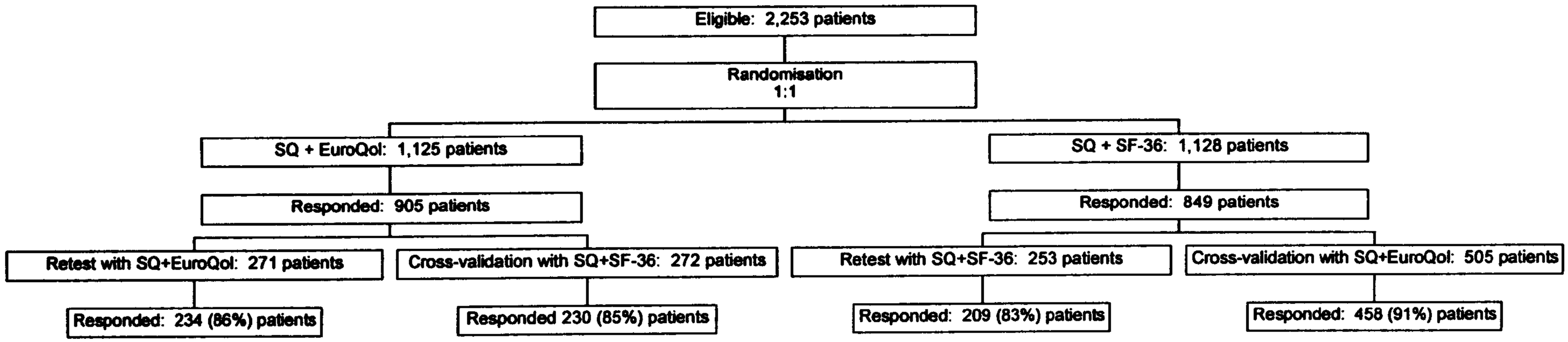


Figure 8.2: A flow diagram illustrating how the simple questions could be used to classify patients into one of three hierarchical groups

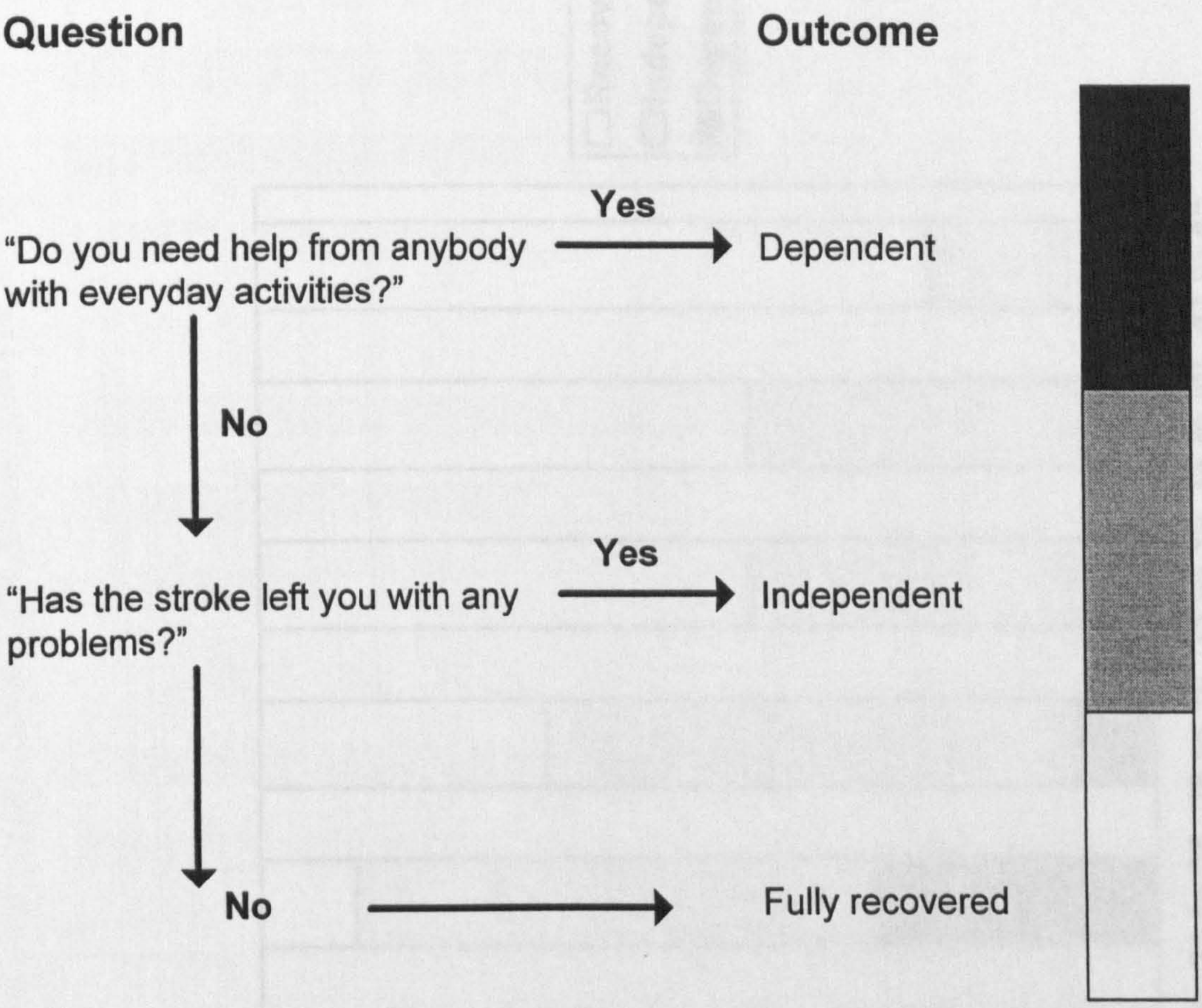


Figure 8.3: Distribution of responses to the Frenchay Activities Index in patients grouped by responses to the modified simple questions (LSR Series)

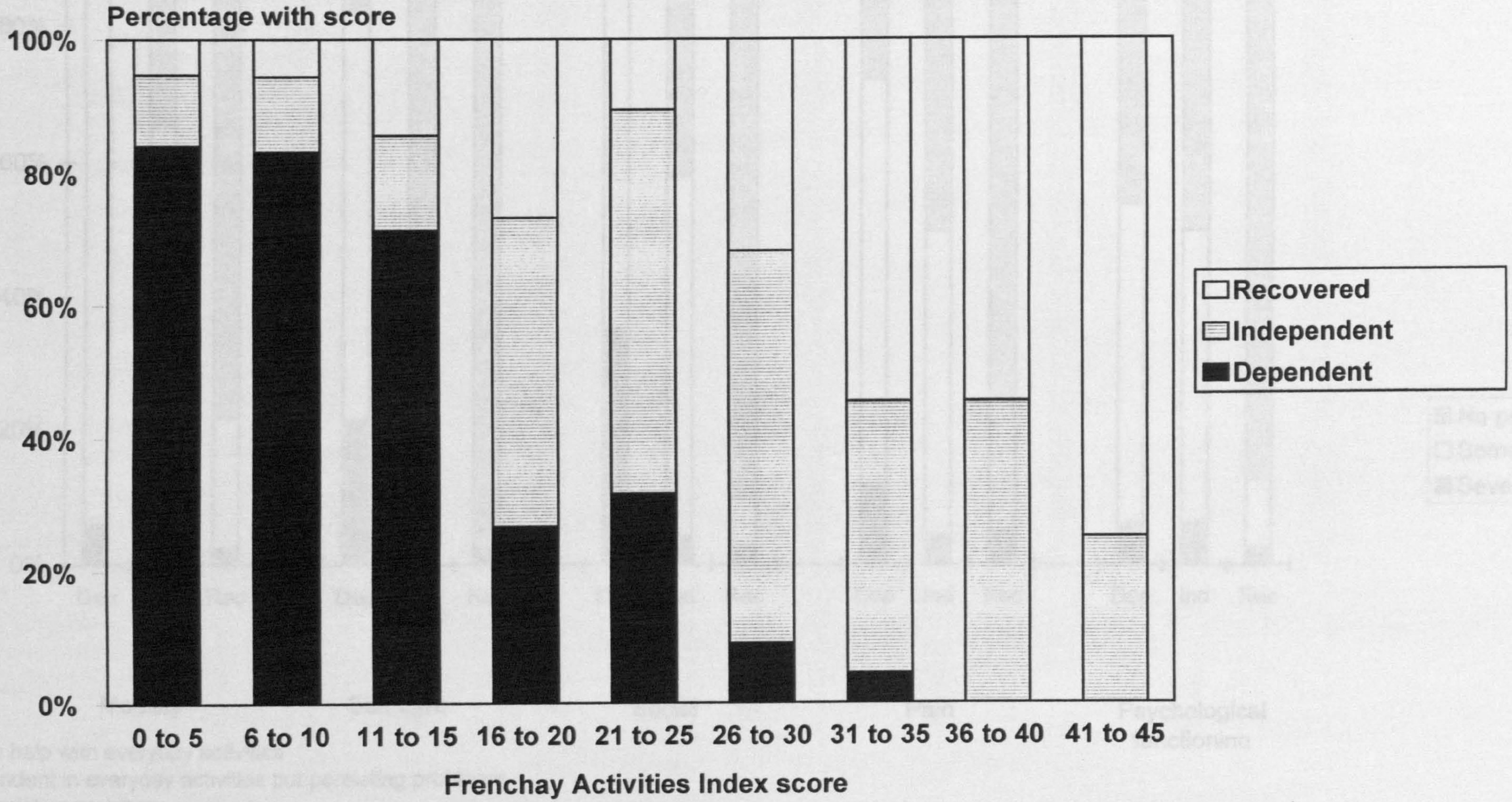
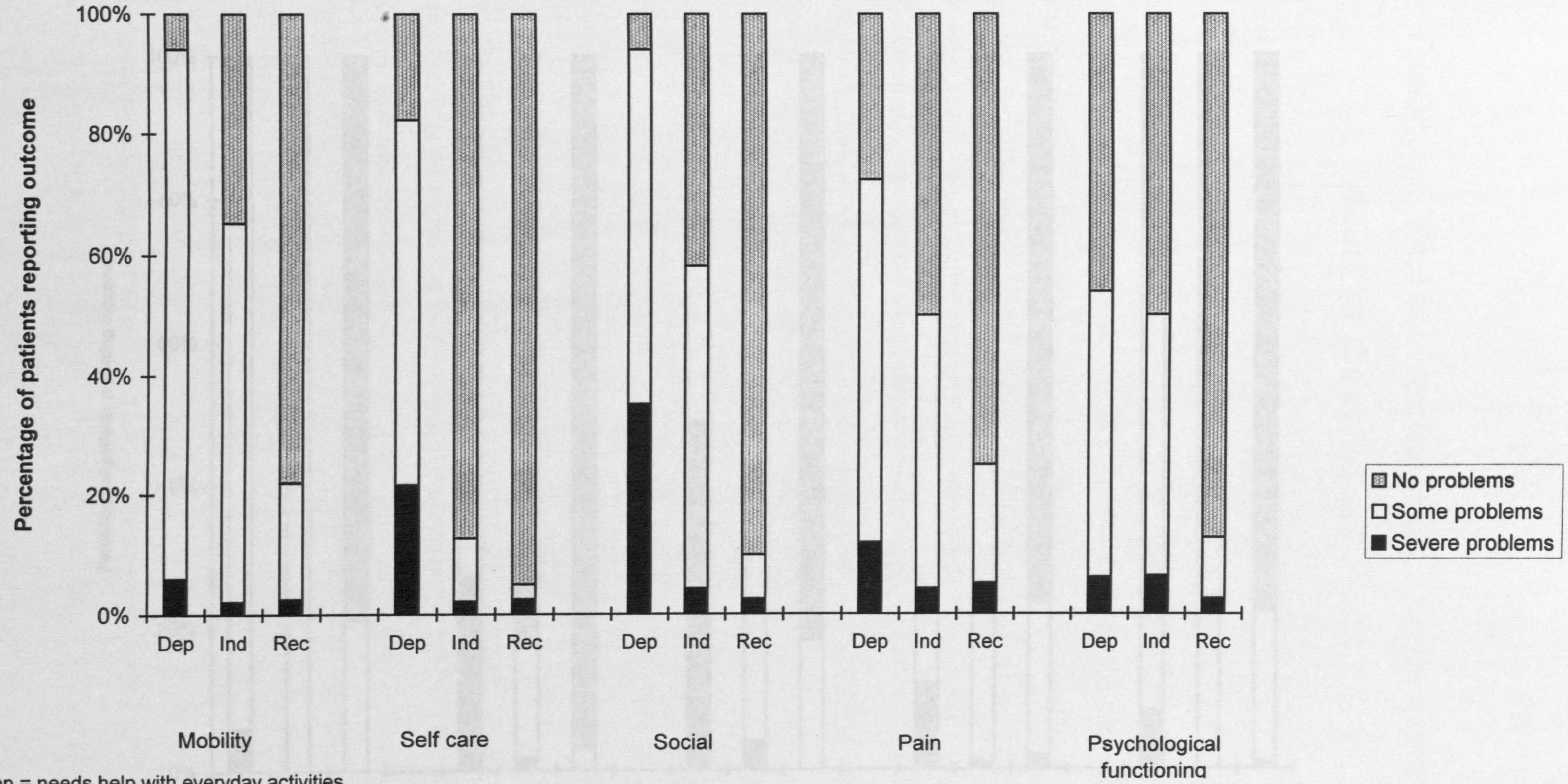
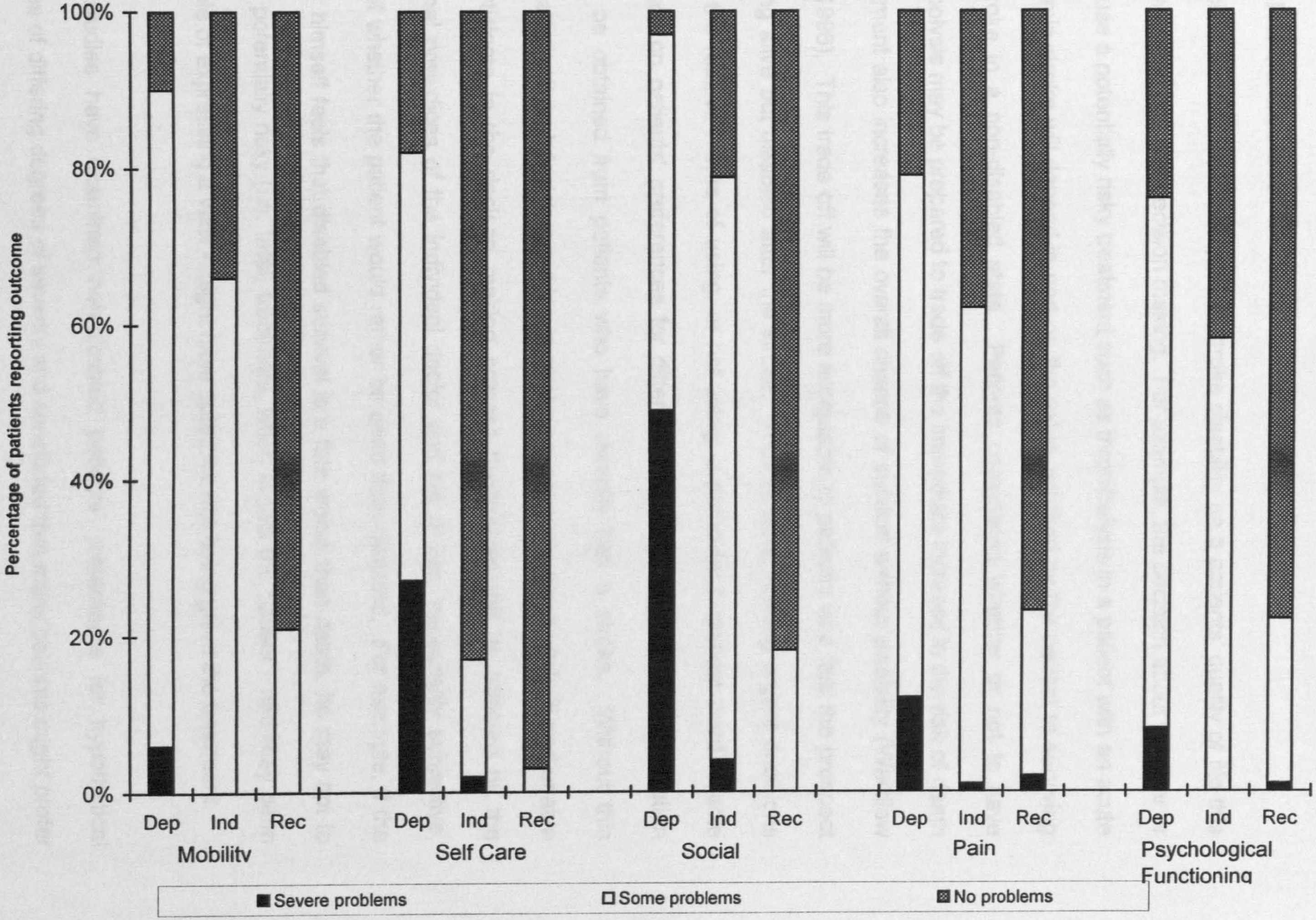


Figure 8.4: Health related quality of life after stroke with the EuroQol for patients defined by their responses to the modified simple questions (LSR Series, n=147)



Dep = needs help with everyday activities
Ind = independent in everyday activities but persisting problems
Rec = no persisting problems

Figure 8.5: Health related quality of life with the EuroQol for patients defined by their responses to the modified simple questions (IST Series)



9 Quality of life after stroke: do patients prefer death or disabled survival?

9.1 Introduction

If clinicians understand the impact of stroke disability on a patients' quality of life this may help inform clinical decision making. For example, the decision about whether or not to use a potentially risky treatment such as thrombolysis in a patient with an acute ischaemic stroke will depend in part on the value ascribed by the patient to surviving the stroke in a non-disabled state. Patients considering whether or not to have thrombolysis may be prepared to trade off the immediate increase in the risk of death if treatment also increases the overall chance of survival without disability (Wardlaw et al. 1996). This trade off will be more acceptable to patients who fear the prospect of being alive but disabled after the stroke. Thus doctors, making urgent decisions about the relative merits of using, or not using, a particular treatment need reliable evidence on patients' preferences for different outcomes. Ideally, this information would be obtained from patients who have recently had a stroke. Without this information, clinical decision making in stroke patients who are acutely ill and unable to participate in the decison making process themselves will be affected by the personal prejudices of the individual doctor and his or her, necessarily subjective, view of whether the patient would rather be dead than disabled. For example, if the doctor himself feels that disabled survival is a fate worse than death, he may opt to use a potentially risky (i.e. fatal) treatment, when in fact the patient - had they been capable of expressing a view - might have opted for **not** being given the treatment.

Two studies have examined non-disabled patients' preferences for hypothetical strokes of differing degrees of severity and concluded that many patients might prefer death to survival in a severely disabled condition (Solomon *et al.* 1994; Gage *et al.*

1996). This conclusion may be unjustified. Firstly, the studies examined only the attitudes of non-disabled patients to anticipated physical disability, and not the attitudes of patients actually disabled by stroke (which might well be very different). However, the attitudes of disabled stroke survivors are often impossible to assess directly because neurological deficits limit the patients' ability to complete questionnaires or participate in interviews. Secondly, the studies only examined attitudes to anticipated physical disability and so neglected the possibility that other domains of outcome (e.g. psychological functioning) might determine a patient's overall quality of life. Thirdly, and most importantly, their conclusions seem inconsistent with the report that the majority of long term stroke survivors are satisfied with their lives (King, 1996).

I therefore used the EuroQol, a generic measure of health related quality of life, which I have validated for use after stroke (Chapter Three), to assess the health related quality of life of a large sample of disabled and non-disabled stroke survivors. I investigated the attitudes of severely disabled stroke patients to their disability by comparing assessments of the patient's overall health related quality of life with those of less severely disabled and non-disabled stroke survivors. I also compared the patient (or carer) valuations of the patient's health states with valuations which reflect the preferences of the general public for each of the patient's health states. I finally investigated whether the measurement of health related quality of life provides a more comprehensive picture of outcome after stroke, than measures of disability alone, by investigating which domains of health related quality of life distinguish between stroke survivors and age- and sex- matched community controls.

9.2 Methods

9.2.1 Selection of stroke patients

In a previous study I performed a direct assessment of the feasibility of a postal version of the EuroQol questionnaire after stroke (Chapter Five). I described the methods used to identify the patients, the format of the instruments and the method of questionnaire administration in Chapter Five. I used these patients' responses to the EuroQol questionnaire for the current analyses.

9.2.2 Selection of control patients

We selected control patients from participants in two national population surveys in which the EuroQol had been included. The first of these surveys was commissioned by the Department of Health to investigate the way that the general public values health states (Measurement and Valuation of Health (MVH) Study). In this survey, a representative sample of the United Kingdom population (n=3,235) completed the EuroQol questionnaire as part of a longer face-to-face interview (Gudex *et al.* 1996; Kind *et al.* 1998). The second of these surveys also drew upon a nationally representative sampling frame and was conducted by the then Office of Population and Census Studies (OPCS) and involved 5,962 respondents (Omnibus Survey, 1995). Both studies surveyed non-institutionalised individuals living in the community. We matched stroke patients by age and sex against individuals in the MVH dataset. Once selected as a match, individuals in the MVH dataset were removed from further consideration. We repeated this process using the Omnibus Survey dataset for stroke patients not matched by the MVH data.

9.2.3 Study instruments and definitions

I used the modified dependency question (Do you need help from anybody with everyday activities?) to assess dependency in activities of daily living after stroke, (Chapter Eight). I used a combination of patients' responses to the EuroQol questionnaire and the dependency question to classify patients according to the severity of their physical disability:

Severe physical disability: I classified patients as "severely physically disabled" if they reported severe problems in either the mobility or self care domains (i.e. "I am confined to bed" or "I am unable to wash or dress myself").

Disabled: I classified patients as disabled if they responded with "yes" to the dependency question and did not report severe problems in either the mobility or self care domains of the EuroQol.

Non-disabled: I classified the remainder as non-disabled.

9.2.4 Statistical analysis

I examined the attitudes of stroke survivors, or their carers, to their current health status by plotting the distribution of their estimates of overall health related quality of life against the severity of their current reported physical disability. I compared the overlap between the distributions by assessing the proportion of severely physically disabled or disabled patients who had scores within the range of scores for non-disabled patients. I also examined the attitudes of the general public to the stroke patients' health states by plotting the distribution of EuroQol utilities (Dolan et al. 1995) for patients classified according to the severity of their current physical disability. Dolan and colleagues derived these utilities by asking the 3,235 respondents of the MVH survey to directly value 45 (of the 243) EuroQol health states using a time trade off technique. The remaining health states were valued indirectly by a statistical modelling technique.

I also compared the overall health related quality of life of each stroke survivor with that of individually age- and sex- matched controls. I performed a conditional logistic multiple regression analysis (S-Plus) to investigate which domains of the EuroQol differed between patients with stroke and community controls, i.e. which domains are important in post-stroke outcome.

9.3 Results

905 patients returned a EuroQol questionnaire booklet after one reminder, (Chapter Five). On the basis of their responses to the EuroQol and the dependency question, I classified 187 patients as severely disabled, 476 as disabled and 210 as non-disabled. I was unable to classify 32 patients because they did not complete the dependency question. Only 15 (8%) of the severely disabled patients completed the EuroQol questionnaire themselves; by contrast, 229 (48%) of the disabled patients and 184 (88%) of the non-disabled patients completed the EuroQol questionnaire without help. The baseline distributions of age, sex and stroke syndromes differed amongst the severely disabled, disabled and non-disabled patients, are given in Table 9.1.

There was a significant overlap in the distribution of the estimates of their overall health related quality of life for patients with differing levels of physical disability (Figure 9.1). Fifty six (34%) of the severely disabled patients and 237 (54%) of the disabled patients reported, or were assigned, overall health related quality of life scores which lay between the 5th and 95th percentile of the scores of the non-disabled patients; 393 (90%) of disabled patients and 68 (35%) of the non-disabled patients had overall health related quality of life scores below the 95th percentile of the scores of the severely disabled group. Similarly, there was a significant overlap in

the distribution of estimates of overall health related quality of life between stroke patients and age- and sex- matched community controls (Table 9.2).

By contrast, the overlap in the distribution of utilities for patients with differing levels of physical disability was much less marked, Figure 9.2. None of the severely disabled patients had utilities which lay between the 5th and 95th percentile of the valuations of the non-disabled patients. Only 128 (27%) of the moderately disabled patients and one (<1%) of the non-disabled patients had utilities below the 95th percentile of the severely disabled group. Approximately 47% of the severely disabled patients had utility scores of less than zero (i.e. in the view of the general public their health states might be considered worse than death).

Stroke patients reported significantly worse overall health related quality of life than age- and sex- matched community controls, Table 9.2. In a conditional logistic regression, five of the six domains of the EuroQol (self care, social functioning, pain, psychological functioning, and subjective assessments of overall health related quality of life) contributed independent information which helped separate stroke patients from control subjects, Table 9.3. In this model, stroke patients reported less pain than the control subjects after adjustment for the other variables.

9.4 Discussion

9.4.1 Do stroke survivors prefer death to disabled survival?

There was a marked overlap in the subjective valuations of overall health related quality of life given by, or for, patients who were severely disabled, disabled and non-disabled. Many disabled patients reported, or were assigned, estimates of overall

health related quality of life which were similar to those reported by both the non-disabled stroke patients and the community controls. This suggests that many disabled stroke patients might regard their quality of life as acceptable and would not prefer death to being alive, but physically disabled. The distribution of utility scores (generated by the valuation of hypothetical health states by the general public - see Section 9.2.4) suggests a qualitatively different conclusion. The distribution of the utility scores had a completely different shape to that of the patient, or carer, estimates of overall health related quality of life. The scores of the severely disabled patients did not overlap as much with those of the disabled and non-disabled patients. Moreover, approximately half of the severely disabled patients had utility scores of zero or less which implies that their health state might be considered, at least in the eyes of the general public, worse than death.

These apparently conflicting results might simply reflect a difference in the attitudes of respondents depending on whether they are directly valuing an actual real-life health state or a hypothetical health state. Alternatively, the personal experience of disability may have modified the manner in which an individual perceives his/her or a relative's current health state, i.e. many of the disabled stroke respondents might have also considered death as preferable to disabled survival before they suffered their disabling stroke. Subjects with poor health in a community based study gave higher valuations (i.e. closer to good health) for disordered health states (Kind & Dolan, 1995), and in a study of individual quality of life in patients awaiting unilateral hip replacement, the patients had similar scores to the healthy controls (O'Boyle *et al.* 1992). These experimental conclusions are also consistent with more anecdotal examples of the remarkable capacity with which individuals maintain hope and meaning in life despite adverse circumstances, e.g. victims of concentration camps (Fallowfield, 1990).

Over 40% of disabled stroke survivors reported, or were assigned, estimates of overall health related quality of life which were similar to those reported by the non-disabled patients. However, there was a substantial number of severely disabled patients who were assigned, or reported, overall health related quality of life below this range. It is plausible that a significant proportion of these patients have such poor health related quality of life that they might prefer death to disabled survival. It would be ideal if each of these groups of patients (i.e. those who subsequently achieve acceptable health related quality of life and those who do not) could be identified prospectively at the time they present to the hospital with their stroke. However, this would require very accurate prognostic models for the prediction of both disability and overall health related quality of life. Current models for the prediction of disability after stroke are only accurate enough for predicting outcome in large groups of patients, rather than for individuals (personal communication, Carl Counsell). As yet, no models for the prediction of health related quality of life after stroke have been described. Furthermore, it is unlikely that any such models will ever be sufficiently accurate.

9.4.2 Are the patients (and the carer's) assessments of overall health related quality of life valid?

The validity of our conclusions depend largely on the validity of the patient and carer assessments of health related quality of life with the EuroQol. I have demonstrated the concurrent validity of the EuroQol when administered by either questionnaire or interview after stroke, (Chapter Three). I found that patients' subjective assessments of overall health related quality of life with the EuroQol were best explained by the patients' mood, level of social functioning and pain, and so were likely to be clinically meaningful. In this current study many of the assessments of overall health related quality of life in patients with disability were completed by proxies rather than by the

patients themselves. However, this problem is not unique to this study; physical and cognitive problems prevent the direct assessment of quality of life in most severely disabled patients by any means. Moreover, it is unlikely to be important as I have previously demonstrated assessments by proxy to be moderately accurate, Chapter Four. Furthermore, proxy assessments tend to underestimate rather than overestimate the patients' health, at least for patients with mild to moderate disability, see Section 4.3. Unfortunately, the accuracy of proxy assessments for more severely disabled patients is not testable, see Section 4.4.4.

The observed overlap in health related quality of life between disabled and non-disabled patients might have been exaggerated if the non-disabled patients had sub-optimal overall health related quality of life because of problems in the domains not related to physical functioning. Again, this seems unlikely to be important because I also observed a clear overlap between scores of the stroke patients and those of the age- and sex- matched community controls.

The surprisingly high overall health related quality of life scores of the disabled patients could be interpreted as reflecting the patients' optimism about their prospects for recovery. However, this also seems unlikely, because these assessments were performed after a mean interval of approximately 60 weeks after the index stroke, at a time when most patients' neurological status would have been stable for at least a period of several months.

9.4.3 Are the determinants of health related quality of life after stroke multidimensional?

In this study, stroke patients reported, or were assigned, a wide range of scores of overall health related quality of life. This is consistent with the observation that any

two patients may have very different functional and emotional responses to the same clinical deficit (Guyatt *et al.* 1993). The wide range of estimates of overall health related quality of life for patients with apparently similar severity of physical disability also suggests that physical functioning is just one of several factors which determine the patient's overall health related quality of life. I investigated this further by comparing the health related quality of life of stroke patients with that of age- and sex- matched community controls. A conditional logistic regression indicated that five of the six domains of the EuroQol (self care, pain, psychological and social functioning, and subjective overall health related quality of life, but not mobility) provided useful independent information which helped to distinguish stroke patients from community controls. These findings suggest that any studies on the patients' views about hypothetical strokes, should include descriptions of problems in the relevant domains of health related quality of life, rather than just focusing on patients' views about physical impairments, i.e. Solomon and colleagues should have included psychological problems and pain in their hypothetical scenarios, rather than just neurological impairments. Similarly, any future studies of outcome after stroke should include measures of health related quality of life, rather than just measures of disability, as this will help provide a more global picture of the patients' outcome. Measuring all the domains of health related quality of life may improve the power of trials to detect treatment effects and focus attention on the outcomes which are more relevant to patients (Rothwell *et al.* 1997).

9.4.4 Clinical implications and conclusions

It may well be that healthy, non-disabled individuals with no personal experience of stroke would prefer to die from a hypothetical stroke than to survive in a disabled state. The present study shows that a significant proportion of disabled stroke survivors, or at least their carers, do not view survival in a disabled state as such a

bad thing as one might expect. Clinicians should bear these findings in mind when considering treatments which apparently trade off duration of survival against quality of survival. The EuroQol domains of self care, social functioning, pain, psychological functioning and subjective overall health related quality of life all provided independent information which helped separate stroke patients from community controls. Assessments of health related quality of life may therefore provide a more comprehensive description of patients' functioning after stroke than measures of disability.

9.5 Summary of Chapter Nine

1. Many disabled patients reported, or were assigned, estimates of overall health related quality of life which were similar to those reported by both non-disabled stroke patients and community controls. This suggests that some disabled stroke patients, or their carers, regard their quality of life as acceptable and would not prefer death to being alive, but physically disabled.

2. The utility scores (generated from general public valuations of the health states of severely disabled patients) differed substantially from the patient or carer valuations. In the view of the general public, many of these patients have health states which might be considered worse than death. These contrasting results might simply reflect a difference in the attitudes of respondents with direct experience of disability.

3. Patients with a similar severity of physical problems reported a diverse range of estimates of overall health related quality of life. This suggests that physical functioning is just one of several factors which determine a stroke patient's overall health related quality of life. A conditional logistic regression indicated that five of the six domains of the EuroQol (self care, pain, psychological and social functioning, and subjective overall health related quality of life) provided useful independent information which helped to distinguish stroke patients from community controls.

Table 9.1: Characteristics of patients included

	Severely disabled (n=187)	Disabled (n=476)	Non-disabled (n=210)
Gender			
Males	81 (43%)	241 (51%)	149 (71%)
Females	106 (57%)	235 (49%)	61 (29%)
Initial stroke syndrome			
TACS	78 (42%)	84 (18%)	18 (9%)
PACS	72 (38%)	197 (41%)	97 (46%)
LACS	26 (14%)	154 (32%)	55 (26%)
POCS	11 (6%)	41 (9%)	40 (19%)
Mean weeks from stroke (SD)	59 (31)	60 (32)	59 (30)
Mean age in years (SD)	75 (8.8)	71 (11.4)	68 (11.1)

TACS = total anterior circulation stroke syndromes (most extensive neurological deficit)
PACS = partial anterior circulation stroke syndromes
LACS = lacunar stroke syndromes
POCS = posterior circulation stroke syndromes

Table 9.2: Estimates of overall health related quality of life of severely disabled, mildly disabled and non-disabled stroke survivors compared with age- and sex- matched (non-disabled) community controls

Disability of stroke patients	N	Mean estimates overall HRQoL (95% CI)		Overlap†
		Stroke patients	Controls matched by age and sex for each stratum of stroke disability	n (%)
Severely disabled	166	37 (0 to 76)*	72 (35 to 100)	90 (54)
Disabled	431	53 (21 to 84)*	74 (32 to 100)	375 (87)
Non-disabled	194	73 (42 to 100)*	79 (45 to 100)	188 (97)

N = number of stroke patients and controls with no missing data

*stroke patients have significantly lower estimates of overall HRQoL ($p<0.01$)

† number of stroke patients with scores within the 95% confidence interval of the control patients

Table 9.3: Which domains of HRQoL independently distinguish between stroke patients and age- and sex- matched community controls? A conditional logistic regression.

Domain	*Odds ratio (95% CI)*
Self care	7.41 (4.31 - 12.75)*
Social functioning	6.42 (3.79 - 10.87)*
Pain	0.25 (0.16 - 0.40)
Psychological functioning	1.95 (1.26 - 3.03)*
Subjective overall HRQoL	0.97 (0.96 - 0.98)*

* Increasing dysfunction with the EuroQol increases the probability of predicting a stroke patient, i.e. problems in the self care domain increase the odds of identifying a patient rather than a control by a factor of 7.41. An odds ratio <1 implies that problems in that domain are more frequent in the community controls.

† Mobility was not a significant independent predictor

Figure 9.1: Distribution of patient, or carer, estimates of overall health related quality of life in severely disabled, mildly disabled and non-disabled stroke survivors

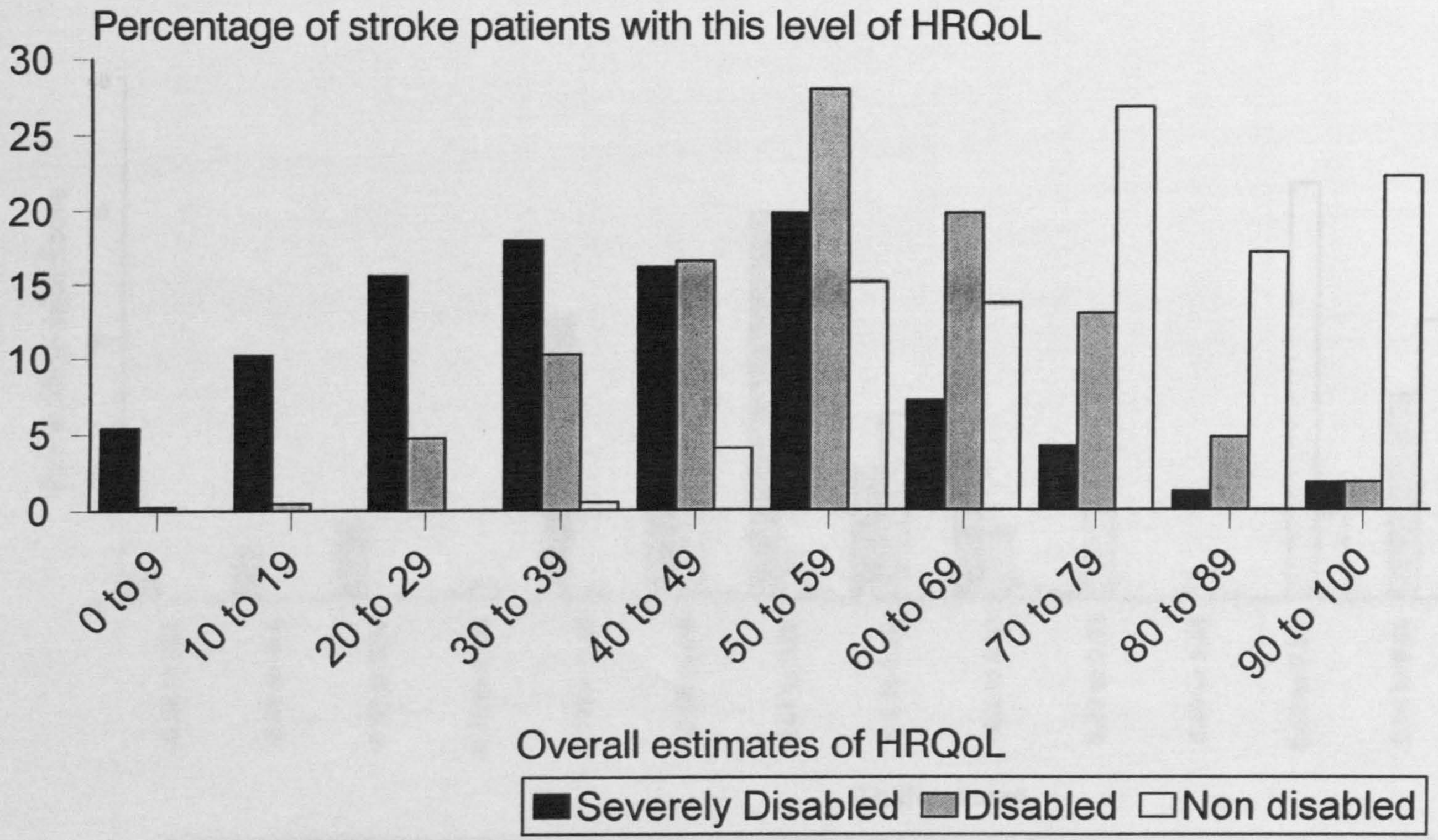
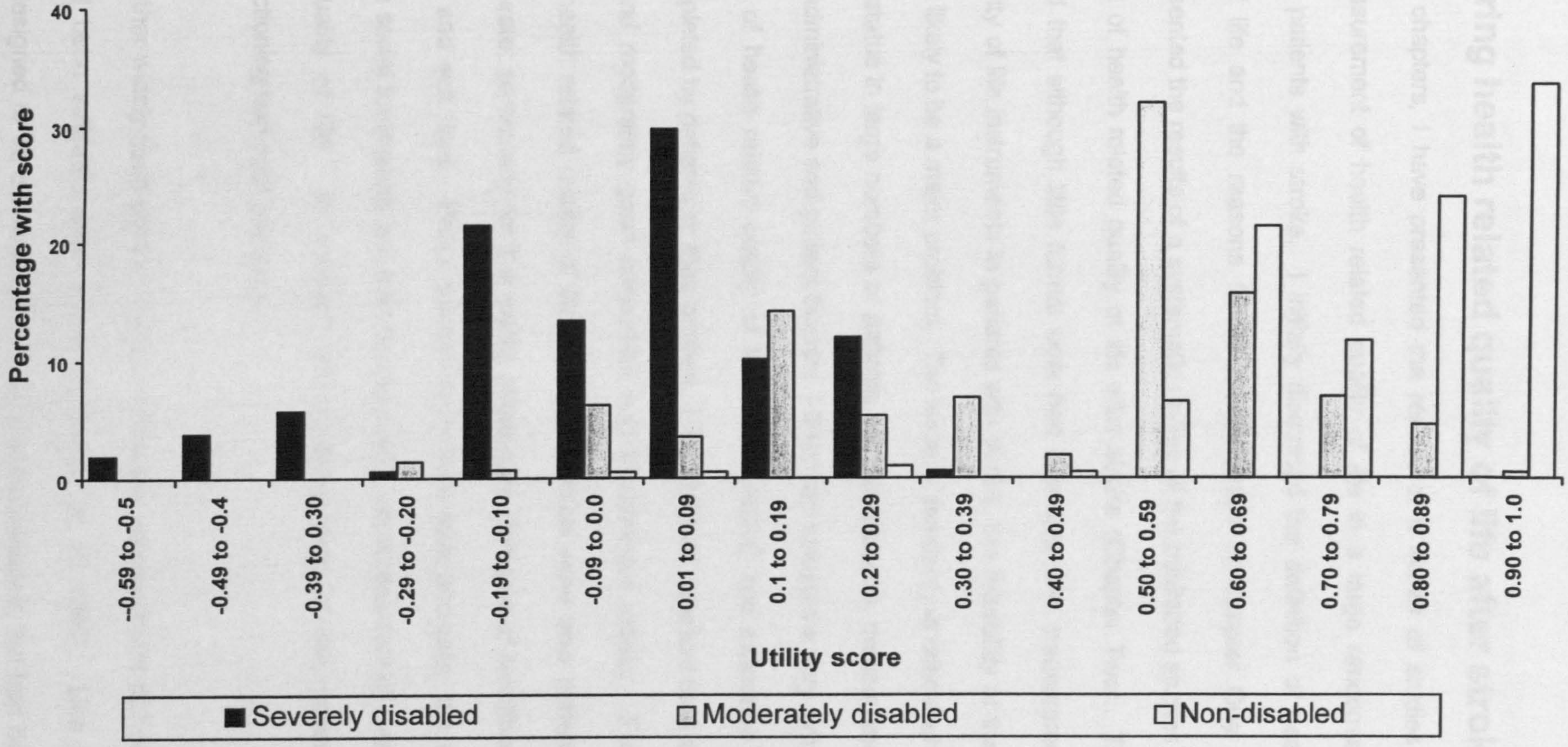


Figure 9.2: Distribution of EuroQol utility scores in the severely disabled, mildly disabled and non-disabled stroke survivors



10 Summary and Conclusions

10.1 Measuring health related quality of life after stroke

In the preceding chapters, I have presented the results of a series of studies to examine the measurement of health related quality of life in a large randomised controlled trial in patients with stroke. I initially discussed the definition of health related quality of life and the reasons for its measurement in Chapter One. I subsequently presented the results of a systematic review of the published studies on the measurement of health related quality of life after stroke (Chapter Two). This review highlighted that although little formal work had examined the measurement properties of quality of life instruments in patients with stroke, the feasibility of these assessments was likely to be a major problem. The issue of feasibility is critical when assessing health status in large numbers of patients. In this situation, measurement tools need a low administrative and patient burden. I therefore selected a very simple generic measure of health related quality of life (the EuroQol) and examined its validity when completed by patients or their proxies. I found it had excellent face and content validity and moderately good concurrent and discriminant validity. Proxy assessments of health related quality of life with the EuroQol were also generally feasible and accurate; particularly for the easily observable domains of functioning such as mobility and self care. Proxy assessments were less accurate, but still acceptable for the social functioning and pain domains and the assessment of overall health related quality of life. In contrast, proxy assessment of the patients' psychological functioning had poor accuracy.

The SF-36 is another widely used generic instrument for the measurement of health status which had been validated after stroke (Anderson *et al.* 1996). Like the EuroQol, it was designed to be delivered by post and self-completed, but had been

administered by interview because of concerns over feasibility in patients with stroke (Anderson *et al.* 1996). I compared the EuroQol with the SF-36 to inform the selection of health related quality of life instrument after stroke. This comparison raised a variety of scientific issues.

10.2 Comparing instruments

The question: "which is the best health related quality of life instrument?" is frequently considered by investigators during the planning phase of clinical studies. Most commentators suggest that investigators should select the instrument with the best measurement properties (validity, reproducibility, and responsiveness) in the particular population of interest. There are problems with this approach. Firstly, interpreting the results of studies which examine measurement attributes is not straightforward. This is partly because there is no "gold-standard" for the measurement of health related quality of life, but also because the results of the tests used to assess these measurement properties require subjective interpretation (for instance, tests of validity do not generally show that it is simply absent or present, but require qualitative interpretation). Consequently, this hinders the comparison of the available instruments.

Most studies have examined the properties of single instruments in isolation, and so most comparisons have been indirect and potentially confounded by differences in case mix between studies. Some studies have compared instruments by administering both instruments sequentially in the same order to all patients (a non-random crossover design). Unfortunately, these studies are vulnerable to biases which arise from ordering effects (e.g. the response to the EuroQol might differ if the SF-36 is administered immediately before). These biases would be eliminated in a randomised comparison, as the process of randomisation aims to ensure that both

groups are similar at baseline. In such a study, it would be valid to administer only one instrument to each group of patients, and so eliminate potential problems with ordering, as well as test both instruments in a more directly relevant way. I therefore performed a direct randomised comparison of the EuroQol and SF-36 in survivors of acute stroke (Chapter Five). This gave a reliable and unbiased comparison of their relative feasibility. I found significantly better response frequency and quality of response in patients allocated to the EuroQol questionnaire. However, I was unable to extend this methodology to compare (in a reliable and quantitative manner) their reproducibility, as different statistical techniques are required to examine the reproducibility of continuous (SF-36) and categorical (EuroQol) data. However, a weaker qualitative comparison suggested their reproducibility was similar (Chapter Six). I was also unable to compare directly the validity of these instruments, as there is no widely accepted gold-standard for the measurement of health related quality of life. I used a randomised cross-over study to examine the relationship between the domains of either instrument. This suggested that both instruments were sampling broadly similar areas of health. Thus, in summary, although the EuroQol had a better quality and frequency of response than the SF-36, it was not clear which is the best instrument overall.

The simple questions were developed to measure dependency and recovery after stroke. Both questions can be used together to classify patients into one of three hierarchical levels of outcome (dependent, independent but not recovered, independent and fully recovered). We modified the original wording of both questions to try and improve their validity and reliability (Chapter Eight). I examined the relationship between patients' responses to these questions and the assessments of health related quality of life. I found the combined use of the modified dependency and recovery questions - to classify patients as dependent, independent, and fully recovered - provided a simple and valid indicator of global health related quality of

life. Moreover, this classification had very good reliability which was equivalent to that observed for the assessments of overall health related quality of life with the EuroQol or the assessments of general health with the SF-36. It was also moderately accurate when assessed by proxies. These questions may prove to be an even more feasible means for the assessment of health related quality of life than the EuroQol. However, there would be several disadvantages to replacing generic health related quality of life instrument assessments with the simple questions. In particular, these relate to the loss of generic information, as well as the loss of information about specific domains of outcome, e.g. psychological functioning. The relative feasibility of the EuroQol and the modified simple questions requires examination.

However, a health related quality of life instrument should not be simply selected on the basis of its measurement attributes. The key issue that the choice of measure must take account of the purpose was highlighted in Section 1.6.1. A generic measure of health related quality of life, such as the EuroQoL, is likely to be applied in three major areas after stroke: randomised controlled trials (and other epidemiological research), routine clinical care and medical audit.

10.3 Applications for quality of life data

10.3.1 Randomised controlled trials

This thesis supports the EuroQol as a useful measure of health related quality of life after stroke. It is particularly suitable for use in large randomised controlled trials, audits and screening projects. The data generated by the EuroQol may be presented in a variety of formats. These include as a health status profile, patient derived estimates of overall health related quality of life or as utilities generated from the

health status profiles. Presenting EuroQol derived outcomes as health status profiles (e.g. Patient JD : Mobility - some problems, Self Care - unable to wash \ dress, Activities - unable, Anxiety \ depression - not, Pain - none) is likely to be immediately comprehensible to doctors, patients, and relatives. By contrast, the clinical significance of a reported change of ***x*** points in the overall health related quality of life estimates or health state utility is less clear.

10.3.1.1 Should clinical trialists measure utilities, multidimensional health profiles and/or clinical events?

Summary scores have the advantage of allowing conflicting changes in the individual domains to be combined into a single assessment (e.g. a novel treatment for stroke might improve physical functioning, but cause psychological adverse effects), as well as providing a means by which deaths may be included in the overall analysis. Summary scores, therefore, facilitate the comparison of health gains across different conditions and patient groups. However, this approach may be criticised as it may obscure important differences between individual dimensions (Fletcher *et al.* 1992), and is very dependent on the methodology used to generate the utilities – see Section 1.6.1. Ideally, trialists should report both domain specific and summary data, as they provide different and potentially complementary information (Revicki & Kaplan, 1993).

Several of the leading workers in this area have suggested that the new generic health related quality of life measures should be used alongside disease-specific instruments to provide a generic core of information which could be common to most randomised controlled trials (Ware, 1993; Williams, 1995). Jenkinson and colleagues compared two generic measures (the EuroQol and SF-36) with disease specific measures in a randomised controlled trial of laser vaporization prostatectomy for

patients with benign prostatic hypertrophy (Jenkinson *et al.* 1997). They found a statistically significant improvement in symptoms between baseline and follow up for both groups with the disease specific measures. The SF-36 revealed a small decline in health for those receiving conservative treatment in two of the domains; however, the effect sizes for these changes with the SF-36 were small. In contrast, the EuroQol utilities and self-assessment thermometer indicated no change for either arm in the trial. Several interpretations of these results are possible: firstly, these results could support the view that generic measures are less relevant, and responsive, measures of outcome than disease specific measures. Alternatively, disease specific measures may have too narrow a focus and the changes observed may not be significant in a wider context. Equally, it may be that the clinical changes were too small to have an impact on the valuation procedure used by the EuroQol. This may be a limitation of the crude nature of descriptive system of the EuroQol (a change of one level in one dimension from perfect health is associated with a mean 16% reduction in utility (Jenkinson *et al.* 1997)). However, the responsiveness of an instrument is likely to be specific to the particular characteristics of the patient group involved. For instance, in patients with rheumatoid arthritis, generic measures were as responsive as the disease-specific measures (Hurst et al. 1997). Similarly, in this thesis, although I did not examine the responsiveness of the EuroQol and SF-36 after stroke, I did find strong correlations between the stroke specific measures of outcome and the generic measures of health related quality of life.

10.3.1.2 Whose values count?

A randomised trial of the effect of antihypertensive therapy was one of the first influential trials to use quality of life assessments as a measure of outcome (Croog *et al.* 1986). It compared the effects of captopril, methyldopa and propanolol on blood pressure control and health related quality of life. It assessed outcome in a variety of

dimensions of health: well being and satisfaction with life, physical state, emotional state, intellectual functioning, and ability to perform social roles. The multi-dimensional nature of this data has important implications for its analysis and interpretation. Firstly, any study which examines multiple outcomes runs the risk that, simply by chance, one or more of the outcomes may spuriously appear significant (Cox *et al.* 1992; Guyatt *et al.* 1993). Trialists are therefore encouraged to select and state their primary outcome(s) and main analysis in the protocol and final report (Begg *et al.* 1996). Therefore, for trials using health related quality of life assessments, trialists would need to select and pre-specify certain domains as the primary measure of outcome. However, this is not necessarily straightforward: we have previously shown (in patients with multiple sclerosis) that doctors' assessments of the relative importance of the domains of health related quality of life differed from those of patients; in general, the clinicians were more concerned than the patients about the physical manifestations of disease and the patients were more concerned with less tangible quantities such as mental health and vitality (Rothwell *et al.* 1997).

Other researchers have investigated the differences between doctors and patients when making judgements about quality of life. Slevin and colleagues concluded that if a reliable and consistent method of measuring quality of life is required, it must come from the patients themselves and not from the doctors and nurses (Slevin *et al.* 1988). Therefore, should the selection of primary outcome domain be based on the clinicians' or patients' opinion? Goodare and Smith argue that patients are by definition the best people to advise on which outcomes should be studied (Goodare & Smith, 1995). It does not necessarily follow, however, that the patient's viewpoint is more important, or valid, than that of the doctor. Doctors will usually have a better understanding of the natural course and possible clinical manifestations of a particular disease, and their opinions will be based on the experience of treating many patients. Nevertheless, doctors should bear in mind that their concerns might not agree with

those of their patients, and should work with patients - with mutual respect - when designing clinical trials and selecting outcomes (Goodare & Smith, 1995; Chalmers, 1995).

10.3.2 Routine clinical care

As yet, health related quality of life measures have been used most in a wide range of group-level research applications, and relatively little attention has been given to their potential role in routine medical care and audit. It has been suggested that incorporation of health related quality of life assessments into routine clinical practice could theoretically serve a variety of purposes (Deyo & Carter, 1992; Lohr, 1992), including: describing patients' overall state; screening for unrecognised problems; assessing needs; setting treatment goals; monitoring disease or response to treatment; improving physician-patient communication; and standardising interactions between health care providers and patients.

The notion, that quality of life assessments might be useful in everyday clinical practice, is supported by evidence that clinicians may be poor at, or not interested in, addressing issues that matter to patients (Slevin *et al.* 1988; Pearlman & Uhlmann, 1997; Rothwell *et al.* 1997). Alternatively, some patients may have difficulty broaching specific areas, such as psychosocial concerns, and the routine clinical application of quality of life scales could serve the purpose of opening these areas up for discussion. It has also been suggested that the clinical use of health related quality of life measures might allow essential information to be communicated quickly and precisely (Thier, 1992). However, some patients might experience form filling as distressing, which may have the effect of distancing physicians from their patients (Thier, 1992).

A recent randomised trial examined the efficacy of the routine use of a questionnaire which screened patients' outcomes in the physical, psychological and social domains (Rubenstein *et al.* 1989). Although clinicians found the information useful, the additional information did not appear to change the clinical management or outcome (Rubenstein *et al.* 1989). One possible explanation for this disappointing result may be that the questionnaire did not have the appropriate measurement properties for screening. Health related quality of life measures used for screening need to be evaluated in terms of their sensitivity (false negative results) and specificity (false positive results) (Fitzpatrick *et al.* 1992a).

McHorney and Tarlov have suggested that if health related quality of life instruments are to be used at the individual-patient level, they should meet several measurement standards, (Table 10.1) (McHorney & Tarlov, 1995). Both the EuroQol and SF-36 meet most of these standards in patients with stroke. However, neither instrument meets the reliability standards (Chapter Six). Therefore, on this basis, neither instrument is likely to be useful for monitoring the health of individual patients. However, this assumes that the instruments will be used in isolation. For instance, many individual components of the neurological examination (e.g. assessment of language or motor functioning) also have only moderate reproducibility (Shinar *et al.* 1985), but are widely used (with other assessments) to monitor individual patients.

In Chapter Nine, I found that patient estimates of overall health related quality of life gave a very different conclusion to valuations which were based on the views of the general public. The question of whose valuations should be used exposes tensions in the concept of quality of life assessment. Embedded within this concept, is a commitment to respecting the autonomy of persons to define their own conceptions of a good life. Therefore valuations of health which are derived from the views of the lay public, although appropriate for the purpose of allocating societal resources, may

not be ethically appropriate when applied to individual patients in routine clinical practice.

10.3.3 Medical audit

The combined pressures of "medical inflation", fiscal constraints, and a shift in attitudes to publicly funded services during the 1980s has made the search for measures of quality, efficiency and effectiveness in health care a government priority (Orchard, 1994). Whilst it has been suggested that three components of the health care system may be audited, the structure, process and outcome (Donabedian, 1988), most emphasis is presently placed on the measurement of outcome.

Of the available measures of outcome, case-fatality statistics are the most widely used (Clinical Outcomes Working Group, 1995). However, as death is only one aspect of the public health burden attributable to stroke (Chapter One), there is some interest in extending the scope of routinely collected data to include aspects of health related quality of life (Working Group on Outcome Indicators for Stroke, 1997). Indeed, the SF-36 resulted from the Medical Outcomes Study, which aimed to develop practical tools for the routine monitoring of patient outcome in clinical practice (Tarlov *et al.* 1989).

Variations in case mix have a crucial influence on the interpretation of outcome data, particularly where such data are used to compare providers (Orchard, 1994). Therefore, to allow meaningful interpretation of these data, audits must try to correct for case mix. Davenport and colleagues have demonstrated how data on survival and disability after stroke could be adjusted for variations in case mix using simple clinical variables measured routinely at the time of the stroke (Davenport *et al.* 1996). Unless health related quality of life data can also be successfully adjusted for variations in

case mix, it is unlikely to be useful for the comparison of provider units (although it may still provide helpful information about local healthcare needs).

10.4 Weaknesses of the study design

This study has several limitations, some of which have been already discussed in the relevant chapters. However, there are three areas that require further discussion. They are; the selection of instruments for the study, the patient selection methods and the possibility of bias.

Perhaps the most fundamental limitation relates to how I selected the health related quality of life instruments for this study. I selected the EuroQol because, in my opinion, it had the best face and content validity of the currently available generic health status measures for the assessment of health status in a large randomised controlled trial. Although I discussed the suitability of the instrument informally with the local stroke research group, I did not quantitatively assess its face and content validity. Moreover, I did not ask stroke survivors, or their carers, to examine its content, relevance or clarity. The importance of seeking the views of patients when selecting trial outcomes has been recently highlighted (Goodare & Smith, 1995).

The methods used to select the patients included in the study might have affected the generalisability of its results. I selected a "convenience sample" of patients on our local hospital stroke register for the study of the validity of the EuroQol, Chapters Three & Four. This group represents a selected population of stroke patients in that it only includes hospital referred patients who survived long enough to fulfil the study eligibility criteria. An "ideal study" might have assessed the validity of the EuroQol in a community-based cohort. The population used for the study of the feasibility and reproducibility of the EuroQol was even more selected in that it only included

surviving patients who had been randomised in the International Stroke Trial. This group is likely to include less than 10% of all hospital referred stroke patients. This is not a major weakness, however, as the primary aim of the study was to examine the attributes of the instruments in a population of patients who might be included in a large randomised trial.

Finally, the study of the validity of the EuroQol (Chapters Three & Four) might have been biased. The study nurse, who conducted the assessments by interview in the patients' homes, was aware of the aims of the study and could have influenced the patients' responses. However, it is unlikely that this has been a major source of bias, as the data from the postal administration of the EuroQol also supported its validity in patients with stroke.

10.5 Future research possibilities

Although the studies described in this thesis are complete, some of the work performed may be used as a foundation for future work. One of the aims of the study was to examine the validity of the EuroQol and compare it with that of the SF-36. One aspect of validity not examined was the predictive validity of either instrument. This aspect of validity refers to the extent to which patients' responses to a scale predict future important events, e.g. do patients' responses to the multiple domains of health related quality of life allow a better prediction of future events than current predictors (e.g. disability alone)? Assessments of predictive validity have rarely been performed because they demand a long period of follow-up. However, the large number of health related quality of life assessments performed in this study could serve as baseline assessments for a subsequent study of the predictive validity of the EuroQol, SF-36, and modified simple questions after stroke.

The studies performed in this thesis were guided by the need to identify the "best" instrument for the measurement of health related quality of life in a planned randomised controlled trial in patients with stroke. Throughout the thesis, I have highlighted some important issues which still need to be addressed. Firstly, it is unclear whether simple measures of health related quality of life (e.g. the EuroQol) are sufficiently responsive to detect clinically important change. Future research studies should examine the sensitivity to change of these measures. Secondly, the natural history of quality of life after stroke has not been well defined and further study is required to inform trialists about when it should be measured. Finally, the development of accurate models for the prediction of health related quality of life after stroke would serve a number of useful purposes, such as case-mix adjustment (see above) and the identification of patients at presentation who might be accepting of risky treatments (see Chapter Nine).

In conclusion, the work described in this thesis adds new and significant information to the current understanding of how health related quality of life assessments should be applied in clinical studies in patients with stroke.

Table 10.1: Proposed measurement standards for the individual-patient use of health related quality of life measures (adapted from (McHorney & Tarlov, 1995))

Criterion	Proposed standard
Practicality	5-15 minutes, and ease of scoring
Breadth of health assessed	Physical, role, social, and mental domains
Floor effects	<15%
Ceiling effects	<15%
Precision (cross-sectional)	0.90 - 0.95 reliability
Precision (longitudinal)	0.90 - 0.95 reliability
Validity	Group level
	Individual level

Appendix 1: The LSR Registration Form

Personal details

Study No. | | | | | | |

*Please PRINT all details in **BLACK** ink*
Use reverse for details or narrative

WGH Hospital No. WG0 _____ **NHS no.** _____

DCN X-ray no. | | | / | | | | | |

Surname _____ Title _____

Forenames _____

Address _____

_____

Postcode _____ Tel. No. _____

Date of birth / / Sex **M** / **F** *(circle)*

Next of kin / Contact person _____

Address _____

_____

Postcode _____ Tel. No. _____

General Practitioner _____

Address _____

_____

Postcode _____ Tel. No. _____

Patient of interest ? **Y** / **N** **Reasons :**
(circle)

(continue overleaf)

Admission details

Name : _____ No. : | | | | | | |

Inpatient : **Y / N** *(circle)* Time ____:____ Date ____/____/____ of admission

Time is by 24 hour clock, dates are dd/mm/yy *(if known)* Date ____/____/____ of discharge

Consultant *(circle)* JLA / REC / MSD / RG / LK / CL / CM°K / JDM / PLP / WHP / TR / PAGS / RS / PFXS / AJS / CPW / DW / IW / RW / other :

Time ____:____ Date ____/____/____ of examination

Examined by *(circle)* RJD / MSD / PD / PAGS / CPW / other :

Summary of this event

Focus of event : **Brain / Eye** *(circle)* Abnormal neurological signs on examination : **Y / N** *(circle)*

Final diagnosis

Stroke (> 24 h) | |

Transient Ischaemic Attack (< 24 h) | |

Retinal Artery Occlusion *(RAO)* | |

Other *(specify):* | |

Code 1 = possible (NOT permitted for RAO)

2 = probable (NOT permitted for RAO)

3 = definite

9 = not applicable

Include events within the last 6 months only. Leave no blanks.

Patient history

Code boxes 1 = Yes, 2 = No, 9 = unassessable, Blank = may be completed later.

Patient able to give adequate history		
Previous Myocardial Infarction		Year *(if known)*
Previous stroke with residual disability		Year *(if known)*
Previous stroke without residual disability		Year *(if known)*
Previous TIA *(specify territories in narrative)*		
Previous carotid endarterectomy		Year *(if known)*
(Code side of CEA 1 = R, 2 = L, 3 = both, 9 = not known)		Side *(if known)*
Hypertension *(history or treatment at any time)*		Non-caucasian *(specify):*
Angina pectoris known before stroke		Alcohol > 2 units daily
Atrial fibrillation known before stroke		Current smoker
Breathless walking on an incline		Ex-smoker > 12 months
Cardiac surgery *(specify):*		Employed until this event
Intermittent claudication		Car driver in past 3 months
Peripheral vascular surgery		Lives alone
Diabetes mellitus known before stroke		Known prior malignancy
Epilepsy known before stroke		
History of migraine with aura		
Oxford Handicap Scale before stroke *(Modified Rankin Scale)*		

Oxford Handicap Scale:

0 = no symptoms

1 = minor symptoms which do not interfere with lifestyle

2 = some restriction to lifestyle, but look after themselves

3 = significant restriction to lifestyle, preventing total independence

4 = severe handicap preventing independent existence but not requiring constant attention

5 = severe handicap, totally dependent, requiring attention night and day

Treatment

Code boxes 1 = Yes, 2 = No, 9 = unassessable, Blank = may be completed later.

	At time of event	Started since event
Antiplatelet		
Anticoagulant		
Antihypertensive		
Anticonvulsants *(if history of epilepsy)*		
Antifailure		
Contra-indications to antithrombotics		

List all drugs in use at examination (in narrative) :

General Examination

Blood pressure

(admission) _____ / _____

(examination) _____ / _____

Cervical Bruits *(specify)* — R L

Noted by referring doctor

Seen at examination

Clinical heart failure *(ie signs of LVF / RVF, not just on Rx, specify)*

Clinical valvular heart disease *(not simple flow murmur < 2/6, specify)*

Peripheral vascular disease *(both foot pulses absent or femoral bruits)*

Irregular pulse

Right handed

Left handed

Brain Symptoms > 24 hours

Skip for patients NOT exhibiting brain symptoms longer than 24h

History of ictus

Time ____:____ Date ____/____/____ symptoms first noticed

Time is by 24 hour clock, dates are dd/mm/yy

Time ____:____ Date ____/____/____ of maximum deficit

Symptoms present on waking

Headache within 2 hours of onset

Vomited since symptom onset

Loss of consciousness at onset

Drowsiness since symptom onset

Seizure(s) since symptom onset

- date ____/____/____ of first
- confirmed seizure ?
- type *(1 = General, 2 = Partial, 9 = uncertain)*
- number *(Use 9 for 9 or more)*

Mental Test Score *(Hodkinson, tick below, score 0-10)*

Age

Time

42 West St. *(ask patient to recall at end)*

Name of Hospital

Year

Recognise 2 people *(eg. Dr. and Nurse)*

Date of birth

Dates of World War I or II

Present Monarch

Count down from 20 to 1

Total

(Code '88' if clinically unassessable)

Stroke diagnosis

Side of brain lesion *(one only)*

1 = right
2 = left
3 = brainstem / cerebellum
4 = uncertain
5 = bilateral

Clinical classification *(one only)*

1 = TACS
2 = PACS
3 = LACS
4 = POCS
5 = uncertain

Clinical prediction of outcome at 1 year *(0 - 6 on Rankin Scale)*

Persistent Neurological Signs

Skip this page for patients NOT exhibiting neurological signs at examination and skip for patients NOT exhibiting brain symptoms longer than 24h

Glasgow Coma Scale *(circle below, score 3-15)*

Eye Opening -	Never	1
	To pain	2
	To sound	3
	Spontaneously	4
Best Motor -	None	1
	Extend to pain	2
	Abn flex to pain	3
	Flex to pain	4
	Localises pain	5
	Normal	6
Best verbal -	None	1
	Noises only	2
	Inappropriate	3
	Confused	4
	Normal	5
Total		

(circle)

Dysphasia ☐ Fluent / Non-fluent / Other *(specify)* :

Dysarthria ☐

(circle)

Other cortical signs ☐ Dyspraxia / Neglect / Sensory inattention / Visuospatial dysfunction

Code boxes 1 = Yes, 2 = No, 9 = unassessable, Blank = may be completed later.

	R	L
Hemianopia	☐	☐
Visual inattention	☐	☐
Gaze palsy to this side	☐	☐
Abnormal swallowing	☐	
Motor deficit *(if yes, code next column)*	☐ If 1	
Sensory deficit *(if yes, code next column)*	☐ If 1 →	
Cerebellar deficit *(if yes, code next column)*	☐ If 1	
Truncal ataxia	☐	
Unable to sit independently	☐	
Unable to stand independently	☐	
Unable to walk independently	☐	
Incontinence since stroke	☐	
Bilateral extensor plantars	☐	
Neck stiffness	☐	
Definite brainstem signs	☐	

Deficit Severity Codes R L

Motor deficit code:

1 = no deficit, 2 = mild, 3 = moderate, 4 = severe

	R	L
Face	☐	☐
Arm	☐	☐
Drift	☐	☐
Hand	☐	☐
Fine finger movements	☐	☐
Leg	☐	☐

Sensory and cerebellar abnormalities code:

1 = normal, 2 = reduced, 3 = severely impaired / absent

Sensation - proprioception

Arm / hand	☐	☐
Leg	☐	☐

Sensation - spinothalamic (pain and touch)

Face	☐	☐
Arm / hand	☐	☐
Leg	☐	☐

Cerebellar function and co-ordination

Arm	☐	☐
Leg	☐	☐

Eye Symptoms / Brain Symptoms lasting < 24h

Skip this page for patients ONLY exhibiting brain symptoms lasting longer than 24 hours

Code sides : 2 = Probable, 3 = Definite, 9 = none. Probable is NOT accepted for RAO.

Type			Date of first	Date of last	Duration of longest (hh:mm)	Total number
RAO	R		___/___/___	___/___/___	n/a	n/a
	L		___/___/___	___/___/___	n/a	n/a
A Fx	R		___/___/___	___/___/___	:	
	L		___/___/___	___/___/___	:	
Cortical	R		___/___/___	___/___/___	:	
	L		___/___/___	___/___/___	:	
LACS	R		___/___/___	___/___/___	:	
	L		___/___/___	___/___/___	:	
POCS	R		___/___/___	___/___/___	:	
	L		___/___/___	___/___/___	:	
	M		___/___/___	___/___/___	:	

Registration

Code boxes 1 = Yes, 2 = No, Blank = may be completed later.

Enter patient into Register ☐

Enter patient into Follow up ☐

Trials

	Eligible	Randomised
IST	☐	☐
IST 2	☐	☐
CAPRIE	☐	☐
MAST	☐	☐
NASCET	☐	☐
CAVATAS	☐	☐

Studies

MRS	☐
ULTRASOUND	☐
SECONDARY INSULTS	☐
PICH	☐
Other CRI imaging	☐

Investigations

Test	Ordered	Date done	Results	
Haemoglobin	☐	___/___/___	☐☐●☐	g/dl
Haematocrit	☐	___/___/___	☐☐●☐	%
Platelets	☐	___/___/___	☐☐☐☐	x 10^9/l
ESR	☐	___/___/___	☐☐☐☐	mm/hr
Urea	☐	___/___/___	☐☐●☐	mmol/l
Glucose	☐	___/___/___	☐☐●☐	mmol/l
Cholesterol	☐	___/___/___	☐☐●☐☐	mmol/l

ECG	☐	___/___/___
Doppler *(but not in study)*	☐	___/___/___
Trans-thoracic echocardiogram	☐	___/___/___
Trans-oesophageal echocardiogram	☐	___/___/___
CT	☐	___/___/___

If 1 →

Atrial fibrillation	☐
Bundle branch block	☐
ST segment change	☐
LVH	☐
Acute MI	☐
Old MI	☐
Normal	☐

Appendix 2: The EuroQol Questionnaire Booklet

HEALTH OUTCOMES SURVEY

Confidential

If any of these details are incorrect, please change below:

_____

_____

_____

You were recently admitted to hospital, and we would like to know how you are now. We need to know what you have **actually managed** to do since leaving hospital, not what you used to do, or would like to do.

Please tick one box on each line

	YES	NO
Has the stroke left you with any problems?	☐	☐
Do you need help **from anybody** with everyday activities?	☐	☐

Do you live? *(please tick one box)*

On your own	☐
With your partner or relatives	☐

Where do you live? *(please tick one box only)*

In your own home	☐
In a residential home	☐
In a nursing home	☐
In hospital	☐

Please turn over and fill in the following questions ➔

By placing a tick (✓) in one box in each group below, please indicate which statements best describe your own health state today.

Mobility

I have no problems in walking about ☐

I have some problems in walking about ☐

I am confined to bed ☐

Self-Care

I have no problems with self care ☐

I have some problems with washing or dressing myself ☐

I am unable to wash or dress myself ☐

Usual Activities

I have no problems with performing my usual activities ☐
(eg work, study, housework, family or leisure activities)

I have some problems with performing my usual activities ☐

I am unable to perform my usual activities ☐

Pain/discomfort

I have no pain or discomfort ☐

I have moderate pain or discomfort ☐

I have extreme pain or discomfort ☐

Anxiety/depression

I am not anxious or depressed ☐

I am moderately anxious or depressed ☐

I am extremely anxious or depressed ☐

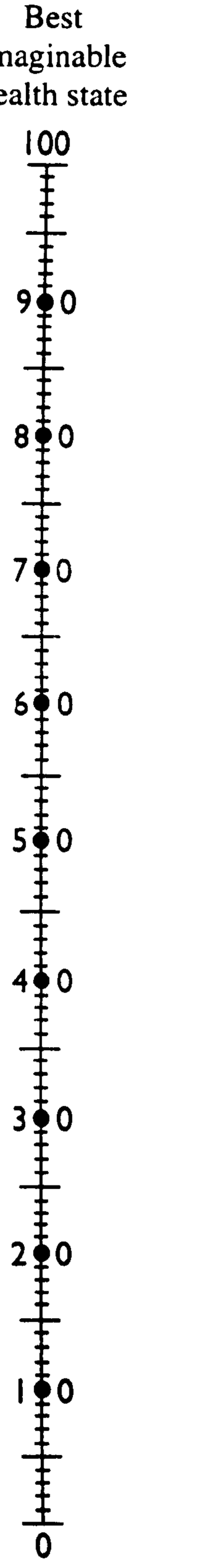
Compared with my general level of health over the past 12 months, my health state today is: ***(please tick one box)***

Better ☐ Much the same ☐ Worse ☐

To help people say how good or bad a health state is, we have drawn a scale (rather like a Thermometer) on which the best state you can imagine is marked by 100 and the worst state you can imagine is marked by 0.

We would like you to indicate on this scale how good or bad is your health today, in your opinion. Please do this by drawing a line from the box below to whichever point on the scale indicates how good or bad your current health state is.

Your own health state today



Please turn over and fill in the following questions →

Did you find filling in this questionnaire: (*please tick one box*)

Very Difficult ☐ Fairly Difficult ☐ Fairly Easy ☐ Very Easy ☐

Did you complete this form yourself? *(please tick one box)*

YES ☐ NO, it was completed for me by a relative or friend ☐

Name of person filling in form _____

Signature _____ Date _____

Thank you very much for your help with this survey. Please return this questionnaire using the FREEPOST envelope enclosed (no stamp is required).

Appendix 3: The SF-36 questionnaire booklet

HEALTH OUTCOMES SURVEY

Confidential

If any of these details are incorrect, please change below:

_____

_____

_____

You were recently admitted to hospital, and we would like to know how you are now. We need to know what you have **actually managed** to do since leaving hospital, not what you used to do, or would like to do.

Please tick one box on each line

	YES	NO
Has the stroke left you with any problems?	☐	☐
Do you need help **from anybody** with everyday activities?	☐	☐

Do you live? *(please tick one box)*

On your own ☐

With your partner or relatives ☐

Where do you live? (*please tick one box only*)

In your own home ☐

In a residential home ☐

In a nursing home ☐

In hospital ☐

Please turn over and fill in the following questions →

Now you have completed the first page, we would like to ask 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 number)*

Excellent	Very Good	Good	Fair	Poor
1	2	3	4	5

2. Compared to one year ago, how would you rate your health in general now? *(circle one number)*

Much better than 1 year ago	**1**
Somewhat better now than 1 year ago	**2**
About the same as 1 year ago	**3**
Somewhat worse than 1 year ago	**4**
Much worse than 1 year ago	**5**

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 **half a mile**	1	2	3
i. Walking **one hundred yards**	1	2	3
j. Bathing or dressing yourself	1	2	3

4. During the <u>past 4 weeks</u>, have you had any of the following problems with your work or other regular daily activities <u>as a result of your physical health</u>? *(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

5. During the <u>past 4 weeks</u>, have you had any of the following problems with your work or other regular daily activities <u>as a result of any emotional problems</u> (such as feeling depressed or anxious)? (*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

6. During the <u>past 4 weeks</u>, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbours, or groups? (*circle one number*)

Not at all	Slightly	Moderately	Quite a bit	Extremely
1	2	3	4	5

7. How much <u>bodily</u> pain have you had during the <u>past 4 weeks</u>? (*circle one number*)

None	Very Mild	Mild	Moderate	Severe	Very Severe
1	2	3	4	5	6

8. During the <u>past 4 weeks</u>, how much did <u>pain</u> interfere with your normal work (including both work outside the home and housework)? (*circle one number*)

Not at all	A little bit	Moderately	Quite a bit	Extremely
1	2	3	4	5

9. These questions are about how you feel and how things have been with you <u>during the past 4 weeks</u>. 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 <u>past 4 weeks</u>. (*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 life?	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 low?	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

10. During the <u>past 4 weeks</u>, how much of the time has your <u>physical health or emotional problems</u> interfered with your social activities (like visiting friends, relatives etc)? *(circle one number)*

All of the time	Most of the time	Some of the time	A little of the time	None of the time
1	2	3	4	5

11. Please choose the answer that best describes how TRUE or FALSE <u>each</u> of the statements is for you. *(circle one number on each line)*

	Definitely True	Mostly True	Don't Know	Mostly False	Definitely False
a. I seem to get ill more easily 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

Did you find filling in this questionnaire: (*please tick one box*)

Very Difficult ☐ Fairly Difficult ☐ Fairly Easy ☐ Very Easy ☐

Did you complete this form yourself? *(please tick one box)*

YES ☐ NO, it was completed for me by a relative or friend ☐

Name of person filling in form _____

Signature _____ Date _____

Thank you very much for your help with this survey. Please return this questionnaire using the FREEPOST envelope enclosed (no stamp is required).

Appendix 4: The OPCS Locomotion Subscale

L1	Cannot walk at all	11.5
L2	Can only walk a few steps without stopping or severe discomfort / Cannot walk up and down one step	9.5
L3	Has fallen 12 or more times in the last year	7.5
L4	Always needs to hold on to something to keep balance	7.0
L5	Cannot walk up and down a flight of 12 stairs	6.5
L6	Cannot walk 50 yards without stopping or severe discomfort	5.5
L7	Cannot bend down far enough to touch knees and straighten up again	4.5
L8	Cannot bend down and pick something up from the floor and straighten up again	4.0
L9	Cannot walk 200 yards without stopping or severe discomfort/ Can only walk up and down a flight of twelve stairs if holds on and takes a rest/ Often needs to hold on to something to keep balance / Has fallen 3 or more times in the last year	3.0
L10	Can only walk up and down a flight of 12 stairs if holds on (doesn't need a rest)	2.5
L11	Cannot bend down to sweep up something from the floor and straighten up again	2.0
L12	Can only walk up and down a flight of stairs if goes sideways or one step at a time	1.5
L13	Cannot walk 400 yards without stopping or severe discomfort	0.5

Appendix 5: The Barthel Index

What the patient ACTUALLY DOES ?

1) FEEDING	Independent Needs some help Needs to be fed	= 2 = 1 = 0	☐
2) BATHING	Able to wash all over Needs help	= 1 = 0	☐
3) GROOMING	Totally independent Dependent in some way	= 1 = 0	☐
4) DRESSING	Totally independent Needs help with some items Unable to do any without help	= 2 = 1 = 0	☐
5) BOWELS	No accidents Occasional accident / help with enema Incontinent	= 2 = 1 = 0	☐
6) BLADDER	No accidents Occasional accident / use of device Incontinent, or catheterized and unable to manage alone	= 2 = 1 = 0	☐
7) TOILET	Independent (on/off dressing and wiping) Minor assistance Unable to use	= 2 = 1 = 0	☐
8) TRANSFER	Totally independent Minimal help needed Sit unaided, major help for transfer Unable	= 3 = 2 = 1 = 0	☐
9) AMBULATION	Independent Walk with the help of one person Independent in wheelchair for 50 metres Immobile	= 3 = 2 = 1 = 0	☐
10) STAIRS	Independent Needs physical / verbal support Unable	= 2 = 1 = 0	☐

TOTAL: ☐☐

Appendix 6: The Frenchay Activities Index

In the last 3 months how often have you been : -

PF1) Preparing the main meal?

☐	☐	☐	☐
Never	Less than once p/wk	Once or twice p/wk	Most days

PF2) Washing up?

☐	☐	☐	☐
Never	Less than once p/wk	Once or twice p/wk	Most days

PF3) Washing clothes?

☐	☐	☐	☐
Never	Once or twice every 3 months	Between 3 & 12 times every 3 months	At least weekly

---

PF4) Doing light housework?

☐	☐	☐	☐
Never	Once or twice every 3 months	Between 3 & 12 times every 3 months	At least weekly

---

PF5) Doing heavy housework?

☐	☐	☐	☐
Never	Once or twice every 3 months	Between 3 & 12 times every 3 months	At least weekly

---

In the last 3 months how often have you been : -

PF6) Local shopping?

Never	Once or twice every 3 months	Between 3 & 12 times every 3 months	At least weekly

---

PF7) On Social Outings?

Never	Once or twice every 3 months	Between 3 & 12 times every 3 months	At least weekly

---

PF8) Walking outside for up to 15 minutes?

Never	Once or twice every 3 months	Between 3 & 12 times every 3 months	At least weekly

---

PF9) Actively pursuing a hobby?

Never	Once or twice every 3 months	Between 3 & 12 times every 3 months	At least weekly

---

PF10) Driving a car or traveling on a bus?

Never	Once or twice every 3 months	Between 3 & 12 times every 3 months	At least weekly

---

In the last 6 months have you been : -

PF11) On any outings / car rides?

☐	☐	☐	☐
Never	Once or twice every 3 months	Between 3 & 12 times every 3 months	At least weekly

---

PF12) Gardening?

☐	☐	☐	☐
Never	Light	Moderate	All that is Necessary

PF13) Doing household or car maintenance?

☐	☐	☐	☐
Never	Light	Moderate	All that is Necessary

PF14) Reading books?

☐	☐	☐	☐
None	1 every 6 months	Less than 1 every 2 wks	More than1 every 2 wks

---

PF15) Gainful work?

☐	☐	☐	☐
None	Up to 10 Hours p/wk	Between 10 & 30 hours per week	Over 30 hours per week

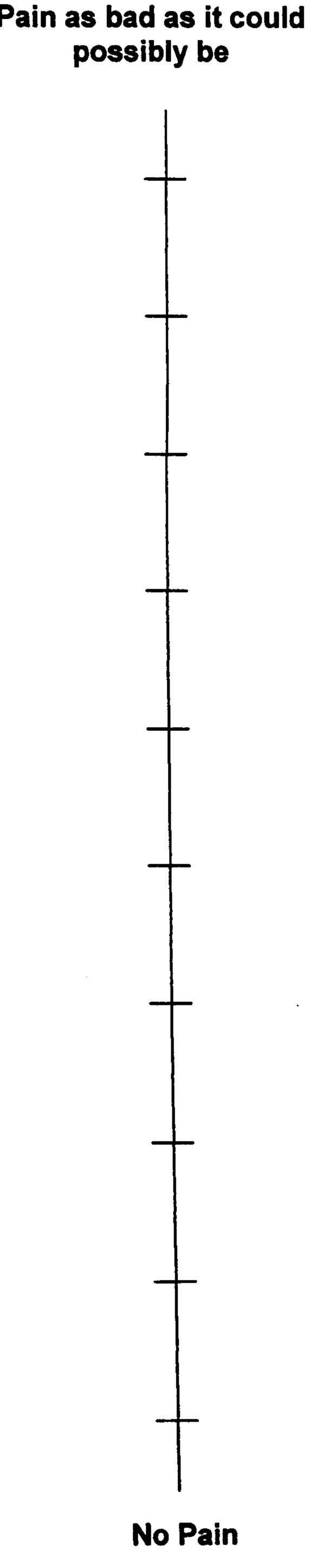
---

Appendix 7: The visual analogue pain scale

To help people say how bad their pain is, we have drawn a scale on which the pain experienced ranges from "no pain" to "pain as bad as it could possibly be".

We would like you to indicate on this scale how bad your pain is today. Please do this by drawing a line from the box below to whichever point on the scale indicates how bad your pain is.

Your pain today



Appendix 8: The Hospital Anxiety and Depression Scale

Read each item and place a firm tick in the box opposite the reply which comes closest to how you have been feeling in the past week.
Don't take too long over your replies: your immediate reaction to each item will probably be more accurate than a long thought out response.

Please answer every question.

PH1) **I feel tense or 'wound up':**

- ☐ Most of the time
- ☐ A lot of the time
- ☐ Time to time, occasionally
- ☐ Not at all

PH2) **I still enjoy the things I used to enjoy:**

- ☐ Definitely as much
- ☐ Not quite so much
- ☐ Only a little
- ☐ Hardly at all

PH3) **I get a sort of frightened feeling as if something awful is about to happen.**

- ☐ Very definitely and quite badly
- ☐ Yes, but not too badly
- ☐ A little, but it doesn't worry me
- ☐ Not at all

PH4) **I can laugh and see the funny side of things:**

- ☐ As much as I always could
- ☐ Not quite so much now
- ☐ Definitely not so much now
- ☐ Not at all

Please answer every question.

PH5) **Worrying thoughts go through my mind**

- ☐ A great deal of the time
- ☐ A lot of the time
- ☐ From time to time, but not too often
- ☐ Only occasionally

PH6) **I feel cheerful**

- ☐ Not at all
- ☐ Not often
- ☐ Sometimes
- ☐ Most of the time

PH7) **I can sit at ease and feel relaxed:**

- ☐ Definitely
- ☐ Usually
- ☐ Not often
- ☐ Not at all

PH8) **I feel as if I am slowed down:**

- ☐ Nearly all the time
- ☐ Very often
- ☐ Sometimes
- ☐ Not at all

PH9) **I get a sort of frightened feeling like 'butterflies' in the stomach:**

- ☐ Not at all
- ☐ Occasionally
- ☐ Quite often
- ☐ Very often

Please answer every question.

PH10) **I have lost interest in my appearance:**

- ☐ Definitely
- ☐ I don't take as much care as I should
- ☐ I may not take as much care
- ☐ I take just as much care as ever

PH11) **I feel restless as if I have to be on the move:**

- ☐ Very much indeed
- ☐ Quite a lot
- ☐ Not very much
- ☐ Not at all

PH12) **I look forward with enjoyment to things:**

- ☐ As much as I ever did
- ☐ Rather less than I used to
- ☐ Definitely less than I used to
- ☐ Hardly at all

PH13) **I get sudden feelings of panic:**

- ☐ Very often indeed
- ☐ Quite often
- ☐ Not very often
- ☐ Not at all

PH14) **I can enjoy a good book or radio or TV programme:**

- ☐ Often
- ☐ Sometimes
- ☐ Not often
- ☐ Very seldom

Appendix 9: My contribution to the work in this thesis

The collection and analyses of the data in this thesis were carried out in the Neurosciences Trials Unit, whilst I was working as a Medical Research Council Clinical Training Fellow in the Department of Clinical Neurosciences in the University of Edinburgh (between December 1994 and September 1997). I was under the supervision of Dr Peter Sandercock and Dr Martin Dennis. Ethical approval was obtained for the International Stroke Trial and the Lothian Stroke Register. We were advised by the Chairman of the Lothian Ethical Committee that additional ethical approval was not required for the work contained within this thesis.

My contribution to the thesis was as follows:

From December 1994 to October 1995, I was the principal assessor for the Lothian Stroke Register. In total, I registered over 200 patients and assessed many more. I also personally randomised over thirty patients into the International Stroke Trial.

I was responsible for the design of all of the studies described in the thesis, writing the protocol for these studies, and obtaining the financial support to fund this work.

I managed the conduct of all the studies. I designed all the study materials and the specifications for the study databases and software. Mr David Perry and Mr Gary Robertson developed the software for data management in these studies.

I organised and participated in the posting and the punching the questionnaire data.

Mrs Fiona Waddell conducted the face to face interviews.

I analysed all the data, apart from the conditional logistic regression in Chapter Nine which was performed by Mr David Signorini. Statistical advice was provided by Mr Jim Slattery and Mr David Signorini. Control patients were obtained from the Measurement and Valuation of Health and Omnibus Surveys, Chapter Nine. They were selected by Dr Paul Kind, Centre for Health Economics, University of York.

Appendix 10: Publications arising from work within this thesis

Papers

Dorman PJ, Slattery JM, Farrell B, Dennis MS, Sandercock PAG and the United Kingdom Collaborators in the International Stroke Trial (IST).(1997)
A randomised comparison of the EuroQol and SF-36 after stroke.
British Medical Journal. **315**: 461 (short report).

Dorman PJ, Waddell F, Slattery J, Dennis MS & Sandercock PAG (1997)
Is the EuroQol a valid measure of health related quality of life after stroke?
Stroke. **28**: 1876-1882.

Dorman PJ, Waddell F, Slattery J, Dennis MS & Sandercock PAG (1997)
Are proxy assessments of health status after stroke with the EuroQol questionnaire feasible, accurate and unbiased?
Stroke. **28**: 1883-1887.

Dorman PJ, Slattery JM, Farrell B, Dennis MS, Sandercock PAG and the United Kingdom Collaborators in the International Stroke Trial (IST).(1998)
A qualitative comparison of the reliability of health status assessments with the EuroQol and SF-36 after stroke.
Stroke. 29: 63-68.

Presentations to learned societies

Dorman P, Farrell B, Dennis MS, & Sandercock PAG (1996)
Health-related quality of life after stroke: a randomised comparison of the EuroQol and SF-36 questionnaires in 2,252 survivors of acute stroke.
Cerebrovascular Diseases. 6(S2): 152.
Presented at the Joint 3rd World Stroke Congress and 5th European Stroke Conference, Munich, 1-4 September 1996.

Dorman PJ, Dennis MS & Sandercock PAG (1997)
Stroke trials: do patients prefer death or disability?
Joint Meeting of the Australian Association of Neurologists and the Association of British Neurologists, Sydney, Australia, 29 April - 2 May 1997.

Dorman PJ, Waddell F, Slattery J, Dennis MS & Sandercock PAG (1997)
Is the EuroQol a valid measure of health related quality of life after stroke?
6th European Stroke Conference, Amsterdam, The Netherlands, May 28-31,1997.
Cerebrovascular Diseases 1997;7(S4):10

References

Aaronson, N.K. (1992) Assessing the quality of life of patients in cancer clinical trials: common problems and common sense solutions. *European Journal of Cancer* **28**, pp 1304-1307.

Adkins, E.R. (1993) Quality of life after stroke: exposing a gap in nursing literature. *Rehabilitation Nursing* **18** , pp 144-147.

Ahlsio, B., Britton, M., Murray, V. & Theorell, T. (1984) Disablement and quality of life after stroke. *Stroke* **15**, pp 886-890.

Altman, D.G. (1993) *Practical statistics for medical research*, First edn. London: Chapman & Hall.

Anderson,R. (1992) *The Aftermath of Stroke*. Cambridge: Cambridge University Press.

Anderson, C., Laubscher, S. & Burns, R. (1996) Validation of the Short Form 36 (SF-36) health survey questionnaire among stroke patients. *Stroke* **27**, pp 1812-1816.

Anderson, C.S., Taylor, B.V., Hankey, G.J., Stewart-Wynne, E.G. & Jamrozik, K.D. (1994) Validation of a clinical classification for subtypes of acute cerebral infarction. *J Neurol Neurosurg Psychiatry* **57**, pp 1173-1179.

Anderson, R.T., Aaronson, N.K. & Wilkin, D. (1993) Critical review of the international assessments of health related quality of life. *Quality of Life Research* **2**, pp 369-395.

Andresen, E.M., Bowley, N., Rothenberg, B.M., Panzer, R. & Katz, P. (1996) Test-retest performance of a mailed version of the Medical Outcomes Study 36-Item short-form health survey among older adults. *Medical Care* **34**, pp 1165-1170.

Angeleri, F., Angeleri, V.A., Foschi, N., Giaquinto, S. & Nolfe,G. (1993) The influence of depression, social activity, and family stress on functional outcome after stroke. *Stroke* **24**, pp 1478-1483.

Anonymous (1902) A fund for the study of chronic disease. *Boston Medical and Surgical Journal* **146**, pp 150-151.

Anonymous (1800a) Report of diseases in an Eastern district of London. *Monthly Magazine* **9**, pp 67-69.

Anonymous (1800b) Report of diseases in Finsbury district. *Monthly Magazine* **10**, pp 79-80.

Astrom, M., Asplund, K. & Astrom, T. (1992) Psychosocial function and life satisfaction after stroke. *Stroke* **23**, pp 527-531.

Bamford, J., Sandercock, P., Dennis, M., et al (1988) A prospective study of acute cerebrovascular disease in the community: the Oxfordshire Community Stroke Project 1981-1986.1. Methodology, demography and incident cases of first ever stroke. *J Neurol Neurosurg Psychiatry* **51**, pp 1373-1380.

Bamford, J., Sandercock, P., Dennis, M., Burn, J. & Warlow, C. (1990) A prospective study of acute cerebrovascular disease in the community: the Oxfordshire Community Stroke Project 1981-1986.2. Incidence, case fatality and overall outcome at one year of cerebral infarction, primary intracerebral haemorrhage and subarachnoid haemorrhage. *J Neurol Neurosurg Psychiatry* **53**, pp 16-22.

Bamford, J., Sandercock, P., Dennis, M., Burn, J. & Warlow, C. (1991) Classification and natural history of clinically identifiable subtypes of cerebral infarction. *Lancet* **337**, pp 1521-1526.

Bamford, J.M., Sandercock, P.A.G., Warlow, C.P. and Slattery, J. (1989) Interobserver agreement for the assessment of handicap in stroke patients. *Stroke* **20** 828

Barham Carter, A. (1961) Anticoagulant treatment in progressing stroke. *BMJ* pp 70-73.

Begg, C., Cho, M., Eastwood, S., et al (1996) Improving the quality of reporting of randomized controlled trials. *JAMA* **276**, pp 637-639.

Bergner, M. (1985) Measurement of health status. *Medical Care* **23**, pp 696-704

Bergner, M. (1989) Quality of life, health status, and clinical research. *Medical Care* **27 (suppl)**, pp S148-S156.

Bland, J.M. & Altman, D.G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* pp 307-310.

Blaxter, M. (1990) *Health & Lifestyles*. London: Routledge.

Bowling, A. (1995) What things are important in people's lives? A survey of the public's judgements to inform scales of health related quality of life. *Soc Sci Med* **41**, pp 1477-62

Bowling, A. (1996) The effects of illness on quality of life: findings from a survey of households in Great Britain. *J Epidemiol Community Health* **50**, pp 149-155.

Brazier, J., Jones, N. & Kind, P. (1993) Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnaire. *Quality of Life Research* **2**, pp 169-180.

Brazier, J.E., Harper, R., Jones, N.M.B., et al (1992) Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ* **305**, pp 160-164.

Brazier, J.E., Walters, S.J., Nicholl, J.P. & Kohler, B. (1996) Using the SF-36 and EuroQol on an elderly population. *Quality of Life Research* **5**, pp 195-204.

Brennan, P. & Silman, A. (1992) Statistical methods for assessing observer variability in clinical measures. *BMJ* **304**, pp 1491-1494.

Brooks, R.G. (1991). *Health Status and Quality of Life Measurement*, Lund, IHE: Swedish Institute for Health Economics

Brooks, R. & with the EuroQol group (1996) EuroQol: the current state of play. *Health Policy* **37**, pp 53-72.

Burvill, P.W., Johnson, G.A., Jamrozik, K.D., Anderson, C.S., Stewart-Wynne, E.G. & Chakera, T.M.H. (1995) Prevalence of depression after stroke. The Perth Community Stroke Study. *British Journal of Psychiatry* **166**, pp 320-327.

Cairns, J. (1996) Measuring health outcomes. *BMJ* **313**, pp 6

Calman, K.C. (1984) Quality of life in cancer patients - a hypothesis. *J Med Ethics* **10**, pp 124

Carey, J.W. (1993) Linking qualitative and quantitative methods: integrating cultural factors into public health. *Qualitative Health Research* **3(3)**, pp 298-318.

Chalmers, I. (1995) What do I want from health research and researchers when I am a patient? *BMJ* **310**, pp 1315-1318.

Chinese Acute Stroke Trial Collaborative Group (1997) CAST: randomised placebo-controlled trial of early aspirin use in 20,000 patients with acute ischaemic stroke. *Lancet* **349**, pp 1641-1649.

Clinical Outcomes Working Group (1995) Clinical Outcome Indicators. Edinburgh: The Scottish Office.

Coste, J., Fermanian, J. & Venot, A. (1995) Methodological and statistical problems in the construction of composite measurement scales: a survey of six medical and epidemiological journals. *Statistics in medicine* **14**, pp 331-345.

Counsell, C., Warlow, C.P., Sandercock, P.A.G., Fraser, H. & van Gijn, J. (1995) The Cochrane Collaboration Stroke Review Group: meeting the need for systematic reviews in stroke care. *Stroke* **26**, pp 498-502.

Counsell, C., McDowall, M., Slattery, J., Dennis, M. and in collaboration with the SEPIVAC and Perth Community Stroke Registries (1996) Prediction of functional outcome following stroke using a validated model. *Cerebovasc Dis* **6(suppl2)**, pp 50(Abstract)

Counsell,C. (1997) Formulating questions and locating primary studies for inclusion in systematic reviews. *Annals of Internal Medicine* **127**, pp 380-387

Counsell, C.E. (1998) The prediction of outcome in patients with acute stroke. University of Cambridge. MD.

Cox, D.R., Fitzpatrick, R., Fletcher, A.E., Gore, S.M., Spiegelhalter, D.J. & Jones, D.R. (1992) Quality-of-life assessment: can we keep it simple? *J R Statistical Society* **155**, pp 353-393.

Croog, S.H., Levine, S., Testa, M.A., et al (1986) The effects of antihypertensive therapy on the quality of life. *N Engl J Med* **314**, pp 1657-1664.

Davenport, R.J., Dennis, M.S. & Warlow, C.P. (1996) Effect of correcting outcome data for case-mix: an example from stroke medicine. *BMJ* **312**, pp 1503-1505.

de Haan, R., Aaronson, N., Limburg, M., Langton-Hewer, R. and van Crevel, H. (1993a) Measuring quality of life in stroke. *Stroke* **(24)**:320-327.

de Haan, R., Horn, J., Limburg, M., van der Meulen, J. & Bossuyt, P. (1993b) A comparison of five stroke scales with measures of disability, handicap, and quality of life. *Stroke* **24**, pp 1178-1181.

de Haan, R.J., Limburg, M., van der Meulen, J.H.P., Jacobs, H.M. & Aaronson, N.K. (1995) Quality of life after stroke: impact of stroke type and lesion location. *Stroke* **26**, pp 402-408.

Dennis, M., Wellwood, I., O'Rourke, S., MacHale, S. & Warlow, C. (1997a) How reliable are simple questions in assessing outcome after stroke? *Cerebrovasc Dis* **7**, pp 19-21.

Dennis, M., Wellwood, I. & Warlow, C. (1997b) Are simple questions a valid measure of outcome after stroke? *Cerebrovasc Dis* **7**, pp 22-27.

Deyo, R.A., Diehr, P. & Patrick, D.L. (1991) Reproducibility and responsiveness of health status measures. *Controlled Clinical Trials* **12**, pp 142S-158S.

Deyo, R.A. & Carter, W.B. (1992) Strategies for improving and expanding the application of health status measures in clinical settings. *Med Care* **30**, pp MS176-MS186.

Dolan, P., Gudex, C., Kind, P. and Williams, A. (1995) A social tariff for EuroQol: results from a UK general population survey. York: University of York.

Donabedian, A. (1988) The quality of care. How can it be assessed? *JAMA* **260**, pp 1743-1748.

Dorman, P.J., Counsell, C. & Sandercock, P.A.G. (1996) Recently developed neuroprotective therapies for acute stroke: a qualitative systematic review of clinical trials. *CNS Drugs* **5**, pp 457-474.

Dorman, P.J. & Sandercock, P.A.G. (1996) Considerations in the design of clinical trials of neuroprotective therapy in acute stroke. *Stroke* **27**, pp 1507-1515.

Duncan, P.W., Samsa, G.P., Weinberger, M., et al (1997) Health status of individuals with mild stroke. *Stroke* **28**, pp 740-745.

Dyken, M. & White, P.T. (1956) Evaluation of cortisone in the treatment of cerebral infarction. *JAMA* **162**, pp 1531-1534.

Ebrahim, S., Barer, D. & Nouri, F. (1986) Use of the Nottingham Health Profile with patients after a stroke. *Journal of Epidemiology and Community Health* **40**, pp 166-169.

Ebrahim, S., Barer, D. & Nouri, F. (1987) Affective illness after stroke. *Br J Pscychiatry* **151**, pp 52-56.

Ebrahim, S. (1995) Clinical and public health perspectives and applications of health related quality of life measurement. *Soc Sci Med* **41**, pp 1383-1394.

Essink-Bot, M.L., Krabbe, P.F.M., Bonsel, G.J. & Aaronson, N.K. (1997) An empirical comparison of four generic health status measures: the Nottingham Health Profile, the Medical Outcomes Study 36-Item Short-Form Health Survey, the COOP/WONCA Charts, and the EuroQol Instrument. *Medical Care* **35**, pp 522-537.

Fallowfield, L. (1990) *The quality of life: the missing measurement in health care*, London: Souvenir Press]

Fallowfield, L. (1996) Quality of quality-of-life data. *Lancet* **348** 421-422.

Fitzpatrick, R., Fletcher, A., Gore, S., Jones, D., Spiegelhalter, D. & Cox, D. (1992a) Quality of life measures in health care.I: Applications and issues in assessment. *BMJ* **305**, pp 1074-1077.

Fitzpatrick, R., Zeibland, S., Jenkinson, C. & Mowat, A. (1992b) Importance of sensitivity to change as a criterion for selecting health status measures. *Quality in Health Care* **1**, pp 89

Fletcher, A., Gore, S., Jones, D., Fitzpatrick, R., Spiegelhalter, D. & Cox, D. (1992) Quality of life measures in health care.II: Design, analysis, and interpretation. *BMJ* **305**, pp 1145-1148.

Folstein, M.F., Maiberger, R. & McHugh, P.R. (1977) Mood disorder as a specific complication of stroke. *J Neurol Neurosurg Psychiatry* **40**, pp 1018-1020.

Freemantle, N., Henry, D., Maynard, A. & Torrance, G. (1995) Promoting cost effective prescribing. *BMJ* **310**, pp 955-956.

Friedman, L.M., Furberg, C.D. & DeMets, D.L. (1984) Assessment of quality of life. In: *Fundamentals of clinical trials*, Littleton, Massachusetts: PSG Publishing Company]

Gage, B.F., Cardinalli, A.B. & Owens, D.K. (1996) The effect of stroke and stroke prophylaxis with aspirin or warfarin on quality of life. *Arch Intern Med* **156**, pp 1829-1836.

General Medical Council (1998) *Good Medical Practice*. London: General Medical Council.

Gill, T.M. & Feinstein, A.R. (1994) A critical appraisal of the quality of Quality-of-Life measurements. *JAMA* **272**, pp 619-626.

Gilson, B.S., Gilson, J.S., Bergner, M., Bobbit, R.A., Kressel, S., Pollard, W.E., & Vesselago, M. (1975) The Sickness Impact Profile: development of an outcome measure of health care. *American Journal of Public Health* **65(12)**, pp 1304-1310.

Gompertz, P.H., Pound, P. & Ebrahim, S. (1993) The reliability of stroke outcome measures. *Clinical Rehabilitation* **7**, pp 290-296.

Goodare, H. & Smith, R. (1995) The rights of patients in research: patients must come first in research. *BMJ* **310**, pp 1277-1278.

Greenwood, M. (1970) Medical statistics from Graunt to Farr. In: Pearson, E.S. and Kendall, M.G. (Eds.) *Studies in the history of statistics and probability*, London: Griffin]

Gruenberg, E.M. (1977) The failures of success. *Milbank Q* **55**, pp 1-24

Gudex, C., Dolan, P., Kind, P. & Williams, A. (1996) Health state valuations from the general public using the visual analogue scale. *Quality of Life Research* **5**, pp 521-531.

Guyatt, G.H., Feeny, D.H. & Patrick, D.L. (1993) Measuring health-related quality of life. *Annals of Internal Medicine* **118**, pp 622-629.

Harwood, R.H., Jitapunkul, S., Dickinson, E. & Ebrahim, S. (1994) Measuring handicap: motives, methods, and a model. *Quality in Health Care* **3**, pp 53-57.

Harwood, R.H., Gompertz, P. & Ebrahim,S. (1994) Handicap one year after a stroke: validity of a new scale. *J Neurol Neurosurg Psychiatry* **57**, pp 825-829.

Hayes, V., Morris, J., Wolfe, C. & Morgan, M. (1995) The SF-36 health survey questionnaire: is it suitable for use with older adults? *Age and Ageing* **24**, pp 120-125.

Henwood, K. & Nicolson, P. (1995) Qualitative research. *The Psychologist* **8**, pp 109-110.

Herzlich, C. (1973) *Health and illness: a social psychological analysis*, London: Academic Press.

Hickey, A.M., Bury, G., O'Boyle, C.A., Bradley, F., O'Kelly, F.D. & Shannon, W. (1996) A new short form individual quality of life measure (SEIQoL-DW): application in a cohort of individuals with HIV/AIDS. *BMJ* **313**, pp 29-33.

Hobart, J.C., Lamping, D.L. & Thompson, A.J. (1996) Evaluating neurological outcome measures: the bare essentials. *J Neurol Neurosurg Psychiatry* pp 127-130.

Holbrook,M. & Skilbeck, C.E. (1983) An activities index for use with stroke patients. *Age Ageing* **12**, pp 166-170.

Hollingworth, W., Mackenzie, R., Todd, C.J. & Dixon, A.K. (1995) Measuring changes in quality of life following magnetic resonance imaging of the knee: SF-36, EuroQol or Rosser index? *Quality of Life Research* **4**, pp 325-334.

House, A. (1987) Depression after stroke. *BMJ* **294**, pp 76-78.

Hunt, S.M., McKenna, S.P., McEwen, J., Williams,J. & Papp, E. (1981) The Nottingham Health Profile: subjective health status and medical consultations. *Soc Sci Med* **15A**, pp 221-229.

Hunt, S.M. (1995) Cross-cultural comparibility of quality of life measures. In: Guggenmoos-Holzmann, I., Bloomfield, K., Brenner, H. and Flick, U. (Eds.) *Quality of Life and Health: Concepts, Methods and Applications*, Berlin: Blackwell Wissenschafts-Verlag]

Hurst, N.P., Kind, P., Ruta, D., Hunter, M. & Stubbings, A. (1997) Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). *British Journal of Rheumatology* **36**, pp 551-559.

International Stroke Trial Collaborative Group (1997) The International Stroke Trial (IST): a randomised trial of aspirin, subcutaneous heparin, both or neither among 19435 patients with acute ischaemic stroke. *Lancet* **349**, pp 1569-1582.

Isard, P.A. & Forbes, J.F. (1992) The cost of stroke to the National Health Service in Scotland. *Cerebrovasc Dis* **2**, pp 47-50.

Jenkinson, C., Coulter, A. & Wright, L. (1993) Short form 36 (SF-36) health survey questionnaire: normative data for adults of working age. *BMJ* **306**, pp 1437-1440.

Jenkinson, C., Gray, A., Doll, H., Lawrence, K., Keoghane, S., & Layte, R. (1997) Evaluation of index and profile measures of health status in a randomised controlled trial. *Medical Care* **35**, pp 1109-1118.

Johansson, B.B., Jadback, G., Norrving, B., Widner, H. & Wiklund, I. (1992) Evaluation of long term functional status in first ever stroke patients in a defined population. *Scand J Rehab Med* **Suppl 26**, pp 105-114.

Johnson, G., Burvill, P.W., Anderson, C.S., Jamrozik, K., Stewart-Wynne, E.G. & Chakera, T.M.H. (1995) Screening instruments for depression and anxiety following stroke. Experience in the Perth community stroke study. *Acta Psychiatrica Scandinavica* **91**, pp 252-257.

Kappelle, L.J., Adams, H.P., Heffner, M.L., Torner, J.C., Gomez, F. & Biller, J. (1994) Prognosis of young adults with ischemic stroke. *Stroke* **25**, pp 1360-1365.

Kay, R., Wong, K., Yu, Y., et al (1995) Low-molecular-weight heparin for the treatment of acute ischemic stroke. *NEJM* **333**, pp 1588-1593.

Kazis, L., Anderson, J. & Meenan, R. (1989) Effect sizes for interpreting changes in health status. *Med Care* **27**, pp S178

Kilgour-Christie, J. & Watt, A.H. (1993) The term outcome is ambiguous. *BMJ* **307**, pp126.

Kind, P. (1988) The Development of Health Indices. In: Teeling-Smith,G (ed). *Measuring Health: A Practical Approach*, London: Wiley & Sons.

Kind, P., Dolan, P., Gudex, C. & Williams, A. (1998) Variations in population health status: results from a United Kingdom national questionnaire survey. *BMJ* **316**, pp 736-741.

Kind, P. & Dolan, P. (1995) The effect of past and present illness experience on the valuations of health states. *Medical Care* **33**, pp AS255-AS263.

King, R.B. (1996) Quality of life after stroke. *Stroke* **27**, pp 1467-1472.

Kwa, V.I.H., Limburg, M. & de Haan, R.J. (1996) The role of cognitive impairment in the quality of life after ischaemic stroke. *J Neurol* **243**, pp 599-604.

Landis, J.R. & Koch, G.G. (1977) The measurement of observer agreement for categorical data. *Biometrics* **33**, pp 159-174.

Lawrence, L. & Christie, D. (1979) Quality of life after stroke: a three year follow up. *Age and Ageing* **8**, pp 167-172.

Leplege, A. & Hunt, S. (1997) The problem of quality of life in medicine. *JAMA* **278**, pp 47-50.

Levine, S. (1995) The meanings of health, illness and quality of life. In: Guggenmoos-Holzmann, I., Bloomfield, K., Brenner, H. and Flick, U. (Eds.) *Quality of life and health: concepts, methods and applications*, Berlin . Vienna: Blackwell Wissenschafts-Verlag]

Lilienfeld, A. & Lilienfeld, D.E. (1980) *Foundations of epidemiology*, 2nd edn. Oxford: Oxford University Press]

Lindgren, A., Roijer, A., Norrving, B., Wallin, L., Eskilsson, J. & Johansson, B.B. (1994) Carotid artery and heart disease in subtypes of cerebral infarction. *Stroke* **25**, pp 2356-2362.

Lindley, R.I., Warlow, C.P., Dennis, M.S., Slattery, J. & Sandercock, P.A.G. (1993) Interobserver reliability of a clinical classification of acute cerebral infarction. *Stroke* **24**, pp 1801-1804.

Lindley, R.I., Waddell, F., Livingstone, M., et al (1994) Can simple questions assess outcome after stroke? *Cerebrovasc Dis* **4**, pp 314-324.

Lohr, K. (1992) Applications of health status assessment measures in clinical practice. *Med Care* **30**, pp MS1-MS14.

Lydick, E. & Epstein, R.S. (1993) Interpretation of quality of life changes. *Quality of Life Research* **2**, pp 221-226.

Mahoney, F. & Barthel, D. (1965) Functional evaluation: the Barthel Index. *Maryland Medical Journal* **14**, pp 61-65.

Mahoney, J., Drinka, T.J.K., Abler, R., et al (1994) Screening for depression: single question versus GDS. *Journal of the American Geriatric Society* **42**, pp 1006-1008.

Martin, J., Meltzer, H. and Elliot, D. (1988) OPCS surveys of disability in Great Britain. The prevalence of disability among adults. London: Her Majesty's Stationery Office.

Mathew, N.T., Meyer, J.S., Rivera, V.M., Charney, J.Z. & Hartmann, A. (1972) Double-blind evaluation of glycerol therapy in acute cerebral infarction. *Lancet* pp 1327-1329.

McDowell, I. and Newell, C. (1996) *Measuring Health: a guide to rating scales and questionnaires*, Second edn. New York: Oxford University Press.

McHorney, C.A. & Tarlov, A.R. (1995) Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Quality of Life Research* **4**, pp 293-307.

Medical Outcomes Trust (1994) *SF-36 Health Survey: scoring manual for English-language adaptations*, Boston: Medical Outcomes Trust.

Morton, A.P. & Dobson, A.J. (1989) Assessing agreement. *The Medical Journal of Australia* **150**, pp 384-387.

Mulrow, C.D. (1987) The medical review article: state of the science. *Annals of Internal Medicine* **106**, pp 485-488.

Murray, C.J.L. & Lopez, A.D. (1997) Mortality by cause for eight regions of the world: Global Burden of Disease Study. *Lancet* **349**, pp 1269-1276.

Niemi,M-L., Laaksonen,R., Kotila, M., & Waltimo,O. (1988) Quality of life 4 years after stroke. *Stroke* **19**, pp 1101-1107.

Nydevik, I. & Hulter-Asberg, K. (1991) Subjective dysfunction after stroke. A study with Sickness Impact Profile. *Scand J Prim Health Care* **9**, pp 271-275.

Nydevik, I. & Hulter-Asberg, K. (1992) Sickness impact after stroke: a 3 year follow up. *Scand J Prim Health Care* **10**, pp 284-289.

O'Boyle, C.A., McGee, H., Hickey, A., O'Malley, K. & Joyce, C.R.B. (1992) Individual quality of life in patients undergoing hip replacement. *Lancet* **339**, pp 1088-1091.

O'Rourke, S.J. (1996) Psychosocial problems after stroke. University of Edinburgh. PhD.

Orchard, C. (1994) Comparing healthcare outcomes. *BMJ* **308**, pp 1493-1496.

Paterson, C. (1996) Measuring outcomes in primary care: a patient generated measure, MYMOP, compared with the SF-36 health survey. *BMJ* **312**, pp 1016-1020.

Patrick, D.L., Bush, J.W., & Chen,M.M. (1973) Methods for measuring levels of wellbeing for a health status index. *Health Services Research* **8**, pp 229-244.

Patrick, D.L. & Erickson, P. (1993) *Health status and health policy: allocating resources to health care*, Oxford: Oxford University Press

Pearlman, R. & Uhlmann, R. (1997) Patient and physician perceptions of patient quality of life across chronic diseases. *J Gerontol*

Rankin, J. (1957) Cerebral vascular accidents in patients over the age of 60.II: Prognosis. *Scott Med J* **2** 200-215.

Registrar General of England and Wales (1839) First annual report. pp.99

Revicki, D.A. & Kaplan, R.M. (1993) Relationship between psychometric and utility-based approaches to the measurement of health-related quality of life. *Quality of Life Research* **2**, pp 477-487.

Rosser, R. & Watts, V.C. (1972) The measurement of hospital output. *Int J Epidemiology* **1**, pp 361-367.

Rothman, M.L., Hedrick, S.C., Bulcroft, K.A., Hickam, D.H. & Rubenstein, L.Z. (1991) The validity of proxy-generated scores as measures of patient health status. *Med Care* **29**, pp 115-124.

Rothwell, P.M., McDowell, Z., Wong, C.K. & Dorman, P. (1997) Doctors and patients don't agree: cross sectional study of patients' and doctors' perceptions and assessments of disability in multiple sclerosis. *BMJ* **314**, pp 1580-1583.

Rubenstein, L.V., Calkins, D.R., Young, R.T., et al (1989) Improving patient function: a randomised trial of functional disability screening. *Annals of Internal Medicine* **111**, pp 836-842.

Ruta, D.A., Abdalla, M.I., Garratt, A.M., Coutts, A. & Russell, I.T. (1994) SF 36 health survey questionnaire: 1. Reliability in two patient based studies. *Quality in Health Care* **3**, pp 180-185.

Schuling, J., Greidanus, J. & Meyboom-de Jong, B. (1993) Measuring functional status of stroke patients with the Sickness Impact Profile. *Disability and Rehabilitation* **15**, pp 19-23.

Schuling, J., de Haan, R., Limburg,M. & Groenier, K.H. (1993) The Frenchay Activities Index: assessment of functional status in stroke patients. *Stroke* **24**, pp 1173-1177.

Scientific Advisory Committee (1995) Instrument Review Criteria. *Medical Outcome Trust Bulletin* **September**, pp I-IV.

Secretary of State for Health (1992) The Health of The Nation: A Strategy for Health in England. London: HMSO.

Segal, M.E. & Schall, R.R. (1994) Determining functional\health status and its relation to disability in stroke survivors. *Stroke* **25**, pp 2391-2397.

Shinar, D., Gross, C.R., Mohr, J.P., et al. (1985) Interobserver variability in the assessment of neurological history and examination in the stroke data bank. *Arch Neurol* **42**, pp 557-565.

Slevin, M., Plant, H., Lynch, D., Drinkwater, J. & Gregory, W.M. (1988) Who should measure quality of life, the doctor or the patient? *Br J Cancer* **57**, pp 109-112.

Sneeuw, K.C.A., Aaronson, N.K., De Haan, R.J. & Limburg,M. (1997) Assessing quality of life after stroke: the value and limitations of proxy ratings. *Stroke* **28**, pp 1541-1549.

Solomon, N.A., Glick, H.A., Russo, C.J., Lee, J. & Schulman, K.A. (1994) Patients preferences for stroke outcomes. *Stroke* **25**, pp 1721-1725.

Spitzer, W.O. (1987) Quality of life and functional status as target variables for research. *Journal of Chronic Diseases* **40(6)**, pp 465-471.

Streiner, D.L. & Norman, G.R. (1989) *Health measurement scales: a practical guide to their development and use*, Oxford: Oxford University Press]

Svensson, E. (1993) *Analysis of systematic and random differences between paired ordinal categorical data*, Gothenborg: Almqvist and Wiksell International.

Tarlov, A.R., Ware, J.E., Greenfield, S., Nelson, E.C., Perrin, E. & Zubkoff, M. (1989) The Medical Outcomes Study: an application of methods for monitoring the results of medical care. *JAMA* **262**, pp 925-930.

Tate, D.G., Dijkers, M., Johnson-Greene, L. (1996) Outcome measures in quality of life. *Topics in Stroke Rehabilitation*. **2(4)**, pp 1-17

Testa, M.A. & Simonson, D.C. (1996) Assessment of quality-of-life outcomes. *NEJM* **334**, pp 835-840.

The EuroQol Group (1990) EuroQol - a new facility for the measurement of health related quality of life. *Health Policy* **16**, pp 199-208.

The MOS Trust, I. (1992) The MOS 36-Item Short Form Health Survey (SF-36). Update: July 1992 edn,

The Secretary of State for Health (1991) The Health of The Nation: A Consultative Document for Health in England. London: HMSO.

Thier, S.O. (1992) Forces motivating the use of health status assessment measures in clinical settings and related clinical research. *Medical Care* **30**, pp MS15-MS22.

Torrance, G.W. (1987) Utility approach to measuring health-related quality of life. *J Chron Dis* **40**, pp 593-600.

van Agt, H.M.E., Essink-Bot, M.L., Krabbe, P.F.M. & Bonsel, G.J. (1994) Test-retest reliability of health state valuations collected with the EuroQol questionnaire. *Soc Sci Med* **39**, pp 1537-1544.

Van Dalen, H., Williams, A. & Gudex, C. (1994) Lay peoples's evaluations of health: are there variations between different subgroups. *J Epidemiol Community Health* **48**, pp 248-253.

van Gijn, J. (1992) Measurement of outcome in stroke prevention trials. *Cerebrovasc Dis* **2**, pp 23-34.

van Gijn, J. & Warlow, C. (1992) Down with stroke scales. *Cerebrovasc Dis* **2**, pp 244-246.

Vickrey, B.G., Hays, R.D., Graber, J. & et al (1992) A health related quality of life instrument for patients evaluated for epilepsy surgery. *Medical Care* **30**, pp 299-319.

Vickrey, B.G., Hays, R.D., Myers, L.W. & Ellison, G.W. (1995) A health related quality of life measure for multiple sclerosis. *Quality of Life Research* **4**, pp 187-206.

Viitanen, M., Fugl-Meyer, K.S., Bernspang, B. & Fugl-Meyer, A.R. (1988) Life satisfaction in long term survivors after stroke. *Scand J Rehab Med* **20**, pp 17-24.

Visser, M.C., Koudstaal, P.J., Erdman, R.A.M., et al (1995) Measuring quality of life in patients with myocardial infarction or stroke: a feasibility study of four questionnaires in The Netherlands. *J Epidemiol Community Health* **49**, pp 513-517.

Wade, D.T., Legh-Smith, J. & Hewer, R.L. (1985) Social activities after stroke: measurement and natural history using the Frenchay Activities Index. *Int Rehabil Med* **7**, pp 176-181.

Wade, D.T., Legh Smith, J. & Hewer, R.A. (1987) Depressed mood after stroke. A community study of its frequency. *Br J Pscychiatry* **151**, pp 200-205(S).

Wade, D.T. (1992) *Measurement in neurological rehabilitation*, Oxford: Oxford University Press]

Wade, D.T. & Langton Hewer, R. (1987) Functional abilities after stroke: measurement, natural history and prognosis. *J Neurol Neurosurg Psychiatry* **50**, pp 177-182.

Warbuton, A. and Long, A. (1994) A classification of the literature for stroke outcomes. Leeds: UK Clearing House for Information on the Assessment of Health Outcomes.

Wardlaw, J.M., Dennis, M.S., Lindley, R.I., Sellar, R. & Warlow, C.P. (1996) The validity of a simple clinical classification of acute ischaemic stroke. *J Neurol* **243**, pp 274-279.

Wardlaw, J., Yamaguchi, T., del Zoppo, G. & Hacke, W. (1996) Thrombolysis in acute ischaemic stroke. In: Warlow, C.P., Sandercock, P.A.G. and Van Gijn, J. (Eds.) *Cochrane Database of Systematic Reviews Acute Stroke Module*, London: BMJ Publishing]

Ware, J.E. (1984) Conceptualizing disease impact and treatment outcomes. *Cancer* **53 (suppl)**, pp 2316- 2323.

Ware, J.E. (1992) *Measures for a new era of health assessment*, Durham and London: Duke University Press.

Ware, J.E. (1993) Measuring patients' views: the optimum outcome measure. *BMJ* **306**, pp 1429-1430.

Ware, J.E. & Donald Sherbourne, C. (1992) The MOS 36-Item Short-Form Health Survey (SF-36). 1. Conceptual framework and item selection. *Med Care* **30**, pp 473-483.

Weinberger, M., Oddone, E.Z., Samsa, G.P. & Landsman, P.B. (1996) Are health-related quality-of-life measures affected by the mode of administration? *J Clin Epidemiol* **49**, pp 135-140.

Wellwood, I., Dennis, M. & Warlow, C.P. (1995) A comparison of the Barthel Index and the OPCS Disability Instrument used to measure outcome after acute stroke. *Age and Ageing* **24**, pp 54-57.

White, S.J. & Freedman, L.S. (1978) Allocation of patients to treatment groups in a controlled clinical trial study. *Br J Cancer* **37**, pp 849-857.

Wilkins, R. & Adams, O.B. (1983) Health expectancy in Canada, late 1970s: demographic, regional and social dimensions. *Am J Public Health* **73**, pp 1073-1080

Williams, A. (1995) *The Measurement and Valuation of Health: A Chronicle*. York: University of York.

Wilson, I.R. & Cleary, P.D. (1995) Linking clinical variables with health related quality of life: a conceptual model of patient outcomes. *JAMA* **273(1)** pp 59-65

Wood-Dauphinee, S. (1995) Quality of Life Research in Neurologic Rehabilitation. In: Guggenmoos-Holzmann, I., Bloomfield, K., Brenner, H. and Flick, U. (Eds.) *Quality of Life and Health: Concepts, Methods and Applications*, Berlin: Blackwell Wissenschafts-Verlag]

Woods, K.L. (1995) Mega-trials and management of acute myocardial infarction. *Lancet* **346**, pp 611-614.

Working Group on Outcome Indicators for Stroke (1997) Stroke Health Outcomes: Report to the Department of Health. Department of Health.

World Health Organisation (1958) Constitution of the World Health Organisation, Annex 1. In: *The First Ten Years of the World Health Organisation*. Geneva: World Health Organisation.

World Health Organisation (1967) International Classification of Diseases: manual of the international statistical classification of diseases, injuries and causes of death. Geneva: World Health Organisation.

World Health Organisation (1980) *International Classification of Impairments, Disabilities, and Handicaps*, Geneva:

Yoon,H. (1997) Factors affecting quality of life of the Korean aged stroke patients. *Intl J Aging and Human Development* **4(3)**, pp 167-181.

Zigmond, A.S. & Snaith, R.P. (1983) The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica* **67**, pp 361-370.