

Patient-Reported

**PATIENT-
REPORTED
OUTCOME
MEASUREMENT
GROUP, OXFORD**

**A STRUCTURED REVIEW
OF
PATIENT-REPORTED
OUTCOME MEASURES
(PROMs)
FOR STROKE**

**Report to the Department of
Health, 2009**



O

U

T

C

O

M

E

Measures

A STRUCTURED REVIEW OF PATIENT-REPORTED OUTCOME MEASURES IN RELATION TO STROKE

Crispin Jenkinson
Elizabeth Gibbons
Ray Fitzpatrick

*Patient-reported Outcome Measurement Group
Department of Public Health
University of Oxford*

2009

Copies of this report can be obtained from:

Elizabeth Gibbons
Senior Research Officer
Patient-reported Outcome Measurement Group

tel: +44 (0)1865 289402
e-mail: elizabeth.gibbons@dphpc.ox.ac.uk

Alternatively, it can be downloaded free of charge from the PROM group website:

<http://phi.uhce.ox.ac.uk/>

Contents

A STRUCTURED REVIEW OF PATIENT-REPORTED MEASURES IN RELATION TO STROKE: An update 2009

Executive summary	1
1. Introduction, methods, search terms and results	3
2. Generic PROMs	6
3. Disease Specific PROMs	8
4. Discussion and recommendations	11
Appendix A - Sources for PROM Bibliography and Search Strategy	15
Appendix B - Methodology of Psychometric Evaluation	17
Appendix C - Details of Generic PROMs	20
Appendix D - Details of Stroke Specific PROMs	26
Appendix E - Licensing requirements and costs	32
Appendix F: Methods of working, membership and conclusions of multi-disciplinary panel	33
References	40

A STRUCTURED REVIEW OF PATIENT-REPORTED OUTCOME MEASURES IN RELATION TO STROKE 2009

EXECUTIVE SUMMARY

Aims of the report

The aims of this report are to review the evidence of patient-reported outcome measures (PROMs) for people with stroke and to provide recommendations to the Department of Health of PROMs for stroke that could potentially be used on a large scale population basis, combining good measurement properties with the likelihood of modest burden to respondents in order not to jeopardise response rates. A literature review of relevant PROMs resulted in the identification of a short-list of both generic and stroke-specific instruments which were then presented to a multidisciplinary panel for discussion. The literature review of the evidence-base and the discussions of the multi-disciplinary panel underpin final recommendations to the Department of Health.

The Patient-reported Outcome Measurement Group previously submitted a report to the Department of Health of evidence of Patient-reported Outcome Measures (PROMs) for chronic conditions (Fitzpatrick et al., 2006)¹. The report included a review of evidence regarding PROMs for stroke with some recommendations.

The methods of the review are described and the results of the search including sources and search terms used to identify specific published research. Details of this updated evidence are presented firstly for generic PROMs evaluated with people with stroke, followed by stroke-specific PROM results. Two newly identified stroke-specific measures are also briefly discussed. The report concludes with discussion and recommendations.

Results

The 2006 review identified PROMs with some evidence of measurement performance and operational characteristics. Details of the content, domains and scoring are outlined in Appendices C and D.

From the previous review, the most comprehensively evaluated generic PROM was the SF-36. However, the available evidence presented something of a mixed picture, suggesting that the instrument had satisfactory psychometric properties, but poor evidence for the acceptability (to patients) of the instrument. Evidence for the EuroQol EQ-5D was less extensive, but indicated that it had reasonable measurement characteristics and appeared reasonably acceptable to patients.

In carrying out the current update of evidence, further evidence was identified for the SF-36 and the EQ-5D.

¹ A structured review of PROMs for Stroke, 2006 can be downloaded at <http://phi.uhce.ox.ac.uk/>

There was some evidence to support the use of three stroke specific measures in the previous review (2006): the Stroke Impact Scale (SIS), the Stroke Specific Quality of Life Scale (SS-QOL) and the Subjective Index of Physical and Social Outcome (SIPSO). However, the amount of information was fairly limited and it was suggested more research was needed to determine which is the most appropriate.

Well established rehabilitation measures, though not specifically quality of life instruments, fared well in terms of their psychometric properties. Three instruments were suggested: The Self Complete Barthel Index, the Frenchay Activities Index (FAI) and the Nottingham Extended ADL Scale.

In carrying out the current update of evidence, further evidence was identified for the Stroke Impact Scale (SIS), the Stroke Specific Quality of Life Scale (SS-QOL) and Barthel Index.

Two more measures are included for review in this update: the Burden of Stroke Scale and the Stroke and Aphasia Quality of Life Scale.

Based on volume of evaluations and good measurement and operational characteristics, the following PROMs were presented to the multi-disciplinary panel for discussion:

- SF-36
- SF-12
- EQ-5D
- Stroke Impact Scale
- Barthel Questionnaire
- Frenchay Activities Index
- Nottingham Extended ADL Scale

The final recommendations of PROMs for the measurement of outcomes of stroke were agreed as follows:

- EQ-5D for the evaluation of general health
- The **Stroke Impact Scale V3 (SIS)** was high-lighted as a promising PROM for individuals with stroke, with considerable support also coming from the multi-disciplinary panel. There was not extensive supporting evidence for the SIS in the context of the UK. Moreover its length may be a problem. However a pilot study would provide an opportunity further to examine its measurement properties and feasibility. The main reservation of the multidisciplinary panel had been that it was not appropriate in the period immediately following a stroke when short disability measures such as the Barthel Index is preferred.

These two measures used together will provide complementary evidence of health status of people with stroke in the context of potential population-level applications in the NHS. In making the final selection of PROMs considered suitable for piloting in the NHS, the DH will consider salient factors in addition to the evidence and multidisciplinary panel comment

1. INTRODUCTION

Patient-reported outcome measures (PROMs) offer enormous potential to improve the quality and results of health services. They provide validated evidence of health from the point of view of the user or patient. They may be used to assess levels of health and need in populations, and in users of services, and over time they can provide evidence of the outcomes of services for the purposes of audit, quality assurance and comparative performance evaluation. They may also improve the quality of interactions between health professionals and individual service users.

Lord Darzi's Interim Report on the future of the NHS recommends that patient-reported outcome measures (PROMs) should have a greater role in the NHS (Darzi 2007). The new Standard NHS Contract for Acute Services, introduced in April 2008, includes a requirement to report from April 2009 on patient-reported outcome measures (PROMs) for patients undergoing Primary Unilateral Hip or Knee replacements, Groin Hernia surgery or Varicose Vein. Furthermore, Lord Darzi's report 'High Quality Care for All' (2008) outlines policy regarding payments to hospitals based on quality measures as well as volume. These measures include PROMs as a reflection of patients' experiences and views. Guidance has now been issued regarding the routine collection of PROMs for the selected elective procedures (DH, 2009).

The Patient-reported Outcome Measurement Group previously submitted a report to the Department of Health of evidence of Patient-reported Outcome Measures (PROMs) for chronic conditions, carer impact and patient perceptions of quality (Fitzpatrick et al., 2006). The report included a review of evidence regarding PROMs for stroke with some recommendations.

The aim of this report is two-fold: to provide an update of more recently published evidence for PROMs in stroke and to provide as clear recommendations as possible to Department of Health of PROMs that could be used on a potentially large scale population basis to assess health status of people with stroke to provide evidence relevant to determining the quality of services in the NHS. Recommended instruments would need to combine good measurement properties with the likelihood of low burden to respondents in order not to jeopardise responses rates. An additional consideration would be the availability of a PROM which yielded preference values derived from a UK source. As widely recommended, a strategy of combining a generic measure with a condition-specific measure was considered the most appropriate way of assessing complementary aspects of health status.

This current update review and recommendations draws on the existing evidence for each PROM up to 2006 but only provides fuller descriptive details of measurement and operational evidence which has emerged since 2006.

The full body of evidence was presented to a multidisciplinary panel for discussion. Details of their discussion and views are reported in Appendix F. The PROMs review group considered the combination of the full review of evidence and the multidisciplinary panel's views before reaching its own conclusions and recommendations (Chapter 4).

Structure of the report

The methods of the review update are described and the results of the search, including sources and search terms used to identify specific published research. Details of this updated evidence are presented firstly for generic PROMs evaluated with people with stroke, followed by stroke-specific PROM results. The report concludes with discussion and recommendations for short-listed PROMs.

Methods for the update review

Methods adopted were as documented in previous reviews performed by the PROM group, Oxford. Comprehensive searches were conducted; articles retrieved were assessed for relevance and checked by another reviewer; and evidence of measurement performance and operational characteristics abstracted for each PROM identified. Assessment and evaluation of the PROMs was performed independently by two reviewers adapting the London School of Hygiene and Tropical Medicine appraisal criteria outlined in their review (Smith et al, 2004). These criteria were modified for our reviews (Appendix C). The final short listing of promising PROMs to formulate recommendations was based on these assessments and discussion between reviewers. The most promising PROMs were then presented to a multi-disciplinary panel for final agreement.

Search terms and results: identification of articles

The methods for searching were conducted using the following sources:

- The PROM group bibliography which included over 16, 000 processed records up to 2006 (online version) and over 14, 000 unprocessed records 2006 to 2007.
- Pubmed and hand searching of journals covered the period January 2006 to August 2008.

The primary search strategy, using the term 'stroke' generated 144 records, as shown in Table 1.

All abstracts were reviewed. When assessed against the review inclusion criteria, 45 articles were retrieved and reviewed in full. Of these, 8 articles were included in the review. A supplemental search, using the names of the instruments reviewed in this and the previous review, uncovered a further 76 potential papers, of which 8 were retrieved and one included. Supplementary searches which included hand searching of titles from 2006 to 2008 of the following key journals:

Health and Quality of Life Outcomes

Medical Care

Quality of Life Research

Stroke

<i>Source</i>	<i>Results of search</i>	<i>No. of articles considered eligible</i>	<i>Number of articles included in review</i>
PROM database: search (up to 2008). Unprocessed= 14,296	144	45	8
Supplementary search	76	8	1
TOTAL	220	53	9
2006 review			54
TOTAL			63

2. RESULTS: Generic PROMs

In the previous review six generic PROMs were identified: the SF-36 (and related measure, the SF-12), the SF6-D, EuroQol EQ-5D, Health Utilities Index and the Nottingham Health Profile

SF-36/SF-12:

The SF-36 is the most widely validated measure of subjective health status in stroke. In the initial review fifteen papers were found evaluating the SF-36 in stroke in English speaking populations.

Construct validity of the SF-36V (a modified version of the SF-36 with Likert response categories in place of dichotomous responses) was assessed against the Stroke Impact Scales (SIS) in a telephone survey. The authors conclude that the SIS telephone survey has superior validity over the SF-36V, and has better ability to differentiate between levels of severity of stroke, at least in the context of telephone administration (Kwon et al, 2006).

Construct validity of the physical functioning domains of the SF-36 was assessed in patients post stroke (n=189). Rasch analysis was undertaken which indicated that the 10 items on the physical functioning domain formed a unidimensional scale, supporting the use of the total score as a measure of physical functioning in stroke (Dallmeijer, et al, 2007).

Convergent validity was assessed in subjects with stroke assessed one year (n=490) and three years post stroke (n=342). A graded positive relationship was found on all SF-36 domains and categories on the (clinically completed) FAI and Barthel Index at both time points (Patel et al, 2006). Further evidence of convergent validity was provided in a prospective study of stroke survivors, comparing results from the Burden of Stroke Scale (BOSS) with results from the SF-36. The Mental Component Summary Score was found to be highly correlated with the Psychological distress domain of the BOSS, and the Physical Component Summary highly correlated with the BOSS physical limitations subscale (Doyle, et al, 2007).

The sensitivity to change of the SF-36 was assessed in a longitudinal study of younger stroke patients (mean age 50 years, range 18-60) two years after 6-8 weeks hospital rehabilitation. The only significant improvement reported in the study was for the Role Physical domain. The pattern of change indicated by this and other measures (the FIM and the EuroQol EQ-5D) is difficult to interpret. Changes in the EQ-5D Index are significant, but not so for the EQ-5D thermometer. Similarly, the FIM indicates significant changes for the Overall Score and Cognitive dimension, but no for any other domain. The sample size for this study is small (n=52) and no account has been taken for multiple comparisons (Olsson et al, 2007).

SF-6D

The SF-6D is a utility index derived from the SF-36. It was included in the previous review because the items and scoring can be gained from any dataset including the SF-36. No further papers were discovered for this supplemental review.

EuroQol EQ-5D

The previous review evaluated six papers which included the EuroQol EQ-5D. Only one further paper was found for this update.

The sensitivity of the EuroQol EQ-5D was evaluated in a small scale study of younger stroke patients over a two year period (Olsson et al, 2007). The EQ-5D Index indicated change over time, but this was not reflected in the EQ-5D Thermometer.

In terms of convergent validity, EQ-5D Index scores were found to be associated with FIM Cognitive and Physical Scales (EQ-5D) (Olsson et al, 2007).

Health Utility Index

The previous review included six papers which evaluated the HUI in stroke. No further papers were found for this supplemental review.

Nottingham Health Profile (NHP)

The previous review included two papers which evaluated the NHP in stroke. No further papers were found for this supplemental review.

3. Stroke- specific PROMs

Seven patient completed questionnaires were included in the previous review. Subsequent literature is reported here. A further two measures are also included since the previous review: the Burden of Stroke Scale (BOSS) and the Stroke and Aphasia Quality of Life Scale (SAQLS-39).

Stroke Impact Scale (SIS)

In the previous review seven papers were found documenting the development and use of the instrument in the North American context, and one Australian study. This update reports a further two papers, once again originating from America.

The convergent validity of the SIS was assessed in a prospective study of stroke survivors comparing results from the Burden of Stroke Scale (BOSS) with comparable scales on the SIS (Doyle, et al, 2007). High and significant correlations were found between all relevant scales (SIS Mobility and BOSS Mobility, SIS ADL and BOSS Self-care, SIS Communication and BOSS Communication, SIS Emotion and BOSS Mood, SIS-16 Score and BOSS Physical Limitations, SIS Communication and Memory scales and BOSS Cognitive Limitations, SIS Emotion and BOSS Psychological Distress, and SIS Total and Boss Summary Scale).

Kwon et al (2006) evaluated the construct validity of the SIS for use in telephone administration. The SIS telephone survey was found to have superior convergent validity, and was better at differentiating between groups of stroke patients with different levels of disability than the Functional Independence Measure and the SF-36. No evidence was found of floor or ceiling effects.

The Stroke Impact Scale forms the basis of the SAQOL-39 (see below).

Stroke Specific Quality of Life Scale

In the previous review three papers were judged suitable for inclusion. One further paper is included in this review. It is also worth noting that the questionnaire has been validated in a number of European non-English languages since the previous review.

Williams et al (2006) evaluated the feasibility of the SS-QOL in terms of proxy report. They found that proxy respondents systematically report greater dysfunction in multiple aspects of health related quality of life than stroke patients. Agreement between patient and proxy score was found to be modest at best. Such differences may make proxy report on the SS-QOL inappropriate in research settings.

Subjective Index of Physical and Social Outcome (SIPSO)

In the previous review three papers were judged suitable for inclusion. No further papers were judged suitable for inclusion in this update. However, it is worth noting that the questionnaire has been validated in a number of European non-English languages since the previous review.

The Barthel Index

In the previous review seven papers were judged suitable for inclusion. A conference abstract is included for the update.

Self report Barthel was administered to 64 patients with a median (and interquartile range) age range of 45.5 (39-61) years (O'Connor 2006). Age, sex and method of scoring (no data provided on the latter two variables) were used to examine for differential item functioning (DIF) using Rasch analysis. Initial analysis indicated poor fit. By reordering the thresholds of the mobility item and removing bowel and bladder and toilet use items, which demonstrated DIF, it was possible to develop a unidimensional model. It may be possible to develop an interval level unidimensional measure with a sub-set of the items on the self report Barthel.

Frenchay Activities Index

In the previous review five papers were judged suitable for inclusion. No further papers were judged suitable for inclusion in this update.

Nottingham Extended ADL Scale

In the previous review three papers were judged suitable for inclusion. No further papers were judged suitable for inclusion in this update.

London Handicap Scale

In the previous review two papers were found which were suitable for inclusion in the review. No further papers were judged suitable for inclusion in this update.

The Burden of Stroke Scale (BOSS)

The Burden of Stroke Scale was omitted from the original 2006 as only one paper reporting either its use or psychometric properties was found. However, a recent paper documents its validation and hence its inclusion here. The development of the Burden of Stroke Scale was initially documented in Doyle et al (2004). The questionnaire was developed on the basis of literature reviews, expert panels, stroke survivors and healthcare providers. The instrument contains 64 questions measuring Mobility, Self Care, Communication, Cognition, Swallowing, Social Relations, Energy and Sleep, Negative Mood and Positive Mood. There are three composite scales (Physical Limitations, Psychological Distress, and Cognitive

Limitations) and a summary 'Burden of Stroke' scale. The questionnaire is designed for interviewer based completion.

Construct validity was assessed by comparing results from healthy controls with stroke survivors. Across all dimensions results indicated statistically significant differences with stroke survivors reporting greater activity limitations, and psychological distress. BOSS scores were found to be highly correlated with relevant domains on the SF-36 and Stroke Impact Scales (Doyle et al, 2004, 2007).

Internal consistency reliability of domains on the instrument have been found to be high (ranging from 0.74 - 0.91) (Doyle, et al, 2004).

The measure was found to be sensitive to changes over a one year period.

To date, all testing of the questionnaire has been undertaken by the group that developed the measure, in the United States of America. No other evidence was found for the use of the instrument.

Stroke and Aphasia Quality of Life Scale-39 item (SAQOL-39)

The SAQOL-39 was derived from the Stroke Specific Quality of Life Scale and four additional items specifically targeted at patients with aphasia, covering four domains: Physical, Psychosocial, Communication and Energy. At the time of the initial review only one paper was found documenting the validation of this measure, and hence it was not included. A further paper was found for this review.

Internal reliability of the domains of the questionnaire were found to be high ($\alpha = 0.74 - 0.94$) (Hilari et al, 2003). Test-re-test reliability was also found to be good ($ICC=0.89$) (Hilari et al, 2003).

Construct validity was assessed evaluating the extent to which social support was associated with SAQOL-39 overall score/ Social Companionship and Informational Support domains of the MOS Social Support Scale (Sherbourne and Stewart, 1991) were found to be associated with SAQOL-39 overall score ($n=83$) (Hilari, et al, 2006).

Acceptability of the SAQOL-39 was demonstrated by minimal missing data.

4. Discussion and recommendation

This document updates the previous review of quality of life measures in stroke. For the most part, the picture remains largely unchanged. Two more measures are included for review (the Burden of Stroke Scale and the Stroke and Aphasia Quality of Life Scale). However, the evidence for these latter measures is scant, and they are only included as they fulfil the inclusion criteria for the review (more than one validation paper published). The following recommendations are based upon the results of both the previous review and this update.

Generic Instruments used in stroke research

In terms of generic measures the overwhelming amount of evidence (in terms of number of papers published) available suggests the SF-36 as an appropriate tool. However, whilst the evidence suggests that the instrument has good psychometric properties in this patient group it is worth noting that that it fared less well in terms of acceptability. The previous review noted that a number of authors indicated that response rates were low. Feasibility was also assessed with the possibility of proxy completion, but the levels of agreement between family members and those with stroke were low. There is some evidence to suggest it may not be as sensitive to changes as some clinical measures.

The EQ-5D is the next most validated quality of life measure in stroke. The instrument has been found to have good test re-test characteristics, and appears to be marginally more acceptable to patients, at least in terms of response rates. However, the weighting scheme generally applied to responses to the items is strikingly at variance from weights gained when they are derived from stroke patients. A summary of the assessment of the psychometric qualities of the generic measures reviewed in this document and the original review can be found in Table 2.

Table 2: Appraisal of psychometric and operational performance of generic PROMs for stroke

	SF-36	SF-12	SF-6D	EQ-5D	HUI	NHP
Reproducibility	++	0	0	++	+	-
Internal consistency	++	++	0	0	0	0
Validity: Content	0	0	0	0	0	0
Construct	++	++	+	+	+	++
Responsiveness	+	+	+	++	++	0
Interpretability	+	0	0	+	0	0
Floor/ceiling/precision	+	0	0	+	0	0
Acceptability	-	+	0	+	-	0
Feasibility	-	0	0	+	-	0

Disease specific measures used in stroke research

In the previous review the Stroke Impact Scale (SIS) and the Subjective Index of Physical and Social Outcomes (SIPSO) were cited as potential candidates for a disease specific PROM in stroke. Further evidence for the SIS has been provided in this update, suggesting this is the most likely candidate for further research. Indeed, it only fails to get a more wholehearted recommendation because the available evidence derives from America and Australia. Evidence for the validity of the measure is required in England/UK before a more wholehearted recommendation can be given. At the present stage of development no single multi-dimensional outcome tool has sufficient information available to recommend it wholeheartedly.

Although not strictly regarded as PROMs, well established rehabilitation measures fared reasonably in terms of their psychometric properties. The Barthel Index, Frenchay Activities Index and Nottingham Extended ADL Scale were all primarily designed to evaluate rehabilitation outcomes. They are not strictly multi dimensional health outcome/quality of life instruments, but all measure important aspects of health status. The Barthel Index is a measure of independence, and was not initially designed for self completion, but versions of the instrument exist that can be completed by patients. Self completion and interview versions of the instrument have been found to have good reliability and validity, although the sensitivity of the instrument to change is a matter of debate. The Frenchay Activities Index was designed for interview administration, and is a measure of social activities and lifestyle following stroke. The instrument is generally used in interview settings, and there is evidence that the interviewer agreement on items can vary, albeit not dramatically. Available evidence suggests the instrument has good validity, and is amongst the easier measures for stroke patients to complete. The Nottingham Extended ADL Scale has been found to be reliable, and valid in concurrent validation with other instruments. Furthermore it appears sensitive to changes, and appears acceptable to patients.

Rehabilitation measures are widely used in the arena of stroke, are well understood by physicians and consequently provide useful and interpretable data. It is hard not to suggest a place for such instruments in evaluation of stroke. In this regard the best validated is the Barthel Index. One potential criticism of such instruments is that they are typically designed on the basis of clinical judgement and may not reflect issues of importance to patients. Consequently, it seems that such instruments might reasonably be used in conjunction with other quality of life measures. It seems that, at least for the time being, interview and self completion versions of the Barthel Index, Frenchay Activities Index and Nottingham Extended ADL Scale would appear the most appropriate condition specific instruments.

A summary of the assessment of the psychometric qualities of the disease specific measures reviewed in this document and the original review can be found in Table 3.

Table 3: Appraisal of psychometric and operational performance of stroke specific PROMs

	SIS	SS-QOL	SIPSO	Barthel	FAI	Nottingham ADL	LHS	BOSS	SAQOL-39
Reproducibility	+	0	0	+	+	+	+	0	+
Internal consistency	+	+	++	+	0	0	+	+	+
Validity: Content	+	0	0	+	0	0	0	0	0
Construct	++	+	++	++	+	+	+	+	+
Responsiveness	+	0	0	+	0	+	0	+	0
Interpretability	0	0	0	+	0	0	0	0	0
Floor/ceiling/precision	+	+	+	+	0	+	0	0	0
Acceptability	+	+	+	++	+	+	-	0	+
Feasibility	+	0	0	0	+	0	0	0	0

Summary of evidence

Based on the evidence presented in this review, no single measure of quality of life could be recommended without qualification. Unfortunately, one of the most widely cited quality of life instruments in this field (the Stroke Adapted Sickness Impact Profile) has never been validated in an English speaking patient group, hence it is not included in this or the previous review. Of the instruments included in this review it would seem timely to undertake validation of the Stroke Impact Scale in the UK context. However, until it has been validated in the UK it cannot be recommended as an appropriate instrument for routine evaluation of outcomes in stroke. At the current stage of development a combination of EuroQol EQ-5D and some disability index would seem appropriate in terms of measurement properties and brevity, the latter essential in this patient group. However, based on volume of evaluations and good measurement and operational characteristics, the following PROMs were presented to a multi-disciplinary panel for discussion:

- SF-36
- SF-12
- EQ-5D
- Stroke Impact Scale (SIS)
- Barthel Index
- Frenchay Activities Index
- Nottingham Extended ADL Scale

Details of consensus methods, multi-disciplinary members and minutes of the meeting are found in Appendix F

Recommendations

The final recommendations of PROMs for the measurement of outcomes of stroke were agreed as follows:

- EQ-5D for the evaluation of general health
- The **Stroke Impact Scale V3 (SIS)** was high-lighted as a promising PROM for individuals with stroke, with considerable support also coming from the multi-disciplinary panel. There was not extensive supporting evidence for the SIS in the context of the UK. Moreover its length may be a problem. However a pilot study would provide an opportunity further to examine its measurement properties and feasibility. The main reservation of the multidisciplinary panel had been that it was not appropriate in the period immediately following a stroke when short disability measures such as the Barthel Index is preferred.

The simplicity and the brevity of the EQ-5D, make it likely that it will not adversely influence response rates. The fact that it yields UK-derived preference values, makes it an attractive generic measure providing complementary evidence on health status alongside the Stroke Impact Scale..

APPENDIX A

Sources for PROM Bibliography and Search Strategy

I. Sources for PROM Bibliography

1. AMED: Allied and Complimentary Medicine Database
2. Biological Abstracts (BioAbs)
3. BNI: British Nursing Index Database, inc. RCN Journals Database
4. CINAHL: Cumulative Index to Nursing and Allied literature
5. EconLit (American Economic Association)
6. EMBASE (Elsevier)
7. MEDLINE (US National Library of Medicine)
8. PAIS: Public Affairs Information Service
9. PsychInfo (formerly PsychLitt (American Psychological Association))
10. SIGLE: System for Information on Grey Literature in Europe
11. Sociofile: Cambridge Scientific Sociological Abstracts Database
12. In addition, all records from the journal 'Quality of Life Research' are downloaded via Medline

II. Search Strategy

a. records to December 2005 (downloads 1-12)

((acceptability or appropriateness or (component\$ analysis) or comprehensibility or (effect size\$) or (factor analys\$) or (factor loading\$) or (focus group\$) or (item selection) or interpretability or (item response theory) or (latent trait theory) or (measurement propert\$) or methodol\$ or (multi attribute) or multiattribute or precision or preference\$ or proxy or psychometric\$ or qualitative or (rasch analysis) or reliabilit\$ or replicability or repeatability or reproducibility or responsiveness or scaling or sensitivity or (standard gamble) or (summary score\$) or (time trade off) or usefulness\$ or (utility estimate) or valid\$ or valuation or weighting\$)

and

((COOP or (functional status) or (health index) or (health profile) or (health status) or HRQL or HRQoL or QALY\$ or QL or QoL or (qualit\$ of life) or (quality adjusted life year\$) or SF-12 or SF-20 or SF?36 or SF-6) or ((disability or function or subjective or utilit\$ or (well?being)) adj2 (index or indices or instrument or instruments or measure or measures or questionnaire\$ or profile\$ or scale\$ or score\$ or status or survey\$))))

or

((bibliograph\$ or interview\$ or overview or review) adj5 ((COOP or (functional status) or (health index) or (health profile) or (health status) or HRQL or HRQoL or QALY\$ or QL or QoL or (qualit\$ of life) or (quality adjusted life year\$) or SF-12 or SF-20 or SF?36 or SF-6)

or ((disability or function or subjective or utilit\$ or (well?being)) adj2 (index or indices or instrument or instruments or measure or measures or questionnaire\$ or profile\$ or scale\$ or score\$ or status or survey\$))))

b. records from January 2006 (download 13)

((acceptability or appropriateness or component\$ analysis or comprehensibility or effect size\$ or factor analys\$ or factor loading\$ or feasibility or focus group\$ or item selection or interpretability or item response theory or latent trait theory or measurement proper\$ or methodol\$ or multi attribute or multiattribute or precision or preference\$ or proxy or psychometric\$ or qualitative or rasch analysis or reliabilit\$ or replicability or repeatability or reproducibility or responsiveness or scaling or sensitivity or valid\$ or valuation or weighting\$)

and

(HRQL or HRQoL or QL or QoL or qualit\$ of life or quality adjusted life year\$ or QALY\$ or disability adjusted life year\$ or DALY\$ or COOP or SF-12 or SF-20 or SF-36 or SF-6 or standard gamble or summary score\$ or time trade off or health index or health profile or health status or ((patient or self\$) adj (rated or reported or based or assessed)) or ((disability or function\$ or subjective or utilit\$ or well?being) adj2 (index or indices or instrument or instruments or measure or measures or questionnaire\$ or profile\$ or scale\$ or score\$ or status or survey\$))))

or

((bibliograph\$ or interview\$ or overview or review) adj5 (HRQL or HRQoL or QL or QoL or qualit\$ of life or quality adjusted life year\$ or QALY\$ or disability adjusted life year\$ or DALY\$ or COOP or SF-12 or SF-20 or SF-36 or SF-6 or standard gamble or summary score\$ or time trade off or health index or health profile or health status or ((patient or self\$) adj (rated or reported or based or assessed)) or ((disability or function\$ or subjective or utilit\$ or well?being) adj2 (index or indices or instrument or instruments or measure or measures or questionnaire\$ or profile\$ or scale\$ or score\$ or status or survey\$))))

APPENDIX B

Methodology of Psychometric Evaluation

The methods used for assessing the performance of PROMs were developed and tested against multi-disciplinary consensus and peer review. They focus on explicit criteria to assess reliability, validity, responsiveness, precision, acceptability and feasibility. A pragmatic combination of the criteria developed and used in previous reports to DH by the Oxford and LSHTM groups was used.

The appraisal framework focuses on psychometric criteria and PROMs must fulfil some or all to be considered as a short-listed instrument. Practical or operational characteristics are also assessed (acceptability and feasibility).

Once evidence has been assessed for eligibility, records considered as inclusions will be assembled for each PROM identified. Measurement performance and operational characteristics will be appraised using the following rating scale independently by two reviewers and inter-rater reliability calculated.

Psychometric evidence:

0	<i>not reported</i>
+	<i>some limited evidence in favour</i>
++	<i>some good evidence in favour, but some aspects do not meet criteria</i>
+++	<i>good evidence in favour</i>
-	<i>evidence available does not support criteria</i>

PROMs for which there are strong psychometric properties will be judged in terms of operational characteristics and clinical credibility.

Appraisal of psychometric and operational performance of PROMs for Stroke

PROM	Reproducibility	Internal consistency	Validity: Content	Construct	Responsiveness	Interpretability	Floor/ceiling/precision	Acceptability	Feasibility

Appraisal criteria (adapted from Smith et al., 2005 and Fitzpatrick et al., 1998; 2006)

Appraisal component	Definition/test	Criteria for acceptability
Reliability		
Test-retest reliability	The stability of a measuring instrument over time; assessed by administering the instrument to respondents on two different occasions and examining the correlation between test and re-test scores	Test re-test reliability correlations for summary scores 0.70 for group comparisons
Internal consistency	The extent to which items comprising a scale measure the same construct (e.g. homogeneity of items in a scale); assessed by Cronbach's alpha's and item-total correlations	Cronbach's alphas for summary scores ≥ 0.70 for group comparisons Item-total correlations ≥ 0.20
Validity		
Content validity	The extent to which the content of a scale is representative of the conceptual domain it is intended to cover; assessed qualitatively during the questionnaire development phase through pre-testing with patients. Expert opinion and literature review	Qualitative evidence from pre-testing with patients, expert opinion and literature review that items in the scale represent the construct being measured Patients involved in the development stage and item generation

Construct validity	Evidence that the scale is correlated with other measures of the same or similar constructs in the hypothesised direction; assessed on the basis of correlations between the measure and other similar measures	High correlations between the scale and relevant constructs preferably based on a priori hypothesis with predicted strength of correlation
	The ability of the scale to differentiate known-groups; assessed by comparing scores for sub-groups who are expected to differ on the construct being measured (e.g a clinical group and control group)	Statistically significant differences between known groups and/or a difference of expected magnitude
Responsiveness	The ability of a scale to detect significant change over time; assessed by comparing scores before and after an intervention of known efficacy (on the basis of various methods including t-tests, effect sizes (ES), standardised response means (SRM) or responsiveness statistics	Statistically significant changes on scores from pre to post-treatment and/or difference of expected magnitude
Floor/ceiling effects	The ability of an instrument to measure accurately across full spectrum of a construct	Floor/ceiling effects for summary scores <15%
Practical properties		
Acceptability	Acceptability of an instrument reflects respondents' willingness to complete it and impacts on quality of data	Low levels of incomplete data or non-response
Feasibility/burden	The time, energy, financial resources, personnel or other resources required of respondents or those administering the instrument	Reasonable time and resources to collect, process and analyse the data.

APPENDIX C

Details of Generic PROMs

SF-36: Medical Outcomes Study 36-item Short Form Health Survey (Ware and Sherbourne, 1992; Ware et al. 1994; Ware, 1997)

The Medical Outcomes Study (MOS) Short Form 36-item Health Survey (SF-36) is derived from the work of the Rand Corporation during the 1970s (Ware and Sherbourne, 1992). It was published in 1990 after criticism that a previous instrument, the SF-20, was too brief and insensitive. The SF-36 is intended for application in a wide range of conditions and with the general population and is intended to capture both mental and physical aspects of health. International interest in this instrument is considerable, and it is by far the most widely evaluated measure of health status.

Items were derived from several sources, including extensive literature reviews and existing instruments (Ware and Sherbourne, 1992; Jenkinson and McGee 1998). The original Rand MOS Questionnaire (245 items) was the primary source, and several items were retained from the SF-20. The 36 items assess health across eight domains, namely bodily pain (BP: two items), general health perceptions (GH: five items), mental health (MH: five items), physical functioning (PF: ten items), role limitations due to emotional health problems (RE : three items), role limitations due to physical health problems (RP: four items), social functioning (SF: two items), and vitality (V: four items), as shown in Table 1. An additional health transition item, not included in the final score, assesses change in health. All items use categorical response options (range: 2-6 options). Scoring uses a weighted scoring algorithm (although only one item is weighted) and a computer-based programme is recommended. Eight domain scores give a health profile; scores are transformed into a scale from 0 to 100 scale, where 100 denotes the best health. Scores can be calculated when up to half of the items are omitted. Two component summary scores for physical and mental health (MPS and MCS, respectively) can also be calculated. A version of the SF-36 plus three depression questions has been developed and is variously called the Health Status Questionnaire (HSQ) or SF-36-D.

The SF-36 can be self-, interview-, or telephone-administered.

SF-12: Medical Outcomes Study 12-item Short Form Health Survey (Ware et al., 1995)

In response to the need to produce a shorter instrument that could be completed more rapidly, the developers of the Medical Outcomes Study (MOS) 36-item Short Form Health Survey (SF-36) produced the 12-item Short Form Health Survey (SF-12) (Ware et al., 1995).

Using regression analysis, 12 items were selected that reproduced 90% of the variance in the overall Physical and Mental Health components of the SF-36 (Table 1). The same eight domains as the SF-36 are assessed and categorical response scales are used. A computer-based scoring algorithm is used to calculate scores: Physical Component Summary (PCS) and Mental (MCS) Component Summary scales are generated using norm-based methods. Scores

are standardised to have a mean value of 50, standard deviation (SD) 10, where scores above or below 50 are above or below average physical or mental well-being, respectively.

SF-6D

The SF-6D is a preference based algorithm based on a sub-set of items from the SF36, developed by Brazier et al (2002). The measure was designed to be used in health economic analyses.

EuroQol (The EuroQol Group, 1990; revised 1993)

The European Quality of Life instrument (EuroQol) was developed by researchers in five European countries to provide an instrument with a core set of generic health status items (The EuroQol Group, 1990; Brazier et al., 1993). Although providing a limited and standardized reflection of HRQL, it was intended that use of the EuroQol would be supplemented by disease-specific instruments. The developers recommend the EuroQol for use in evaluative studies and policy research; given that health states incorporate preferences, it can also be used for economic evaluation. It can be self or interview-administered.

Existing instruments, including the Nottingham Health Profile, Quality of Well-Being Scale, Rosser Index, and Sickness Impact Profile were reviewed to inform item content (The EuroQol Group, 1990). There are two sections to the EuroQol: the EQ-5D and the EQ thermometer. The EQ-5D assesses health across five domains: anxiety/depression (AD), mobility (M), pain/discomfort (PD), self-care (SC), and usual activities (UA), as shown in Table 1. Each domain has one item and a three-point categorical response scale; health 'today' is assessed. Weights based upon societal valuations of health states are used to calculate an index score of -0.59 to 1.00, where -0.59 is a state worse than death and 1.00 is maximum well-being. A score profile can be reported. The EQ thermometer is a single 20 cm vertical analogue scale (sometimes referred to as the EQ VAS) with a range of 0 to 100, where 0 is the worst and 100 the best imaginable health.

Health Utilities Index

The Health Utilities Index (HUI) was designed as a comprehensive measure of health status and health related quality of life. The Health Utilities Index (Mark 3) is a system composed of a health status classification which defines 972,000 discrete health states, and a preference, or utility, function which can be used to calculate the desirability for each health state. The HUI3 health status classification was developed by Feeny et al (1995) to assess capacity on eight dimensions: vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain/discomfort. The utility function reflects community preferences and scores each unique health state on a scale ranging from 0 (death) to 1 (perfect health). The HUI3 is a development of the Health Utilities Index containing a sub-set of items which constituted the HUI2. This report summarises data for the most recent version of the HUI (i.e. the HUI3).

Nottingham Health Profile (Hunt et al., 1980)

The Nottingham Health Profile (NHP) was developed in the UK during the 1970s for use in the evaluation of medical or social interventions (Hunt et al., 1980). Instrument content was derived from over 2000 statements given by 768 patients with a variety of chronic ailments and other lay people.

Part I of the instrument has 38 items across six domains: bodily pain (BP), emotional reactions (ER), energy (E), physical mobility (PM), sleep (S), and social isolation (SI), as shown in Table 1. All items are statements that refer to departures from normal functioning, and relate to feelings and emotional state rather than change in behaviour. Respondents answer 'yes' or 'no' according to whether or not they feel the item applies to them in general. Positive responses are weighted and summed to give six domain scores between 0 and 100, where 100 denotes maximum limitation.

Part II of the NHP is less widely used and provides a brief indicator of handicap. The instrument may be self- interview-, or telephone-administered.

Generic patient-reported outcome measures - Summary of features

<i>Instrument</i>	<i>Domains (no. items)</i>	<i>Response options</i>	<i>Score</i>	<i>Completion (time in minutes)</i>
SF-36: MOS 36-item Short Form Health Survey (36)	Bodily pain (BP) (2), General health (GH) (5) Mental health (MH) (5), Physical functioning (PF) (10) Role limitation-emotional (RE) (3), Role limitation-physical (RP) (4), Social functioning (SF) (2), Vitality (V) (4)	2-6 options Recall: standard 4 weeks, acute 1 week	Algorithm Domain profile (0-100, 100 best health) Summary: Physical (PCS), Mental (MCS) (mean 50, sd 10)	Interview (mean values 14-15) Self (mean 12.6)
SF-12: MOS 12-item Short Form Health Survey (12)	Bodily pain (BP) (1), Energy/Vitality (V) (1), General health (GH) (1), Mental health (MH) (2), Physical functioning (PF) (2), Role limitation-emotional (RE) (2), Role limitation-physical (RP) (2), Social functioning (SF) (1)	2-6 options Recall: standard 4 weeks, acute 1 week	Algorithm Domain profile (0-100, 100 best health) Summary: Physical (PCS), Mental (MCS) (mean 50, sd 10)	Interview or self
EuroQol- EQ5D) (5+1)	EQ-5D Anxiety/depression (1), Mobility (1), Pain/discomfort (1), Self-care (1), Usual activities (1) EQ-thermometer Global health (1)	EQ-5D 3 options to each of the five items <i>EQ-thermometer</i> VAS	EQ-5D Summation: domain profile Utility index (-0.59 to 1.00) <i>Thermometer</i> VAS (0-100)	Interview or self
Health Utility Index 3 (Feeny et al, 1995) (8)	Vision, Hearing, Speech, Ambulation, Dexterity, Emotion, Cognition, Pain	Four domains have five response options and five have six response options	Global Utility index and single attribute utility scores for the eight separate dimensions	Self report, face to face and telephone interview
SF-6D: (6)	Bodily pain (BP) (1), Energy/Vitality (V) (1), Mental health (MH) (1), Physical functioning (PF) (1), Role limitation (1), Social functioning (SF) (1)	4-6 response options. Recall: last 4 weeks	Utility Index (0 - 1)	Interview or self

<i>Instrument</i>	<i>Domains (no. items)</i>	<i>Response options</i>	<i>Score</i>	<i>Completion (time in minutes)</i>
NHP	Part I of the instrument has 38 items across six domains: bodily pain (BP), emotional reactions (ER), energy (E), physical mobility (PM), sleep (S), and social isolation (SI). Part II contains 6 unweighted single item questions measuring aspects of disability	Yes/No for all items	Algorithm 0-100 for domain scores 0/1 for single item questions	Interview, self or telephone

Summary of generic instruments: health status domains

<i>Instrument (no. items)</i>	<i>Instrument domains (after Fitzpatrick et al., 1998)</i>							
	Physical function	Symptoms	Global judgement	Psychol. well-being	Social well-being	Cognitive functioning	Role activities	Personal construct
SF-36 (36)	x	x	x	x	x		x	
SF-12 (12)	x	x	x	x	x		x	
SF-6D (6)	x	x		x	x		x	
EuroQol(5+1)	x	x	x	x	x		x	
HUI (8)	x			x		x		
NHP (38 +6)	x	x		x	x			

APPENDIX D

Details of Stroke Specific PROMs

Stroke Impact Scale Versions 2 and 3

The developers of the Stroke Impact Scale noted that many instruments, such as the SIP and SF-36, exhibited ceiling and floor effects in stroke populations. Consequently, these measures had limited ability to evaluate stroke outcomes over time. Consequently, they decided to develop a stroke specific measure that may overcome such problems (Lai et al 2003). The instrument content was derived from input from stroke patients, caregivers and health professionals with experience in the field of stroke. It contains 59 items across eight domains (strength, hand function, ADL/IADL, Mobility, Emotion, Memory, Communication and Social Participation). A related measure to the SIS is the SIS-16 which was designed to assess physical functioning and be more sensitive to differences than existing measures of physical function. The SIS-16 contains 16 items from the SIS measuring ADL/IADL, mobility and hand function (Edwards and O'Connell, 2003). The SIS Version 3 contains minor modifications but consists of the same items and domains as the SIS Version 2. Note: Version 1 of the SIS is reported only in unpublished literature.

Stroke Specific Quality of Life Scale (SS-QOL)

At the time of the development of the SS-QOL the authors argued that there were no stroke specific health related quality of life measures available. Consequently Williams et al (1999) set about devising a stroke specific QOL measure developed from interviews with patients. Thirty four survivors of ischemic stroke were interviewed to identify common themes that affect stroke patients' quality of life. Subjects included in the interviews were identified from stroke clinics one to six months after stroke and no significant cognitive or language impairment. Patients were asked to identify three areas most affected by their stroke. Twelve commonly affected domains were identified: energy, family roles, language, mobility, mood, personality, self-care, social roles, thinking, upper extremity function, vision, work/productivity. The final instrument contains 49 items measuring these concepts.

Subjective Index of Physical and Social Outcomes (SIPSO)

The SIPSO is an outcome tool that was designed to measure people's social integration rather than their abilities per se. It contains 10 items giving an overall score as well as Physical and Social Component scores.

The definition of social integration used in the work initially incorporated environment, activities as well as social integration. However, during test development the items relating to Environment were omitted as they failed to fulfil the criteria necessary for inclusion. The authors claim that the main aim of rehabilitation should be to reintegrate the patient into as normal a lifestyle as possible. Interviews with patients and carers were undertaken covering three aspects of their life: (1) pre stroke, (2) life since stroke and (3) perceptions of change since stroke.

Content analyses were undertaken on this data. (Trigg et al, 1999). On the basis of the interviews a questionnaire was developed and tested (Trigg and Wood, 2000). The authors claim that the SIPSO measures the ability of an individual to reintegrate to his or her own satisfaction.

An overall score can be calculated together with Physical and Social subscale scores.

The Barthel Index

The Barthel Index was originally developed for use in clinical practice as a means of assessing the degree of independence in patients with neurological and neuromuscular limitations. Strictly speaking the instrument is neither stroke specific nor developed for completion by patients. However, it is widely used in the field of rehabilitation and patient completed versions of the instrument have been developed.

The original Barthel Index consists of ten items, each of which is rated in terms of the patient's ability to undertake the task. Patients are classified into one of dependent, performs task with help and independent. In the original index there were ten areas covered (Bowel control, Bladder control, Grooming, Toilet use, Feeding, Transfer (from bed to chair), Mobility, Dressing, Stairs, Bathing). There have been a number of modifications to this original formulation, including a version with fifteen areas covered called the Modified Barthel Index (Granger et al, 1979), and one developed by Wade and Collin (1988) which uses simplified scoring algorithms.

Frenchay Activities Index

The Frenchay Activities Index (FAI) was developed as a means of measuring social activities and lifestyle following stroke, to supplement the more basic functional activities of daily living assessed by measured such as the Barthel Index. The FAI was designed from the outset to be an instrument that would be administered by the clinician to the patient in the clinical interview (Holbrook and Skilbeck, 1983; Wade et al, 1985).

Nottingham Extended Activities of Daily Living (ADL) Scale

The Nottingham Extended ADL Scale was developed and evaluated as a questionnaire for postal use (Nouri and Lincoln, 1987). It assesses the ability to carry out functional tasks, such as using public transport, housework, social life and hobbies. Scores in four areas: mobility, kitchen tasks, domestic activities and leisure activities can be added to give a summary score out of 22. Respondents are asked whether they do the activity rather than if they can do it, in order to assess level of activity rather than capability.

London Handicap Scale

The London Handicap Scale (LHS) was developed in response to the need for measures of morbidity to complement mortality statistics in the evaluation of health care interventions and services (Harwood, et al , 1994). Handicap is the disadvantage experienced by an individual patient because of ill-health. The developers adopt a definition of handicap developed by the World Health Organisation and claim that it

can be classified according to disadvantages in each of six dimensions: mobility, physical independence, occupation, social integration, and economic self sufficiency. The LHS contains one item for each of these dimensions. A single index score is gained by summing and weighting responses to these items. The measure was designed for use in rehabilitation, hence its inclusion in this review as a stroke specific measure. However, although it has been primarily used in stroke patients it could be used in other serious illness where patients undergo rehabilitation.

The Burden of Stroke Scale (BOSS)

The Burden of Stroke Scale was omitted from the original 2006 as only one paper reporting either its use or psychometric properties was found. However, a recent paper documents its validation and hence its inclusion here. The development of the Burden of Stroke Scale was initially documented in Doyle et al (2004). The questionnaire was developed on the basis of literature reviews, expert panels, stroke survivors and healthcare providers. The instrument contains 64 questions measuring Mobility, Self Care, Communication, Cognition, Swallowing, Social Relations, Energy and Sleep, Negative Mood and Positive Mood. There are three composite scales (Physical Limitations, Psychological Distress, and Cognitive Limitations) and a summary 'Burden of Stroke' scale. The questionnaire is designed for interviewer based completion.

Stroke and Aphasia Quality of Life Scale-39 item (SAQOL-39)

The SAQOL-39 was derived from the Stroke Specific Quality of Life Scale and four additional items specifically targeted at patients with aphasia, covering four domains: Physical, Psychosocial, Communication and Energy (Hilari, et al, 2003). At the time of the initial review only one paper was found documenting the validation of this measure. A further paper was found for this review.

Stroke-specific patient-reported health instruments

<i>Instrument (no. items)</i>	<i>Domains (no. items)</i>	<i>Response options</i>	<i>Score</i>	<i>Administration/ Completion (time)</i>
Stroke Impact Scale (SIS)	Strength, Hand Function, ADL/IADL, Mobility, Emotion, Memory, Communication, Social Participation (SIS version 3 contains 59 items in total)	5 point scales	0-100 for all dimensions and aggregate 'Physical Domain'	Interview Self completion
Stroke Specific Quality of Life Scale	Energy (3), Family roles (3), Language (5), Mobility (6), Mood (5), Personality (3), Self-care (5), Social roles (5), Thinking (3), Upper extremity function (5), Vision (3), Work (3)	5 point scale	Unweighted averages of items per domain (0-5) Overall score 0-60	Interview
Subjective Index of Physical and Social Outcome (SIPSO)	Overall score (10) Physical component (5) Social component (5)	5 point scales	Mean score 0-40 (overall) 0-20 for Physical and Social Component scores	Self completion
Barthel Index (10)	Bowels (1) Bladder (1) Grooming (1) Toilet use (1) Feeding (1) Transfer (1); Mobility (1); Dressing (1); Stairs (1); Bathing (1)	Categorical: 2-4 options	0-100 (0-20 with simplified scoring)	Measure initially designed for completion by clinician, but interview and self completion versions have been developed
Modified Barthel Index (15)	Drinking from a cup (1) Eating (1) Dressing - upper body (1); Dressing - lower body (1); Putting on brace or artificial limb (1); Grooming (1) Getting in and out of chair (1); Toilet use (1); Getting in and out of tub or shower; Walking 50 yards (1); Walking up/down one flight of stairs (1); If not walking: pushing a wheelchair .	Categorical: 2-4 options	-2 - 100 Self care functions: -2 - 53 Mobility: 0-47.	Clinician, interview and self completion

<i>Instrument (no. items)</i>	<i>Domains (no. items)</i>	<i>Response options</i>	<i>Score</i>	<i>Administration/ Completion (time)</i>
London Handicap Scale (6)	Handicap (6)	6 options per question	Index of handicap	Interview or self
Frenchay Activities Index (FAI) (15)	Single Index Scores (15) 15 items are Work; Driving, Hobby, Preparing meals, Local shopping, Reading books, Gardening, Washing up, Washing clothes, walking outside for longer than 15 minutes, Light housework, Heavy housework, Household/car maintenance, Social occasions, Travel outings	4 point scales	0 - 45 (or 15 to 60) point Index score Sub dimensions: Domestic Activities; Work and Leisure; Outdoors and Other	Interview Self/proxy completion
Nottingham Extended Activities of Daily Living Scale (24)	Mobility (6); Kitchen Tasks (5); Domestic tasks (5); Leisure activities (6)	4 point scales	Total score, Mobility, Kitchen, Domestic and Leisure scores	Interview Self completion
Burden of Stroke Scale (64)	Mobility, Self-care, Communication, Cognition, Swallowing, Social Relations, Energy and Sleep, Negative Mood and Positive Mood. Burden of Stroke summary scale	5 point scales	0-100 for all scales	Interview
39 item Stroke and Aphasia Quality of Life Scale	Physical (17); Psychosocial (11); Communication (7); Energy (4) Overall score	5 point scales	0-5 for all dimensions and summary scores	Interview

Summary of stroke specific instruments: health status domains (measures included in this supplemental review)

<i>Instrument (no. items)</i>	Physical function	ADL/ Self care	Emotions	Sleep	Social/ Interpersonal	Cognitive functioning	Communication	Role Functioning	Fatigue	Vision
SIS	X	X			X		X			
SS-QOL	X	X	X		X		X	X	X	X
SIPSO	X				X					
Barthel Index	X	X								
FAI					X			X		
Nottingham Extended ADL Scale	X	X								
London Handicap Scale								X		
BOSS	X	X	X	X	X	X	X		X	
SAQOL- 39	X		X				X		X	

APPENDIX E - Licensing requirements and costs

In order to use some of the questionnaires listed in this report a license will be required and a fee may be payable. Licences are required for the following:

The Nottingham Health Profile (www.galen-research.com)

The SF-36/SF-12 (www.qualitymetric.com)

SF-6D (www.shef.ac.uk/scharr/sections/heds/mvh/sf-6d)

EuroQol (www.euroqol.org)

The Nottingham Health Profile requires a license but charges are only levied on commercial/pharmaceutical organisations.

Licenses are required for the SF-36/SF-12, and fees are negotiated with the copyright holder (Quality Metric Inc). These charges are typically fairly nominal to not for profit organisations.

A license is required to use the SF-6D, but no fee is payable by non commercial organisations.

A license is required for the EuroQol. A charge is made for it's use, but this is negotiated directly with organisations using it. No indication of the fee structure is published by the EuroQol Group.

No evidence was found that the other measures reviewed in this report require a license or fees for use.

APPENDIX F: Methods of working, membership and conclusions of multi-disciplinary panel

Members of the multi-disciplinary panel were invited to participate based on their clinical experience of stroke and special interest in Patient-reported Outcome Measures. Other members included a patient representative, researchers and a health-economist. A meeting was held in December 2008.

The panel were given agenda papers which included:

- A structured review of patient-reported outcome measures for stroke: An update 2008
- A structured review of patient-reported outcome measures for stroke (2006)
- Copies of the PROMs short-listed for discussion.

Following presentation of each questionnaire and discussion, the group judged the suitability of the questionnaire for use in the NHS for the evaluation of services using responses of 'not at all suitable' (score = 0); 'to some extent suitable' (score=1); 'uncertain' (score=2); 'to some extent suitable' (score 3); 'very suitable' (score 4). Scores for each questionnaire were ranked in order of preference. The Total maximum score=36.

The results were then discussed amongst the group and members given the opportunity to change their vote.

Patient-reported Outcome Measure Rating Scale

1. On the basis of the review of evidence and your personal experience, is this questionnaire suitable for the measurement of the quality and outcomes of services for people with stroke? (please tick one box)

Not at all suitable To some extent unsuitable Uncertain To some extent suitable Very suitable

Do you have another questionnaire you could suggest?

Any additional comments

STROKE OUTCOMES CONSENSUS GROUP MEMBERS.

Rhoda Alison, Stroke Physiotherapist, Professional Executive Committee: Service Redesign, Devon PCT, Stroke Unit, Newton Abbot Hospital.

John Brazier, Professor of Health Economics, School of Health and Related Research, University of Sheffield.

Diana Day, Research Nurse in Stroke and Neurology, Department of Clinical Neurosciences, University of Cambridge.

Paul Dorman, Consultant Neurologist, Newcastle General Hospital.

Sarah Easton, Speech and Language Therapist, Portsmouth Community Stroke Rehabilitation Team.

Damian Jenkinson, Clinical Director, Stroke Unit, The Royal Bournemouth and Christ Church Hospitals NHS Trust.

Jonathan Mant, Professor of Primary Care Research, General Practice and Primary Care Research Unit, University of Cambridge.

Stephen Theobald, Patient Representative, Oxford.

Caroline Watkins, Professor of Stroke and Older People's Care, Department of Nursing, University of Central Lancashire.

Research members

Ray Fitzpatrick, Professor of Public Health and Primary Care, Deputy Head, Department of Public Health, University of Oxford.

Elizabeth Gibbons, Senior Research Officer, Department of Public Health, University of Oxford.

Crispin Jenkinson, Professor of Health Services Research, Department of Public Health and Harris Manchester College, University of Oxford.

Summary of multi-disciplinary meeting December 2008

1. All invited members present: Rhoda Alison, John Brazier, Diana Day, Paul Dorman, Sarah Easton, Ray Fitzpatrick, Elizabeth Gibbons, Crispin Jenkinson, Damian Jenkinson, Jonathan Mant, Stephen Theobald, Caroline Watkins.
2. CJ presented an overview of the purpose of the meeting- to formulate recommendations for the most suitable PROM for routine implementation in the NHS to measure the quality and outcomes of services for people with stroke.
3. Each PROM identified in the review with the most promising evidence base was presented to the group for discussion. It was noted that despite the fact that many questionnaires which were specific to stroke have been developed, very little evidence of performance has been reported particularly in English speaking populations. It was therefore judged to be appropriate to include disability measures which may have originally been developed for clinician assessment. Patient-reported versions or method of administration of these measures made them acceptable for inclusion in the review.

SF-36

The SF-36 is the most widely evaluated health status measure in people with stroke. JB reported that the SF-6D index can be derived from the SF-36 and the SF-12 but that there are copyright issues with the use of the SF-6D as a stand alone instrument.

Discussion points:

- The length of the questionnaire was considered to be a barrier to completion
- PD reported more missing data on SF-36 than EQ-5D in an RCT with stroke patients
- Interpretation and communication of data scores was highlighted as a problem
- Floor effects were considered with particular reference to Role Functioning
- ST expressed difficulties with some items-for example, people who have suffered a stroke may not have difficulty climbing stairs but coming down the stairs was more difficult. This would not be captured by the item related to stair climbing. Also, the point in time of completion of the questionnaire will depend on health state at the time.
- It was considered that completion required a high cognitive process with reference to phraseology which may be a problem for people who have suffered a stroke.
- The instructions were different for each question.

SF-12

- The shortened version was considered to be more acceptable to patients but that there still remained a level of complexity of questioning
- The items seemed to be heavily weighted for mental health constructs.

EQ-5D

It was reported that the EQ-5D was currently undergoing further development. In light of complexity with the VAS, it was suggested that the group judged the instrument solely on the questionnaire items.

-ST evaluated the questionnaire and considered the first two items to be satisfactory.

However, the Pain item refers to moderate 'pain or discomfort', which did not reflect his own experiences. ST suggested that 'some pain or discomfort' would be suitable wording, (as used in previous questions on the EQ-5D).

-The Usual Activities item did not reflect being able to carry out activities with some help.

-Problems with response shift were noted with people with stroke

Barthel Index

The Barthel index self-report has been evaluated in several studies and despite being primarily a disability index, has been included in the review

-The narrow focus and heavy weighting to disability of the items was judged to be limiting

-Ceiling effects were highlighted

-The absence of emotional aspects was deemed to be restrictive

-Difficulty with detecting change was also reported

Frenchay Activities Index

-JM referred to own publication of the comparison of postal and interview administered versions

-The postal version has more explanation per question

-It was noted that there is inconsistency in the format of recall, sometimes referring to the last 3 months and sometimes the last 6 months.

NEADL

-ST reported that all items were simple, straight forward and relevant but that the 'car' question may not be applicable. The absence of emotional items was highlighted as well as 'work' and 'washing'. Some questions depended on context for example, taking a drink from one room to another may not be a problem but to carry a drink upstairs may be difficult for some people.

Stroke Impact Scale

The SIS has been comprehensively developed with the involvement of patients specifically for self-report- this meets the FDA requirements for approval of PROMs.

-No British study was identified in the review

-Despite the length of the questionnaire, all aspects are covered which are deemed relevant to people with stroke

-The language, despite the use of American English terminology in some questions was considered to be good. It was recommended that the questionnaire be Anglicised.

-Consideration of a shortened version was suggested

4. Following presentation of each questionnaire and discussion, the group judged the suitability of the questionnaire for use in the NHS for the evaluation of services using the following responses:

- ‘not at all suitable’ (score 0);
- ‘to some extent unsuitable’ (score 1);
- ‘uncertain’ (score 2);
- ‘to some extent suitable’ (score 3);
- ‘very suitable’ (score 4).

Scores for each questionnaire were ranked in order of preference. The following table reports the scores and ranking. Total maximum score=36

Generic		
EQ-5D	23	Five people voted ‘to some extent suitable’ or ‘very suitable’
SF-36	18	Five people voted ‘to some extent suitable’, 3 ‘to some extent unsuitable’ and 1 person voted ‘not at all suitable’.
SF-12	12	Four voted ‘to some extent unsuitable’, 2 ‘very unsuitable’, 2 ‘to some extent suitable’ and one person was uncertain
Stroke-specific		
Stroke Impact Scale	34	Eight voted ‘very suitable’, 1 person was uncertain
NEADL	25	Six voted ‘to some extent suitable’
Barthel	23	Five voted ‘to some extent suitable’, one ‘very suitable’ and two were uncertain.
FAI	20	Two voted ‘to some extent suitable’, others were uncertain

Following voting and discussion, no person wished to change their scores. The following questionnaires were considered suitable for the measurement of quality and outcomes of services for people with stroke:

- EQ-5D for the evaluation of general health
- Stroke Impact Scale (SIS) as a condition-specific outcome measure, although more research on this questionnaire is required in the UK context. However, the SIS was considered inappropriate for use in the first few days/weeks following a stroke and that it may be more appropriate for a disability measure to be used such as the Barthel Index. It was therefore suggested that the Barthel be used in the early period following a stroke.

Commentary

There was some discussion about the suitability of using the EQ-5D and SIS for the evaluation of service quality and outcomes. RF explained that in addition to patient-reported health measures, patient experience and views of services are likely to be obtained from another questionnaire.

The impact on carers was also discussed- carer-specific measures were considered more appropriate.

Proxy rating was highlighted as often the only method of obtaining responses and the suitability of the PROMs in that context was questioned. The divergence in perceptions between carers and patients was identified as a barrier.

It was considered that a different questionnaire may be required for people with communication difficulties, such as The Stroke and Aphasia Questionnaire.

PD referred to '2 simple questions' which could possibly be used for case adjustment.

One important factor identified by the group was that a pre-stroke assessment could not be made but it was agreed that instruments could be used pre and post –interventions and for longitudinal evaluation. The SIS was considered inappropriate for use in the first few days/weeks following a stroke and that it may be more appropriate for a disability measure to be used such as the Barthel Index. It was therefore suggested that the Barthel be used in the early period following a stroke.

The SIS is more useful for medium to long-term evaluation.

REFERENCES

- 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(3):169-80.
- Brazier J, Roberts J, Deverill M. (2002) The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics*; 21(2):271-92.
- Dallmeijer A, deGroot V, Roorda L, Schepers VPM, Lindeman E, van den Berg L, Beelen A, Dekker J. (2007) Cross-diagnostic validity of the SF-36 Physical Functioning scale in patients with stroke, multiple sclerosis and amyotrophic lateral sclerosis: a study using Rasch analysis. *J Rehabil Med* 2007; 39: 163-69.
- Darzi A. Our NHS Our Future: NHS Next Stage Review Interim Report. Department of Health, London, October 2007.
- Darzi A. Our NHS Our Future: High Quality Care for All. NHS Next Stage Review Final Report. Department of Health, London, June 2008.
- Department of Health (2008) Guidance on the routine collection of Patient Reported Outcome Measures (PROMs)
- Doyle PJ, McNeil MR, Mikolic JM, Prieto L, Hula WD, Lustig AP, et al. (2004) The Burden of Stroke Scale provides valid and reliable score estimates of functioning and well-being in stroke survivors with and without communication disorders. *Journal of Clinical Epidemiology*; 57: 997-1007.
- Doyle PJ, McNeil M, Mikolic J, Prieto L, Hula W, Lustig A, Ross K, Wambaugh J, Gonzalez-Rothi, Elman R. (2007) The Burden of Stroke Scale provides valid and reliable score estimates of functioning and well-being in stroke survivors with and without communication disorders. *Journal of Clinical Epidemiology* 2007; 57: 997-10007.
- Doyle PJ, McNeil M, Bost J, Ross K, Wambaugh J, Hula W, Mikolic J. (2007) The Burden of Stroke Scale (BOSS) provided valid, reliable, and responsive score estimates of functioning and well-being during the first year of stroke. *Quality of Life Research*; 16: 1389-98.
- Edwards B., O'Connell B. (2003) Internal consistency and validity of the Stroke Impact Scale 2.0/SIS 2.0 and SIS-16 in an Australian sample. *Quality of Life Research*; 20:1127-35.
- EuroQol Group (1990) EuroQol--a new facility for the measurement of health-related quality of life. *The EuroQol Group. Health Policy*; 16(3):199-208.
- Feeny D, Furlong W, Boyle M, Torrance GW. (1995) Multi-attribute health status classification systems. *Health Utilities Index. Pharmacoeconomics*; 7(6):490-502.

Fitzpatrick R, Bowling A, Gibbons E, Haywood K, Jenkinson C, Aackintosh A, Peters M. (2006) A Structured Review of Patient-Reported Measures in Relation to Selected Chronic Conditions, Perceptions of Quality of Care and Carer Impact. Oxford: National Centre for Health Outcomes Development.

Granger CV, Albrecht GL, Hamilton BB. (1979) Outcome of comprehensive medical rehabilitation: Measurement by PULSES Profile and the Barthel Index. *Archives of Physical Medicine and Rehabilitation*; 60: 145-54.

Harwood RH, Gompertz P, Ebrahim S. (1994) Handicap one year after a stroke: validity of a new scale. *Journal of Neurology, Neurosurgery and Psychiatry*; 57: 825-9.

Hilari K, Byng S, Lamping D, Smith SC. (2003) Stroke and Aphasia Quality of Life Scale-39/SAQOL-39: Evaluation for acceptability, reliability and validity. *Stroke*; 34: 1944-50.

Hilari K, Northcott S. (2006) Social support in people with chronic aphasia. *Aphasiology*; 20 (1): 17-36.

Holbrook M, Skilbeck C. (1983) An activities index for use with stroke patients. *Age and Ageing*; 12: 166-70.

Hunt S, McKenna SP, McEwen J, Backett EM, Williams J, Papp E. (1980) A quantitative approach to perceived health status: a validation study. *Journal of Epidemiology and Community Health*; 34(4): 281-6.

Jenkinson C, McGee H. (1998) *Measuring Health Status: A Brief but Critical Introduction*. Oxford: Radcliffe Medical press.

Kwon S, Duncan P, Studentski S, Perera S, Lai S M, Reker D. (2006) Measuring stroke impact with SIS: Construct validity of SIS telephone administration. *Quality of Life Research*; 15: 367-76.

Lai S, Perera S, Duncan W, Bode R. (2003) Physical and social functioning after stroke: Comparison of the Stroke Impact Scale and the SF-36. *Stroke*; 34: 448-493.

O'Connor RJ. (2006) The self-report Barthel Index: is it a unidimensional scale? *Clinical Rehabilitation*; 20: 637.

Nouri FM, Lincoln NB. (1987) An extended activities of daily living index for stroke patients. *Clinical Rehabilitation*; 1: 301-305.

Olsson BG, Sunnerhagen KS. (2007) Functional and Cognitive Capacity and health-related quality of life 2 years after day hospital rehabilitation for stroke: A prospective study. *Journal of Stroke and Cardiovascular Diseases*; 16: 208-15.

Patel MD, Tilling K, Lawrence E, Rudd AG, Wolfe CDA, McKeivitt C. (2006) Relationships between long term stroke disability, handicap and health related quality of life. *Age and Ageing*; 35: 273-79.

Sherbourne C, Stewart A. (1991) The MOS Social Support Survey. *Social Science and Medicine*; 32: 705-14.

Smith S, Cano S, Lamping D, Staniszewska S, Browne J, van der Lewsey J, Cairns J, Black N. (2005) Patient-Reported Outcome Measures (PROMs) for routine use in Treatment Centres: recommendations based on a review of the scientific evidence. Report to the Department of Health. London: London School of Hygiene and Tropical medicine (University of London).

Trigg R, Wood VA. (2000) The Subjective Index of Physical and Social Outcome/SIPSO: a new measure for use with stroke patients. *Clinical Rehabilitation*; 14:288-99.

Trigg R, Wood V, Hewer R. (1999) Social integration after stroke: the first stages in the development of the Subjective Index of Physical and Social Outcome (SIPSO). *Clinical Rehabilitation*; 13: 341-353.

Wade DT, Collin C. (1988) The Barthel ADL Index: a standard measure of disability? *International Disability Studies*; 10: 64-7.

Ware J, Sherbourne, C. (1992) The MOS 36-Item Short-Form Health Survey 1: Conceptual Framework and Item Selection. *Medical Care*, 30, 473-83.

Williams L, Bakas T, Brizendine E, Plue L, Tu W, Hendrie H, Kroenke K. (2006) How valid are family proxy assessment of stroke patients quality of life? *Stroke* 2006; 37: 2081-2085.