Measuring Outcomes for Carers for People with Mental Health Problems

---

Report for the National Co-ordinating Centre for NHS Service Delivery and Organisation R & D (NCCSDO)

September 2005

Prepared by

Kate Harvey*

Amy Langman†

Hannah Winfield†

Jocelyn Catty†

Sarah Clement‡

Sarah White†

Ellen Burns§

Tom Burns¶

*Department of Psychology and Clinical Language Sciences, University of Reading, Reading, †Division of Mental Health, St George's, University of London, London, ‡Faculty of Health and Social Care, London South Bank University, London, §freelance researcher and ¶Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford

Correspondence to Dr Kate Harvey, Lecturer, School of Psychology and Clinical Language Science, University of Reading, Earley Gate, Whiteknights, PO Box 238, Reading RG6 6AL
Tel: 0118 378 7524; fax: 0118 378 6715; e-mail: k.n.harvey@reading.ac.uk

Contents

Acknowledgements

This study was made possible by the 86 carers, representatives of carer organisations and service providers who gave up their time to participate in interviews or complete a conference questionnaire. Thanks are due to Age Concern Lambeth, Alzheimer Society, Asian Health Agency, Carer Support Merton, Carer UK, Crossroads Care, Cumberland Day Centre, Effra Day Centre, Family Welfare Organisation Building Bridges, Jewish Care, Lambeth Carers, Lambeth Social Services, Lewisham Carers, Mind (National Association for Mental Health), Princess Royal Trust for Carers, Sainsbury Centre for Mental Health and Wandsworth Carers, who provided us with access to carers and carer representatives, and in particular Rethink, who also distributed a questionnaire at their conference on our behalf. Thanks are also due to Sarah Turner and Paul Farmer for their support throughout the project.

This study was enabled by a grant from the National Co-ordinating Centre for NHS Service Delivery and Organisation.

Executive Summary

Objectives

We aimed to review available outcome measures and, on the basis of consultation with a diverse range of carers, generate data with which to advise the National Co-ordinating Centre for NHS Service Delivery and Organisation (SDO) programme of:

- instruments that could be used to measure the outcomes of services from carers' perspectives;
- the further work required to develop these instruments so that they are robust;
- any outcomes considered important by carers for which measures have not been developed.

Methods

Consultation exercise

A consultation exercise was conducted with carers, representatives of carer organisations and service providers, to determine which outcomes they considered important. This comprised one-to-one interviews and focus groups with all three groups of stakeholders, and a questionnaire survey of the views of delegates at a carers' conference.

Literature review

We conducted a literature review to identify papers describing instruments measuring outcomes for carers for people with mental health problems and dementia. This comprised electronic searches of Medline (1966–) and PsycINFO (1999–), hand-searching of key journals and obtaining unpublished measures from carer organisations. To be included, papers needed to describe an outcome measure used with carers for people with mental health problems and dementia.

Instrument assessment

To be included in the instrument assessment, measures needed to have been used in studies assessing carer outcomes and either (a) to have been used three times or more, not including by the instruments' authors, or (b) to have been developed in the 10 years prior to the review. Psychometric data on acceptability, appropriateness, reliability, validity, responsiveness, precision and interpretability were extracted and, where appropriate, rated according to pre-determined benchmarks. In addition, the importance to carers for the construct measured and carer involvement in the development of the measure were assessed and rated on the basis of the consultation-exercise findings and the data presented by the authors respectively.

Results

Consultation exercise

Forty-four carers, eight carer representatives and four service providers were consulted individually or through focus groups, representing a broad range of views. Carers and representatives of carers from minority groups tended to contribute additional information to the discussion, such as the importance of addressing language issues when designing questionnaires, rather than different views about which outcomes were important. There was little consensus about important outcomes, most participants considering most outcomes important and favouring the assessment of a broad range of outcomes to capture carers' holistic experience. There was greater consensus among participants that carers disliked closed-question questionnaires, especially when sent by post, and preferred semi-structured interviews.

Literature review

We found around 500 published papers, referring to 241 outcome measures used with carers. Forty-nine met inclusion criterion (a), used three times or more, not including by the instruments' authors, and 15 met inclusion criterion (b), developed in the 10 years prior to the review (1994 or later), providing a total of 64 measures eligible for review. Six unpublished measures were provided by carer organisations.

Instrument assessment

Most measures (39) had been developed for use with populations other than carers. The degree and quality of the psychometric testing conducted varied, although there was a trend for measures developed for non-carer populations to be the better validated. Carers had rarely been involved in item generation, even when the measure was developed specifically for them, although their involvement was more common in the recently developed measures.

Conclusions

Several instruments with good psychometric properties were found to measure outcomes that carers considered important. These (a) identified carers whose own health was at risk, (b) assessed aspects of the carers' well-being, namely their physical and mental health, their psychological well-being, their coping and their quality of life or (c) assessed aspects of the caregiving experience, namely the effect of caregiving on the family, the caregiving activities the carer undertakes, how the carer feels about caregiving, the positive aspects of caregiving, the carer's sense of competence and the burden the carer experiences, or the carer's perception of the professional services they have received. In each of these three areas there is at least one instrument that could be recommended. Overall the Involvement Evaluation Questionnaire stands out as the most established, comprehensive instrument which measures both the experience of caregiving and carer well-being. All of the instruments require some further work to fully establish their psychometric properties in particular work on appropriateness for, and acceptability to, carers and precision, responsivity and interpretability.

The Report

Section 1 Background

1.1 Informal caregiving

Over recent years, the contribution that family and friends make to caring for people with mental health problems and their need for greater support have been increasingly recognised by Government (Department of Health, 1998; 1999a; 1999b; 2001). In a recent comprehensive review of studies evaluating services for carers for people with mental health problems, it was noted that, whereas three-quarters of studies used standardised outcome measures to assess effectiveness, and there was some commonality between them in relation to the outcome domains assessed, a plethora of different instruments had been employed (Arksey *et al.*, 2002). It has, in fact, been an ongoing criticism of research in this field that no standard, robust outcome measure has been developed to assess carers' experiences (Platt, 1985; Maurin and Barmann Boyd, 1990; Schene *et al.*, 1994; Harvey, 2000). We estimate that at least 200 different measures have been used to assess carer outcomes and most lack data on reliability, validity or sensitivity, particularly in relation to change over time, an essential capacity of instruments measuring the effectiveness of services (Harvey, 2000).

There is a clear need to ascertain the scope of existing outcome measures and their quality in terms of both psychometric properties and their acceptability to carers themselves. The identification of an instrument or instruments that can robustly measure carer outcomes or, if necessary, their development, would be vital steps towards providing carers with better support. Key to any instrument's appropriateness would be its ability to measure the effectiveness of services from the perspective of the carer. The neglect of carers' perspectives in the development of standard outcome measures may explain the apparent discrepancy between quantitative and qualitative findings in this field. The former often fail to confirm benefits of a service, such as support groups, that are found beneficial by the latter (Arksey *et al.*, 2002). That instruments are based on carers' perspectives is particularly crucial for their application to the experiences of carers in more diverse situations (e.g. carers from ethnic minority groups, young carers, carers in same-sex relationships and those caring from a distance). Few instruments are available for use with non-English-speaking carers and few, if any, have been developed to assess the specific experiences of carers in other diverse situations. It is likely that the concerns and needs of these carers in particular have been neglected by previous researchers and service providers.

1.2 Psychometric assessment

Traditionally, psychometric assessment has focused on three main criteria: reliability, validity and responsiveness. Broader approaches have contributed the concepts of acceptability and feasibility. In a recent structured review on the evaluation of patient-based outcome measures for use in clinical trials, Fitzpatrick and colleagues (1998) have widened this framework for assessment further still, to include three additional criteria that have received increasing attention in the recent methodological literature, namely appropriateness, precision and interpretability. This eight-criteria framework, shown in Table 1, was the product of a detailed and carefully considered review, which included assessment by and incorporation of the views of an expert panel with diverse interests and approaches.

Table 1 Framework for assessment of psychometric properties (Fitzpatrick *et al.*, 1998)

Criterion	Definition
Conventional	
Reliability	Does the outcome measure produce results that are reproducible and internally consistent?
Validity	Does the outcome measure measure what it claims to?
Responsiveness	Does the outcome measure detect changes over time that matter to the population for whom it was intended?
Acceptability	Is the outcome measure acceptable to the population for whom it is intended?
Feasibility	Is the outcome measure easy to administer and process?
Additional	
Appropriateness	Is the content of the outcome measure appropriate to the questions to which the evaluation is intended to address?
Precision	How precise are scores on the outcome measure?
Interpretability	How interpretable are scores on the outcome measure?

Four central points made by Fitzpatrick and colleagues are worth reiterating here. Firstly, for none of the eight criteria they give are there absolutely explicitly defined and universally accepted understandings, and there are uncertainties and differences of view in many areas. Secondly, it is not possible from the evidence to rank order the relative importance of the eight criteria in relation to decisions about the selection of measures to include in an evaluation. Thirdly, there will sometimes be a trade-off between one criterion and another: for example, a scale with a large number of relevant items may have high validity, but at the possible cost of lower acceptability and feasibility. Finally, several properties of measures are context-specific: the reliability of a measure, for instance, is not a fixed property but is dependent upon the context and population studied, whereas validity can only be assessed in relation to a specific purpose and setting. Given these four considerations, Fitzpatrick and colleagues conclude that the selection of measures remains to some extent as much an art as a science. We envisaged that the current study, in mapping the properties of carer outcome measures, would aid researchers, providers and policy-makers in performing this art.

1.3 Terminology

- The term carers will be used as a proxy for carers for people with mental health problems or dementia, unless otherwise specified.

- The term mental health problems will be used for all mental health problems other than dementia. People with dementia will be referred to as such. We recognise that, in practice, dementia is considered a mental health problem, and have simply used this distinction here for clarity.
- The term caring from a distance will be used to refer to carers who do not live geographically close to the care recipient. Rather than attempting to specify a geographical distance, this category was applied to carers who perceived themselves as caring from a distance.
- Carers from ethnic minority groups, young carers, gay and lesbian carers and carers caring from a distance will be referred to as carers from minority groups.
- The term service is used to refer to either voluntary or statutory services.
- The people who staff such services are referred to as health-care workers.

Section 2 Aims

We aimed to review available outcome measures and, on the basis of consultation with a diverse range of carers, generate data with which to advise the National Co-ordinating Centre for NHS Service Delivery and Organisation (SDO) programme of:

- instruments that could be used to measure the outcomes of services from carers' perspectives;
- the further work required to develop these instruments so that they are robust;
- any outcomes considered important by carers for which measures have not been developed.

These aims were achieved by:

- Conducting a consultation exercise with carers, representatives from carers' organisations and service providers to determine outcomes considered important by carers. This consultation exercise comprised both quantitative and qualitative methods.
- Conducting a comprehensive literature review using multiple methods to identify existing instruments.
- Assessing each of the instruments identified to establish acceptability, appropriateness, reliability, validity, responsiveness, precision, interpretability and feasibility.

Section 3 Consultation exercise

Carers, representatives from carer organisations and service providers were consulted in two ways. First, interviews were conducted, as either one-to-one interviews or focus groups. Second, a questionnaire was distributed to delegates at a conference organised by Rethink in March 2003.

3.1 One-to-one interviews and focus groups

3.1.1 Methods

A consultation exercise with carers, carer representatives and service providers was conducted to determine the outcomes that they perceived to be important. To maximise diversity, carers were identified via carer organisations using purposive sampling and snowballing. They were then invited to contribute by either participating in a focus group with other carers or discussing their views in a one-to-one interview with a researcher (A.L.).

To draw on the substantial expertise of carer organisations, we sought the views of their representatives in one-to-one interviews. This was designed to maximise the diversity of perspectives gathered: representatives were selected purposively so that they could comment on the particular needs of carers from groups we were unable to access directly. This included individuals identified as having particular expertise or playing a key role in developing and running support services for young carers, gay and lesbian carers and carers from ethnic minority groups. In view of the importance of service providers in influencing the choice of instruments used in service evaluations, their views were sought on the outcomes they considered it important to assess, and the practicalities of assessing services.

In total, nine carers and six representatives from carer organisations were interviewed individually. Four focus groups were run, involving a combined total of 35 carers and two organisational representatives. Four service providers were interviewed: one manager of an adult community mental health team (CMHT), one senior clinical psychologist in a CMHT providing services for older people, one service manager for the adult mental health service and a clinical governance manager of a mental health trust. Interviews took place in the interviewees' homes, or other place convenient to them; service providers were interviewed in a private room in their workplace; three focus groups were held in meeting rooms at the carers' centres and one was held in a conference room provided by the research team.

The topic guide for the interviews and focus groups is given in Appendix 1. At the outset of the study it was thought that simply asking participants which outcomes they considered important would be adequate. After a few interviews, however, it became clear that participants found it difficult to think about services in this way, or to articulate what was important to them in terms of which outcomes should be measured. To address this difficulty, the researcher therefore amended her approach. Participants were encouraged to talk about the aspects of their caregiving experience that they found difficult, and the aspects of services they thought important. They were then asked them to consider whether these were things they thought should be assessed when services were evaluated. Once spontaneously arising outcomes had been discussed, participants were shown a list of outcomes used in previous evaluations and asked to comment on which, if any, they thought were important. The list was based on that used in the conference questionnaire (described below) and is given in Appendix 2. The outcomes listed were as follows:

- the effect caregiving has on your family life,
- the effect caregiving has on your daily life,
- the sorts of caregiving activities that you do,
- the effect caregiving has on your social life,
- your well-being,
- your mental health (for example, whether you are depressed or anxious),
- your ability to cope,
- your physical health,
- your quality of life,
- how you feel about caregiving,
- amount of time spent providing care,
- how burdened you feel by caregiving,
- your satisfaction with the service,
- your confidence in providing care,
- your relationship with the person you care for,
- your knowledge of the person’s illness.

This task often resulted in additional outcomes being mentioned that were not on the list. In most cases interviews were lengthy and ranged beyond the topic guide. The data reported here relate only to participants’ views about outcomes and evaluating services.

Interviews and focus groups were recorded and transcribed. They were then analysed to identify common themes. Initially, two researchers (A.L. and K.H.) independently read each transcript and developed categories into which the data would be coded. They did this by separately recording each new response and developing categories that interlinked them. These categories were changed and refined throughout the process as understanding increased and improved. A.L. and K.H. then compared their categories and discrepancies were discussed until the final coding frame was agreed. The interviews were then coded using this framework. Rigour was improved by having two researchers involved throughout the process of analysis and by attaining a degree of triangulation with the use of three methods (interviews, focus groups and questionnaires).

Details of participants’ demographic and social characteristics and, where applicable, characteristics relating to their caregiving situation were collected using a short, self-report questionnaire administered during the interview (Appendix 3).

3.1.2 Results

Participants

In total 56 carers, representatives of carer organisations and service providers participated in the interviews: 19 participated in one-to one interviews and the remaining 37 participated in four focus groups. Their characteristics are shown in Table 2. While the majority of participants were white (including 15 Jewish participants) around a quarter came from ethnic minority groups. One of the participants, a representative of a carer organisation, represented young carers and one carer had been a young carer. One carer representative worked for the gay and lesbian sub-section of a carer organisation. At least four carers were known to be caring from a distance.

Table 2 Characteristics of interview participants (*n*=56)

Characteristic	**Carers (*n*)**	**Carer representatives (*n*)**	**Providers (*n*)**	**Total, *n* (%)**
Number of participants	44	8	4	56
Mean age in years (±S.D.)[a]	62 (±13.84)	40 (±14.74)	42 (±9.20)	57 (±16.22)
Gender				
Female	36	7	4	47 (83.9%)
Male	8	1	0	9 (16.1%)
Self-reported ethnicity				
White	34	6	3	43 (76.8%)
Black-Caribbean	3	1	1	5 (8.9%%)
Black-African	2	0	0	2 (3.6%)
Black-other	1	0	0	1 (1.8%)
Indian	0	1	0	1 (1.8%)
East African Asian	1	0	0	1 (1.8%)
Sri Lankan	2	0	0	2 (3.6%)
Russian	1	0	0	1 (1.8%)

[a] *Data missing for 17 carers and three carer representatives.*

Characteristics of the carers who participated in the interviews are shown in Table 3. Most were the care recipient's spouse/partner, parent/parent-in-law, or child. Of those who provided information, most were (living as) married, had face-to-face contact with the care recipient every day, and cared for someone with dementia. Of those who provided information, just over half lived with the care recipient. It should be noted that a substantial number of carers did not provide full information. All except one of the carers who participated in the interviews were a member of a carer organisation. Four also worked for a carer organisation.

Table 3 Characteristics of carers (*n*=44)

Characteristic	***n***
***Relationship to care recipient*[a]**	
Parent/parent-in-law	14
Child	12
Sibling	1
Spouse/partner	19
Friend	1
Marital status	
Single	5
(Living as) married	13
Separated/divorced/widowed	5
Missing	21
Face-to-face contact with the care recipient in the preceding 4 weeks	
Every day	16
Nearly every day	3
More than weekly	3
Weekly (four times)	1
Less than weekly (two to three times)	1
Once	1
Missing	19
Telephone contact with the care recipient in the preceding 4 weeks	
Every day	5
Nearly every day	0
More than weekly	3
Weekly (four times)	0
Less than weekly (two to three times)	1
Once	0
Missing	35
Living with care recipient	
No	11
Yes	15
Missing	18
***Care recipient's diagnosis*[b]**	
Dementia	20
Schizophrenia	8
Anxiety disorder	3
Bipolar/affective disorder	8

Missing	9
Member of a carer organisation	
No	0
Yes	42
Missing	2
Carer organisation worker	
No	26
Yes	4
Missing	14

[a]Total=47 because three carers cared for both a child and a spouse/partner.

[b]Total=48 because one carer cared for a person with anxiety disorder and bipolar illness, one carer cared for two people (one with schziophrenia and another person with depression) and one carer cared for two people with schizophrenia.

What outcomes should be assessed?

Value of evaluating services

Evaluating services was seen as important by both carers and providers, although it was noted by the young carer representative that the importance of doing so was not always obvious to young carers. Most carers recalled having been asked to complete questionnaires, usually postal, in the past. Few, however, recalled ever being invited to evaluate a service but they were keen to do so, perceiving it as an opportunity to make their voices heard. Providers recognised that evaluations could inform them about how to improve services by enabling them to understand carers' experiences better. Providers were aware, however, of the complexity of evaluating services and the need for evaluations to produce findings that could realistically be applied in practice.

Identifying outcomes

As discussed earlier, most participants found it difficult to articulate which outcomes they thought were important. A few carers argued that as long as they were asked about their experience, the specific outcome assessed was unimportant. By contrast, providers noted the need for focused outcomes that could be operationalised.

The complexity of carers' experiences was commented upon, in particular the fact that caregiving could be positive as well as negative and that experiences could seem contradictory: a carer could feel both burdened and happy, for example. Several participants noted that a carer's perception of a service is likely to vary over time depending on factors external to the service such as the carer's expectations or the care recipient's state of health. Participants thought it important that outcome measures accommodate these complexities.

Providers commented that carers' views on which outcomes are important were likely to differ from their own. One provider noted that it was sometimes difficult to determine what tangible benefits carers are receiving from a service, despite carers themselves being adamant that the service was of benefit to them.

Some participants identified outcomes specific to minority groups that should be assessed, for example whether professionals were sensitive to minority carers' particular needs and whether those needs were met, whether minority carers had difficulty accessing the service, and whether they would prefer integrated or specialist support services.

Participants were recruited to represent a broad range of views, namely carers for people with mental health problems, carers for people with dementia, carers from various ethnic minority groups, young carers, gay and lesbian carers and those caring from a distance. Relatively few differences were found according to these groups. Where differences specific to certain groups did occur, they are commented upon.

Listed outcomes

The majority of carers considered all the listed outcomes important and relevant. Several participants thought that the list was comprehensive. Providers noted, however, that resource constraints would make it impossible to assess all of the outcomes listed. One provider suggested that it would be more realistic to assess the specific outcome that the service aimed to change rather than attempt to assess broad improvements in less concrete concepts. Generally, however, participants preferred the idea of a broad outcome measure, such as quality of life, that provided an assessment of carers' holistic experience. Several participants suggested giving individual carers a list of potential outcomes and inviting them to assess those they considered important.

- The most commonly mentioned outcome was health. The majority of participants considered carers' mental health an important outcome, and many considered physical health important.
- Most participants considered the effect of caregiving on family life, daily life and social life important outcomes.
- The majority of participants considered carers' knowledge of the care recipient's illness an important outcome.
- Many participants considered carers' well-being, quality of life and ability to cope important outcomes. Indeed, several commented that quality of life would be an important outcome to assess because of its breadth.
- Several participants considered how carers feel about caregiving and their confidence in providing care important outcomes.
- Participants had mixed views on carers' relationship with the care recipient as an outcome, with some considering it important and others considering it unimportant.
- Participants' views on burden were mixed. Several participants considered it an important outcome, but others thought it too subjective. It was noted by several participants that many carers, especially those from minority groups, may not perceive caregiving as burdensome and may find the language of burden inappropriate.
- Participants were divided in their opinions about caregiving activities, with some considering the amount and type of caregiving activities undertaken an important outcome and others considering them unimportant. Although the division was not absolute, it was carers for people with dementia or their representatives that tended to consider caregiving activities important rather than carers for people with mental health problems or their representatives.

- Relatively few participants considered carers' needs an important outcome.
- Only one participant considered the amount of time devoted to caregiving important. Others considered it an unimportant outcome and several thought it unquantifiable.
- The majority of providers considered carers' satisfaction with the service an important outcome, but only one carer mentioned its importance. Nevertheless, many carers were keen to tell the interviewer how satisfied they were with the services they had received, which may indicate that satisfaction is an important outcome. On the other hand, several participants commented that the lack of satisfaction with services made it an unimportant outcome. One participant thought that assessing satisfaction with services alone would be inadequate.

The majority of participants thought that many of the listed items were closely related, and it was noted that young carers might be unable to discriminate between several of them. Many of the participants thought that several listed outcomes were related to quality of life, and several thought that all were encompassed within quality of life.

Non-listed outcomes

The most commonly mentioned outcomes not on the list were related to carers' views about the service. It appears that participants made a distinction between assessing the specific aspects of the service, and rating their satisfaction with services. The former was something they considered important, whereas they considered it unimportant to rate satisfaction with services. It is difficult to determine, however, whether this is a distinction that participants made, or a contradiction in their views.

- Many participants considered the attitude of health-care workers towards carers an important outcome; for example, whether they are sensitive to carers' needs, whether they are proactive, and whether carers are able to criticise services without fear of negative repercussions.
- Many participants also considered practical aspects of the service important, in particular whether the service is available when required, often early on in the caregiving experience; whether the service is accessible, both geographically and out-of-hours; and whether it is flexible. Other issues participants thought were important were the reliability of health-care workers and carers' confidence in them, and the extent and quality of communication between health-care workers and carers.
- Several participants thought carers should evaluate how effective the service was in terms of how the service benefited them and whether the service provided the support that they required.

Other outcomes mentioned as important by several participants included:

- whether the carer has received adequate information about the care recipient's illness and the services available (both statutory and voluntary),
- the adequacy of the support the carer receives, particularly emotional support,
- the effect of caregiving on the carer's finances.

Numerous other outcomes were mentioned as important by individual carers. The variability of these outcomes makes them difficult to group, and as they arose spontaneously it was not possible to establish to what extent other carers considered them important. These outcomes were:

- whether the carer feels reassured,
- whether the carer feels a sense of achievement,
- whether the carer feels empowered,
- whether the carer is better able to accept help,
- whether the carer has gained caregiving skills,
- whether the carer has gained more respect for himself/herself,
- whether the carer is able to take a break and have time to himself/herself,
- whether the carer is in contact with a support service,
- whether the carer receives recognition for caregiving,
- whether the carer feels his/her expertise is acknowledged,
- whether the carer is at risk from the care recipient,
- the effect of caregiving on the family,
- the effect of caregiving on children in the family,
- the effect of caregiving on the carer's relationships.

Most important outcome

There was no consensus among participants about the single most important outcome. Many participants considered it impossible to identify one single outcome as most important because it would vary according to the individual carer, the carer's circumstances and the service being evaluated. Two participants commented that the particular outcome evaluated should be that which the service aims to address.

Other participants did identify a single most important outcome, although few were identified as most important by more than one participant. The first five were listed outcomes, while the remaining six were outcomes identified by participants:

- the effect of caregiving on the carer's life, including family life,
- the carer's mental health,
- the carer's ability to cope,
- the carer's relationship with the care recipient,
- the carer's knowledge of the care recipient's illness,
- whether or not the carer wants to continue caring,
- whether the service provides emotional support,
- when the carer last had an uninterrupted night's sleep,
- when the carer last missed a meal,
- whether the carer feels recognised,
- whether the carer is in contact with support services.

How should outcomes be assessed?

Participants tended to distinguish between two methods of assessment: interviews, by which they meant open-ended questions to which they respond verbally, and questionnaires, by which they meant paper forms containing both closed and open-ended questions. The majority of participants preferred interviews to questionnaires, although it is possible that our sample were biased towards carers who preferred interviews because this is the way in which we collected our data.

Interviews would preferably be face to face, although some participants mentioned that they would be happy with telephone interviews. Participants were divided as to whether they preferred one-to-one or group interviews. Examination of participants' comments indicates that they were referring to focus groups, but since group interviews was the term used by participants it is adopted here. Two carers expressed no preference regarding the method of evaluation, while one provider acknowledged there were advantages and disadvantages to all methods. Two representatives suggested that it would depend on the individual carer, since some may not have time for an interview while others would value the opportunity to talk. The young carer representative noted that young carers need a different approach to service evaluation, the details of which are discussed below.

Assess outcomes using questionnaires

Participants agreed that carers are most commonly asked to evaluate services through questionnaires, and while some carers object to completing them, most do so. Questionnaires were generally unpopular, however, especially those that comprised closed questions. Several participants felt strongly that carers do not have the time to complete questionnaires and perceive them as yet another form to fill in. Some participants commented that carers particularly object to receiving questionnaires by post and three suggested that postal questionnaires are not completed because carers do not feel engaged in the evaluation process.

The most commonly mentioned limitation of questionnaires, particularly those containing closed questions, was that they fail to capture carers' experiences adequately. Participants commented that they do not allow carers to provide enough detail; the information obtained is inaccurate because carers' experiences do not correspond to the response options provided; and they do not cover the issues considered most important by carers. Several participants thought it likely that carers would be happy to make additional comments on questionnaires comprising closed questions, although two thought they would be unwilling to take the time to do so and one wondered whether any notice of these comments would be taken. Some participants perceived there to be a lack of confidentiality with questionnaires, and were concerned that expressing criticism would lead to negative repercussions.

Only one participant, a carer of a person with mental health problems, seemed to prefer a short questionnaire comprising closed questions and only one participant, a service provider, mentioned any advantages of questionnaires, namely that they were convenient and standardised.

Several participants, all either representatives or providers, considered the design and wording of questionnaires an important issue. Most agreed that questionnaires need to be short (one or two sides of A4 paper), broadly spaced and large print. The language needs to be informal ('everyday'), culturally sensitive and assumption-free. The questions need to engage carers, follow on from one another logically and avoid repetition. A representative of African-Caribbean carers noted that questionnaires raised literacy and language issues that need to be addressed. Several participants commented that response rates to postal questionnaires can be enhanced with the use of stamped-addressed return envelopes and proactive follow-up.

Assess outcomes using one-to-one interviews

Participants perceived several advantages of one-to-one interviews. Many thought that interviews were more personal, and that carers would find it easier to express themselves during interviews, particularly in relation to sensitive topics. A few commented that the information obtained would be richer and more meaningful. One participant noted that carers' thoughts are often disorganised, but in interviews carers can rely on the interviewer to interpret what they are saying. A few carers commented that interviews are more therapeutic and help with isolation, and several thought that interviewing the carer at home had the advantage of providing additional information about their situation. Several participants thought that the information obtained during interviews would be more influential than the information obtained from questionnaires.

Only service providers commented on the disadvantages of one-to-one interviews. One provider mentioned the financial cost, and two expressed concern that they were time-consuming and difficult to interpret.

Assess outcomes using group interviews (focus groups)

Several participants commented on the advantages of group interviews. Providers tended to note that they were a good compromise, providing rich data in a less resource-intensive way than one-to-one interviews. Several participants thought that the broader range of views characteristic of group interviews meant that a wider range of issues was discussed. Many participants thought that carers would be more comfortable in a group setting and more confident about expressing negative views because they could rely on the support of other group members.

Some disadvantages of group interviews were mentioned. It was thought that some carers might not feel at ease in a group, especially if it were a one-off meeting, and that they might feel inhibited about expressing their views. One participant noted that negativity can snowball in group interviews, resulting in the exclusion of any positive comments. The representative for gay and lesbian carers commented on the advantages and disadvantages of having groups specifically for gay and lesbian carers. She was clear that there was an advantage in having gay- and lesbian-only groups, in that some gay or lesbian carers would feel more comfortable expressing their views in such a group. She was also aware, however, that other gay or lesbian carers might feel that such a group would over-emphasise their sexuality when it was their caregiving role that was more relevant. She was not able to reach a conclusion about the relative merits of these two positions.

Assess outcomes using a choice or combination of methods

Several participants, all carers or representatives, thought that carers should be offered a choice of how to evaluate a service. It was thought that by doing so, more carers would be likely to respond. Other participants noted that a combination of methods could be used. For example, carers may be happy to complete brief questionnaires if they are combined with an interview, or the drawbacks of questionnaires could addressed by administering them in a face-to-face setting. Alternatively, questionnaires could be used to cover a broad range of issues and then interviews conducted to obtain greater depth of information about specific topics.

The young carer representative commented that young carers need a different approach to evaluation that takes into consideration their age and maturity. He noted that it was important that young carers be encouraged to spend any free time they have enjoying themselves rather than evaluating services. He thought that young carers might be willing to complete very short questionnaires, but that alternatives to conventional response categories, such as smiley faces, would be better. He also commented that older young carers might be happier with questionnaires conducted by e-mail. He suggested that more creative approaches to evaluations, such as group work and games, would be more successful than interviews and questionnaires.

Who should evaluate the service?

It was generally providers who expressed most concern about how to recruit carers to evaluate a service. A few representatives objected to random sampling and suggested that all carers should be involved in evaluation, but providers tended to see this as an unrealistic goal. Providers considered it best practice to obtain a representative sample, but were unsure how to achieve this. Only providers and representatives were asked about sampling issues, but one carer commented on them spontaneously. She urged that carers should be recruited through a variety of sources and that carer-support groups should not be relied upon since the majority of carers do not attend them and those who do not attend may be those experiencing most difficulty. By contrast, other participants thought that current support groups would be appropriate forums for evaluation of services.

Who should conduct the evaluation?

Participants' views about who should conduct evaluations were mixed. Some thought that it should be someone independent of services so that the carer can speak honestly and the findings are not biased. By contrast, others thought it should be conducted by someone whom the carer knows and trusts, and who understands the service and the carer's situation.

Feedback following the evaluation

Several participants mentioned the importance of feedback following evaluation. It was noted by one participant that carers are rarely given feedback following their participation in evaluations or forums. It was generally agreed by participants that carers would like rapid action following their evaluation as well as rapid feedback about how the information they have given is being used.

Professionals' approach to evaluating services

Several participants noted that professionals' approach to evaluation was important. The people evaluating services need to be sensitive and culturally aware, they need to clearly acknowledge the value of carers' input, they need to have good communication skills and an understanding of carers' experiences, and they need to engage with carers, perhaps by visiting them at home or attending their support group meetings. A few participants commented on the importance of anonymity and confidentiality if carers were to be able to evaluate services honestly.

3.2 Conference questionnaire

3.2.1 Methods

A short, self-report questionnaire was distributed to delegates at a conference run by Rethink in March 2003 (see Appendix 4). The questionnaire asked respondents to identify, first, all the important outcomes that should be measured from a given list and, second, the most important outcome from the list. The list was generated by examining the outcomes used in approximately 200 previous studies investigating services to support carers for people with mental health problems and selecting those most commonly assessed. Approximately 150 delegates attended the conference, and 30 questionnaires were completed and returned. Of these, two were excluded from the analyses because they were completed by delegates who were neither carers nor representatives of carer organisations.

3.2.2 Results

Respondents

Questionnaire data were available for 28 conference delegates (Table 4). Of these, 24 were carers and four were not. All of the four respondents who were not carers worked for a carer organisation.

Table 4 Characteristics of conference-questionnaire respondents (*n*=28)

Characteristic	***n***
Carers	***(24)***
Mean age in years (±S.D.)[a]	59.9 (±9.3)
Female[b]	19
Ethnicity[a]	
White	22
Chinese	1
Relationship to care recipient	
Parent	17
Child	1
Spouse/partner	2
Friend	2
Other	1
Care recipient's diagnosis	
Dementia	2
Schizophrenia	17
Depression	3
Bipolar/affective disorder	6
Living with care recipient	8
Carer-organisation member[a]	18

Carer-organisation worker[c]	11
Representatives of carer organisations	***(4)***
Mean age in years (±S.D.)	46.3 (±15.6)
Female	3
Ethnicity: white	4

[a] *Data missing for one respondent*

[b] *Data missing for two respondents*

[c] *Data missing for three respondents*

All important outcomes

As shown in Table 5, most of the 19 listed outcomes were endorsed as important by the majority of respondents. Only satisfaction with life was markedly less endorsed, with less than half of respondents identifying it as important.

Table 5 Outcomes considered important by conference delegates (*n*=28)

Questionnaire item	**Important (*n*)[a]**
The effect caregiving has on your daily life	25
Your satisfaction with the service	25
Your knowledge of the person's illness	25
The effect caregiving has on your family life	24
Your ability to cope	24
Your quality of life	23
Your physical health	23
The amount of time that you spend providing care	21
The effect caregiving has on you social life	21
Your needs	21
The relationship with the person you care for	20
The sorts of caregiving activities that you do	20
Your confidence in providing care	20
Whether you are depressed	19
Your morale	19
How you feel about caregiving	18
How burdened you feel by caregiving	18
Whether you are anxious	17
Your satisfaction with life	13

[a]*Data missing for two respondents.*

Most important outcome

There was little consensus regarding the most important outcome (Table 6). Seven respondents considered satisfaction with the service as most important. A further 10 items were endorsed as most important by between one and four respondents. The remaining eight items were not considered most important by any respondents.

Table 6 Outcome considered most important by conference delegates (*n*=28)

Questionnaire item	Most important (*n*)[a]
Your satisfaction with the service	7
Your knowledge of the person's illness	4
The relationship with the person you care for	3
Your ability to cope	3
Your quality of life	2
The effect caregiving has on your daily life	2
Your needs	2
The effect caregiving has on you social life	1
The effect caregiving has on your family life	1
How you feel about caregiving	1
Your physical health	1
The sorts of caregiving activities that you do	0
The amount of time that you spend providing care	0
Your satisfaction with life	0
Whether you are depressed	0
How burdened you feel by caregiving	0
Your confidence in providing care	0
Whether you are anxious	0
Your morale	0

[a] *Data missing for one respondent.*

3.3 Summary of outcomes identified in the consultation exercise

Summarising qualitative data inevitably results in a loss of richness. To achieve our aim, however, it was necessary to condense the rich data obtained in the consultation exercise into a manageable set of outcomes that reflected participants' views (shown in Table 7).

3.4 Development of 'importance to carers' criterion

The primary aim of the consultation exercise was to determine the outcomes considered important by carers, so that the instrument assessment exercise could be informed by their views. To achieve this, it was necessary to map the outcomes identified in the consultation exercise on to the constructs that the instruments were designed to measure. This task proved challenging. Consultation-exercise participants and researchers may have used different language to describe similar concepts; conversely they may have used similar language but have had different meanings in mind. Nevertheless, by close examination of the terms that participants used, and the context in which they used them, the two elements were synthesised to create Criterion A in the framework for assessing the properties of outcome measures. Details of particular decisions relating to the mapping exercise are discussed below.

Table 7 Summary of outcomes identified as important in the consultation exercise

Outcome
Outcomes relating to the carer
Carer's physical and mental health
Carer's needs
Carer's ability to cope
Carer's quality of life
Carer's morale
Whether the carer receives emotional support
Carer's relationship with the care recipient
Outcomes relating to caregiving
Effect of caregiving on carer's daily life (for example when the carer last missed a meal or had an uninterrupted night's sleep), family life, social life and finances
Amount and type of caregiving activities the carer undertakes
How the carer feels about caregiving, including how burdened he/she feels
Carer's confidence in providing care
Carer's willingness to continue caring
Carer's knowledge of the care recipient's illness
Outcomes relating to services
Whether the carer has received adequate information about the support services available
Whether the carer is in contact with support services
Attitudes of health-care workers
Accessibility and availability of the service
Effectiveness of the service
Carer's satisfaction with the service

The outcomes identified from the interviews were somewhat different to those identified from the conference questionnaire. Moreover, unlike the interviews, the conference questionnaire required respondents to rate the importance of suggested outcomes rather than generate important outcomes themselves. For this reason, the findings from the interviews and the conference questionnaire have been kept separate, both here and in the indicators for Criterion A.

We thought it informative to indicate the extent to which carers and carer representatives considered each outcome important, and so have reflected this in the benchmarks for Criterion A by distinguishing met from partially met outcomes.

Table 8 Criterion A

Indicator	Benchmark	Important outcomes from the consultation exercise
i Data from interviews with cares and carer representatives	Met	Carer's mental health Carer's ability to cope Carer's relationship with the care recipient Carer's willingness to continue caring Whether the service provides emotional support When the carer last had an uninterrupted night's sleep When the carer last missed a meal Whether the carer feels recognised Whether the carer is in contact with support services
	Partially met	Effect of caregiving on the carer's family life Effect of caregiving on the carer's social life Amount of caregiving activities the carer undertakes How burdened the carer feels How the carer feels about caregiving Carer's confidence in providing care Carer's needs, physical health, quality of life Carer's knowledge of the care recipient's illness Attitudes of health-care providers Accessibility and availability of the service Effectiveness of the service Whether the carer has received adequate information about the support services available Whether the carer receives emotional support The effect of caregiving on the carer's finances
ii Data from conference questionnaires	Met	Carer's ability to cope Carer's relationship with the care recipient Effect of caregiving on the carer's family life Effect of caregiving on the carer's social life Carer's confidence in providing care Carer's needs Carer's physical health Carer's quality of life Carer's knowledge of the care recipient's illness Effect of caregiving on the carer's daily life Carer's satisfaction with the service Amount of time the carer spends providing care Sort of caregiving activities the carer undertakes

	Partially met	Carer's depression, anxiety and morale How burdened the carer feels How the carer feels about caregiving Carer's satisfaction with life

For data collected in the interviews we distinguished between outcomes identified by at least one participant as *most* important and considered by no participant as unimportant, and outcomes identified as important by two or more participants. The two categories of outcome were designated as being met and partially met respectively. For data collected in the conference questionnaire we distinguished outcomes considered important by more than three-quarters of respondents (met) from outcomes considered important by between half and three-quarters of respondents (partially met). The results of this exercise are shown in Table 8. Several decisions made when mapping the findings of the consultation exercise on to the instrument assessment require elucidation. First, it was clear from the transcripts that when participants described carers' mental health as an important outcome they were referring to a broad range of constructs that included 'affect' and 'psychological well-being', as well as mental health problems. For this reason, when assessing instruments we treated all these constructs as corresponding to carers' mental health.

Second, another construct considered important by participants was the effect of caregiving on the carer's family life. In the consultation exercise, participants mainly mentioned this construct in relation to instrumental effects, such as whether the family could go on holiday. Broader effects, such as family members' relationships with one another and tension within the family, were not commented on in this context. They were, however, mentioned in other contexts. In the assessment exercise we therefore treated instruments that assess family functioning, environment and satisfaction as corresponding to the construct effect of caregiving on carer's family life.

Third, one participant-generated outcome was 'when the carer last has an uninterrupted night's sleep'. Superficially this appeared to map on to the Pittsburgh Sleep Quality Index, but on closer inspection it is clear that this instrument assesses insomnia and its causes, whereas in the consultation exercise participants discussed sleep in the context of care recipients' behaviour causing disturbance. We did not, therefore, treat this instrument as corresponding to the criterion.

Fourth, several instruments measured strain, stress relating to caregiving and appraisal of caregiving. We treated all these as corresponding to the construct how the carer feels about caregiving. We judged this construct distinct from the notion of burden, however, which was also measured by several instruments.

Finally, one instrument measured mastery (Mastery Scale). While at a conceptual level, mastery is arguably different from coping, for the purpose of this instrument assessment it was treated as corresponding to the construct 'carer's ability to cope'.

Some instruments, such as the Involvement Evaluation Questionnaire, measured more than one construct. Where this was the case, for the purpose of the instrument assessment exercise, Criterion A was treated as met if one or more of the constructs was important.

Section 4 Review of the literature

4.1 Methods

We conducted a literature review to identify papers describing instruments that had been used to measure outcomes for carers for people with mental health problems or dementia. Both published and unpublished papers were eligible for inclusion. Since our aim was to review all relevant measures and make recommendations for further development, we did not restrict our search to studies focusing on methodological issues but included studies that reported the use of a measure without methodological evaluation. The *Services to Support Carers for People with Mental Health Problems Scoping Exercise* (Arksey *et al.*, 2002) was used as a starting point, but its search strategy was extended to include other studies in which measures of carer outcome were employed but in which a service was not evaluated.

To be eligible for selection, papers needed to describe an outcome measure used with carers for people with mental health problems, or describe a study using such a measure. To be eligible for inclusion in the instrument review, outcome measures needed to have been used in studies of carers, whether or not they had originally been developed for them, and either (a) to have been used three times or more, not including by the instruments' authors or (b) to have been developed in the 10 years prior to the review (i.e. in 1994 or later). The latter criterion was included to ensure that we did not exclude reliable and valid instruments that had been developed only recently, and had not yet been used widely. Measures in categories (a) and (b) are reviewed separately.

We conducted an electronic search of Medline (1966–) and PsycINFO (1999–) using the following combination of search terms for caregiving and mental health: 'Caregivers'/ all subheadings OR (care$ ADJ2 giv$) OR carer$ OR informal care$ OR befriend$ OR (home ADJ2 care) OR ((famil$ OR neighbour$ OR relative$) ADJ2 (care OR caring)) OR ((famil$ OR neighbour$ or relative$) ADJ2 (support))) AND (explode 'Mental-Disorders'/ all subheadings OR 'Mental-Health'/ all subheadings OR (mental$ ADJ2 disorder$) OR mental health problem$ OR (mental$ ADJ2 ill$) OR (psych$ ADJ2 (disorder$ OR problem$)) OR dement$ OR schizophreni$ OR alzheimers OR (mental$ ADJ2 disease$) OR (psych$ ADJ2 ill$) OR (mental ADJ2 impairment) OR (eating ADJ2 disorder$) OR cognitive disorder$ OR (mood ADJ2 disorder$) OR (anxiety ADJ2 disorder$) OR personality disorder$ OR anorexia OR bulimia OR explode 'Substance-Related-Disorders'/ all subheadings OR hysteri$ OR neuros$ OR psychos$ OR manic depression OR (bipolar ADJ2 disorder$).

The full database-specific search terms are given in Appendix 5.

For reasons of time, the search was restricted to English-language items and, for PsycINFO only, to the last 5 years.

Key journals, *The Gerontologist* and *Ageing and Mental Health*, were hand-searched for the period 2001–2. A search of relevant internet sites – the Princess Royal Trust for Carers, the UK Government website for carers, Carersnet, the UK Department of

Health and the World Fellowship for Schizophrenia – was conducted. The key organisations through which the consultation exercise was arranged were approached to identify unpublished material. The expertise of the research team was also utilised to ensure the comprehensiveness of the literature review.

The instruments identified from the review were categorised according to the types of carer for whom they were developed and the constructs on which they were based; for example, physical health, caregiver strain or coping. To be included in the review, an instrument had to measure an aspect of carers' well-being. We applied this criterion in its broadest possible sense, and a wide range of different outcomes was assessed by the instruments we included. Our criterion did exclude some instruments, for example those that measure expressed emotion (such as Vaughn and Leff, 1976), because although they assess aspects of caregiving, they do not measure an aspect of carers' well-being.

4.2 Results

The reference list from the NCCSDO scoping exercise (Arksey *et al.*, 2002) was referred to first and yielded 361 papers, of which 203 were deemed relevant. Of those, 135 were read; the remaining 68 were not retrievable because they were not available from any London library.

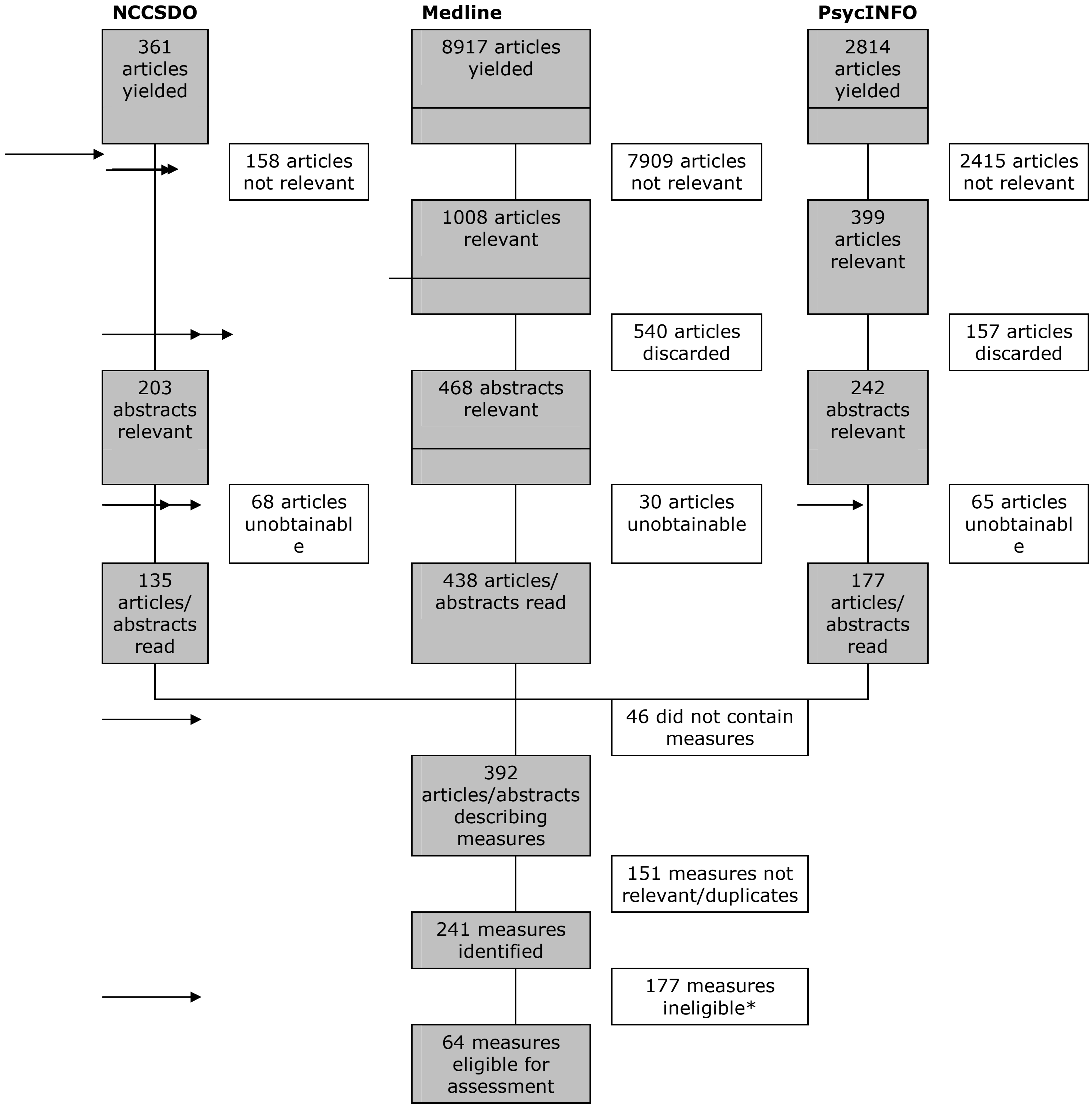
The Medline search yielded 8917 items, of which 1008 were deemed likely to be appropriate studies and saved in a Reference Manager database. On inspection of the abstracts, 395 were discarded for not using outcome measures with carers, 87 were discarded because they were reviews and 58 were discarded because they were comment articles. Of the 468 that remained, 123 contained details of measures used in their abstract so were not retrieved, but the information was extracted directly from the abstract; 315 were read in full; and the remaining 30 were not retrievable, either because they were not available in any London library or because of the time frame of the study.

The search of PsycINFO (last 5 years only) yielded 2814 items, of which 399 were deemed likely to be appropriate studies. On inspection of the abstracts, another 157 were discarded for not using outcome measures with carers. Of the 242 that remained, 54 contained details of measures used in their abstract so were not retrieved, but the information was extracted directly from the abstract; 123 were read in full. The rest were not available within the time frame of the study.

The total number of articles read which contained measures included in the review was 392. The total number of measures identified was 241; 53 of these had appeared more than three times so were eligible for psychometric assessment and are reviewed below. Four of them were ineligible for assessment and are presented in Appendix 6 (see also References 6). The 49 eligible measures used three or more times and 15 scales used fewer than three times and developed in the past 10 years are reviewed separately below. The process of identifying eligible published instruments is shown in Figure 1.

Searching internet sites produced one further measure, and requesting measures from carer organisations produced five further measures.

Figure 1 Process of identifying instruments eligible for psychometric assessment



**Measure used less than three times and developed more than ten years ago or did not measure an aspect of carer well-being.*

Section 5 Instrument assessment

5.1 Methods

5.1.1 Framework for assessment

Our assessment of the properties of outcome measures takes the eight criteria proposed by Fitzpatrick and colleagues (1998) as its basic framework (Table 9; Criteria C–J), but extends it by adding two further criteria central to our research remit of identifying outcome measures considered important by carers themselves. Given the orientation of this review, we have placed them at the top of the table. Criterion A is the extent to which carers or carer representatives view the construct being measured as important. Criterion B is the extent to which carers or carer representatives were involved in the development of the measure. The items generated in the consultation exercise for use as benchmarks for Criterion A are given above (Table 8).

5.1.2 Criteria, indicators and benchmarks

For each criterion, a number of indicators were identified. Where appropriate, we identified benchmarks for use in deciding whether on each indicator the criterion was met, partially met or not met. Decisions about which indicators to use, and where to set the benchmarks, were guided by the existing literature (Fitzpatrick *et al.*, 1998), discussion within the research team and the need to select indicators and benchmarks that were relatively unambiguous (for criteria and benchmarks, see Table 9). There is inevitably some subjectivity in the selection of the benchmarks, and what is deemed met or partially met. For several indicators, such as the number of items, it was not appropriate to set benchmarks. Whether or not the number of items is unfeasibly great, for instance, will depend upon the overall context of the study and the nature of the measure. How many measures are being used in the study would be a relevant factor, while the appropriateness of having a larger number of items might depend on whether or not it was covering several domains of outcome. In such cases, the factual information (such as number of items) is presented. For a small number of indicators there is no partially met benchmark, as the indicator is an all-or-nothing variable (e.g. short form available).

Table 9 Framework for assessing properties of outcome measures

Criterion	Definition	Indicators	Benchmarks (where appropriate)

A Importance to carers	Is the construct that the instrument measures important to carers?	(i) Data from interviews with carers and carer representatives (ii) Data from conference questionnaire	(i) Met=identified as the most important outcome by at least one carer, and no carers considered it unimportant. Partially met=identified as an important outcome by two or more carers. (ii) Met=majority of carers (≥75%) view this construct as of importance. Partially met=50–74% of carers view this construct as important.
B Carer involvement in outcome measure development	Were carers involved in the development of the outcome measure?	(i) The outcome measure was devised for use with carers specifically (ii) Carer involvement in overall project to develop measure (iii) Carer involvement in item generation	(i) Met=the outcome measure was devised for use with carers specifically. (ii) Met=carer-led research; partnership research with carer organisation; members of research team are carers; or other significant ongoing carer involvement. Partially met=some other involvement, not covered above. (iii) Met=carers generated some of the items. Partially met=carers feedback on already generated items.
C Acceptability[a]	Is the outcome measure acceptable to carers?	(i) Acceptability to carers (ii) Acceptability to other populations (iii) Response rates (iv) Time to complete (v) Short form available (vi) Translated measure available	(i) Met=carers report that it has good acceptability. Partially met=carers report that it has some acceptability. (ii) Met=other populations report that it has good acceptability. Partially met=other populations report that it has some acceptability. N/A=it is a carer-specific measure. (iii) N/A (iv) N/A (v) N/A (vi) N/A

D Appropriateness[b]	Is the content of the outcome measure appropriate to the questions to which the evaluation is intended to address?	(i) Whether carers were included in the piloting of the measure (ii) Whether the reliability or validity of the measure has been established on a carer population	(i) N/A (ii) N/A
E Reliability	Does the outcome measure produce results that are reproducible and internally consistent?	(i) Split-half reliability (ii) Cronbach's alpha (iii) Correlation of each item with rest of scale (iv) Internal consistency assessed by factor-analytic or similar techniques (v) Test-retest reliability (vi) Inter-rater reliability	(i) Met=correlation of 0.7–0.9. Partially met=correlation of 0.50–0.69 or over 0.9. (ii) Met=alpha of 0.7–0.9. Partially met= alpha of 0.50–0.69 or over 0.9. (iii) N/A (iv) N/A (v) N/A (vi) N/A
F Validity[c]	Does the outcome measure measure what it claims to?	(i) Face validity: the measure seems to measure what its title suggests. (ii) Content validity: the measure covers the full range of what is to be measured. (iii) Construct validity: the measure behaves in a way that is consistent with what is known about the construct or theoretical propositions about it.	(i)–(iii) N/A
G Responsiveness	Does the outcome measure detect changes over time that matter to carers?	(i) Used to measure change over time? (ii) Evidence that change over time correlates with other measures? (iii) Are the changes over time found significant to carers?	(i) Met=used to measure change over time. (ii) N/A (iii) N/A

H Precision	How precise are scores on the outcome measure?	(i) Type of response categories (ii) Ceiling and floor effects (iii) Precision of scales (iv) Sources of potential bias in scoring	(i)–(iv) N/A
I Interpretability	How interpretable are scores on the outcome measure?	(i) Minimal difference considered important by people the outcome measure will be used on is presented	(i)–(vii) N/A
		(ii) Norms presented (iii) Cut off (e.g. for caseness) presented (iv) Calibrated against other life events or experiences (v) Number of times used in this review (vi) Existence of a meaningful overall score (vii) Number of subscales	
J Feasibility[d]	Is the outcome measure easy to administer and process?	(i) Mode of administration (ii) Number of items (iii) Ease of explaining measure to study participants (iv) Amount of researcher training required (v) Weighting used in scoring[e] (vi) Freedom of usage (vii) Cost	(i)–(vii) N/A

[a] *Indicators (iii)–(vi) are also, to some extent, feasibility indicators.*

[b] *Appropriateness, as defined by Fitzpatrick et al. (1998), is whether the content of the measure is appropriate to the questions to which the evaluation is intended to address. Criterion A has covered one aspect of this already. It is not possible to include here a generic judgement about which outcomes are appropriate to the evaluation of which types of carer interventions, as the answer is specific to the particular intervention. The remaining aspect of appropriateness, addressed here, is whether the measure is appropriate to the population for whom it is intended.*

[c] *Criteria A, B, C(i) and C(ii) all overlap, to some degree with face, content and construct validity.*

[d] *Indicators (i) to (iii) are also, to some degree, acceptability indicators.*

[e] *This could also be seen as an indicator of precision, although Fitzpatrick et al. (1998) state that weighted items do not necessarily have better precision as they may give an unjustified illusion of precision.*

5.1.3 Data extraction and assessment

One researcher (K.H. for Criterion A and H.W. for Criteria B–J) located and compiled data on the properties of each measure. The data for Criterion A (importance of the construct to carers) were the primary data collected by means of the consultation exercise. They were not collected to be specific to each measure; rather, the importance of the construct to carers was rated in view of the findings of the consultation exercise. The data for all the remaining criteria (B–J) were secondary data derived from the existing literature. The main sources of data about the properties of the measures assessed in Criteria B–J were the references located in the literature searches undertaken; relevant data were found here in original papers describing the development of the measure, in validation papers, and in other methodological papers that referred to the measure. A compendium of measures (Burns *et al.*, 1999) also provided further details of psychometric properties. A further search was also conducted in Medline and PsycINFO for each measure (by putting its name in the title field), to check for the existence of short or translated versions or additional articles describing its psychometric properties.

Where applicable, a rating was given (by J.C.) for each indicator of each criterion of whether the criterion was met, partially met or not met, or whether there was insufficient evidence. Areas of ambiguity were discussed in the research group, with psychometric advice from S.C. and statistical advice from S.W. Where no benchmarks were applicable, the data are presented.

5.2 Results: measures used three or more times

The psychometric data presented below are predominately those presented in the main or original paper describing the development of the measure (articles listed in References 2), although in some cases additional papers were found in the search, or additional information was found on websites (listed in References 3). Where we have commented on the absence of psychometric testing, it is always possible that such testing has been conducted and the results not found by our search. How often the measure has been used, however, and whether it has subsequently been validated for carers (when they were not its original target) has been recorded on the basis of the total findings of the review.

In the 49 tables that follow, ratings are only given when appropriate (see Table 9). Where statistical values (usually correlation coefficients) are given for several items, the range and mean value is given. The rating (partially) met should then be taken to indicate that some values fall within the partially met range and others within the met range; similarly, unmet–partially met should be taken to indicate that some values fall within the partially met range and some within the unmet range.

5.2.1 Published measures developed for carers for people with mental health problems or dementia

Table 10 Caregiver Burden Interview (Zarit et al., 1980)

The Caregiver Burden Interview was developed to measure the burden experienced by carers for a spouse with dementia. It comprises 29 statements that the authors judge relevant to carers, for example 'I feel uncomfortable when I have friends over', 'because of my involvement with my spouse, I don't have enough time for myself'. Carers were not themselves involved in the development of the measure and no evidence of its acceptability to them is presented. No reliability testing is reported, nor face or content validity. No testing of reliability or validity was performed.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: caregiver burden		
A Importance to carers		
(i) Data obtained from interviews	How burdened the carer feels	Partially met
(ii) Data obtained from questionnaires	How burdened the carer feels	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Yes	Met
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No (questions based on clinical experience and prior studies)	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	N/A	N/A
(iii) Response rates	Not reported	-
(iv) Time to complete	25 min (estimated in Burns *et al.*, 1999).	-
(v) Short form available	12-item version (O'Rourke and Tuokko, 2003). Exploratory and confirmatory factor analyses support a two-factor structure of responses. Subsequent to control for demographic variables, dementia illness features, and baseline depressive symptoms at baseline, responses to this brief Caregiver Burden Interview provide a significant increase to prediction of depressive symptoms at time 2 (R^2=0.24, P<0.01) with no additional variance provided by the 10 remaining items from the complete Caregiver Burden Interview (ΔR^2=0, not significant).	-

(vi) Translated measure available	Japanese (Arai *et al.*, 1997)	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	Yes	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	No	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Not reported	Unmet/no evidence
(iii) Correlation of each item with rest of scale	Not reported	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	Not reported	-
(v) Test-retest reliability	Not reported	-
(vi) Inter-rater reliability	Not reported	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	Not reported	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	Likert scale: not reported how many points	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	-

(ii) Norms presented	Not reported	-
(iii) Cut off (e.g. for caseness) presented	Not reported	-
(iv) Calibrated against other life events or experiences	Not reported	-
(v) No. of times used in review	93	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	None	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	29	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	None	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 11 Caregiver Reaction Assessment (Given et al., 1992)

The Caregiver Reaction Assessment (CRA) was developed to assess the reactions of family members caring for elderly persons with physical impairments, Alzheimer’s disease and cancer. The validity of the scale is good, and there is some evidence regarding its reliability.

Criterion indicator	**Data**	**Rating (where applicable)**
Constructs covered: the response to caregiving		
A Importance to carers		
(i) Data from interviews	How the carer feels about caregiving	Partially met
(ii) Data from questionnaire	How the carer feels about caregiving	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Yes	Met
(ii) Carer involvement in overall project to develop measure	Not reported	Unmet/no evidence
(iii) Carer involvement in item generation	Interviews with caregivers were used (as well as a literature review) to identify a set of recurring dimensions of the caregiving experience. These dimensions were used as guides for researchers to generate the 111 items.	Unmet

C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	N/A	N/A
(iii) Response rates	Not reported	-
(iv) Time to complete	Not reported	-
(v) Short form available	No	-
(vi) Translated measure available	Dutch	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	Yes	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Yes	-
E Reliability		
(i) Split-half reliability	Not tested	Unmet/no evidence
(ii) Cronbach's alpha	Alpha coefficient for the subscales of the CRA are as follows: Schedule (impact on daily schedule): 0.82 Esteem (whether caregiving is enjoyable and rewarding or whether it causes resentment): 0.9 Lack of family support: 0.85 Health (impact on caregiver health): 0.8 Finance (impact on caregiver's finance): 0.81	Met
(iii) Correlation of each item with rest of scale	During the development of the scale, variables with an MSA (measure of sampling adequacy) coefficient of less than 0.7 were excluded on the grounds that they showed little communality with the other variables and thus would not load on any common factors. Individual correlations were not reported.	
(iv) Internal consistency assessed by factor-analytic or similar techniques	After exploratory factor analysis, the CRA comprised 24 indicator items forming five distinct subscales. Confirmatory factor analysis allowed for a test of factorial invariance across relevant comparison groups (diseases of care recipient: Alzheimer's vs cancer, and patient–caregiver relationships (spouses vs non-spouses).	-
(v) Test-retest reliability	Not tested	-
(vi) Inter-rater reliability	N/A	-
F Validity		

(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	The five subscales of the CRA were correlated with the number of patient dependencies in activities of daily living and caregivers' levels of depression. The authors expected the depression measure to be more strongly related to the CRA subscales than the measure of Activities of Daily Living (ADL) dependencies. This is so because the effect of ADL dependencies is mediated by a host of other variables; including the relationship of the caregiver to the patient, the gender of the caregiver and the amount of support provided by the family. On the other hand, the CRA subscales measure psychological states which ought to be more closely related to caregivers' depression. Depression was measured by the Centre for Epidemiological Studies Depression Scale (Radloff, 1977). Ten dichotomous indicators were employed to measure patient ADL dependencies, including need for assistance in eating, dressing, grooming, bathing, toileting, etc. The correlations with the CRA subscales were as follows: With ADL dependencies: Esteem: -0.11 Finance: 0.25 Family support: 0.20 Health: 0.29 Schedule: 0.52 With caregiver depression: Esteem: -0.23 Finances: 0.34 Family support: 0.39 Health: 0.57 Schedule: 0.46 These correlations are all in the expected direction and have the expected magnitude.	-
G Responsiveness		
(i) Used to measure change over time?	The researchers wished to produce a scale that would be appropriate for the analysis of change in caregivers' attitudes. Longitudinal factor models were based on the hypothesis that factor loadings and error variances would be invariant over time. The factor model was chosen which had factor loadings that were highly significant, large and stable over time. Because the subscales of the CRA exhibit highly stable factor structures, this makes them ideally suited for the measurement of change in caregiver reactions.	Met

(ii) Evidence that change over time correlates with other measures?	No	-
(iii) Are the changes over time found significant to carers?	Unknown	-
H Precision		
(i) Type of response categories	Five-point Likert scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	No	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of uses in review	3	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	Five	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	24	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 12 Caregiving Hassles Scale (Kinney and Stephens, 1989)

The Caregiving Hassles Scale was designed to assess the daily hassles of caring for a family member with Alzheimer's disease. It differs from other caregiving stress and burden scales in its focus on the minor events comprising day-to-day caregiving, rather than caregiving events or caregiver responsibilities occurring over longer periods of time. It has good reliability and validity.

Criterion+indicators	**Data**	**Rating (where applicable)**
Constructs covered: the day-to-day experience of caring for someone with dementia		
A Importance to carers		
(i) Data from interviews	Effect of caregiving on the carer's family life, effect of caregiving on the carer's social life, amount of caregiving activities the carer undertakes, how the carer feels about caregiving	Partially met
(iii) Data from questionnaire	Effect of caregiving on the carer's family life, effect of caregiving on the carer's social life, effect of caregiving on the carer's daily life, sort of activities the carer undertakes, how the carer feels about caregiving	Met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Yes	Met
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	N/A	N/A
(iii) Response rates	Not reported	-
(iv) Time to complete	Not reported	-
(v) Short form available	No	-
vi) Translated measure available	No references to a translated version were found	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	Yes	-

(ii) Whether the reliability or validity of the measure has been established on a carer population	Yes	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Full scale: 0.91 Basic Activities of Daily Living subscale: 0.79 Instrumental Activities of Daily Living subscale: 0.75 Cognitive status of care-recipient subscale: 0.82 Behaviour of care-recipient subscale: 0.89 Social network of caregiver: 0.74	Partially met Met Met Met Met Met
(iii) Correlation of each item with rest of scale	The original 110-item Caregiving Hassles Scale was reduced to create the current version. Pearson correlations were computed between individual items and each subscale total. The analysis indicated that each item correlated most strongly with the subscale in which it had been placed rationally. Items sharing less than 25% of the variance with the subscale total were deleted from the subscale. Intercorrelations among the five subscales ranged from -0.1 to 0.67 (median, -0.37). All intercorrelations were not reported.	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	The response format of the scale precluded the use of factor analysis because hassle items were rated only if the event had occurred during the previous week. Events that either did not occur, or occurred but were not appraised as a hassle, received a score of 0.	-
(v) Test-retest reliability	Caregivers' responses on the initial administration were compared to their responses provided on the second administration (a day later) using Pearson product-moment correlation coefficients: Full scale: 0.83 Basic Activities of Daily Living subscale: 0.86 Instrumental Activities of Daily Living subscale: 0.71 Cognitive status of care-recipient subscale: 0.80 Behaviour of care-recipient subscale: 0.87 Social network of caregiver: 0.66	-
(vi) Inter-rater reliability	N/A	-
F Validity		

(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-

(iii) Construct validity	Construct validity was assessed in two ways. Caregivers' reports of care recipients' impairment as assessed by the London Psychogeriatric Rating Scale (Hersch *et al.*, 1978) were correlated with the corresponding subscales of the Caregiving Hassles Scale. Correlations indicated significant associations between objective measures of impairment and hassles associated with basic ADL (r=0.44) and behaviour hassles (r=0.31), although the association between cognitive impairment and cognitive hassles failed to reach statistical significance (r values not reported). Similar measures were not available for instrumental ADL hassles or social network hassles. Second, construct validity was assessed by examining the relationship between hassles and measures of well-being representing the theorized outcomes of stress, measured by the Caregiver Social Impact Scale (Poulshock and Deimling, 1984) and the SCL-90-R (Derogatis, 1983). Neither hassles associated with basic ADL nor those associated with assistance in instrumental ADL were significantly related to any index of well-being. In general, behavior hassles demonstrated the strongest associations with caregiver well-being, followed by cognitive and social network hassles. ***Caregiver Social Impact Scale Correlations:*** With Full Hassles Scale: Activity restriction: -0.18 Negative relations: 0.51 With Basic ADL subscale: Activity restriction: -0.12 Negative relations: -0.15 With Instrumental ADL subscale: Activity restriction: -0.13 Negative relations: 0.23 With cognitive status subscale: Activity restriction: -0.22 Negative relations: 0.47 With behaviour of care-recipient subscale: Activity restriction: -0.16 Negative relations: 0.68 With social network subscale: Activity restriction: 0.9 Negative relations: 0.35 ***SCL-90-R Correlations:*** With Full Hassles Scale: Anxiety: 0.32 Hostility: 0.31	-

G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	Yes/no to whether the behaviour has occurred in the last week, and then a 4-point Likert scale on how much the behaviour has been a hassle (not at all, 0, to a great deal, 4)	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	-
(ii) Norms presented	For a sample of 60 primary caregivers providing care to family members diagnosed with probable Alzheimer's disease (mean±S.D.): Total 31.8±20 Behaviour 9.3±8.7 Cognitive 8.9±6.2 Basic ADL 6.4±5.6 Instrumental ADL 4.8±4.2 Social network 2.4±3.2	-
(iii) Cut off (e.g. for caseness) presented	N/A	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of uses in review	3	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	5	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	42	-

(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 13 Experience of Caregiving Inventory (Szmukler et al., 1996)

The Experience of Caregiving Inventory (ECI) was designed specifically for carers for people with serious mental illness to assess their appraisal of caregiving. Respondents are asked to rate how often they have thought about 66 items, for example 'feeling unable to tell anyone of the illness'. The items were generated by carers themselves, who reported the instrument's acceptability. Its validity and reliability have been shown to be good, although there is no evidence of responsiveness to change over time.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: appraisal of caregiving		
A Importance to carers		
(i) Data obtained from interviews	How the carer feels about caregiving	Partially met
(ii) Data obtained from questionnaires	How the carer feels about caregiving	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Yes	Met
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	Yes (items generated through interviews with carers, singly and in groups)	Met
C Acceptability		
(i) Acceptability to carers	Authors report that carers' feedback revealed it to be readily acceptable.	Met
(ii) Acceptability to other populations	N/A	N/A
(iii) Response rates	Not reported for final 66-item version	-
(iv) Time to complete	Not reported	-
(v) Short form available	Not found	-
(vi) Translated measure available	Not found	-
D Appropriateness		

(i) Whether carers were included in the piloting of the measure	Yes	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Yes	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Reported for subscales: Difficult behaviours: 0.91 Negative symptoms: 0.89 Stigma: 0.82 Problems with services: 0.90 Effects on family: 0.82 Need to backup: 0.76 Dependency: 0.74 Loss: 0.79	 Partially met Met Met Met Met Met Met Met
	Positive personal experiences: 0.86 Good aspects of relationship: 0.82	
(iii) Correlation of each item with rest of scale	Item-scale correlations are reported in subscales: Difficult behaviours: range 0.65–0.76 Negative symptoms: range 0.60–0.77 Stigma: range 0.51–0.69 Problems with services: range 0.61–0.74 Effects on family: range 0.53–0.61 Need to backup: 0.35–0.58 Dependency: 0.42–0.54 Loss: 0.46–0.58 Positive personal experiences: 0.55–0.71 Good aspects of relationship: 0.49–0.65	For subscales; (Partially) met (Partially) met Partially met (Partially) met Partially met Unmet/partially met Unmet/partially met Unmet/partially met (Partially) met Unmet/partially met
(iv) Internal consistency assessed by factor-analytic or similar techniques	Measure was developed partially through factor-analytic techniques. The end result was 10 factors of 5–8 items each, which map on to the subscales detailed above.	-
(v) Test-retest reliability	Not reported	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Content based on reports from 626 carers	-

(ii) Content validity	Not reported	-
(iii) Construct validity	The ECI was based on a stress-coping model. To determine construct validity, multiple regression analyses were performed to test the extent to which the ECI, coping (MacCarthy and Brown, 1989), and mastery (Pearlin and Schooler, 1978) predicted: psychological distress (GHQ-28; Goldberg and Hillier, 1979), positive and negative affect (PANAS; Watson and Clark, 1988), self-rated physical health and number of GP visits in the preceding 6 months. The ECI alone accounted for 24% of the variance in psychological distress. Mastery added 6% and coping 10%. The finding for negative affect was almost identical, with the combination accounting for 49% of the variance. ECI accounted for 17% of the variance in positive affect. Mastery did not contribute but coping added a further 9%. ECI-neg (the 8 negative subscales) explained 11% of the variance in physical health, mastery a further 16%, and coping further 5%. ECI-neg explained 14% of the variance in GP visits, but coping made no significant further contribution. The authors report other evidence for good construct validity. Further analyses using new and independent data (Joyce *et al.*, 2000) revealed ECI scores alone accounting for 27% of the variance in psychological distress, a similar result to the original findings.	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	5-point Likert scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	-

(ii) Norms presented	Not reported	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	6	-
(vi) Existence of a meaningful overall score	Scores of the subscales can be summed into the ECI-positive and ECI-negative scores. Two subscales are positive and eight negative.	-
(vii) Number of subscales	10	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	66	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Available from the authors	-
(vii) Cost	Free	-

Table 14 Family Burden Interview Schedule (Tessler et al., 1992)

The Family Burden Interview Schedule is a self-report measure of burden developed for people caring for people with severe mental illness. It has 100 items, divided into 'care', 'control' and 'worry'. No evidence is presented of its acceptability to carers or their involvement in its development. Its internal consistency is good. Other types of reliability and validity data may be presented in the manual, but we were unable to get a copy of this from the University of Massachusetts within the time frame of the current study.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: burden of care		
A Importance to carers		
(i) Data obtained from interviews	How burdened the carer feels	Partially met
(ii) Data obtained from questionnaires	How burdened the carer feels	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Yes	Met
(ii) Carer involvement in overall project to develop measure	Not reported	Unmet/no evidence

(iii) Carer involvement in item generation	Not reported	Unmet/no evidence
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	N/A	N/A
(iii) Response rates	Not reported	-
(iv) Time to complete	1 h	-
(v) Short form available	The Family Burden Interview Schedule – Short Form (FBIS-SF; Gamache, 1996)	-
(vi) Translated measure available	Not found	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	Yes	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Yes	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	A shorter version of the scale has alphas reported as follows: Care subscale (over 3 waves): 0.76, 0.70, 0.78 Control subscale (over 3 waves): not reported, as they failed to meet a standard set by the authors of 0.70. This was not unexpected because issues of control occurred far more sporadically than problems in daily living within the 30-day measurement period. Worry subscale (over 3 waves): 0.77, 0.81, 0.85	Met Not 'Met' (cannot assess whether partially met) Met
(iii) Correlation of each item with rest of scale	Not reported	Unmet/no evidence
(iv) Internal consistency assessed by factor-analytic or similar techniques	Internal consistency information is reported in the manual. A factor analysis supported the 3 subscales of care, control and disruption.	-
(v) Test-retest reliability	Not reported	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	Construct validity information is reported in the manual.	-
G Responsiveness		

(i) Used to measure change over time?	There is some information about the instrument's sensitivity to change in the manual.	Could not be assessed
(ii) Evidence that change over time correlates with other measures?	See above	-
(iii) Are the changes over time found significant to carers?	See above	-
H Precision		
(i) Type of response categories	3-point Likert scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Unknown	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	5	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	3	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	100	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	0	-
(vi) Freedom of usage	Manual available from the Social and Demographic Research Institute, University of Massachusetts, Amherst, MA, USA	-
(vii) Cost	Unknown	-

Table 15 Involvement Evaluation Questionnaire – European Version (Schene and van Wijngaarden, 1992; van Wijngaarden, 2003)

The Involvement Evaluation Questionnaire (IEQ) was developed specifically for use with carers for people with mental illness. As in the original, the most recent version, the IEQ – European Version (IEQ-EU), comprises a core module that assesses the consequences of caregiving, which is typically used in conjunction with other modules assessing socio-demographic status, financial consequences, use of professional help and psychological distress via the 12-item General Health Questionnaire. Carers rate the demands and effects of caregiving, for example 'how often during the last four weeks have you helped the person to take proper care of her/himself?' or 'how often during the last four weeks have you been annoyed by the person's behaviour?'. The instrument's validity and reliability are good, and there is some evidence indicating its sensitivity to change.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: consequences of caregiving		
A Importance to carers		
(i) Data obtained from interviews	Carer's mental health, carer's physical health, effect of caregiving on the carer's family life, effect of caregiving on carer's social life, amount of caregiving activities the carer undertakes, how burdened the carer feels, how the carer feels about caregiving, the effect of caregiving on the carer's finances	Met
(ii) Data obtained from questionnaires	Effect of caregiving on the carer's family life, carer's physical health, effect of caregiving on carer's social life, carer's mental health (depression and anxiety), how the carer feels about caregiving, how burdened the carer feels	Met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Yes	Met
(ii) Carer involvement in overall project to develop measure	Not reported	Unmet/no evidence
(iii) Carer involvement in item generation	Three separate studies were conducted with carers. In the first, with Dutch carers, no items were considered to be missing. In the second, also with Dutch carers, the items considered to be missing were beyond the scope of the questionnaire. In the third, with UK, Spanish and Danish carers, items were considered to be missing but were not added because they were culturally specific.	Partially met
C Acceptability		
(i) Acceptability to carers	Feedback regarding acceptability was established through three separate studies with carers.	Met
(ii) Acceptability to other populations	N/A	N/A

(iii) Response rates	70–89% (Schene and van Wijngaarden, 1993; Tennakoon *et al.*, 2000; van Wijngaarden *et al*., 1996, 2000; Magne-Ingvar and Öjehagen, 2002; Wolthaus *et al.*, 2002)	-
(iv) Time to complete	The IEQ-EU core module takes 10 min to complete. With the addition of other modules the entire set takes 20 min to complete.	-
(v) Short form available	No	-
(vi) Translated measure available	The instrument was developed in Dutch and has been translated into English, Spanish, Danish and German.	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	Yes	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Yes	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Cronbach's alpha for the four subscales ranged from:	
	0.81–0.84 for tension;	Met
	0.64–0.78 for supervision;	(Partially) met
	0.81–0.84 for worrying; and	Met
	0.77–0.80 for urging.	Met
	Cronbach's alpha for the overall score ranged from 0.89 to 0.90	Met
(iii) Correlation of each item with rest of scale	Not reported	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	IEQ was developed through factor-analytic techniques. The end result was four factors of 6–9 items each, which accounted for 59.6% of the total variance.	-
(v) Test-retest reliability	The test-retest period was 6–22 days and intraclass correlation coefficients was 0.83–0.90 for the four subscales and the overall score.	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Found to be good in three separate studies.	-

(ii) Content validity	Found to be good in two separate studies with Dutch carers. In a third study with UK, Spanish and Danish carers, items were considered to be missing but were not added because they were culturally specific.	-
(iii) Construct validity	A study using the IEQ and the Experience of Caregiving Inventory indicates good convergent validity (Szmukler *et al.*, 1996), as does a study using the IEQ and the Burden on the Family Interview (Kramer, 1998). Further evidence for construct validity is demonstrated by several studies demonstrating the IEQ-EU's fit with the stress-appraisal-coping model of caregiving.	-
G Responsiveness		
(i) Used to measure change over time?	Yes	Met
(ii) Evidence that change over time correlates with other measures?	In a quasi-experimental study of the effectiveness of family interventions IEQ scores were found to have dropped for the experimental group after one year but remained unchanged for the control group (Stam and Cuipers, 2001). Two further studies demonstrated change over time after 3 months and after 8 weeks (van Wijk and Haveman, 1992; Dekker, 1996; van Wijngaarden *et al.*, 1996).	-
(iii) Are the changes over time found significant to carers?	Not reported	-
H Precision		
(i) Type of response categories	Mostly 5-point Likert scales	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	-
(ii) Norms presented	Not reported	-
(iii) Cut off (e.g. for caseness) presented	Not reported	-
(iv) Calibrated against other life events or experiences	Not reported	-
(v) No. of times used in review	8	-
(vi) Existence of a meaningful overall score	Yes	-

(vii) Number of subscales	IEQ-EU core module comprises 4 subscales, the entire set includes 4 additional modules.	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	Entire set comprises 80 items	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	Not reported	-
(vi) Freedom of usage	Available from authors	-
(vii) Cost	Free	-

Table 16 The Neuropsychiatric Inventory Caregiver Distress Scale (Kaufer et al., 1998)

The Neuropsychiatric Inventory Caregiver Distress Scale (NPI-D) is an adjunct to the Neuropsychiatric Inventory and was developed to assess the impact of neuropsychiatric symptoms in Alzheimer's disease patients on caregiver distress. It has good validity but good reliability has not wholly been demonstrated.

Criterion+indicators	**Data**	**Rating (where applicable)**
Constructs covered: caregiver distress caused by neuropsychiatric symptoms of care recipient		
A Importance to carers		
(i) Data obtained from interviews	Carer's mental health	Met
(ii) Data obtained from questionnaire	Carer's depression and anxiety	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Yes	Met
(ii) Carer involvement in overall project to develop measure	No	Unmet/no evidence
(iii) Carer involvement in item generation	No	Unmet/no evidence
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	N/A	N/A
(iii) Response rates	Not reported	-
(iv) Time to complete	Not reported	-
(v) Short form available	Yes	-

(vi) Translated measure available	No	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Yes	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Not reported	Unmet/no evidence
(iii) Correlation of each item with rest of scale	Not reported	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	Not reported	-
(v) Test-retest reliability	Test-retest administrations of the NPI-D in 20 caregivers showed a correlation of *r*=0.92 with a mean duration between test and re-test interviews of 4.5 days.	-
(vi) Inter-rater reliability	The intraclass correlation between two raters of the NPI-D in 16 caregivers was 0.96 and was similar across spouse and children caregivers.	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	The NPI-D was assessed by comparison with an abridged version of the Relatives' Stress Scale (RSS; Greene *et al*., 1982) with 69 caregivers. Caregiver NPI-D distress ratings were correlated significantly with the RSS (*r*=0.6). RSS ratings correlated strongly with NPI-D scores (*r*=0.64) even after controlling for degree of cognitive impairment based on the Mini-Mental State Exam (MMSE) score (*r*=0.61). MMSE scores showed a moderate correlation to RSS ratings (-0.3) but this association was markedly attenuated when controlling for the degree of neuropsychiatric disturbance based on the NPI-D score (*r*=-0.14).	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet

(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	6-point Likert scale (not at all to extremely distressing)	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	No	-
(ii) Norms presented	Total (mean±S.D.): 8.7±6.9 (range=0–36) Male caregiver scores: 8.2±8.7 Female caregiver scores: 9.0±6.2 Spouse caregiver scores: 8.2±6.2 Child caregiver scores: 9.7±8.1 Lives with care recipient: 9.0±7.2 Lives apart from care recipient: 8.1±6.2	-
(iii) Cut off (e.g. for caseness) presented	N/A	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of uses in review	3	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	None	-
J Feasibility		
(i) Mode of administration	Interview, interviewer-rated	-
(ii) Number of items	11	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	Not reported	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 17 Pearlin Caregiving Measures for Carers for People with Alzheimer's (Pearlin et al., 1990; Skaff and Pearlin, 1992)

The Pearlin Caregiving Measures are 15 self-report scales developed for carers for people with Alzheimer's disease. Carers were not involved in their development, nor is evidence of their acceptability to carers presented. There is evidence of good internal consistency among the scales, but no evidence of validity is reported.

Criterion+indicators	Data	Rating (where applicable)
Constructs covered: various caregiving concepts related to burden such as overload, role conflict, economic strains		
A Importance to carers		
(i) Data from interviews	How burdened the carer feels	Partially met
(ii) Data from questionnaire	How burdened the carer feels	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Yes	Met
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No. The items were developed both to map on to the conceptual scheme of the authors and also in response to exploratory research among spouses and adult children who care for relatives with Alzheimer's disease.	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	Not reported	Unmet/no evidence
(iii) Response rates	Not reported	-
(iv) Time to complete	Not reported	-
(v) Short form available	Not found	-
vi) Translated measure available	Spanish	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	Yes	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Yes	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence

(ii) Cronbach's alpha	For the scales as follows: A Cognitive status of patient: 0.86 (8 items) B Problematic behaviour: 0.79 (14 items) C Overload: 0.8 (4 items) D Relational deprivation: (i) Deprivation of intimate exchange: 0.77 (3 items) (ii) Deprivation of goals and activities: 0.67 (3 items) E Family conflict: (i) Issues of Seriousness/Safety: 0.8 (4 items) (ii) Attitudes and Actions Toward Patient: 0.86 (4 items) (iii) Attitudes and Action Toward Caregiver: 0.84 (4 items) F Job-caregiving conflict: 0.75 (5 items) G Economic strains: alpha not reported (5 items) H Role captivity: 0.83 (3 items) I Loss of self: 0.76 (2 items) J Caregiving competence: 0.74 (2 items) K Personal gain: 0.76 (4 items) L Management of situation: alpha not reported (5 items) M Management of meaning: (i) Reduction of expectations: 0.48 (3 items) (ii) Making positive comparisons: 0.63 (3 items) (iii) Construction of larger sense of illness: 0.49 (3 items) N Management of distress: alpha not reported (8 items) O Expressive support: 0.87 (8 items)	Met Met Met Met Met Partially met Met Met Met Met Unmet/no evidence Met Met Met Met Unmet/no evidence Unmet Partially met Unmet Unmet/no evidence Met
(iii) Correlation of each item with rest of scale	This is not reported for the scales	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	No	-
(v) Test-retest reliability	Not reported	-
(vi) Inter-rater reliability	Not reported	-
F Validity		
(i) Face validity	Not reported	-

(ii) Content validity	Not reported	-
(iii) Construct validity	Not reported	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	Ranges from 3- to 4-point Likert scales	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	-
(ii) Norms presented	Not reported	-
(iii) Cut off (e.g. for caseness) presented	Not reported	-
(iv) Calibrated against other life events or experiences	Not reported	-
(v) No of uses in this review	19	-
(vi) Existence of a meaningful overall score	The scales each yield a total score	-
(vii) Number of subscales	There are 15 scales. The Relational Deprivation scale has two subscales; the Family Conflict scale has three subscales; and the Management of Meaning scale has three subscales. None of the other scales have any subscales.	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	There are 15 different scales, with a total of 95 items overall. There is an average of 6.3 items per scale.	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	None	-

(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 18 Relatives' Stress Scale (Greene et al., 1982)

The Relatives' Stress Scale is designed to measure stress in people caring for elderly relatives with dementia. Examples of items that relatives are asked to rate include 'do you ever feel that you need a break?' and 'how much has your social life been affected?'. It was designed specifically for carers and piloted with them, and the reliability and validity testing, although limited, was specifically for this population.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: stress related to caregiving		
A Importance to carers		
(i) Data obtained from interviews	How the carer feels about caregiving	Partially met
(ii) Data obtained from questionnaires	How the carer feels about caregiving	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Yes	Met
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	N/A	Unmet/no evidence
(iii) Response rates	Not reported	-
(iv) Time to complete	5–10 min (estimate)	-
(v) Short form available	Not found	-
(vi) Translated measure available	Not found	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	Yes	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Yes	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	0.85 (3 subscales 0.72, 0.80 and 0.88).	Met

(iii) Correlation of each item with rest of scale	Not reported	Unmet/no evidence
(iv) Internal consistency assessed by factor-analytic or similar techniques	Factor-analytic techniques extracted three factors which map on to the three subscales; Personal Distress, Domestic Upset and Negative Feelings	-
(v) Test-retest reliability	Tested but not reported	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Reported as good	-
(ii) Content validity	Not reported	-
(iii) Construct validity	Measures of the care recipient's cognitive level, their self-care abilities and their behavioural disturbance were correlated with the subscale scores of the Relatives' Stress Scale	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	No	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	5-point scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not available	-
(ii) Norms presented	Not reported	-
(iii) Cut off (e.g. for caseness) presented	Not reported	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	14	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	3	-
J Feasibility		
(i) Mode of administration	Self-report	-

(ii) Number of items	15	-
(iii) Ease of explaining measure to study participants	Easy	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 19 Revised Memory and Behaviour Problems Checklist (Teri et al., 1992)

The Revised Memory and Behaviour Problems Checklist is a self-administered measure. Carers first rate the frequency of the care recipient's memory and behaviour problems. Carers then rate their reaction to these problems. It was developed specifically for carers for people with dementia, although its acceptability to them is not recorded. Its reliability and validity are good, although test-retest and inter-rater reliability evidence is not reported.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: carers' rating of and reaction to care recipients' memory and behaviour problems		
A Importance to carers		
(i) Data obtained from interviews	Not a construct mentioned by carers	Unmet
(ii) Data obtained from questionnaires	Not a construct mentioned by carers	Unmet
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Yes	Met
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	N/A	N/A
(iii) Response rates	Not reported	-
(iv) Time to complete	10 min	-
(v) Short form available	Not found	-
(vi) Translated measure available	Not found	-
D Appropriateness		

(i) Whether carers were included in the piloting of the measure	Yes	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Yes	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	0.84 for patient behaviour and 0.90 for caregiver reaction; subscale alphas ranged from 0.67 to 0.89	Met/met/ partially met
(iii) Correlation of each item with rest of scale	Items had to have a correlation of at least 0.50 with the total score in order to be included	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	Factor analysis confirmed 3 first-order factors consistent with the subscales of memory-related, depression and disruptive behaviours, and 1 general factor of behavioural disturbance	-
(v) Test-retest reliability	Not reported	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-

(iii) Construct validity	The validity of frequency items was investigated through Pearson correlations between the subscales and the Hamilton Depression Rating Scale (HDRS; Hamilton, 1967) and Mini Mental State Examination (Folstein *et al*., 1975) as well as depression and dementia diagnoses. The Depression subscale was positively associated with HDRS scores (0.44) and depression diagnosis (0.36) but not MMSE (-0.04) or dementia diagnosis (0.05) as predicted. The Memory-Related Problems subscale was significantly negatively correlated with the MMSE (-0.48) and positively with a dementia diagnosis (0.45) but not with the HDRS (0.00) or depression diagnosis (-0.01). Construct validity of the Disruption subscale could not be evaluated because no similar measures were available. Memory-Related Problems items also occurred more frequently in subjects with dementia than subjects without (P=0.001, t=5.56), whereas depression and disruption were not significantly different in the two groups. Thus discriminant validity was also demonstrated. Construct validity of the Reaction scale was evaluated with measures of caregiver depression (Centre for Epidemiological Studies Depression Scale; Radloff, 1977) and burden (Caregiver Stress Scale; Deimling and Bass, 1984). For caregiver depression reliability coefficients were 0.29, 0.31 and 0.26 for Memory-Related Problems, Depression and Disruption; for burden reliability coefficients were 0.32, 0.42 and 0.41.	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	5-point Likert scales	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		

(i) Minimal difference considered important by people the outcome measure will be used on is presented	Unknown	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	40	-
(vi) Existence of a meaningful overall score	1 total score for observable behavioural problems in the patient, 3 subscale scores and parallel scores for caregiver reaction	-
(vii) Number of subscales	3	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	24	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	None	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 20 Screen for Caregiver Burden (Vitaliano et al., 1991)

The Screen for Caregiver Burden is designed to measure distressing experiences associated with caring for a spouse with Alzheimer's Disease. It measures both the prevalence of caregiving experiences (objective burden) and carers' appraisal of them (subjective burden). Some reliability and validity testing has been conducted, with moderately good results, and it has been used to measure change over time.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: objective burden and subjective burden		
A Importance to carers		
(i) Data obtained from interviews	How burdened the carer feels	Partially met
(ii) Data obtained from questionnaires	How burdened the carer feels	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Yes	Met

(ii) Carer involvement in overall project to develop measure	None	Unmet
(iii) Carer involvement in item generation	Generated all items	Met
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	N/A	-
(iii) Response rates	Not reported	-
(iv) Time to complete	20 minutes	-
(v) Short form available	Not found	-
(vi) Translated measure available	French, German, Italian, Spanish (www.qolid.org/public/SCB.html); evidence exists that it has been translated into Chinese, but details unclear (Chou *et al.*, 2003).	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	Not reported	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Yes	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	0.85 for objective burden, 0.89 for subjective burden	Met
(iii) Correlation of each item with rest of scale	Not reported	Unmet/no evidence
(iv) Internal consistency assessed by factor-analytic or similar techniques	Not reported	-
(v) Test-retest reliability	Objective burden, 0.64; subjective burden, 0.70	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	5 measures of care-recipient functioning were related to objective and subjective burden. There was a much stronger association of caregiver distress with subjective burden than with objective burden.	-
G Responsiveness		

(i) Used to measure change over time?	Yes	Met
(ii) Evidence that change over time correlates with other measures?	Changes in scores over time were correlated with changes in care recipient and caregiver variables.	-
(iii) Are the changes over time found significant to carers?	Unknown	-
H Precision		
(i) Type of response categories	5-point scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Unknown	-
(ii) Norms presented	Available for carers	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	10	-
(vi) Existence of a meaningful overall score	2 scores	-
(vii) Number of subscales	2 (Objective burden and Subjective burden)	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	25	-
(iii) Ease of explaining measure to study participants	Not difficult	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 21 Social Behaviour Assessment Schedule (Platt et al., 1980)

The Social Behaviour Assessment Schedule (SBAS) is a semi-structured interview to assess the social functioning of people with mental illness and its impact on significant others. The elements of relevance to carers are: objective burden, distress arising from the patient's behaviour, distress arising from the patient's social performance and distress arising from objective burden. It was designed specifically for this group, although carers were not involved in its development and its acceptability to them is not reported. Inter-rater reliability was good where tested, but evidence for other forms of reliability and validity is not presented.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: psychiatric patients' social functioning and its impact on carers		
A Importance to carers		
(i) Data obtained from interviews	How burdened the carer feels and how the carer feels about caregiving	Partially met
(ii) Data obtained from questionnaires	How burdened the carer feels and how the carer feels about caregiving	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Yes	Met
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	
(ii) Acceptability to other populations	Not reported	
(iii) Response rates		-
(iv) Time to complete	45–75 min	-
(v) Short form available	Not found	-
(vi) Translated measure available	Spanish (Salvador-Carulla *et al.*, 1998)	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Not reported	Unmet/no evidence

(iii) Correlation of each item with rest of scale	Not reported	Unmet/no evidence
(iv) Internal consistency assessed by factor-analytic or similar techniques	Not reported	-
(v) Test-retest reliability	Not reported	-
(vi) Inter-rater reliability	Four raters assessed the ratings of 2 interviewers on 4 and 5 interviews respectively. The intraclass correlation coefficients of the 3 major subscales were: Behaviour – objective, 0.96; Social Performance – objective, 0.98; Adverse Effects – objective, 0.99; Behaviour – distress, 0.94; Social Performance – distress, 0.93; Adverse Effects – distress, – 0.92. (Inter-rater reliability of the subscales of background information, concurrent events and support to informant/informant's housing situation was not calculated).	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	Not reported	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	No	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	Varies from 3- to 5-point Likert scales. Most items are rated on at least three different rating scales.	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	To assess the extent of bias the authors examined the differences between mean scores of the 4 raters used to assess inter-rater reliability. An analysis of variance was carried out to estimate the extent to which the various raters were using the scales differently. It revealed that mean squares for raters were not significant, thus ruling out the existence of systematic bias between individual raters.	-
I Interpretability		

(i) Minimal difference considered important by people the outcome measure will be used on is presented	No	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	'Concurrent events' is a subscale in the interview. It is a record of important events affecting the life of the informant, patient or relatives and friends of the informant. It is considered to be essential in order to assess whether the distress and adverse effects rated in the schedule could have been influenced by events in the informant's life which have no logical connection with the patient.	-
(v) No. of times used in review	14	-
(vi) Existence of a meaningful overall score	No – aggregate scores can be calculated for the 6 subscales.	-
(vii) Number of subscales	6	-
J Feasibility		
(i) Mode of administration	Structured interview with patient's most closely involved relative or friend.	-
(ii) Number of items	Unclear from article	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	Requires interviewer practice and training – amount not reported. A combined training manual and rating guide is necessary.	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Unknown. May be unavailable as it is no longer sold by NFER-Nelson and it was not possible to contact the authors.	-
(vii) Cost	Unknown	-

Table 22 Social Satisfaction Scale (George and Gwyther, 1986)

The Social Satisfaction Scale is a self-report measure of carers' social participation and their satisfaction with it. It was devised specifically for carers for people with dementia, although its acceptability to them is not reported. Little reliability and validity data are presented.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: social participation and satisfaction with social participation		
A Importance to carers		
(i) Data obtained from interviews	Effect of caregiving on the carer's social life	Partially met

(ii) Data obtained from questionnaires	Effect of caregiving on the carer's social life	Met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Yes	Met
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	N/A	N/A
(iii) Response rates	89%	-
(iv) Time to complete	Not reported	-
(v) Short form available	Not found	-
(vi) Translated measure available	Not found	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	Yes	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Yes	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	0.79	Met
(iii) Correlation of each item with rest of scale	Not reported	Unmet/no evidence
(iv) Internal consistency assessed by factor-analytic or similar techniques	No	-
(v) Test-retest reliability	Not reported	-
(vi) Inter-rater reliability	Not reported	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	Not reported	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-

H Precision		
(i) Type of response categories	Not reported	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	Calibrated against patient–caregiver relationship and patient living arrangements	-
(v) No. of times used in review	3	-
(vi) Existence of a meaningful overall score	2 scores – amount of social participation, and a global score for satisfaction with time spent in social activities	-
(vii) Number of subscales	None	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	7	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	None	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 23 Strain Scale (Gilleard, 1984)

The Strain Scale is a self-report measure of subjective strain from caring for an elderly relative with dementia. Examples of questions are 'do you feel frustrated with your situation?' and 'would you say that your health has suffered from looking after your relative?' It was designed for this group originally, although no evidence is presented here of its acceptability to them or items being generated from carers themselves. No reliability or validity data are presented, nor evidence of its responsiveness to change over time.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: subjective strain		
A Importance to carers		

(i) Data obtained from interviews	How the carer feels about caregiving	Partially met
(ii) Data obtained from questionnaires	How the carer feels about caregiving	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Yes	Met
(ii) Carer involvement in overall project to develop measure	Not reported	Unmet/no evidence
(iii) Carer involvement in item generation	Not reported	Unmet/no evidence
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	N/A	N/A
(iii) Response rates	Not reported	-
(iv) Time to complete	Not reported	-
(v) Short form available	Not found	-
(vi) Translated measure available	Not found	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	Yes	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Not reported	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Not reported	Unmet/no evidence
(iii) Correlation of each item with rest of scale	Not reported	Unmet/no evidence
(iv) Internal consistency assessed by factor-analytic or similar techniques	Not reported	-
(v) Test-retest reliability	Not reported	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	Not reported	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-

(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	3-point Likert scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	-
(ii) Norms presented	Not reported	-
(iii) Cut off (e.g. for caseness) presented	Not reported	-
(iv) Calibrated against other life events or experiences	Not reported	-
(v) No. of times used in review	9	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	None	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	13	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	None	-
(vi) Freedom of usage	Available in Burns *et al*. (1999)	-
(vii) Cost	Free	-

5.2.2 Published measures developed for non-carer populations

Table 24 Affect Balance Scale (Bradburn, 1969)

The Affect Balance Scale is a self-report measure of psychological well-being, divided into positive and negative affect subscales. Examples of questions include ‘during the past few weeks did you ever feel pleased about having accomplished something?’ (positive affect) or ‘did you ever feel depressed or very unhappy?’ (negative affect). It was developed for use with the general population and has not been specifically validated for carers. Reliability and validity testing for the general population was limited, apart from test-reliability, which was good. Construct validity was tested by using ridit and gamma values to examine relationships between positive or negative affect scores and ratings of three single-item measures of happiness and life satisfaction (general happiness, 'getting the things you want out of life' and whether or not the respondent wanted aspects of their life to change).

Criterion+indicators	Data	Rating (if applicable)
Constructs covered: psychological well-being		
A Importance to carers		
(i) Data obtained from interviews	Carer's mental health	Met
(ii) Data obtained from questionnaires	Carer's depression and anxiety	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Unknown	Unmet/no evidence
(ii) Acceptability to other populations	Unknown	Unmet/no evidence
(iii) Response rates	Unknown	-
(iv) Time to complete	Not reported	-
(v) Short form available	Not found	-
(vi) Translated measure available	Cantonese, Castilian Spanish, Catalan, Dutch, French, German, Laotian, Vietnamese (www.qolid.org/public/ABS.html)	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Not reported	Unmet/no evidence
iii) Correlation of each item with rest of scale	Not reported	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	Cluster analysis supported the concept of two independent dimensions of affect – positive and negative.	-

(v) Test-retest reliability	Positive affect, 0.83; negative affect, 0.81; whole scale, 0.76; 3-day interval	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	Scores correlated with self-report happiness and life satisfaction ratings (average ridit value, 0.48; gamma value, 0.32).	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	Binary	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	There is a slight retest effect tending to shift the scale scores in the more positive direction.	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Unknown	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	Yes – the author examined the effect of significant aspects of his respondents' lives in relation to scores on the ABS, including marriage, work, and standard of living.	-
(v) No. of times used in review	17	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	2 (positive and negative affect)	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	10	-
(iii) Ease of explaining measure to study participants	Unknown	-

(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	None	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 25 Beck Depression Inventory (Beck et al., 1961)

The Beck Depression Inventory (BDI) was developed to measure depression in psychiatric patients, but also for use in other populations. It enquires about symptoms and attitudes in 21 categories, for example mood, guilt and social withdrawal. It can be administered by trained researchers or used as a self-report measure. The measure has been reasonably well validated, including its responsiveness to change over time, although it is of concern that Kruskal–Wallis tests had been used to assess the correlation of each item with the rest of the scale (see section Eiii in Table 25). It has not been validated specifically for carers, and its acceptability to carers has not been assessed.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: depression		
A Importance to carers		
(i) Data obtained from interviews	Carer’s mental health	Met
(ii) Data obtained from questionnaires	Carer’s depression	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	Not reported	Unmet/no evidence
(iii) Response rates	Not reported	-
(iv) Time to complete	20 min	-

(v) Short form available	Card form (May *et al.*, 1969; cited in Groth-Marnat, 1990), computerised forms also available. 13-item short form; similar reliabilities found for 13-item short form (Groth-Marnat, 1990). Correlations for reliability between the 21-item and 13-item short form range from 0.89 to 0.97 indicating that the short form is an acceptable substitute (Beck *et al.*, 1974). However, possible 'sins' of short-form development are noted (Smith *et al.*, 2000). BDI-11 (Beck *et al.*, 1996; see Steer *et al.*, 2000 for information on the clinical utility of the BDI-11).	-
(vi) Translated measure available	Persian, Arabic, Cambodian, Chinese, Dutch, French, German, Italian, Japanese, Korean, Persian, Polish, Portuguese, Serbo-Croatian (Roman script), Spanish, Turkish, Xhosa (the Victorian Transcultural Psychiatry Unit, Victoria, Australia) www.vt.pu.org.au/resouces/ translated_instruments/mi/mi-tbdi.html)	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Pearson=0.86; Spearman=0.93	Met/partially met
(ii) Cronbach's alpha	25 studies reported for nine psychiatric populations, alpha range=0.76–0.95, mean=0.86; for 15 non-psychiatric populations, alpha range=0.73–0.92, mean=0.81.	Met/partially met
(iii) Correlation of each item with rest of scale	$P>0.001$ (Kruskall–Wallis test; correlations not reported)	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	13 factor analytic studies carried out. Number of factors ranged from 3 to 7 (depends on extraction procedure). Recent studies suggested that it represents one syndrome which can be decomposed into three highly inter-correlated factors representing differing clinical populations.	-

(v) Test-retest reliability	Considered inappropriate. If there was too long a period between two tests, the consistency would be lowered because of the fluctuations in the intensity of depression that occur in psychiatric patients. Too short – memory effect. Instead, a clinical estimate was made by one of the psychiatrists at the same time as the administration and re-administration of the inventory. Interval varied from 2 to 6 weeks. Was found that changes in the score tended to parallel changes in clinical depth of depression.	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Constructed from a clinical consensus about depressive symptoms displayed by psychiatric patients. Over the years have been some changes in content of BDI to reflect changes in Diagnostic and Statistical Manual (DSM) criteria. Authors based items on how they found the pattern of symptoms to be in clinical practice.	-
(iii) Construct validity	35 studies which reported correlations between BDI and variety of concurrent measures of depression, e.g. correlation between BDI scores and clinical ratings were 0.55–0.96 for psychiatric patients (mean=0.72) and 0.55–0.73 for non-psychiatric patients (mean=0.6). Showed good sensitivity and specificity with respect to detecting major depression (Turner and Romano, 1984) and some discriminant validity between depressive disorders. Much more evidence of construct validity – biological, suicide, alcoholism, adjustment, anxiety and others.	-
G Responsiveness		
(i) Used to measure change over time?	Yes	Met
(ii) Evidence that change over time correlates with other measures?	See section on test-retest. Change in score on BDI correlated with change in clinical rating of patient by psychiatrist.	-
(iii) Are the changes over time found significant to carers?	Effect sizes not reported	-
H Precision		
(i) Type of response categories	4-point scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-

(iv) Sources of potential bias in scoring	It is cautioned that high scores for university students should not be interpreted as indicative of depression as BDI is highly correlated with other measures of psychopathology in student populations. May just represent maladaptive functioning.	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not available	-
(ii) Norms presented	Available for non-carers	-
(iii) Cut off (e.g. for caseness) presented	Available for non-carers	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	34 (including 4 uses of short form)	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	None	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	21	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	Not reported	-
(v) Weighting used in scoring	No	
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 26 Brief Symptom Inventory (Derogatis and Melisaratos, 1983)

The Brief Symptom Inventory is a brief self-report instrument assessing nine primary psychological symptoms (somatization, obsessive-compulsive, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, psychotic symptoms). It was developed from its longer parent instrument, the SCL-09-R (Derogatis, 1983). It is designed to assess the psychological symptom status of psychiatric and medical patients, as well as individuals who are not patients. Only a limited amount of reliability and validity testing had been performed, and not specifically for carers. Its acceptability to carers has not been assessed.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: psychological symptoms		
A Importance to carers		
(i) Data obtained from interviews	Carer's mental health	Met
(ii) Data obtained from questionnaires	Carer's depression and anxiety	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	None	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	Not reported	Unmet/no evidence
(iii) Response rates	Not reported	-
(iv) Time to complete	Less than 10 min	-
(v) Short form available	Is a short form of the SCL-90-R instrument	N/A
(vi) Translated measure available	Arabic, Chinese, English, French, French Canadian, Hindi, Korean, Norwegian, Spanish, Turkish (http://psychtest.com), Polish (Aroian *et al.*, 1995)	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	Unmet/ no evidence
E Reliability		
(i) Split-half reliability	Not reported	Unmet/ no evidence
(ii) Cronbach's alpha	Average (9 subscales)=0.78	Met
(iii) Correlation of each item with rest of scale	Not reported	Unmet/ no evidence
(iv) Internal consistency assessed by factor-analytic or similar techniques	Factor analytic techniques extracted 9 factors, lending weight to construct of nine symptom dimensions measured by BSI.	-

(v) Test-retest reliability	Average=0.81; range over 9 subscales=0.68–0.91	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported, but it is the brief form of the SCL-09-R, and reflects the 9 symptom dimensions of that scale.	-
(iii) Construct validity	Convergent validity – coefficients of greater than 0.30 with the clinical scales of the, the Wiggins Content Scales of the MMPI and the Tryon Cluster Scores. Also good convergence with subscales of these measures (although not as good as in its larger form, SCL-90-R).	-
G Responsiveness		
(i) Used to measure change over time?	Not used	Unmet/ no evidence
(ii) Evidence that change over time correlates with other measures?	Not reported	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	5-point scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	Unmet/ no evidence
(ii) Norms presented	Norms have been published for non-carers but were not reported in the article.	-
(iii) Cut off (e.g. for caseness) presented	Not available	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	40	-
(vi) Existence of a meaningful overall score	3 different scores can be calculated – the General Severity Index, the Positive Symptoms Distress Index and the Positive Symptom Total.	-

(vii) Number of subscales	9	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	53	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	Not reported	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 27 Caregiver Burden Scale (Montgomery et al., 1985)

The Caregiver Burden Scale is a 14-item self-report scale devised for use with carers for elderly dependent relatives. Carers were not involved in the project to develop the measure or in the item generation. Good evidence is presented for its internal consistency but not its validity, precision, responsiveness or interpretability.

Criterion+indicators	**Data**	**Rating (where applicable)**
Constructs covered: caregiver burden		
A Importance to carers		
(i) Data from interviews	Caregiver burden	Partially met
(ii) Data from questionnaire	Caregiver burden	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Yes	Met
(ii) Carer involvement in overall project to develop measure	Not reported	Unmet/no evidence
(iii) Carer involvement in item generation	Not reported	Unmet/no evidence
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	N/A	N/A
(iii) Response rates	Not reported	-
(iv) Time to complete	Not reported	-
(v) Short form available	Not found	-
(vi) Translated measure available	Hungarian, Israeli, Italian, Polish, Russian, Spain, Chinese	-
D Appropriateness		

(i) Whether carers were included in the piloting of the measure	No (carers for frail elderly people, not mental health)	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	No (carers for frail elderly people, not mental health)	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Objective burden subscale, 0.92 Stress burden subscale, 0.84 Demand burden subscale, 0.88	Partially met Met Met
(iii) Correlation of each item with rest of scale	Objective burden subscale range, 0.67–0.84 Stress burden subscale range, 0.6–0.74 Demand burden subscale range, 0.72–0.76	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	Confirmatory analysis indicated an excellent goodness of fit for three factors representing the three subscales of the measure.	-
(v) Test-retest reliability	Not reported	-
(vi) Inter-rater reliability	Not reported	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	Not reported	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	5-point Likert scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	-
(ii) Norms presented	Not reported	-

(iii) Cut off (e.g. for caseness) presented	Not reported	-
(iv) Calibrated against other life events or experiences	Not reported	-
(v) No. of times used in review	9	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	3	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	14	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	None	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 28 Caregiver Strain Index (Robinson, 1983)

The Caregiver Strain Index (CSI) is a self-report measure of strain, developed specifically for carers for people over 65 discharged from hip surgery or heart patients. The 13 items comprise things carers may find difficult when assisting someone discharged from hospital, for example 'sleep is disturbed', 'it is a physical strain'. Items were generated from the carers themselves, although no evidence is reported of its acceptability to them. Its internal consistency is good, but other kinds of reliability were not tested. Its construct validity appears to be good. Its reliability and validity have not apparently been validated with carers for people with mental health problems.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: caregiver strain		
A Importance to carers		
(i) Data obtained from interviews	How the carer feels about caregiving	Partially met
(ii) Data obtained from questionnaires	How the carer feels about caregiving	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet

(iii) Carer involvement in item generation	Adult children's interviews at 3 time-points over a 5-year period were reviewed systematically to identify the most common stressors in caring for an elderly parent. Ten stressors were identified and became the basis for the initial questionnaire. After a review of the relevant literature 3 new stressors were added. An 'other' category was added for caregivers to mention areas of strain that might have been missed. Only 4 people responded by naming another stressor so this item was dropped.	Met
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	N/A	N/A
(iii) Response rates	97%	-
(iv) Time to complete	Not reported	-
(v) Short form available	14-item Caregiver Burden Tool (Ferrell *et al*., 1995). Authors modified the CSI by rewriting and adding several items to assess the perceived impact of the care recipient's pain and other illness symptoms on caregiver burden, e.g. whether the patient's pain disturbs the caregiver's sleep or causes emotional adjustments. In addition, a 6-point rating scale was introduced to replace the dichotomous response format of the CSI. Authors state that further validation is required. Moreover, the Caregiver Burden Tool may be limited in its application to different palliative care patient groups by virtue of its primary focus on pain. (www.ircm.qc.ca/bioethique/english/publications/sample/article6.html)	-
(vi) Translated measure available	Spanish, Dutch, Japanese, Chinese (Chen and Hu, 2002)	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	Yes	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Yes	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	0.86	Met
(iii) Correlation of each item with rest of scale	Not reported	Unmet/no evidence

(iv) Internal consistency assessed by factor-analytic or similar techniques	Not reported	Unmet/no evidence
(v) Test-retest reliability	Not reported	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-

(iii) Construct validity	Construct validity was examined by analysing the relationships between CSI scores and a number of criterion variables that theoretically would reflect strain. This include the physical and psychological disabilities of the care recipients, a negative perception by caregivers of the impact of caregiving on their life style, their perceptions of undertaking a major involvement in caregiving, and a negative perception of the affective quality of the caretaking relationship. *Care recipient characteristics (correlations with CSI scores):* Age of care recipient: 0.18 Bradburn (ABS) overall happiness: -0.14 Bradburn (ABS) affect balance: -0.14 (Bradburn, 1969) Profile of Mood States anxiety: 0.31 Performs activities of daily living: -0.46 (Katz *et al*., 1970) Mental impairment: 0.39 (short portable mental-status questionnaire; Pfeiffer, 1975) Satisfaction with progress: -0.26 Rehospitalized within 2 months: 0.19 *Subjective perception of the caregiving relationship and caregivers' emotional status (correlations with CSI scores):* Caregiver is very involved in helping, 0.28 Caregiver perceives it is hard to give help, 0.67 Caregiver says could use help in caregiving, 0.50 Caregiver perceives helping as having changed lifestyle, 0.56 Lifestyle changed, but little, -0.42 Lifestyle changes were difficult, 0.43 Lifestyle changes were in a negative direction, 0.56; caregiver is satisfied with care recipient's understanding of problems in helping, -0.43 Caregiver gets along with care recipient, -0.20 Bradburn overall happiness, -0.24 Profile of Mood States anxiety, 0.31 Profile of Mood States depression, 0.20 Profile of Mood States hostility, 0.37	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-

(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	Binary	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	-
(ii) Norms presented	Not reported	-
(iii) Cut off (e.g. for caseness) presented	Not reported	-
(iv) Calibrated against other life events or experiences	Not reported	-
(v) No. of times used in review	6	-
(vi) Existence of a meaningful overall score	Yes: computed by summing the 0 and 1 responses for the 13 items.	-
(vii) Number of subscales	None	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	13	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 29 Centre for Epidemiological Studies Depression Scale (Radloff, 1977)

The Centre for Epidemiological Studies Depression Scale (CES-D) is a self-report measure of depression, designed for use in the general population as a tool for epidemiological studies. Respondents are asked to rate the presence of 20 symptoms, for example ‘I did not feel like eating; my appetite was poor’ and ‘I felt sad’. Principal components analysis established that it measures four factors: depressed affect, positive affect, somatic and retarded activity and interpersonal factors. Reliability and validity were reasonably good, although test-retest reliability was only modest after a few weeks and even lower after long intervals. Its acceptability to carers is unknown and it has not been validated for use with this group.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: depression		
A Importance to carers		
(i) Data obtained from interviews	Carer’s mental health	Met
(ii) Data obtained from questionnaires	Carer’s depression	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Unknown	Unmet
(ii) Acceptability to other populations	Author reports it as 'acceptable to the respondent'. Some items were unanswered in the piloting.	Met
(iii) Response rates	Average of 71.5% over four samples (different geographical areas).	-
(iv) Time to complete	5 min	-
(v) Short form available	CESD-10 (10-item): reliability and validity tested in sample of healthy older adults; good predictive accuracy compared to full-length version ($\kappa=0.97$; $P<0.001$). Cut-off scores for depressive symptoms were =16 for the full-length questionnaire and =10 for the 10-item version. Positive correlation with poorer health status scores ($r=0.37$) and strong negative correlation with positive affect ($r=-0.63$). Retest correlations comparable to those in other studies ($r=0.71$). We administered the CESD-10 again after 12 months, and scores were stable with strong correlation of $r=0.59$ (Andresen *et al.*, 1994).	-

(vi) Translated measure available	Cambodian, Dutch, Italian, Japanese, Spanish, Swedish (The Victorian Transcultural Psychiatry Unit, Victoria, Australia; www.vtpu.org.au/resources/ translated_instruments/mi/mi-cfesd.html), Dutch, French, Danish, German, Greek, Italian, Portuguese, Afrikaans, Spanish, Swedish (MAPI Institute, Quality of Life Instruments Database www.mapi-researchinst.com/translation02.asp? PA_ID=14)	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Average over four samples was 0.79 (range, 0.76–0.85). Highest split-half reliability was for the patient sample	Met
(ii) Cronbach's alpha	Average over four samples: 0.86 (range, 0.84–0.90). Highest reliability coefficient was for the patient sample.	Met
(iii) Correlation of each item with rest of scale	Range, 0.30–0.79; all over 0.20	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	Principal components analysis conducted. For all samples there were four eigenvalues greater than 1, which together accounted for a total of 45% of the variance. The pattern of factor loadings is quite consistent across the samples. Including items with loadings above 0.40, the factors are readily interpretable as: depressed affect; positive affect; somatic and retarded activity; and interpersonal.	-
(v) Test-retest reliability	4 mail-back samples were re-tested at 2, 4, 6 or 8 weeks, with correlations of 0.51, 0.67, 0.59 and 0.59 respectively. Four interview samples were re-tested at 3, 6 and 12 months, with correlations of 0.48, 0.54 and 0.49 respectively.	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	The items which comprise the scale are clinically relevant.	-

(iii) Construct validity	The CES-D scores discriminated well between psychiatric inpatient and general population samples, and moderately well among levels of severity within patient groups. 70% of a psychiatric inpatient sample but only 21% of a general population sample scored at and above an arbitrary cut-off score of 16. The correlations of the CES-D with the Hamilton Clinician's Rating Scale (Hamilton, 1960) and with the Raskin Rating scale (Raskin *et al.*, 1969) were moderate (0.44–0.54) at admission but after 4 weeks of treatment they were substantially higher (0.69–0.75).	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	4-point scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Low negative correlations with the Marlowe–Crowne scale of social desirability which suggests that there may be some general response set. The author considered the bias small, however, and it does not mask relationships with other variables.	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	No	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	Available for non-carers	-
(iv) Calibrated against other life events or experiences	The author investigated effect of negative life events on test-retest correlations. Those with no negative life events in the interval between test and retest had the highest correlations (0.54 with no significant life events in the previous year to both the first and the second administration, compared to 0.31 with significant life events prior to both the first and second administration). The more negative the event, the higher the depression score of those who experienced it.	-

(v) No. of times used in review	91	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	None	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	20	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 30 Coping Responses Inventory (Moos, 1988)

The Coping Responses Inventory is a self-report or structured interview measure. It measures 'approach coping responses', divided into 'logical analysis', 'positive reappraisal', 'seeking support and information' and 'taking problem-solving action'; and 'avoidance coping responses', divided into 'cognitive avoidance', 'acceptance or resignation', 'seeking alternative rewards' and 'emotional discharge'. It has two forms, one for adults and one for young people, and the psychometric information quoted above is for the adult form of the scale only. Its validity seems good and there is some evidence of internal consistency and test-retest reliability. It was developed for use with healthy adults, psychiatric patients or medical patients, and has not been validated specifically for use with carers.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: coping responses		
A Importance to carers		
(i) Data obtained from interviews	Carer’s ability to cope	Met
(ii) Data obtained from questionnaires	Carer’s ability to cope	Met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet

C Acceptability		
(i) Acceptability to carers	Unknown	Unmet/no evidence
(ii) Acceptability to other populations	Not reported	Unmet/no evidence
(iii) Response rates	Not reported	-
(iv) Time to complete	15 min	-
(v) Short form available	Not found	-
(vi) Translated measure available	Spanish version of initial Coping Responses Inventory (Perez, 1989)	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Average alpha reported=0.65 for women and 0.67 for men	Partially met
(iii) Correlation of each item with rest of scale	Not reported	Unmet/no evidence
(iv) Internal consistency assessed by factor-analytic or similar techniques	Not reported	-
(v) Test-retest reliability	The scale is moderately stable over 1 year: women, 0.43; men, 0.45	-
(vi) Inter-rater reliability	Not reported	-
F Validity		
(i) Face validity	The authors report that both content and face validity were built into the Coping Responses Inventory by formulating definitions of specific domains, preparing items to fit the construct definitions, and selecting items that were conceptually related to a dimension and were shown empirically to belong to that dimension.	-
(ii) Content validity	See above	-

(iii) Construct validity	The Coping Responses Inventory scale and earlier versions of it discriminate between problem and non-problem drinking adults, early- and late-onset problem drinkers, alcoholic patients and case-controls, and depressed patients and case controls. Among community samples, more reliance on approach coping and less on avoidance coping is associated with less concurrent and future distress and with stable functioning among individuals who have experienced a high level of stressors.	-
G Responsiveness		
(i) Used to measure change over time?	Authors claim that the scale can be used to depict an individual's current coping strategies and to describe shifts over time in how an individual copes with a major stressor. They do not report any empirical testing of this.	Unmet/no evidence
(ii) Evidence that change over time correlates with other measures?	No	-
(iii) Are the changes over time found significant to carers?	Unknown	-
H Precision		
(i) Type of response categories	4-point Likert scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Unknown	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	13	-
(vi) Existence of a meaningful overall score	No – the result is a profile of responses to a situation rather than a score.	-
(vii) Number of subscales	8	-
J Feasibility		
(i) Mode of administration	May be administered as a self-report inventory or a structured interview	-
(ii) Number of items	58	-

(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	Not reported	-
(v) Weighting used in scoring	Unknown – scoring information in manual	-
(vi) Freedom of usage	Must be ordered by Psychological Assessment Resources Inc.	-
(vii) Cost	A professional manual is $32, an introductory kit (with booklets) is $124	-

Table 31 Daily Hassles and Uplifts Scale (Kanner et al., 1981)

The Daily Hassles and Uplifts Scale is a self-report measure of the hassles and uplifts of everyday life, developed for the general population. Its scales cover the areas of work, health, family, friends, the environment, practical considerations and chance occurrences. Its test-retest reliability is good, but internal consistency was not reported. Construct validity is good but not assessed for carers specifically.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: the hassles and uplifts of everyday life		
A Importance to carers		
(i) Data obtained from interviews	Not a construct mentioned by carers	-
(ii) Data obtained from questionnaires	Not a construct mentioned by carers	-
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	N/A	N/A
(iii) Carer involvement in item generation	N/A	N/A
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	Not reported	Unmet/no evidence
(iii) Response rates	Not reported	-
(iv) Time to complete	Not reported	-
(v) Short form available	Not found	-

(vi) Translated measure available	The scales have been used in South Africa. French Canadian (Voyer and Vezina, nd; www.psy.ulaval.ca/~johnny/art/voyer.pdf); Dutch (Huizink *et al.*, 2001; www.library.uu.nl/Digiarchief/dip/diss/1933819/c6.pdf).	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Not reported	Unmet/no evidence
(iii) Correlation of each item with rest of scale	Not reported	Unmet/no evidence
(iv) Internal consistency assessed by factor-analytic or similar techniques	No	-
(v) Test-retest reliability	The scales were administered to a community sample once a month over a 9-month period. The average correlation of each monthly administration with every other one is as follows: Hassles Frequency, mean *r*=0.79 Hassles Intensity, mean *r*=0.48 Uplifts Frequency, mean *r*=0.72 Uplifts Intensity, mean *r*=0.60 (Ranges were not provided.)	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-

(iii) Construct validity	A correlation matrix was generated relating 9 months of hassles frequency to concurrent 9 months of negative affect (measured by the Affect Balance Scale; Bradburn, 1969) and the mean of these correlations computed. The same was done for uplifts frequency, and uplifts intensity with positive affect (also measured by the Affect Balance Scale) and hassles intensity with negative affect. Uplifts frequency was significantly correlated with positive affect (mean *r*=0.25) and so was uplift intensity (mean *r*=0.33). Hassles frequency was significantly correlated to negative affect (mean *r*=0.34) but not hassles intensity (mean *r*=0.11) Ranges were not given. The Hopkins Symptom Checklist (HSCL; Derogatis *et al.*, 1970, 1971) was administered during the second and tenth months of interviewing. Hassles and Uplifts intensity scores and the HSCL were unrelated. Averaging over 9 months of hassles, hassles and month 2 HSCL were correlated: *r*=0.6 for the total sample, 0.55 for men and 0.66 for women. For month 10, parallel results were 0.49, 0.41 and 0.60. Authors claim this provides construct validation for the Hassles scale via its relationship to the significant adaptational outcome of psychological symptoms.	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	3-point Likert scales	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	-

(ii) Norms presented	9-month means were as follows (with gender and age scores collapsed): Hassles frequency: 20.50 Hassles intensity: 1.47 Uplifts frequency: 49.50 Uplifts intensity: 1.77 Norms split by gender and age group are also available in the article.	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	3	-
(vi) Existence of a meaningful overall score	No. It is not strictly appropriate to create a summed score from the two scales. Various scores can be computed. Frequency of hassles or uplifts can be calculated by simply counting the numbers checked. Cumulated severity is the sum of the 3-point severity ratings, and intensity is the cumulated severity divided by the frequency. The authors used only the frequency and the intensity scores in their analyses, as the correlations between frequency and cumulated severity was very high for both scales (*r*=0.95 in both cases).	-
(vii) Number of subscales	2	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	117-items for the Hassles Scale	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 32 Family Adaptability and Cohesion Scale II (Olson et al., 1983)

The Family Adaptability and Cohesion Scale II (FACES II) is a self-report measure of families' adaptability and cohesion. Adaptability relates to the family's ability to be flexible whereas cohesion relates to the extent to which family members are separated or connected. Each family member is asked to rate 30 items, for example 'our family does things together', and 'family members feel very close to each other'. It was developed for use in the general population, with couples and families. Adolescents participated in its development. Its validity and reliability are good but have not apparently been studied in a carer population. A further revised version, FACES III, is available, but authors recommend using FACES II as it is more psychometrically sound.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: family adaptability and cohesion		
A Importance to carers		
(i) Data obtained from interviews	Effect of caregiving on the carer's family life	Partially met
(ii) Data obtained from questionnaires	Effect of caregiving on the carer's family life	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	Not reported	Unmet/no evidence
(iii) Response rates	Not reported	-
(iv) Time to complete	Not reported	-
(v) Short form available	FACESKGIV-16, a 16-item Japanese translation (also 8- and 32-item versions; Tatsuki, 1999); FACES III (Olson *et al.*, 1987), 20 items including family adaptability and cohesion evaluation scales	-
(vi) Translated measure available	Japanese (FACESKGIV-8, -16, -32; Tatsuki, 1999), Spanish (FACES III; Katerndahl, 2002)	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Cohesion subscale, 0.87; adaptability subscale, 0.78; total scale, 0.90	Met

(iii) Correlation of each item with rest of scale	Not reported	Unmet/no evidence
(iv) Internal consistency assessed by factor-analytic or similar techniques	Scale was reduced from 50 to 30 items on the basis of factor analysis. 2–3 items for each of the content areas. Factor analysis also established that the global score is the most empirically valid.	-
(v) Test-retest reliability	0.83 for cohesion and 0.80 for adaptability subscales at a 4–5-week interval.	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Reported to be 'very good'	-
(ii) Content validity	Reported to be 'very good'	-
(iii) Construct validity	Concurrent validity with the Dallas Self-Report Family Inventory: 0.93 with the cohesion subscale, 0.79 with the adaptability subscale.	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	5-point Likert scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Correlations with social desirability (authors have not reported which social desirability scale has been used): 0.35 for the cohesion subscale and 0.00 for the adaptability subscale.	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Unknown	-
(ii) Norms presented	Yes	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	4	-

(vi) Existence of a meaningful overall score	Yes – gives a 'Family Type' score	-
(vii) Number of subscales	2	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	30	-
(iii) Ease of explaining measure to study participants	'Very easy'	-
(iv) Amount of researcher training required	People using the scale are encouraged to read the manual carefully first.	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Permission must be sought from authors, who will send the manual.	-
(vii) Cost	Free	-

Table 33 Family Assessment Measure version III (Skinner et al., 1995)

The Family Assessment Measure version III (FAM-III) is a self-report instrument that provides quantitative indices of family strengths and weaknesses. It was devised for use with the general population. Carers were not involved in its development and its acceptability to them is unknown. It is based on the Process Model of Family Functioning that integrates different approaches to family therapy and research. It consists of three components: a General Scale, which focuses on the family as a system; a Dyadic Relationships Scale, which examines relationships between specific pairs; and a Self-Rating Scale, which measures the individual's perception of his or her functioning within the family. Its validity and precision are good, and there is some evidence regarding its reliability.

Criterion+indicators	**Data**	**Rating (where applicable)**
Constructs covered: family strengths and weaknesses		
A Importance to carers		
(i) Data from interviews with carers and carer representatives	Effect of caregiving on the carer's family life	Partially met
(ii) data from conference questionnaires	Effect of caregiving on the carer's family life	Met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		

(i) Acceptability to carers	Unknown	Unmet/no evidence
(ii) Acceptability to other populations	Unknown	Unmet/no evidence
(iii) Response rates	Not reported	-
(iv) Time to complete	30 min	-
(v) Short form available	Yes	-
vi) Translated measure available	The FAM-III is available in English and Spanish from the Mental Health Foundation at www.mentalhealth.org.uk; researchers may have translated it into other languages.	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence

(ii) Cronbach's alpha	*General Scale*	*Adults*	*Children*	
	Overall rating (35 items)	0.93	0.94	Partially met
	Subscales (5 items each)			Partially met
	Task Accomplishment	0.67	0.60	(Partially) met
	Role Performance	0.73	0.64	Met
	Communication	0.73	0.70	Met
	Affective Expression	0.74	0.71	Met
	Involvement	0.78	0.75	(Partially) met
	Control	0.71	0.63	(Partially) met
	Values and Norms	0.70	0.62	Met
	Social Desirability (7 items)	0.87	0.87	(Partially) met
	Defensiveness (8 items)	0.65	0.70	
	Dyadic Relationships			Partially met
	Overall rating (42 items	0.95	0.94	Met
	Subscales (6 items each)			Met
	Task Accomplishment	0.74	0.73	Met
	Role Performance	0.82	0.71	Partially met
	Communication	0.77	0.77	Partially met
	Affective Expression	0.59	0.55	(Partially) met
	Involvement	0.64	0.59	(Partially) met
	Control	0.72	0.68	
	Values and Norms	0.72	0.66	Met
	Self-Rating			Partially met/Unmet
	Overall rating	0.89	0.86	
	Subscales (6 items each)			Partially met/Unmet
	Task Accomplishment	0.51	0.40	Partially met
	Role Performance	0.53	0.27	Partially met
	Communication	0.67	0.58	Unmet
	Affective Expression	0.64	0.55	Unmet
	Involvement	0.44	0.44	Partially met/Unmet
	Control	0.39	0.39	
	Values and Norms	0.60	0.46	
(iii) Correlation of each item with rest of scale	Not reported			-

<table>
<tr><td>(iv) Internal consistency assessed by factor-analytic or similar techniques</td><td>Inter-correlations among subscales were worked out on a sample of 251 adults from the Family Interaction Project at the University of Pittsburgh. The results indicate a generally high correlation among the different subscales. When the correlation matrices were submitted to a principal components factor analysis, the variance accounted for by the first (principal axis) factor is substantial, relative to the second and remaining factors. This indicates that there is a large general factor underlying the FAM subscales. Nevertheless, the various subscales do provide some reliable unique variance, which warrant their separate use and interpretation.</td><td>-</td></tr>
<tr><td>(v) Test-retest reliability</td><td>In a study by Jacob (1995) a community sample of families completed the FAM on one occasion and then were sent a packet of booklets in the post and asked to complete their forms independently. On average, the time between completion of the two questionnaires was 12 days. The test-retest reliabilities were as follows (items are as for E(ii)).
<table>
<tr><th></th><th>Mothers</th><th>Fathers</th><th>Children</th></tr>
<tr><td>TA</td><td>0.48</td><td>0.56</td><td>0.45</td></tr>
<tr><td>RP</td><td>0.70</td><td>0.67</td><td>0.64</td></tr>
<tr><td>Com</td><td>0.58</td><td>0.61</td><td>0.61</td></tr>
<tr><td>AE</td><td>0.49</td><td>0.54</td><td>0.66</td></tr>
<tr><td>Inv</td><td>0.59</td><td>0.63</td><td>0.72</td></tr>
<tr><td>Control</td><td>0.48</td><td>0.46</td><td>0.69</td></tr>
<tr><td>V&N</td><td>0.57</td><td>0.56</td><td>0.72</td></tr>
</table>
</td><td>-</td></tr>
<tr><td>(vi) Inter-rater reliability</td><td>Not reported</td><td>-</td></tr>
<tr><td><i>F Validity</i></td><td></td><td></td></tr>
<tr><td>(i) Face validity</td><td>Not reported</td><td>-</td></tr>
<tr><td>(ii) Content validity</td><td>Not reported</td><td>-</td></tr>
</table>

(iii) Construct validity	The manual reports a large number of studies that have used the FAM with a large variety of subjects. A sample of them is presented here. To test family functioning against one member having a mental health problem, Jacob (1991) investigated 48 families that contained a father who was classified as clinically depressed. The mean scores obtained by fathers, mothers and children were high, indicating problems with family functioning. Levene (1991) examined 52 subjects with psychosis obtained from a sample of inpatient psychiatric units in 12 hospitals in and around Toronto, Canada. Those of their families who exhibited high expressed emotion had FAM scores showing poorer family functioning than those families with low expressed emotion. Garfinkel *et al*. (1983) examined several families that had a child with anorexia nervosa. Patients and their mothers completed the FAM and in general showed elevated scores relative to the norms. Scores of the patients were especially high. Skinner *et al.* (1983) examined the diagnostic power of the FAM-III General Scale. The sample included 'problem' families (defined as those having one or more family members receiving professional help for psychiatric/ emotional problems, alcohol/drug problems, school-related problems, or major legal problems) and 'non-problem' families. A multiple discriminant function analysis was conducted to determine whether the FAM subscales would significantly differentiate between the groups. Problem families in general reported more family dysfunction in the areas of Role Performance and Affective Involvement. Non-problem families had a slight tendency to score higher in Social Desirability and Defensiveness. The FAM was effective in differentiating the problem families from the non-problem families. Evidence shows that FAM scores correlate with scores from other family scales. Bloomquist and Harris (1984) administered the FAM-III General Scale and MMPI special family scales to their sample of undergraduates. They found a strong relationship between MMPI special family scales and FAM subscale scores. Jacob (1995) administered the FAM, the Family Environment Scale (FES; Moos, 1974; Moos and Moos, 1981), the Family Adaptability and Cohesion Scales (FACES; Olson *et al.*, 1983) and the Family Assessment Device (Epstein *et al.*, 1982). With the FACES, correlations with Cohesion were high (-0.39 to -0.55) but with Adaptability they were low (0.03–0.10). With the FES, correlations were high with Cohesion and Conflict (-0.33 to -0.63, 0.34 to 0.58) moderate with Expressiveness (-0.25 to -0.36), Intellectual-cultural Orientation (-0.21 to -0.32), Active-recreational Orientation (–0.15 to –0.24) and Organisation (-0.29 to -	-

G Responsiveness		
(i) Used to measure change over time?	Blackman *et al.* (1986) used the FAM to study treatment effectiveness in a group of 31 families with 'emotionally disabled' adolescents between 11 and 17 years old. The majority of the adolescents had either adjustment disorders, conduct disorders, affective disorders, psychoses, or anxiety disorders. The FAM was administered to both the adolescents and their parents, pre- and post-treatment. On analysis of the results, no significant differences were noted. Grizenko and Sayegh (1990) assessed 23 consecutive admissions to a psychodynamically oriented day treatment program using a pre-/post-test design. The FAM was completed by children, mothers, and fathers. In children, the FAM detected a significant improvement over the duration of the program (P<0.001). Improvement for mothers approached significance (P=0.05) but was not significant for fathers (P=0.10).	Met
(ii) Evidence that change over time correlates with other measures?	Not presented	-
(iii) Are the changes over time found significant to carers?	Unknown	-
H Precision		
(i) Type of response categories	4-point Likert scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Subscale scores can be checked against the scores in Social Desirability and Defensiveness, which are also included as subscales.	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	No	-

<table>
<tr><td>(ii) Norms presented</td><td>Norms presented for parents of children with cystic fibrosis, developmentally disabled children, children in foster families, for families with an alcoholic father, with a clinically depressed father, for parents of children with a severe degree of mental handicap, of schizophrenic patients, for anorexic patients and their mothers, and for bulimic patients, their mothers and their fathers. The mean raw scores obtained from a sample of undergraduate students are presented here:
<table>
<tr><th></th><th>Mean Raw</th><th>SD</th></tr>
<tr><td>TA</td><td>6</td><td>2.5</td></tr>
<tr><td>RP</td><td>6.4</td><td>5.3</td></tr>
<tr><td>Com</td><td>5.7</td><td>4.5</td></tr>
<tr><td>AE</td><td>5.5</td><td>2.7</td></tr>
<tr><td>Inv</td><td>4.5</td><td>6.4</td></tr>
<tr><td>Con</td><td>6</td><td>7.4</td></tr>
<tr><td>VN</td><td>5.3</td><td>2.3</td></tr>
</table>
These means are slightly higher than those obtained from the community normative sample on which the FAM QuikScore™ forms are based (not presented in the manual). Percentile information from the community sample is presented on each Profile Form, used to sum the answers.</td><td>-</td></tr>
<tr><td>(iii) Cut off (e.g. for caseness) presented</td><td>No</td><td>-</td></tr>
<tr><td>(iv) Calibrated against other life events or experiences</td><td>No</td><td>-</td></tr>
<tr><td>(v) No. of times used in review</td><td>3</td><td>-</td></tr>
<tr><td>(vi) Existence of a meaningful overall score</td><td>Yes</td><td>-</td></tr>
<tr><td>(vii) Number of subscales</td><td>7 in the Dyadic Relationships Scale (looks at specific pairs within the family) 7 in the self-rating scale (focuses on the individual's perception of his or her functioning within the family, and 9 in the General Scale (as it includes Social Desirability and Defensiveness as well).</td><td>-</td></tr>
<tr><td><i>J Feasibility</i></td><td></td><td></td></tr>
<tr><td>(i) Mode of administration</td><td>Self-report</td><td>-</td></tr>
<tr><td>(ii) Number of items</td><td>General Scale, 50 items; Dyadic Relationships Scale, 42 items; Self-Rating Scale, 42 items</td><td>-</td></tr>
<tr><td>(iii) Ease of explaining measure to study participants</td><td>Not reported</td><td>-</td></tr>
</table>

(iv) Amount of researcher training required	None (but some study of the manual beforehand is necessary)	-
(v) Weighting used in scoring	None	-
(vi) Freedom of usage	The FAM-III is available from the Mental Health Foundation at www.mentalhealth.org.uk	-
(vii) Cost	The FAM-III Specimen Set (which includes Manual, 5 General Scale QuikScore™ Forms, 10 Dyadic Relationship Scale QuikScore™ Forms, 5 Self-Rating Scale QuikScore™ Forms, 1 FAM-III ColorPlot™ Profile of Family Perceptions and 1 Progress ColorPlot) costs $47.	-

Table 34 Family Caregiving Inventory (Archbold and Stewart, 1986)

The Family Caregiving Inventory is a structured interview assessing carer role strain. Carers rate role strain arising from nine domains, for example strain arising from direct care or strain arising from economic burden. They are also asked about their feelings of mutuality, preparedness for caregiving and the amount of direct care they provide. Examples of items carers are asked to rate include ‘has assisting your family member increased the stress in your relationship with him/her?’ and ‘how much do you worry about your own ability to continue taking care of your family member because of your own health?’. It was developed for use with carers for frail elderly people, but evidence of its acceptability to them is not presented, nor whether they were involved in item generation. Reliability and validity are good.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: carer role strain		
A Importance to carers		
(i) Data obtained from interviews	How the carer feels about caregiving	Partially met
(ii) Data obtained from questionnaires	How the carer feels about caregiving	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Yes	Met
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	Not reported	Unmet/no evidence
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	N/A	N/A
(iii) Response rates	Not reported	-

(iv) Time to complete	2 h with carer, 1 h with care recipient	-
(v) Short form available	Not found	-
(vi) Translated measure available	Not found	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	Yes	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Yes	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Given for subscales as follows. The first alpha is 6 weeks after hospital discharge of the care recipient, the second is 9 months after hospital discharge.	Met except for
	Cognitive impairment of care recipient – N/A	
	Functional impairment of care recipient – N/A	
	Mutuality as perceived by carer – 0.91, 0.91	Partially met
	Preparedness for caregiving – 0.72, 0.71	Met
	Strain from direct care – not computed because of missing responses	
	Strain from lack of resources – 0.77, 0.78	Met
	Strain from worry – 0.84, 0.82	Met
	Strain from role conflict – not computed because of missing responses	
	Strain from economic burden – 0.74, 0.77	Met
	Strain from mismatched expectations – 0.53, 0.45	Unmet
	Strain from increased tensions – 0.86, 0.91	(Partially) met
	Strain from feelings of being manipulated – 0.89, 0.94	(Partially) met
	Global strain – 0.76, 0.78	Met
(iii) Correlation of each item with rest of scale	Not reported	Unmet/no evidence
(iv) Internal consistency assessed by factor-analytic or similar techniques	Not reported	-

(v) Test-retest reliability	The correlation between scores at 6 weeks after care recipients' discharge from hospital and 9 months after discharge from hospital are as follows. Cognitive impairment of care recipient – 0.60 Functional impairment of care recipient – 0.52 Mutuality – 0.79 Preparedness for caregiving – 0.57 Strain from direct care – 0.60 Strain from lack of resources – 0.64 Strain from worry – 0.80 Strain from role conflict – 0.70 Strain from economic burden – 0.68 Strain from mismatched expectations – 0.67 Strain from increased tensions – 0.78 Strain from feelings of being manipulated – 0.67 Global strain – 0.62	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	The authors predicted that mutuality and preparedness would ameliorate some aspects of role strain, and a regression analysis confirmed this to be the case. Strain from direct care, increased tension, and global strain all appeared lower when higher levels of mutuality and preparedness were reported by caregivers. However, strain from economic burden was not reduced by either mutuality or preparedness, and neither was strain from worry, or from role conflict. There was a limited effect on strain from lack of resources.	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	Mostly 4-point scales	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-

(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	-
(ii) Norms presented	Not reported	-
(iii) Cut off (e.g. for caseness) presented	Not reported	-
(iv) Calibrated against other life events or experiences	Not reported	-
(v) No. of times used in review	4	-
(vi) Existence of a meaningful overall score	No – the inventory contains 7 predictor variables (preparedness, mutuality, gender of caregiver, etc.) and 9 measures of strain. An overall score is not appropriate with an inventory structured in this way. The subscales have been used by themselves.	-
(vii) Number of subscales	16	-
J Feasibility		
(i) Mode of administration	Structured interview	-
(ii) Number of items	148	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	Training required but amount not specified	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Available from authors	-
(vii) Cost	Free	-

Table 35 Family Environment Scale (Moos and Moos, 1981)

The Family Environment Scale is a self-report measure of the social-environmental characteristics of families. It comprises three dimensions: Relationship, Personal Growth and System Maintenance. Three alternative versions are available: the Real Form, which measures people's perceptions of their family environments, the Ideal Form, which rewords items to assess individuals' perceptions of their ideal family environment, and the Expectations Form, which collects what the respondent imagines the environment would be like in the event of anticipated family changes. Reliability data are presented for the Real Form only and are only moderately good; validity is reported to be good. It has not been explicitly validated for carers. Information on this form was gleaned from the description of the measure on the website (www.mindgarden.com), but more psychometric data may be available from the manual.

Criterion+indicators	**Data**	**Rating (if applicable)**

Constructs covered: social-environmental characteristics of families

A Importance to carers		
(i) Data obtained from interviews	Effect of caregiving on the carer's family life	Partially met
(ii) Data obtained from questionnaires	Effect of caregiving on the carer's family life	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Unknown	Unmet/no evidence
(ii) Acceptability to other populations	Not reported	Unmet/no evidence
(iii) Response rates	Not reported	-
(iv) Time to complete	15–20 min	-
(v) Short form available	Child Version of the Family Environment Scale (CVFES-C) (Pino *et al.,* 1984). The CVFES contains 30 items with 3 items for each of 10 dimensions. Children's perceptions of family functioning are assessed through pictorial representations of 3 differing interactions between mother, father and children. Children rate their families on subscales encompassing cohesion, expressiveness, conflict, independence, achievement, intellectual-cultural orientation, active-recreational orientation, moral-religious emphasis, organisation, and control. The subscale *t* scores are used to obtain a categorical description of the child's perception of the family based on a hierarchical system. These criteria should be able to characterise approx. 90% of families. (www.calib.com/nccanch/otherpubs/childassessment/assessmeasures.cfm).	-
(vi) Translated measure available	Spanish, Korean, Chinese (www.nnfr.org/eval/bib_ins/MOOS.html), German (Engfer *et al.*, 1977)	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-

(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Alphas range from 0.61 to 0.78 for the subscales of Form R. Alphas for the Ideal Form and the Expectations Form were not reported.	(Partially) met
(iii) Correlation of each item with rest of scale	Not reported	Unmet/no evidence
(iv) Internal consistency assessed by factor-analytic or similar techniques	Not reported	-
(v) Test-retest reliability	Reliability for the Form R subscales is reported as ranging from 0.52 to 0.91 for 2-, 3- and 12-month intervals (the separate values for each were not reported).	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	It is reported that face and content validity of the instrument are supported by clear statements about family situations that relate to subscale domains.	-
(ii) Content validity		
(iii) Construct validity	Evidence of construct validity is reported in the manual through comparative descriptions of distressed and normal family samples; comparisons of parent responses with those of their adolescent children; descriptions of responses by families with 2–6 or more members; and descriptions of families with a single parent, of minority families, and of older families.	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	Not reported	-

(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	No	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	3	-
(vi) Existence of a meaningful overall score	Based on scores over all the subscales, families are grouped into 1 of 3 family environment typologies based on their most salient characteristics.	-
(vii) Number of subscales	10	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	90	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	Unknown	-
(vi) Freedom of usage	Must be ordered from www.mindgarden.com	-
(vii) Cost	$40 for a sampler set, $120 for a duplication set	-

Table 36 Family Satisfaction Scale (Olson and Wilson, 1982)

The Family Satisfaction Scale is a self-report measure, developed for the general population. Respondents are asked to rate their satisfaction with their families' adaptability and cohesion. Its internal consistency and test-retest reliability are good, and construct validity has also been tested. It has not apparently been validated specifically for carers.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: satisfaction with family adaptability and cohesion		
A Importance to carers		
(i) Data obtained from interviews	Effect of caregiving on the carer's family life	Partially met
(ii) Data obtained from questionnaires	Effect of caregiving on the carer's family life	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet
(ii) Acceptability to other populations	Not reported	Unmet
(iii) Response rates	54%	-
(iv) Time to complete	Not reported	-
(v) Short form available	Not found	-
(vi) Translated measure available	Spanish (Barraca *et al.*, 2000)	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Cohesion subscale, 0.85; Adaptability subscale, 0.84; total score, 0.92	(Partially) met
(iii) Correlation of each item with rest of scale	Range, 0.5–0.74; mean, 0.63 (all are over 0.20)	-

(iv) Internal consistency assessed by factor-analytic or similar techniques	Every item loaded more than 0.50 on the first principal component. When 2 factors were retained, all but 2 of the 14 items loaded more than 0.40 on the first rotated factor. This suggests that the scale is unidimensional and therefore the total score, rather than the adaptability and cohesion scores, is the most empirically valid.	-
(v) Test-retest reliability	Cohesion subscale, 0.76; Adaptability subscale, 0.67; total score, 0.75 (5-week interval)	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	The authors felt that the factor analysis demonstrated the scale's construct validity. As this was the first scale of its kind, concurrent validity could not be demonstrated with any other scale.	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	5-point Likert scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Unknown	-
(ii) Norms presented	Yes	-
(iii) Cut off (e.g. for caseness) presented	Not presented	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	3	-
(vi) Existence of a meaningful overall score	Yes	-

(vii) Number of subscales	8 cohesion subscales (1 item each) and 6 adaptability subscales (1 item each)	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	14	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	None	-
(vi) Freedom of usage	Permission required from authors	-
(vii) Cost	Free	-

Table 37 General Health Questionnaire (60-, 30-, 28- and 12-item; Goldberg, 1978)

The widely used General Health Questionnaire (GHQ) provides a general measure of psychological distress. It was developed for use in community and non-psychiatric settings as a screening device to detect potential psychiatric ‘caseness’. The original 60-item questionnaire has been shortened to 30-, 28- and 12-item versions. The questionnaire was not developed for carers and its acceptability to them has not been established. It has, however, been used widely with carers. In this review, the 28-item version had been used most frequently, whereas the 60-item version had not been used. Reliability and validity has been tested in numerous studies.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: psychological symptoms		
A Importance to carers		
(i) Data obtained from interviews	Carer’s mental health	Met
(ii) Data obtained from questionnaires	Carer’s depression and anxiety	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	Not reported	Unmet/no evidence

(iii) Response rates	Not reported	-
(iv) Time to complete	Not reported	-
(v) Short form available	30-, 28- and 12-item (MAPI Research Institute, Quality of Life Instruments Database; www.qolid.org/public/GHQ.html)	-
(vi) Translated measure available	At least 36 languages (MAPI Research Institute, Quality of Life Instruments Database; www.qolid.org/public/GHQ.html)	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	GHQ-60=0.95; GHQ-30=0.77.	(Partially) met
(ii) Cronbach's alpha	One study reports GHQ-60 alpha=0.92; five studies report GHQ-30 alpha range=0.84–0.93, mean=0.87; one study reports GHQ-12 alpha range=0.82–0.90, mean=0.85.	(Partially) met
(iii) Correlation of each item with rest of scale	Not reported	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	20 principal component analyses of one or other versions of the GHQ have been conducted. GHQ-60 (9 studies), 7–19 significant components (median=11) accounting for 50–92% of the variance. GHQ-30 (9 studies), 4–10 significant components (median=8) accounting for 49–71% of the variance. GHQ-12 (3 studies), 2–3 significant components accounting for 44–62% of the variance.	-
(v) Test-retest reliability	Test-retest reliability co-efficients vary widely according to the population assessed and the duration of time between assessments. Reliability coefficients range between +0.51 (GHQ-60 administered to the general population at a 12-month interval) and +0.90 (GHQ-28 administered to stroke patients at an 8-month interval).	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Instrument construction was based on previous investigations of a representative non-hospital population and the clinical experience of several psychiatrists.	-

(iii) Construct validity	Results from 22 studies demonstrate a median correlation between the GHQ and a criterion interview of +0.70. GHQ-60 (16 studies) sensitivity was 55–91% with a median of 79%. Specificity was 76%–97% with a median of 87%. Variance-weighted mean sensitivity=78%, variance-weighted mean specificity=87%. GHQ-30 (29 studies) sensitivity was 48–100% with a median of 81%. Specificity was 51–94% with a median of 80%. Variance-weighted mean sensitivity=74%, variance-weighted mean specificity=82%. GHQ-28 (12 studies) sensitivity was 44–100% with a median of 86%. Specificity was 74–93% with a median of 82%. Variance-weighted mean sensitivity=84%, variance-weighted mean specificity=82%. GHQ-12 (6 studies) sensitivity was 71–91% with a median of 86%. Specificity was 71–93% with a median of 80%. Variance-weighted mean sensitivity=89%, variance-weighted mean specificity=80%.	-
G Responsiveness		
(i) Used to measure change over time?	Not reported	Unmet/no evidence
(ii) Evidence that change over time correlates with other measures?	Not reported	-
(iii) Are the changes over time found significant to carers?	Not reported	-
H Precision		
(i) Type of response categories	4-point scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not available	-
(ii) Norms presented	Not reported	-
(iii) Cut off (e.g. for caseness) presented	Available for non-carers	-

(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	54	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	GHQ-28 comprises 4 scales: somatic symptoms; anxiety/insomnia; social dysfunction; severe depression.	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	60, 30, 28 or 12 items	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Available from NFER-Nelson	-
(vii) Cost	GHQ-60, £19.30 per 25 copies; GHQ-30 and GHQ-25, each £12.95 per 25 copies; GHQ-12, £30.50 per 100 copies.	-

Table 38 Geriatric Depression Scale (Yesavage et al., 1983)

The Geriatric Depression Scale is a 30-item self-report measure of depression in elderly people. Respondents are asked to rate items such as 'are you basically satisfied with your life?' and 'do you frequently feel like crying?' It was not developed specifically for a caregiving population and its acceptability to such a group is not known. Its reliability and validity are good, but have not been tested for a caregiving population. Authors report that they cannot vouch for the accuracy of any of the translations, but state that most come from major institutions and have been used in clinical practice.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: depression in elderly people		
A Importance to carers		
(i) Data obtained from interviews	Carer's mental health	Met
(ii) Data obtained from questionnaires	Carer's depression	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet

(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Unknown	Unmet/no evidence
(ii) Acceptability to other populations	Unknown	Unmet/no evidence
(iii) Response rates	Unknown	-
(iv) Time to complete	5–10 min	-
(v) Short form available	Geriatric Depression Scale (Short Form), 15-items (www.stanford.edu/~yesavage/GDS.html)	-
(vi) Translated measure available	Chinese, Danish, Dutch, French, French Canadian, German, Greek, Hebrew, Hindi, Hungarian, Icelandic, Italian, Japanese, Korean, Lithuanian, Malay, Portuguese, Romanian, Russian, Russian Ukrainian, Spanish, Swedish, Thai, Turkish, Vietnamese, Yiddish (www.stanford.edu/~yesavage/GDS.html)	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	0.94	Partially met
(ii) Cronbach's alpha	0.94	Partially met
(iii) Correlation of each item with rest of scale	Median correlation, 0.675; range, 0.47–0.83 (i.e. all items had a correlation of over 0.2; individual item correlations were not reported).	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	Median correlation between individual items and corrected item total score (median 0.56, range 0.32–0.83); average inter-correlation among the scale's individual items (0.36).	-
(v) Test-retest reliability	20 subjects completing the questionnaire twice, one week apart; correlation=0.85.	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-

(iii) Construct validity	Validated against Hamilton Rating Scale for Depression and Zung Self-rating Depression Scale. Normal subjects received lowest Geriatric Depression Scale scores whereas severely depressed subjects scored highest.	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	Binary	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Unknown	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	Yes – not for carers.	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	16	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	None	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	30	-
(iii) Ease of explaining measure to study participants	'Simple to administer'	-

(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	None	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 39 Hamilton Rating Scale for Depression (Hamilton, 1960)

The Hamilton Rating Scale for Depression (HDRS) is an interview measure for people already diagnosed with affective disorder of the depressive type. It has not been validated with a caregiving population. Little reliability and validity evidence is presented in this paper and the authors state that the shorter version, the Hamilton Depression Inventory, is more reliable because it uses multiple questions for many items.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: depression		
A Importance to carers		
(i) Data obtained from interviews	Carer’s mental health	Met
(ii) Data obtained from questionnaires	Carer’s depression	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	Not reported	Unmet/no evidence
(iii) Response rates	Not reported	-
(iv) Time to complete	Unknown	-
(v) Short form available	Hamilton Depression Inventory (HDI): 23-item Full Scale HDI or the 9-item HDI-Short Form (Reynolds and Kobak, 1995). Administration: individual or group, takes 10 min, self-report format; screens for some additional symptoms (Psychological Assessment Resources; www.parinc.com/Product.cfm?ProductID=191). 6-, 17- and 21-item HDRS (Spanish version; Bobes *et al.*, 2003).	-

(vi) Translated measure available	Turkish (test-retest reliability coefficient based on a 5-day interval, 0.85; Cronbach alpha coefficient, 0.75; split-half reliability coefficient, 0.76; inter-rater reliability coefficients 0.87–0.98; correlation between HDRS and Beck Depression Inventory, 0.48; between HDRS and Clinical Global Impression, 0.56. Principal Components Analysis yielded six factors. The correlation (-0.13) between the control and patient groups indicates that the HDRS assesses depression very well.), Chinese, Danish	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Not reported	Unmet/no evidence
(iii) Correlation of each item with rest of scale	Not reported	Unmet/no evidence
(iv) Internal consistency assessed by factor-analytic or similar techniques	Factor analysis carried out, with inconclusive results that do not relate well to clinical syndromes.	-
(v) Test-retest reliability	Not reported	-
(vi) Inter-rater reliability	For two raters, the correlation between summed scores for the first 10 patients was 0.84. Adding 10 patients successively at a time, the correlation changed to 0.84, 0.88, 0.89, 0.89, 0.90 and 0.90. The last correlation is therefore for a total of 70 patients.	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	Not reported	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	No	-

(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	Varies between 3-, 4- or 5-point scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not available	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	Yes	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	12	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	None	-
J Feasibility		
(i) Mode of administration	Structured interview	-
(ii) Number of items	17	-
(iii) Ease of explaining measure to study participants	'OK'	-
(iv) Amount of researcher training required	Unknown	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 40 Hospital Anxiety and Depression Scale (Zigmond and Snaith, 1983)

The Hospital Anxiety and Depression Scale (HADS) is a self-report measure to detect states of depression and anxiety. It was designed for use in hospital medical outpatient clinics and has not been validated for a caregiving population. Reliability and validity are good for the original population.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: depression and anxiety		
A Importance to carers		
(i) Data obtained from interviews	Carer's mental health	Met
(ii) Data obtained from questionnaires	Carer's depression and anxiety	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	'Acceptable', according to manual	Met
(iii) Response rates	98% (in original)	-
(iv) Time to complete	Manual reports it takes 'only minutes' to complete.	-
(v) Short form available	Not found	-

(vi) Translated measure available	Arabic, Cantonese, Chinese, Dutch, French, German, Italian, Japanese, Nigerian, Polish, Spanish, Urdu (Victorian Transcultural Psychiatry Unit, Victoria, Australia; www.vtpu.org.ac/rE:sources/translated_instruments/mi/mi-haads.html), Iranian (Montazeri *et al.*, 2003), Pakistani (Mumford *et al.*, 1991), Norwegian (Mykletun *et al.*, 2001), Thai (Nilchaikovit *et al.*, 1996; www.mahidol.ac.th/mahidol/ra/rapc/vol1.html#ghq). HADS was translated into Thai and administered to a sample of 60 inpatients with cancer to test the validity and reliability of the Thai version of HADS (Thai HADS). Semi-structured clinical psychiatric interview was used as a gold standard. Thai HADS had good reliability and validity for both anxiety and depression subscales. At the cut-off point of >11, the sensitivity of anxiety and depression subscales of Thai HADS were 100 and 85.71% respectively, whereas the specificities were 86.0% for anxiety and 91.3% for depression. Both subscales also showed good internal consistencies with Cronbach's alpha coefficient of 0.8551 for anxiety subscale and 0.8259 for depression subscale. All translations must be purchased from the publisher in order to guarantee reliable comparability of studies. They have all been done using careful back-translation (according to the manual) to ensure that the sense of the original wording of the items has been retained.	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Cronbach's alpha 0.93 for anxiety; 0.90 for depression (Moorey *et al.*, 1991).	(Partially) met
(iii) Correlation of each item with rest of scale	The ranges of item–total correlations are reported as follows: anxiety, 0.76–0.41; depression, 0.60–0.30. Individual item–total correlations are not reported. All item–total correlations are above 0.20.	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	A factor analysis was carried out (Moorey *et al.*, 1991). Two independent factors emerged which accounted for the 53% of the variance, and these factors were replicated on subsamples. In each case, items loaded as expected with the exception of item 7 which loaded on both factors.	-

(v) Test-retest reliability	Test-retest reliability may only be established in healthy respondents, since emotional states of people suffering from depression and anxiety disorders vary too widely from day to day. The test retest data taken from within a healthy sample indicated significant correlations of 0.92 for the D-scale and 0.89 for the A-scale (authors' unpublished study).	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	'Good', according to manual	-
(ii) Content validity	Not reported	-
(iii) Construct validity	Concurrent validation was assessed by comparison with five-point psychiatric rating scales of anxiety and depression for 100 medical outpatients. This yielded significant correlations of 0.54 for the A-scale and 0.79 for the D-scale. Further concurrent validity data for HADS has been reported in psychiatric patients (Bramley *et al.*, 1988), in a heterogeneous group of patients with physical illness (Aylard *et al.*, 1987), in patients attending a genitourinary clinic (Barczak *et al.*, 1988) and in patients with cancer where the validity of the HADS as a screening instrument for psychological distress was confirmed (Ibbotson *et al.*, 1989).	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	No	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	5-point scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Unknown	-
(ii) Norms presented	No	-

(iii) Cut off (e.g. for caseness) presented	Available for non-carers	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	5	-
(vi) Existence of a meaningful overall score	2 scores	-
(vii) Number of subscales	2 (anxiety and depression)	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	14	-
(iii) Ease of explaining measure to study participants	'Respondents find it easy to complete' according to manual	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Test company	-
(vii) Cost	Yes	-

Table 41 Interpersonal Support Evaluation List (Cohen et al., 1984)

The Interpersonal Support Evaluation List (ISEL) is a self-report measure of perceived availability of social support. Respondents are asked to rate whether 40 items are true or false, for example 'there is at least one person I know whose advice I really trust', and 'I don't often get invited to do things with others'. It was developed for the general population and has not apparently been validated for carers. Overall reliability and validity are good, although responsiveness to change over time is not presented.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: perceived social support		
A Importance to carers		
(i) Data obtained from interviews	Not a construct mentioned by carers	Unmet
(ii) Data obtained from questionnaires	Not a construct mentioned by carers	Unmet
B Carer involvement in outcome measure development		

(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Unknown	Unmet
(ii) Acceptability to other populations	Unknown	Unmet
(iii) Response rates	Unknown	-
(iv) Time to complete	Unknown	-
(v) Short form available	Not found	-
(vi) Translated measure available	Spanish (showed good internal consistency)	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Alpha vlaues for the general population ISEL are 0.88–0.9. Ranges for general population subscales are:	Met for whole score. For subscales:
	0.70–0.82 for appraisal,	Met
	0.62–0.73 for self-esteem,	(Partially) met
	0.73–0.78 for belonging,	Met
	0.73–0.81 for tangible support.	Met
(iii) Correlation of each item with rest of scale	Not reported	Unmet/no evidence
(iv) Internal consistency assessed by factor-analytic or similar techniques	Subscale independence was maximised by selecting items (from a larger pool) which were highly correlated with items in their own subscale and at the same time minimally correlated with other subscales.	-

(v) Test-retest reliability	There was a 6-week interval between retests of the general ISEL scale, with test-retest correlations of 0.70. The subscale correlations were 0.63 for appraisal, 0.65 for belonging, 0.68 for self-esteem and 0.69 for tangible support.	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	Criterion validity was established by comparing the ISEL with other social support measures. A student version of the ISEL (which had 48 items instead of 40) was found to correlate 0.46 with the Inventory of Socially Supportive Behaviours (Barrera *et al.*, 1981). The general population version of the ISEL (40 items) correlated 0.30 with the total score of the Family Environment Scale (Moos and Moos, 1981). The general population ISEL also correlated 0.31 with the Partner Adjustment Scale (Mermelstein *et al.*, 1983), a measure of the quality of marital or living partner relationships. Some studies examined the relationship between ISEL scores and psychological symptomatology. The correlation with the Centre of Epidemiological Studies Depression Scale was -0.52, with the Hopkins Symptom Checklist was -0.57 and with the Kobassa Symptom Scale was -0.60.	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	Binary: 'probably true' or 'probably false'	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not correlated with the Crowne-Marlowe Social Desirability Scale (Crowne and Marlowe, 1960).	-
I Interpretability		

(i) Minimal difference considered important by people the outcome measure will be used on is presented	Unknown	-
(ii) Norms presented	Unknown	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	6	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	4 (10 items each)	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	40	-
(iii) Ease of explaining measure to study participants	Unknown	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	None	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 42 Inventory of Socially Supportive Behaviours (Barrera et al., 1981)

The Inventory of Socially Supportive Behaviours (ISSB) is a self-report measure of the social support a person has received. It was designed as a 'tool for understanding natural helping processes' and developed with college students. Respondents are asked to rate how often they have received support in the preceding month, for example how often someone has expressed interest and concern in their well-being, or how often someone has provided them with a place where they could get away for a while. Its reliability and validity are good, but have not been tested specifically for carers.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: social support received		
A Importance to carers		
(i) Data obtained from interviews	Not a construct mentioned by carers	Unmet
(ii) Data obtained from questionnaires	Not a construct mentioned by carers	Unmet
B Carer involvement in outcome measure development		

(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Unknown	Unmet/no evidence
(ii) Acceptability to other populations	Unknown	Unmet/no evidence
(iii) Response rates	Unknown	-
(iv) Time to complete	Unknown	-
(v) Short form available	ISSB-SF (Short form)	-
(vi) Translated measure available	Not found	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	0.93 at first administration and 0.94 at second administration of the scale.	Partially met
(iii) Correlation of each item with rest of scale	Range 0.08–0.71: 7 items correlated with the total at less than 0.30. All of these items were rated as occurring infrequently. They were kept in the scale as it was felt that these items would be more frequently endorsed by subsequent samples.	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	No	-
(v) Test-retest reliability	For individual items test-retest correlations ranged from 0.44 to 0.91	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Empirical research, literature reviews and discussion articles related to social support were surveyed for actual items.	-

(iii) Construct validity	ISSB total scores were significantly correlated with both available (0.42) and actual (0.32) social support-network size. The ISSB was also positively correlated with the Family Evaluation Scale Cohesion subscale (Moos *et al.*, 1974) at 0.36, so subjects who reported the greatest frequency of socially supportive interactions with family members also tended to perceive their families as highly cohesive.	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	5-point Likert scale for frequency of occurrence of items	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Unknown	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	5	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	0	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	40	-
(iii) Ease of explaining measure to study participants	Unknown	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-

(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 43 Jalowiec Coping Scale (Jalowiec, 1979)

The Jalowiec Coping Scale is a self-report measure of coping behaviours, developed for use with hypertensive and emergency-room patients. Reliability and validity are good, although responsiveness to change over time is not tested, but have not been tested with carers.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: coping behaviours		
A Importance to carers		
(i) Data obtained from interviews	Carer’s ability to cope	Met
(ii) Data obtained from questionnaires	Carer’s ability to cope	Met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	Not reported	Unmet/no evidence
(iii) Response rates	Not reported	-
(iv) Time to complete	Not reported	-
(v) Short form available	Not found	-
(vi) Translated measure available	Has been used in Swedish, Norwegian and Cantonese.	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence

(ii) Cronbach's alpha	0.86 (based on coping data from 141 subjects (hypertensive and emergency-room patients, and a general population)). Murphy (1982) obtained an alpha value of 0.85 on coping data from 150 dialysis patients.	Met
(iii) Correlation of each item with rest of scale	Not reported	Unmet/no evidence
(iv) Internal consistency assessed by factor-analytic or similar techniques	Using principal factor extraction and varimax rotation, the authors found that a four-factor solution was the most viable. Factor I contains most of the items of the problem-oriented subscale and has an alpha value of 0.86. Factor II contained mainly affective items, and had an alpha value of 0.73. Factor III was hard to characterise, and contained mainly affective coping behaviours. For the most part, it seemed to relate to tension-releasing mechanisms, physical and emotional types of withdrawal and helplessness. It had an alpha value of 0.75. Factor IV suggests going outside of the self in order to cope, including items such as blaming, and seeking comfort from others. It had an alpha value of 0.55. The factor analysis does seem to undermine the cohesiveness of the affective-oriented subscale.	-
(v) Test-retest reliability	28 subjects from the general population were retested after two weeks. Spearman's rank-ordering of the test-retest data yielded significant ($P<0.001$) reliability coefficients of 0.79 for total coping scores, 0.85 for problem-oriented scores and 0.86 for affective scores. Using a 1-month interval and 30 subjects, Langner (1983) obtained reliability coefficients of 0.78 for total scores, 0.84 for problem-oriented scores and 0.83 for affective scores.	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Authors conclude that content validity is substantiated by the manner of tool development – the 40 items were culled from a comprehensive review of the literature on stress, coping and adaptation – by the large number of items used, and by the inclusion of diverse coping behaviours.	-
(iii) Construct validity	Authors believe that the alpha coefficients reported and factor analysis results provide sufficient evidence for construct validity.	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet

(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	5-point Likert scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	3	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	2	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	40	-
(iii) Ease of explaining measure to study participants	Authors report that test instructions are clear and uncomplicated.	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 44 Life Orientation Test (Scheier and Carver, 1985; Scheier et al., 1994)

The Life Orientation Test (LOT) assesses individual differences in generalised optimism versus pessimism. It is a self-report measure not designed for a carer population. Although the LOT has been widely used, it authors report that it has some problems, particularly in that its original items did not all focus as explicitly on expectations for the future as theory dictated. A 10-item revised version (LOT-R) is available which addresses these problems. Reliability and validity are reasonably good but not tested for carers.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: dispositional optimism		
A Importance to carers		
(i) Data obtained from interviews	Not a construct mentioned by carers	Unmet
(ii) Data obtained from questionnaires	Not a construct mentioned by carers	Unmet
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	None	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	Not reported	Unmet/no evidence
(iii) Response rates	The subjects were all students, and participation was in partial fulfillment of a psychology research requirement. Therefore the response rate was 100%.	-
(iv) Time to complete	Not reported	-
(v) Short form available	Unknown	-
(vi) Translated measure available	Spanish (LOT-R; Perczek *et al.*, 2000). Back translation was performed. Correlations between English and Spanish versions in bilingual samples were above 0.72. Alpha reliabilities of the Spanish versions were comparable with those of the English versions. Correlations among measures in a sample of cancer patients were similar across languages. French (LOT-R; Sultan and Bureau, 1999), Chinese (LOT-R; Lai *et al.*, 1988)	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		

(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	0.78	Met
(iii) Correlation of each item with rest of scale	Range, 0.43–0.63; mean, 0.53	Partially met
(iv) Internal consistency assessed by factor-analytic or similar techniques	The 6 LOT-R items yielded one factor accounting for 48.1% of the variance. All items loaded at least 0.58 on this factor. The mean factor loading was 0.69. Optimism/pessimism is theorised as one dimension, and this is rather a moderately loaded factor. 2 factors load as well, which is a bit at odds with the construct.	-
(v) Test-retest reliability	4 months – 0.68; 1 year – 0.62; 2 years – 0.56; 28 months – 0.79.	-
(vi) Inter-rater reliability	Not reported	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	Correlation between revised LOT and original LOT was high for both men and women (0.95 for both). Against the Mastery Scale (Pearlin and Schooler, 1978), correlation was 0.46 for women, 0.51 for men. Against trait form of State-Trait Anxiety Inventory (Speilberger *et al.*, 1974) −0.54 for women, −0.52 for men. Against Rosenberg's (1965) Self-Esteem Scale 0.54 for women, 0.50 for men. Against Emotional Stability subscale of Guildford-Zimmerman Temperament Survey (Guilford *et al.*, 1976) −0.49 for women, −0.40 for men. Against Neuroticism as measured by Eysenck Personality Questionnaire (Eysenck, 1958; Goh *et al*., 1982) −0.36 for women, −0.37 for men. The authors hoped for modest correlations so as to distinguish optimism from these other characteristics.	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	No	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	5-point scale	-
(ii) Ceiling and floor effects	Not reported	-

(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	No	-
(ii) Norms presented	Available for non-carers	-
(iii) Cut off (e.g. for caseness) presented	Not available (not appropriate)	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	3	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	2 (Optimism and pessimism can be separately reported)	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	12 (8 scored items and 4 filler items)	-
(iii) Ease of explaining measure to study participants	Easy	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 45 Life Satisfaction Index (Wood et al., 1969)

The Life Satisfaction Index (LSI-Z) is a self-report measure of 'life satisfaction', not developed for carers specifically or validated for them. It was assessed in a rural ageing population. Its construct validity was tested against the Life Satisfaction Ratings Scale administered by a psychologist, but other types of validity and reliability have not been assessed.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: life satisfaction		
A Importance to carers		
(i) Data obtained from interviews	Not a construct mentioned by carers	Unmet
(ii) Data obtained from questionnaires	Satisfaction with life	Partially met
B Carer involvement in outcome measure development		

(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	Not reported	Unmet/no evidence
(iii) Response rates	Not reported	-
(iv) Time to complete	Not reported	-
(v) Short form available	Not found	-
(vi) Translated measure available	Not found	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	0.79	Met
(ii) Cronbach's alpha	Not reported	Unmet/no evidence
(iii) Correlation of each item with rest of scale	Not reported	Unmet/no evidence
(iv) Internal consistency assessed by factor-analytic or similar techniques	Not reported	-
(v) Test-retest reliability	Not reported	-
(vi) Inter-rater reliability	Not reported	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	Scores were correlated with the Life Satisfaction Ratings Scale (Neugarten *et al.*, 1961) used by a psychologist at 0.57.	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-

H Precision		
(i) Type of response categories	Binary	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	No	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	No	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	Not appropriate	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	13	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	0	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	13	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 46 Mastery Scale (Pearlin and Schooler, 1978)

The Mastery Scale assesses coping. The authors define coping as 'any response to external life-strains that serves to prevent, avoid, or control emotional distress' and had in mind 'normative coping responses to normative life-problems'. It was designed for a general population rather than specifically for carers and not validated for the latter; in fact, little validity or reliability testing had been done at all.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: coping		
A Importance to carers		
(i) Data obtained from interviews	Carer’s ability to cope	Met
(ii) Data obtained from questionnaires	Carer’s ability to cope	Met

B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	Not reported, although items were generated through interviews	Unmet/no evidence
(iii) Response rates	Not reported	-
(iv) Time to complete	Not reported	-
(v) Short form available	Not found	-
(vi) Translated measure available	Not found	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Not reported	-
(iii) Correlation of each item with rest of scale	Not reported	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	Not reported	-
(v) Test-retest reliability	Not reported	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported (concept defined by authors)	-
(iii) Construct validity	High levels of mastery (as measured by this measure) in the role areas of marriage, parenting, household economics and occupation were found to correlate with lower levels of stress caused by the strains of these roles.	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet

(ii) Evidence that change over time correlates with other measures?	No	-
(iii) Are the changes over time found significant to carers?	No	-
H Precision		
(i) Type of response categories	Not reported	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	No	-
(iv) Sources of potential bias in scoring	No	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	No	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) Wideness of usage (no. times used in this review)	9	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	None	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	7	-
(iii) Ease of explaining measure to study participants	Easy	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 47 Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36) (Ware et al., 1992; McHorney et al., 1993, 1994)

The Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36) is a self-report measure of health status, developed with patients presenting to physicians. It measures physical functioning, role limitations because of physical health problems, bodily pain, general health perceptions, vitality (energy/fatigue), social functioning, role limitations because of emotional problems and general mental health (psychological distress and psychological well-being). Its reliability and validity are good, although not tested specifically for carers.

Criterion+indicators	Data	Rating (if applicable)
Constructs covered: health		
A Importance to carers		
(i) Data obtained from interviews	Carer's physical and mental health	Met
(ii) Data obtained from questionnaires	Carer's physical health, depression and anxiety	Met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	Not reported	Unmet/no evidence
(iii) Response rates	Item completeness ranged from 83 to 98%.	-
(iv) Time to complete	5–10 min	-
(v) Short form available	This is a short form of the full Medical Outcomes Study Health Survey.	-

(vi) Translated measure available	Chinese, Dutch, French, German, Polish, Spanish, Swedish (Aaronson *et al.*, 1992). International Quality of Life Assessment – project originally included sponsored investigators from 14 countries: Australia, Belgium, Canada, Denmark, France, Germany, Italy, Japan, the Netherlands, Norway, Spain, Sweden, UK, USA. Researchers from more than 25 other countries are also translating and validating the SF-36 following International Quality of Life Assessment (IQOLA) methods. 4 major stages of activity were included: translation following a standard protocol; formal psychometric tests of scaling and scoring assumptions; validation and norming studies to evaluate the equivalence of interpretations across countries; and documentation of scoring algorithms and interpretation guidelines. Trust-approved language versions of the SF-36 include translations for Germany, Italy, Spain and Sweden, and English-language adaptations for Australia/New Zealand, Canada and the UK. Russia (www.mapi-research-inst.com/pdf/art/qol23sup_0.PDF).	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Physical Functioning (PF) – 0.93	Partially met
	Role limitations because of physical health problems (RP) – 0.84	Met
	Bodily pain (BP) – 0.82	Met
	General health perceptions (GH) – 0.78	Met
	Vitality (energy/fatigue) (VT) – 0.87	Met
	Social functioning (SF) – 0.85	Met
	Role limitations because of emotional problems (RE) – 0.83	Met
	General mental health (psychological distress and psychological well-being) (MH) – 0.90	Met

(iii) Correlation of each item with rest of scale	Correlation between an item and its hypothesised scale exceeded correlations with all other scales by more than two standard errors. All correlations were greater than 0.20. Ranges were: PF – 0.49–0.80 RP – 0.65–0.70 BP – 0.70 GH – 0.44–0.72 VT – 0.69–0.75 SF – 0.74 RE – 0.63–0.73 MH: – 0.65–0.81	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	Components analysis confirmed that there was a substantial general health dimension which had been hypothesised as common to all eight scales. The first principal component accounted for 55% of the total variance and correlated highly with all eight scales (0.67–0.82). Extraction of the second component increased the percentage of total variance explained from 55 to 70%. The total variance in each scale accounted for by the two-factor solution ranged from 0.56 to 0.82 across scales. Rotation of these components confirmed the hypothesised physical and mental dimensions of health.	-
(v) Test-retest reliability	Not reported	-
(vi) Inter-rater reliability	N/A	
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	Patients with serious medical conditions scored significantly lower on all eight scales compared to patients with minor medical conditions. The physical functioning scale was most valid in detecting differences between patients with minor versus serious medical conditions. For clinical comparisons involving the presence or absence of a psychiatric condition the mental health scale proved to be the most valid, followed by the role-emotional and social functioning scales. More examples of discriminant validity of this type are reported.	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-

(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	Likert scale (number of points not reported)	-
(ii) Ceiling and floor effects	Noteworthy floor effects were only observed for RP and RE (24 and 18% respectively). The authors felt this was because these were the coarsest of the scales, representing only 5 and 4 levels of health each. Substantial ceiling effects were also observed for the 2 role-disability scales (37% for RP and 56% for RE) and for SF as well (46%). Other floor and ceiling effects were reported as trivial.	-
(iii) Precision of scales	The SF-36 is a revised version of the SF-20. Items were added to better represent levels and types of limitations between the extremes. Standardised response choices were revised to estimate the severity of each limitation and thereby to increase score precision.	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	-
(ii) Norms presented	Not reported	-
(iii) Cut off (e.g. for caseness) presented	Not reported	-
(iv) Calibrated against other life events or experiences	Not reported	-
(v) No. of times used in review	14	-
(vi) Existence of a meaningful overall score	The subscales measure different concepts but all of the scores together can be used to present a health profile.	-
(vii) Number of subscales	The SF-36 is a multi-item scale measuring 8 health concepts in 8 subscales.	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	36	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	Not reported, but interviews do require training.	-
(v) Weighting used in scoring	No	-

(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 48 Multilevel Assessment Instrument (Lawton et al., 1982)

The Multilevel Assessment Instrument is a structured interview for older people that assesses a broad range of domains, namely the respondent's physical health, cognition, activities of daily living, time use, social interaction, personal adjustment (morale and psychiatric symptoms) and perceived environmental quality. Reliability was reasonably good and some construct validity was assessed, but neither were assessed for a carer population. The Korean translation used Brislin's (1980) three-step approach for cross-cultural research. This included translating the scales from English into Korean, back-translation, and a pilot study to ensure the quality of the procedure.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: physical health, cognition, activities of daily living, time use, social interaction, personal adjustment (morale and psychiatric symptoms), perceived environment		
A Importance to carers		
(i) Data obtained from interviews	Carer's physical health, effect of caregiving on the carer's social life, carer's depression and anxiety	Met
(ii) Data obtained from questionnaires	Carer's physical health, effect of caregiving on the carer's social life, carer's depression and anxiety	Met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	Not reported	Unmet/no evidence
(iii) Response rates	Not reported	-
(iv) Time to complete	50 min	-
(v) Short form available	Not found	-
(vi) Translated measure available	Korean (Kim and Lee, 2003); back-translation performed.	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-

(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Average of 0.81 for the seven subscales (range, 0.71–0.93)	Met
(iii) Correlation of each item with rest of scale	Not reported	Unmet/no evidence
(iv) Internal consistency assessed by factor-analytic or similar techniques	No	-
(v) Test-retest reliability	Average of 0.83 for the seven subscales (range, 0.73–0.95)	-
(vi) Inter-rater reliability	With a sample of 484, interviewers and reader-raters agreed with either a 0- or 1-point discrepancy in 95% of all instances; intraclass correlations ranged from 0.88 (Activities of Daily Living) to 0.58 (social interactions). With an additional sample of 106 all correlations were significant; among the domain ratings the median correlations between interviewer and administrator was 0.51; between the interviewer and clinician was 0.60.	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	Theorised that decreasing levels of competence should be exhibited in independently living older people, high-intensity in-home service recipients, and institutional waiting list clients. These are the groups on which the instrument was piloted, and theory was supported by results.	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	No	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	Varies	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-

(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	-
(ii) Norms presented	Not available	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	28 (different parts)	-
(vi) Existence of a meaningful overall score	It is usually used only in its component parts.	-
(vii) Number of subscales	7	-
J Feasibility		
(i) Mode of administration	Structured interview	-
(ii) Number of items	216	-
(iii) Ease of explaining measure to study participants	Unknown	-
(iv) Amount of researcher training required	3 days	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Full version available from authors	-
(vii) Cost	Free	-

Table 49 NEO Personality Inventory (Costa and McCrae, 1985; Costa et al., 1991)

The NEO Personality Inventory (NEO-PI) is a measure of personality, assessing neuroticism, extraversion, openness, agreeableness and conscientiousness. It is designed to be completed by self-report or by peer, spouse or 'expert', and designed for a general population and not validated specifically for carers. Inter-rater reliability was tested over very long intervals, perhaps because personalities are expected to be stable. It was not designed for a carer population and its reliability and validity have not been tested for carers.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: personality		
A Importance to carers		
(i) Data obtained from interviews	Not a construct mentioned by carers	Unmet
(ii) Data obtained from questionnaires	Not a construct mentioned by carers	Unmet

B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	Not reported	Unmet/no evidence
(iii) Response rates	Not reported	-
(iv) Time to complete	45 min	-
(v) Short form available	The NEO Five Factor Inventory (NEO-FFI): 60-item version, 5-factors, self-report, taking 10–15 min. Correlation with NEO PI-R domain scales: 0.77–0.92; internal consistency values: 0.68–0.86. NEO-FFI subscales show correlations of 0.75–0.89 with the NEO-PI validimax factors. Internal consistency values were 0.74–0.89. NEO-4: 192 items, taking 25–35 min. 4 domains: extraversion, agreeableness, openness to experience and conscientiousness. Normally used in employment and personal counselling settings (e.g. career counselling, career development and training).	-
(vi) Translated measure available	Spanish (240 items – Research Psychologists Press (www.rpp.on.ca) who sell the NEO-PI-R also sell a Spanish version, psychometric information was not available on the web page). Croatian (NEO-PI-R) (Marusic and Bratko, 1998) 240-item inventory consists of the following five scales: Neuroticism (N), Extraversion (E), Openness (O), Agreeableness (A) and Conscientiousness (C), each containing six 8-item facet scales. The alpha reliability coefficients for the 5 scales were 0.91, 0.88, 0.85, 0.88 and 0.90 respectively, i.e. highly comparable to those reported by Costa and McCrae (1991). Chinese (NEO-PI Short form 60-items; Wan *et al.*, 1999), Korean, Croatian and Italian (NEO-PI-R; McCrae *et al.*, 1996), Czech, Polish, Slovak, Philippines, Shona, Turkish, Vietnamese-American, Indian, Portuguese, Russian (edited by McCrae and Allik, 2002).	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-

(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	0.86–0.95 for domain scales, 0.56–0.90 for facet scales	(Partially) met
(iii) Correlation of each item with rest of scale	Not reported	Unmet/no evidence
(iv) Internal consistency assessed by factor-analytic or similar techniques	Factor analytic techniques used to design scale. Correlations between original and revised scales were 0.93–0.95.	-
(v) Test-retest reliability	Short-term test-retest reliability has been found with the NEO-FFI and the NEO-PI-R (more details are not given). For the original version of the scale it ranges from 0.51 to 0.83 after 3, 6 and 7 years for the original scale.	-
(vi) Inter-rater reliability	Not reported	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Content validity was addressed by identifying 6 distinct facets of sample for each domain, and by selecting non-redundant items to measure each facet.	-
(iii) Construct validity	Research Psychologists Press (www.rpp.on.ca) report that the revised version has been validated against other personality inventories but details are not given. Patients in psychotherapy score high on neuroticism and drug abusers score low on agreeableness and conscientiousness, which provides some evidence of criterion validity.	-
G Responsiveness		
(i) Used to measure change over time?	No (inappropriate for a personality inventory)	N/A
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	5-point scale	-
(ii) Ceiling and floor effects	Not reported	-

(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	A series of studies have shown that scores are not strongly influenced by socially desirable responding.	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not appropriate	-
(ii) Norms presented	For college students and adults	-
(iii) Cut off (e.g. for caseness) presented	Not appropriate	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	5	-
(vi) Existence of a meaningful overall score	No	-
(vii) Number of subscales	5 domains, 6 facets per domain	-
J Feasibility		
(i) Mode of administration	Self-report or by peer, spouse or expert ratings	-
(ii) Number of items	243	-
(iii) Ease of explaining measure to study participants	Designed to be easily read and understood	-
(iv) Amount of researcher training required	Training under supervision of qualified psychologist. Adequacy of directions and training required to administer: the NEO PI-R can be easily administered to individuals. The instructions are quite clear and self-explanatory. No particular training is required to present the instrument to subjects. Hand scoring, as noted above, is quite easy and requires no special talents other than basic arithmetic ability. The NEO-PI-R is classified by the publisher as a Level B instrument, requiring of the user at least a BA degree in Psychology or Counselling and relevant training or coursework in the interpretation of psychological tests and measurement at an accredited college or university; http://aac.ncat.edu/newsnotes/y97fall.html.	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Not available – must be ordered and paid for.	-
(vii) Cost	$58 for test item	-

Table 50 Perceived Stress Scale (Cohen et al., 1983)

The Perceived Stress Scale is designed to measure the degree to which situations in one's life are appraised as stressful. Respondents rate 13 items such as 'in the last month, how often have you felt that things were going your way?', 'in the last month, how often have you been upset because of something that happened unexpectedly?'. It can be administered either face to face or via the short telephone version. It showed adequate reliability and construct validity, but not specifically for carers, and its acceptability to carers has not been assessed.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: perceived stress		
A Importance to carers		
(i) Data obtained from interviews	Not a construct mentioned by carers	Unmet
(ii) Data obtained from questionnaires	Not a construct mentioned by carers	Unmet
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	Not reported	Unmet/no evidence
(iii) Response rates	Not reported	-
(iv) Time to complete	Only says 'a few minutes'	-
(v) Short form available	Yes: 4-item version (telephone). Alpha value, 0.72; test-retest 0.55 over 2 months.	-
(vi) Translated measure available	Spanish PSS (10- and 14-item), Mexican Spanish PSS (10-item), Thai PSS (14-item; Cohen *et al.*, 1983). Authors are not aware of its psychometric properties and did not pre-test translations. Chinese (Yam and Shiu, 2003).	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	No	-

E Reliability		
(i) Split-half reliability	Not tested	No evidence
(ii) Cronbach's alpha	Average of 0.85 over 3 samples	Met
(iii) Correlation of each item with rest of scale	Not reported	Unmet/no evidence
(iv) Internal consistency assessed by factor-analytic or similar techniques	Not reported	-
(v) Test-retest reliability	2 days – 0.85; 2 weeks – 0.55	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	Evidence that the scale is a better predictor of symptomatology than life events ($P=0.05$ in a t test) as hypothesised by authors. Also that scale is a better predictor of utilisation of health services than life events ($P=0.007$), and the short-form scores were correlated at 1 month with the average number of cigarettes smoked per day at 1 month with a sample of people trying to give up smoking.	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	Not reported	-
H Precision		
(i) Type of response categories	5-point scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	Met	-

(v) No. of times used in review	14	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	None	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	14	-
(iii) Ease of explaining measure to study participants	Ok	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 51 Pittsburgh Sleep Quality Index (Buysse et al., 1989)

The Pittsburgh Sleep Quality Index is a face-to-face interview with the respondent and their 'bed partner' to measure the respondent’s sleep quality and disturbances over the course of a month. Participants used to develop the measure were 'poor sleepers' who were inpatients and outpatients with major depressive disorder; 'poor sleepers' who had been referred to a sleep clinic; and 'good sleeper' controls from the general population. It was thus not developed for a carer population. Reliability and, to some extent, validity are good, but not tested for carers.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: sleep quality		
A Importance to carers		
(i) Data obtained from interviews	Not a construct mentioned by carers	Unmet
(ii) Data obtained from questionnaires	Not a construct mentioned by carers	Unmet
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	Found it easy to understand and use	Partially met

(iii) Response rates	93.7% gave complete responses to all items.	-
(iv) Time to complete	5–10 min	-
(v) Short form available	Not found	-
(vi) Translated measure available	Spanish (Royuela *et al*., 2002), Norwegian, Swedish (Mapi Research Institute, Quality of Life Instrument Database; (www.Mapi-research-inst.com/result23.asp?free=; Carpenter and Andrykowski, 1998), German (used in study; Centre for Psychiatry, Germany; www.med.uni-giessen.de/psychiat/Arbeitsordner/psychhomepage02/Sites/Englisch/e_kognitionslabor.htm.), Brazilian (Ceolim and Menna-Barreto, 2000).	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Average of 0.83 across 7 component scores (no total)	Met
(iii) Correlation of each item with rest of scale	Range, 0.20–0.66 (100% have 0.20 or above)	Met
(iv) Internal consistency assessed by factor-analytic or similar techniques	Only alphas were computed to test internal consistency (see Eii).	-
(v) Test-retest reliability	Time 1/time 2 correlation, 0.85 for global scores (length of interval not reported; tested on 91 patients)	-
(vi) Inter-rater reliability	Not reported	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	An analysis of variance was used to compare patient groups for Pittsburgh Sleep Quality Index global and component scores. Patients with disorders of initiating and maintaining sleep and depressed patients had significantly higher scores than patients with disorders of excessive somnolence.	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet

(ii) Evidence that change over time correlates with other measures?	No	-
(iii) Are the changes over time found significant to carers?	Effect size not reported	-
H Precision		
(i) Type of response categories	4-point scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	Yes	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	4	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	7	-
J Feasibility		
(i) Mode of administration	Structured interviews with participant and his/her bed partner	-
(ii) Number of items	19	-
(iii) Ease of explaining measure to study participants	Yes	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-

(vii) Cost	Free	-

Table 52 Positive and Negative Affect Schedule (Watson and Clark, 1988)

The Positive and Negative Affect Schedule (PANAS) is a self-report measure designed for a general population and tested on undergraduates. Internal consistency was good but test-retest reliability only moderate. Construct validity was reasonably good but the measure has not been validated for a carer population. The authors report that the Short PANAS can be recommended for use when measures of both positive and negative affect are required.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: positive and negative affect		
A Importance to carers		
(i) Data obtained from interviews	Carer’s mental health	Met
(ii) Data obtained from questionnaires	Carer’s depression and anxiety	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	Not reported	Unmet/no evidence
(iii) Response rates	Unknown	-
(iv) Time to complete	Not reported	-
(v) Short form available	10-item short form (Mackinnon *et al.*, 1999). The existence of two nearly-orthogonal dimensions of positive and negative affect was established using confirmatory factor analytic techniques. Factor structure and factor correlations found to be unchanged with age. A multiple indicators, multiple causes model was used to investigate differences in item responses according to age, sex, education, marital status and financial hardship that could not be accounted for by differences in affect levels between groups. Only 1 item, 'excited' from the Positive Affect scale, was found to elicit differential responses (www.anu.edu.au/perc/anrep/lifespan.htm)	-

(vi) Translated measure available	Used in Argentina, Australia, Austria, Bahrain, Brazil, China, Colombia, Denmark, Egypt, Estonia, Finland, Germany, Ghana, Greece, Guam, Hong Kong, Hungary, India, Indonesia, Italy, Japan, Korea, Lithuania, Nepal, Nigeria, Norway, Pakistan, Peru, Portugal, Puerto Rico, Singapore, Slovenia, South Africa, Spain, Taiwan, Tanzania, Thailand, Turkey, Zimbabwe (Lucas *et al.*, 2000) and reported to be excellent (see Shao, 1997).	-
(vi) Translated measure available continued	Only the Chinese-, Japanese-, Korean- and Spanish-language versions were back-translated. Back translations: French Canadian (Gaudreau, 2000), Korean (Kim *et al.*, 1999). Negative affect subscale only. Convergent and discriminant validity were established by comparing the ratings on the PANAS with those of 5 other short mood scales. Cronbach alpha coefficients, 0.89 at pre-test; 0.85 at post-test.	
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Positive affect subscale, 0.88; negative affect, 0.86	Met
(iii) Correlation of each item with rest of scale	Not reported	Unmet/no evidence
(iv) Internal consistency assessed by factor-analytic or similar techniques	Factor analysis showed good internal consistency.	-
(v) Test-retest reliability	Positive affect, 0.55; negative affect, 0.50	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	Scores in the Hopkins Symptom Checklist (Derogatis *et al.*, 1974) strongly correlated with negative affect (past few weeks, 0.74; today, 0.65). The Beck Depression Inventory (Beck *et al.*, 1961) also correlates with negative-affect scores (past few days, 0.56; past few weeks, 0.58). It also has significant negative correlations with positive affect (past few days, -0.31; past few weeks, -0.36).	-

G Responsiveness		
(i) Used to measure change over time?	Yes – the authors report a study (Watson, 1988) in which it was found that within-subject variations of perceived stress were strongly correlated with fluctuations in negative affect.	Met
(ii) Evidence that change over time correlates with other measures?	See above	-
(iii) Are the changes over time found significant to carers?	Unknown	-
H Precision		
(i) Type of response categories	5-point scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	No	-
(iv) Sources of potential bias in scoring	No	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	No	-
(ii) Norms presented	Available for non-carers	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	6	-
(vi) Existence of a meaningful overall score	2 scores	-
(vii) Number of subscales	2	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	20	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-

(vii) Cost	Free	-

Table 53 Profile of Mood States Inventory (McNair et al., 2003)

The Profile of Mood States Inventory (POMS) is a self-administered questionnaire measuring six identifiable mood or affective states. It was developed for use with the general population in response to increasing attention to mood states and mood changes in the literature on the effects of brief psychotherapies, psychotropic medications, sleep deprivation, emotional stimulation and other experimental operations. The authors felt there was a need for a rapid method of identifying and assessing transient, fluctuating affective states. Carers were not involved in its development and its acceptability to them is unknown. Its reliability, construct validity and responsiveness are good.

Criterion+indicators	**Data**	**Rating (where applicable)**
Constructs covered: affective states		
A Importance to carers		
(i) data from interviews with carers and representatives	Carer's mental health	Met
(ii) data from conference questionnaires	Carer's depression, carer's anxiety	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Unknown	Unmet/no evidence
(ii) Acceptability to other populations	Not reported	Unmet/no evidence
(iii) Response rates	Not reported	-
(iv) Time to complete	3–5 min	-
(v) Short form available	Yes (Shacham, 1983) – 30 items	-
vi) Translated measure available	Only available from the Mental Health Foundation (www.mentalhealth.org.uk) in English. Researchers may have translated it into other languages.	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-

(ii) Whether the reliability or validity of the measure has been established on a carer population	No	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Not reported. A Kuder–Richardson formula was calculated to measure reliability of each of the factors of the scale with two separate samples, the first of 350 male psychiatric outpatients, the second of 650 female psychiatric outpatients:	
	Tension-Anxiety (T): 1 alpha=0.92, 2 alpha=0.90	(Partially) met
	Depression-Dejection (D): 1 alpha=0.95, 2 alpha=0.95	Partially met
	Anger-Hostility (A): 1 alpha=0.92, 2 alpha=0.93	Partially met
	Vigour (V): 1 alpha=0.89, 2 alpha=0.87	Met
	Fatigue (F): 1 alpha=0.94, 2 alpha=0.93	Partially met
	Confusion-Bewilderment (C): 1 alpha=0.87, 2 alpha=0.84	Met
(iii) Correlation of each item with rest of scale	Depression-Dejection (D) items, 0.25–0.69	
	Anger-Hostility (A) items, 0.24–0.66	
	Vigour-Activity (V) items, 0.45–0.80	
	Fatigue-Intertia (F) items, 0.31–0.73	
	Confusion-Bewilderment (C) items, 0.34–0.58	
(iv) Internal consistency assessed by factor-analytic or similar techniques	Six independent factor-analytic studies have been conducted. Studies 1–3 were conducted during the early stages of test development and each involved between 150 and 523 male psychiatric outpatients. All three used the 4-point rating scales and were conducted during the early stages of test development. Studies 4–6 used the 5-point rating scales. These studies indicate that the same six mood factors can be identified, measured reliably and replicated in male psychiatric outpatients, in male college students, and in male and female outpatients at a private teaching institution. The factors appear to be relatively invariant whether the rating period is the immediate present or spans a 1-week period and also regardless of whether the 4- or 5-point scale format is employed.	-

(v) Test-retest reliability	100 patients (10%) in studies 5 and 6 were assessed between intake and pretherapy. Product-moment correlations among their POMS scores were computed at the two time periods. Tension-Anxiety: 0.70 Depression-Dejection: 0.74 Anger-Hostility: 0.71 Vigor: 0.65 Fatigue: 0.66 Confusion-Bewilderment: 0.68 The median time between intake and pre-treatment was 20 days with a range of 3–100 days.	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	A number of reports of controlled clinical drug trials suggest the POMS factors are sensitive to short-term changes associated with mild tranquillisers, e.g. Lorr *et al.* (1964). Significant validity coefficients for outpatient samples were: Factor T(ension) – Manifest Anxiety Scale=0.8, Clinical Rating of Prognosis=0.27; Factor D(epresson) – IMPS Intropunitiveness=0.3, Patient Prediction of Improvement=-0.28; Factor A(nger) – IBI Hostility=0.32, IBI Mistrust=0.31, IBI Passive Dependency=0.21; Factor V(igour) with IBI Exhibitionism=0.21, IBI Sociability=0.18, Observer Rating of Interview Activity=0.29	-
G Responsiveness		
(i) Used to measure change over time?	In several studies one or more of the POMS factor scores have proved sensitive to change associated with psychotherapy. Lorr *et al*. (1961) found highly significant improvement on Tension-Anxiety, Depression-Dejection, Anger-Hostility and Fatigue over an 8-week treatment period. Patients showed a reduction (at a low significance level) in Depressed mood after 12 weeks of time-limited psychotherapy in a study by Haskell *et al.* (1969), and no significant change in Tension and Anger. In an independent study involving 24 time-limited psychotherapy patients at the same clinic, Holstein (1970) found significant improvement on all POMS factors except Vigor. The results suggest that the POMS does not change simply as a function of repeated testing during treatment and that the degree of change is meaningfully related to either duration of treatment, the ending of treatment, or both.	Met

(ii) Evidence that change over time correlates with other measures?	Not presented	-
(iii) Are the changes over time found significant to carers?	Unknown	-
H Precision		
(i) Type of response categories	Five point Likert scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	The Crowne–Marlowe (1960) measure of social desirability was found to have low to moderate correlations with the POMS scores of the 150 patients in study 3. The correlations were as follows: Tension=-0.21 Depression=-0.36 Anger=-0.52 Vigor=0.33 Fatigue=-0.18	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	-

(ii) Norms presented	Data for an adult normative sample (*n*=2360) are as follows: Factor T Female M=12.8, s=7.9, alpha=0.92 Male M=12.3, s=7.0, alpha=0.90 Factor D Female M=10.2, s=10.4, alpha=0.93 Male M=8.3, s=8.7, alpha=0.92 Factor A Female M=9.7, s=9.3, alpha=0.93 Male M=9.2, s=8.3, alpha=0.92 Factor V Female M=14.9, s=6.7, alpha=0.90 Male M=16.3, s=6.3, alpha=0.90 Factor F Female M=8.3, s=6.8, alpha=0.93 Male M=7.0, s=5.7, alpha=0.91 Factor C Female M=7.3, s=5.4, alpha=0.85 Male M=7.0, s=5.7, alpha=0.83 TMD Female M=33.4, s=37.1 Male M=27.2, s=31.8	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	3	-
(vi) Existence of a meaningful overall score	Yes: a profile of scores on states of tension, anger, fatigue, depression, confusion and vigor.	-
(vii) Number of subscales	6	-
J Feasibility		
(i) Mode of administration	Self-administered	-
(ii) Number of items	65	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	None	-

(vi) Freedom of usage	The POMS must be purchased from the Mental Health Foundation (www.mentalhealth.org.uk).	-
(vii) Cost	The POMS manual costs $20, a packet of 100 Quikscore™ forms costs $75.	-

Table 54 Self-Rating Depression Scale (Zung, 1965)

The Self-Rating Depression Scale (SDS) is a self-report measure of depression for use with patients with primary diagnoses of a depressive disorder. Construct validity is good but there is little evidence of reliability and other types of validity testing in the original paper. Psychometric testing for the Greek translation revealed five factors – anxiety-depression, thought content, gastrenterological symptoms, irritability and social-interpersonal functioning – which may be consistent with the original version. It has not apparently been validated for use with carers.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: depression		
A Importance to carers		
(i) Data obtained from interviews	Carer's mental health	Met
(ii) Data obtained from questionnaires	Carer's depression	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	Not reported	Unmet/no evidence
(iii) Response rates	Not reported	-
(iv) Time to complete	Not reported	-
(v) Short form available	Not found	-

(vi) Translated measure available	Chinese, Dutch, French, German, Italian, Polish, Greek. Psychometric properties assessed with depressed patients and comparison 'normal' subjects. Back-translation performed. Clinical diagnosis was reached by consensus of two examiners with the use of the SCAN v.2.0. Sensitivity and specificity exceed 90 at 44/45; Cronbach's alpha for total scale, 0.09. Factor analysis revealed five factors (anxiety-depression, thought content, gastrenterological symptoms, irritability and social-interpersonal functioning). Test-retest reliability satisfactory (Pearson's *R*=0.92). Japanese (Kawada *et al.*, 1999).	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Not reported	Unmet/no evidence
(iii) Correlation of each item with rest of scale	Not reported	Unmet/no evidence
(v) Test-retest reliability	Not reported	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	56 patients who had admitting diagnoses of depressive disorder were administered the SDS. The SDS indices of the group who were admitted and discharged as depressive disorder, before treatment ranged from 0.63 to 0.9 with a mean of 0.74 (the lowest possible index is 0.25, highest possible index is 1). The SDS indices of the group who were admitted with a depressive disorder but discharged with another diagnosis ranged pre-treatment from 0.38 to 0.71, with a mean of 0.53. After treatment the SDS indices of the depressive-disorder group ranged from 0.30 to 0.50, which a mean of 0.39. The SDS indices for the control group (no depressive diagnoses) were 0.25–0.43 with a mean index of 0.33.	-

G Responsiveness		
(i) Used to measure change over time?	Yes – see (Fiii).	Met
(ii) Evidence that change over time correlates with other measures?	SDS indices went down in the group who were treated for depression in hospital and discharged.	-
(iii) Are the changes over time found significant to carers?	Not reported	-
H Precision		
(i) Type of response categories	4-point Likert scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	9	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	None	-
J Feasibility		
(i) Mode of administration	Self-report	
(ii) Number of items	20	
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	None	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 55 Social Support Appraisals Scale (Vaux et al., 1986)

The Social Support Appraisals Scale (SS-A Scale) is a self-report measure developed for use with the general population. Its internal consistency and validity are good, but it has not been validated specifically for carers.

Criterion+indicators	Data	Rating (if applicable)
Constructs covered: subjective appraisals of support		
A Importance to carers		
(i) Data obtained from interviews	Not a construct mentioned by carers	Unmet
(ii) Data obtained from questionnaires	Not a construct mentioned by carers	Unmet
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	Not reported	Unmet/no evidence
(iii) Response rates	Not reported	-
(iv) Time to complete	Not reported	-
(v) Short form available	Not found	-
(vi) Translated measure available	Not found	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence

(ii) Cronbach's alpha	With 5 student samples the mean alpha values were: 0.90 for the total scale, 0.80 for the family subscale and 0.84 for the friend subscale. With 5 community samples the mean alphas were: 0.90 for the total scale, 0.81 for the family subscale and 0.84 for the friend subscale.	Met
(iii) Correlation of each item with rest of scale	Not reported	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	The family and friend subscales were moderately associated for both student samples (mean r=0.51) and community samples (mean r=0.52).	-
(v) Test-retest reliability	Not reported	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-

(iii) Construct validity	Validity of the SS-A Scale was assessed in terms of convergent and divergent validity with other subjective support measures, and an examination of its relationships with theoretically linked antecedents (support network resources) and consequences (psychological distress). Measures completed by respondents therefore fell into 3 major categories: social support appraisals, social support resources, and distress and well-being. For each category the testing was extensive so only 1 example of each is presented here. *SS-A Scale and other support appraisal measures* Correlations between family and friend scores for the Perceived Support Scale (Procidano and Heller, 1983) and SS-A Scale show a pattern indicating moderate convergent and divergent validity. Correlations between respective pairs of family or friend scales (family score correlations for the two student samples, 0.56, 0.82; friend score correlations for the two student samples, 0.53, 0.72) are considerably higher than those between family and friend scales (0.21–0.48). Other examples of convergent and divergent validity with other support appraisal measures are presented by the authors. *SS-A Scale and support resources* Correlations between total network size and total SS-A score were 0.22 for one student sample, and 0.16 for one community sample. Correlations with other types of support resources are also presented by the authors. *SS-A Scale and distress/well-being* The Center for Epidemiological Studies Depression Scale (Radloff, 1977) and SS-A Scale showed significant moderate inverse correlations across four community samples (with total SS-A scores: –0.40, -0.43, -0.55, -0.14, -0.27). SS-A total (0.32, 0.34, 0.21, 0.37), family (0.24, 0.17, 0.35, 0.20) and friend (0.29, 0.30, 0.20, 0.36) scales each showed significant small correlations with positive affect as measured by the Affect Balance Scale (Bradburn, 1969) in 4 of the 5 samples. The SS-A Scale and the UCLA Revised Loneliness Scale (Shaul, 1981) showed significant small to moderate inverse correlations in a student sample (-0.40) and a strong inverse relationship in a community sample (-0.71). Overall the authors conclude that the data on the convergence of the SS-A Scale with other support appraisal measures and its relationship to theoretically linked antecedent and consequent variables yield a picture that provides a good deal of evidence for the scale's validity.	-

G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	4-point Likert scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	5	-
(vi) Existence of a meaningful overall score	3 scores are generally computed: SS-A total (sum), SS-A family (sum of 8 'family' items) and SS-A friends (sum of 7 'friend' items).	-
(vii) Number of subscales	See above. After the friend and the family items, the remaining items refer to 'people' or 'others' in a general way and are not a separate subscale.	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	23	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	None	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 56 Social Support Questionnaire (Sarason et al., 1983)

The Social Support Questionnaire (SSQ) is a self-report measure of social support split into the number of supports and satisfaction with support. It was designed for a general population (tested on undergraduates). Its reliability and validity are good but have not tested for a caregiving population.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: social support		
A Importance to carers		
(i) Data obtained from interviews	Not a construct mentioned by carers	Unmet
(ii) Data obtained from questionnaires	Not a construct mentioned by carers	Unmet
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	Not reported	Unmet/no evidence
(iii) Response rates	Unknown	-
(iv) Time to complete	Unknown	-
(v) Short form available	Social Support Questionnaire – Short Form (6-items; Sarason *et al.*, 1987)	
(vi) Translated measure available	Japanese (Takada, 2002)	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	S score (satisfaction with support), 0.94; N score (number of supports), 0.97	Partially met

(iii) Correlation of each item with rest of scale	N range, 0.51–0.79; S range, 0.48–0.72. Individual item-total correlations were not given. All item-total correlations were greater than 0.20.	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	One factor accounted for 82% of variance of N score, and 72% of S score. Evidence that one strong factor underlies each of the two scores and they represent different dimensions of the same concept.	-
(v) Test-retest reliability	N, 0.90; S, 0.83	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	It was predicted that those high in social support might also be extroverted, have memories of supportive relationships in childhood, show less emotional discomfort, and have a more optimistic outlook about the future. Significant negative correlations for women between the SSQ-N and SSQ-S scores and measures of emotional discomfort and recollections of separation anxiety in childhood. The EPI Extraversion measure (Eysenck and Eysenck, 1968) was positively correlated with SSQ-N, while the Neuroticism measure (Eysenck and Eysenck, 1968) was negatively correlated with SSQ-S only in women. Some significant correlations between some items on the Ladder of Life (Cantril and Roll, 1971) and SSQ scores, and with Bradburn's Affect Balance Scale (Bradburn, 1969).	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	None	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	6-point scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Social desirability was controlled for with the Crowne–Marlowe measure (1960): there was no correlation.	-
I Interpretability		

(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	-
(ii) Norms presented	Not reported	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	7	-
(vi) Existence of a meaningful overall score	2 scores (1 for each subscale)	-
(vii) Number of subscales	2	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	27	-
(iii) Ease of explaining measure to study participants	Easy	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	None	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 57 State-Trait Anxiety Inventory (Spielberger et al., 1974)

The State-Trait Anxiety Inventory is a self-report measure of respondents' anxiety. It distinguishes between respondents' anxiety in response to situations (state) and anxiety as part of their personality (trait). We were unable to obtain information from the publishers within the timescale of this study, and therefore the data provided were gleaned from other sources.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: anxiety		
A Importance to carers		
(i) Data obtained from interviews	Carer's mental health	Met
(ii) Data obtained from questionnaires	Carer's anxiety	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet

(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	Not reported	Unmet/no evidence
(iii) Response rates	Not reported	-
(iv) Time to complete	10–20 min	-
(v) Short form available	Not found	-
(vi) Translated measure available	Used in many countries including The Netherlands, Sweden, Turkey, Germany, Italy, Brazil and Spain. A comprehensive list is not possible as its use is so widespread (a Medline search alone resulted in 1005 hits).	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Not reported	Unmet/no evidence
(iii) Correlation of each item with rest of scale	Not reported	Unmet/no evidence
(iv) Internal consistency assessed by factor-analytic or similar techniques	Not reported	-
(v) Test-retest reliability	The stability of the State-Trait Anxiety Inventory scales was assessed on male and female samples of high school and college students for test-retest intervals ranging from 1 h to 104 days. The magnitude of the reliability coefficients decreased as a function of interval length. For the Trait-anxiety scale the coefficients were 0.65–0.86, whereas the range for the State-anxiety scale was 0.16–0.62. This low level of stability for the State-anxiety scale is expected since responses to the items on this scale are thought to reflect the influence of whatever transient situational factors exist at the time of testing.	-
(vi) Inter-rater reliability	N/A	-
F Validity		

(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	Correlations are presented in the manual between this scale and other measures of trait-anxiety: the Taylor Manifest Anxiety Scale, the IPAT Anxiety Scale, and the Multiple Affect Adjective Check List. These correlations are 0.80, 0.75 and 0.52, respectively.	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	4-point Likert scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	-
(ii) Norms presented	Both percentile ranks and standard (T) scores are available for male and female working adults in three age groups (19–39, 40–49, 50–69), male and female high school and college students, male military recruits, male neuropsychiatric patients, male medical patients and male prison inmates. These are reported in the manual (not available here).	-
(iii) Cut off (e.g. for caseness) presented	Not reported	-
(iv) Calibrated against other life events or experiences	Not reported	-
(v) No. of times used in review	7	-
(vi) Existence of a meaningful overall score	2 overall scores – 1 for trait anxiety and 1 for state anxiety	-
(vii) Number of subscales	2	-
J Feasibility		

(i) Mode of administration	Self-report	-
(ii) Number of items	Not reported	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	Not reported	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Available from www.mindgarden.com	-
(vii) Cost	$30 for a sampler set from www.mindgarden.com	-

Table 58 Ways of Coping Checklist (Folkman and Lazarus, 1985; Vitaliano et al., 1985)

The Ways of Coping Checklist (WOC) is a self-report measure of coping, developed for the general population. Its internal consistency is good, but test-retest reliability and inter-rater reliability are not reported. Construct validity is good. The measure has not apparently been validated for use with carers.

Criterion+indicators	**Data**	**Rating (if applicable)**
Constructs covered: Different ways of coping		
A Importance to carers		
(i) Data obtained from interviews	Carer’s ability to cope	Met
(ii) Data obtained from questionnaires	Carer’s ability to cope	Met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	Not reported	Unmet/no evidence
(iii) Response rates	Not reported	-
(iv) Time to complete	Not reported	-
(v) Short form available	Amended 28-item version developed to be a quicker than the original (MacCarthy and Brown, 1989)	-

(vi) Translated measure available	Used in the following countries: South Africa, Korea, The Netherlands, France, Germany, Poland, Japan, Malawi, Norway, Sweden and Israel.	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	Spouses of patients with Alzheimer's made up one of the samples used to test the psychometric properties of the revised version of the WOC (Vitaliano *et al.*, 1985).	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Unknown	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Vitaliano *et al*. (1985) quote alpha values for their five subscales of the WOC using three samples: medical students, spouses of patients with Alzheimer's, and psychiatric outpatients. They are as follows: Problem-focused: 0.88, 0.85, 0.88 Wishful thinking: 0.85, 0.86, 0.87 Seeks social support: 0.75, 0.79, 0.81 Blamed self: 0.78, 0.8, 0.76 Avoidance: 0.74, 0.73, 0.81 MacCarthy and Brown (1989) quote alpha values for the four subscales resultant from factor analysis, although in analyses they dropped factor 4 and combined factors 1 and 3. The alpha values are as follows:	Met
	1 Problem solving and reorientation: 0.75	Met
	2 Acting out and distraction: 0.69	Partially met
	3 Distancing: 0.70	Met
	4 Wishful thinking and denial: 0.43	Unmet
(iii) Correlation of each item with rest of scale	Not reported	Unmet/no evidence

(iv) Internal consistency assessed by factor-analytic or similar techniques	Vitaliano *et al*. (1985) carried out principal components analysis on the original 68-item WOC with the aim of cutting it down. 6 factors resulted from the analysis, and 46 items. Items were cut which had a loading of less than 0.35. The factor 1 subscale was named 'problem-focused coping', the factor 2 subscale 'blamed self', the factor 3 subscale 'wishful thinking' and the factor 4 subscale 'seeks social support'. Factors 5 and 6 were dropped from the analysis because items loaded equally as high on other factors or did not add much to the explained variance. A fifth subscale, 'avoidance', was created from items which loaded high on factor 2 but which were conceptually different.	-
(iv) Internal consistency assessed by factor-analytic or similar techniques continued	MacCarthy and Brown (1989) performed principal components analysis on their 28-item version of the WOC. It yielded 4 factors with eigenvalues greater than 1. Together the 4 factors accounted for 40.8% of the variance. The fourth factor (which had been provisionally labelled 'wishful thinking and denial') had dubious reliability as it contained a small number of items and failed to achieve an acceptable alpha level, so was excluded. The first and third subscales proved to be moderately highly correlated (r=0.54) and appeared to have some overlap in content. When the items were combined, the subscale achieved an alpha level of 0.79. The composite subscale was labelled 'positive coping' and the second factor subscale was labelled 'maladaptive coping' on the basis of their item contents.	
(v) Test-retest reliability	Not reported	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-

(iii) Construct validity	The construct validity of the original and revised scales was assessed by examining the relationships of coping to the source of the stressor, appraisal, and distress (Vitaliano *et al.*, 1985). Overall there were 10 replicated relationships of the revised scales with appraisal and distress, all of which would have been anticipated theoretically from the transactional model of stress (Lazarus and Launier, 1978) on which the scale is based. They found that the appraisal of the stressful situation as one 'that you could change or do something about' was related to the Seeks social support subscale, which contains problem-solving strategies through social contacts. The problem-focused scale was related to the appraisal of the stressful situation as something 'that you needed to know more about before you could act'. A significant negative association was found between the Problem-Focused scale and depression (measured by the Beck Depression Inventory (Beck *et al.*, 1961) and the Hamilton Depression Scale (Hamilton, 1960)) in both medical students and the spouses of Alzheimer's patients. Other such correlations are reported in Vitaliano *et al*. (1985), although no actual *r* values are reported.	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	5-point Likert scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	No	-

(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	22	-
(vi) Existence of a meaningful overall score	No. The measure has only been used in subscales. As the subscales (in whatever form they take) measure different styles of coping, it would be inappropriate to sum their scores.	-
(vii) Number of subscales	MacCarthy and Brown (1989) analysed their data using two subscales 'positive coping' and 'maladaptive coping.'	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	66-item original version (Folkman and Lazarus, 1985); 28-item version (MacCarthy and Brown, 1989); 48-item version (Vitaliano *et al.*, 1985)	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	The original Ways of Coping Checklist (Folkman and Lazarus, 1985) is in the public domain. The Publishing Psychologists Press (1988) made minor modifications to a few items. Their version was copyrighted and bought by Mind Garden. Permission must be sought for use.	-
(vii) Cost	A sampler set of a revised version is available at www.mindgarden.com costing $30. The original is free.	-

5.3 Results: measures from 1994 onwards used fewer than three times

Of the 184 instruments that had been used fewer than three times, 152 had been cited in articles published from 1994 onwards. These measures were entered into a database.

One listed measure, the Attributional Style Questionnaire, was removed from the database upon discovery that it was cited in an article published in 1988. Another, the Caregiver Activities Checklist, was removed when it was discovered that it was not cited in the reference given, and could not be located in Medline or PsycINFO searches. The Caregiver Burden Questionnaire was found to be comprised of five separate instruments, only two of which originated after 1993. These two additional measures were also entered into the database, resulting in a total of 151 measures to be considered for eligibility.

The articles citing measures from 1994 onwards used fewer than three times were obtained and read, and the article in which the development of the measure is described was obtained if it was published from 1994 onwards.

Of the 151 instruments potentially eligible for review, 25 were found to have been developed from 1994 onwards. Of these, five were discarded because, upon close inspection, they were found to collect data on carers for reasons other than to assess the carers' own outcomes. These instruments (see References 6) included an 'illness perception' questionnaire (Barrowclough *et al*., 2001), a 'social control' scale (Struening *et al.*, 1995), a 'negative social interaction' scale (Rauktis *et al.*, 1995) and a 'responsibility attribution' scale (Provencher and Mueser, 1997). Of the 20 measures remaining, three had not been used with carers for people with mental health problems or dementia, and so were not eligible for inclusion in the instrument review. These measures included the Caregiver Quality of Life Index (CQLI; McMillan, 1994) and Caregiver Burden Scale (Elmståhl *et al.*, 1996), both of which have been used with carers for cancer patients only. We were unable to retrieve four of the remaining 17 articles describing the development of a measure within the time frame of the study, despite attempted contact with their authors, resulting in 13 measures eligible for review.

Two additional measures, the Carers' Assessment of Difficulties Index (CADI) and the Carers' Assessment of Satisfactions Index (CASI) were not developed within the last ten years, but were specifically mentioned as being of interest by the referees who reviewed this report and so are included here, bringing the number of measures considered for eligibility in this section to 15. These 15 measures are reviewed below (see References 4).

5.3.1 Published measures developed for carers for people with mental health problems or dementia

Table 59 Carers' Assessment of Difficulties Index (Nolan and Grant, 1992)

The Carers' Assessment of Difficulties Index (CADI) was developed for use with family caregivers. It can be administered in self-report or interview format. Carers are asked to consider 30 items describing various difficulties that may arise from caregiving, for example 'it causes financial difficulties', to rate whether each problem applies to them and, if so, to rate how stressful they find it. The authors recommend using CADI in conjunction with the Carers' Assessment of Managing Index (CAMI) and the Carers' Assessment of Satisfactions Index (CASI), as the basis for a comprehensive assessment of carers' needs. CADI's acceptability to carers is not reported. Its reliability is good, although validity testing is limited. This may be explained by the fact that the measure is intended to provide a personal profile of individual carers' difficulties, rather than an interpretable overall score.

Criterion+indicators	Data	Rating (where applicable)
Constructs covered: the particular difficulties that a carer faces, and how stressful these are perceived to be		
A Importance to carers		

(i) Data from interviews with carers and carer representatives	Carer's relationship with the care recipient, effect of caregiving on the carer's family life, effect of caregiving on the carer's social life, how burdened the carer feels, how the carer feels about caregiving, carer's confidence in providing care, carer's physical health, attitudes of health-care providers, effectiveness of the service, the effect of caregiving on the carer's finances	Met
(ii) Data from conference questionnaire	Carer's relationship with the care recipient, effect of caregiving on the carer's family life, effect of caregiving on the carer's social life, carer's confidence in providing care, carer's physical health, effect of caregiving on the carer's daily life, carer's satisfaction with the service, sort of caregiving activities the carer undertakes, how burdened the carer feels, how the carer feels about caregiving	Met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Measure designed for family carers for people with various types of disability or illness, including dementia	Met
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	N/A	N/A
(iii) Response rates	2050 questionnaires including the CADI were distributed via the Association of Carers (AOC) newsletter (Nolan and Grant, 1992) with a response rate of 35%. The authors argue that this rate is satisfactory, considering that 250 Associate members of the AOC were non-carers and that follow-up requests were not possible.	-
(iv) Time to complete	Nolan *et al.* (1998) report that, on a self-administered basis, completing CADI 'need only take a matter of minutes', but that as part of an assessment interview it can take 10–20 min to complete	-
(v) Short form available	No	-
(vi) Translated measure available	CADI has been translated into over 15 languages, including most of the major European languages, Japanese, and three languages from the Indian subcontinent. Method of translation not reported. Further details can be obtained from the first author.	-

D Appropriateness		
(i) Whether carers were included in the piloting of the measure	Yes	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Yes	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Total scale, 0.92. Individual factors (see E iv) (Nolan *et al.*, 1998):	Partially met
	Carer-dependent relationships, 0.83	Met
	Reactions to caregiving, 0.84	Met
	Physical demands of caring, 0.73	Met
	Restricted social life, 0.74	Met
	Poor family support, 0.76	Met
	Poor professional support, 0.79	Met
	Financial consequences, 0.71	Met
(iii) Correlation of each item with rest of scale	Not reported	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	Principal components analysis with varimax rotation was conducted to determine the factor structure of CADI. Items with a factor loading of 0.35 or over were included, and where items loaded on to more than one scale, they were only included on the scale with the highest loading. The following 7 factors were identified (Nolan *et al.*, 1998): Carer-dependent relationships: 7 items, factor loadings 0.47–0.68 Reactions to caregiving: 7 items, factor loadings 0.37–0.61 Physical demands of caregiving: 6 items, factor loadings 0.41–0.62 Restricted social life: 3 items, factor loadings 0.45–0.65 Poor family support: 2 items, factor loadings 0.66–0.80 Poor professional support: 2 items, factor loadings 0.54–0.94 Financial consequences: 2 items, factor loadings 0.61–0.93	-
(v) Test-retest reliability	Not reported	-
(vi) Inter-rater reliability	Not reported	-

F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	CADI is based on the transactional model of stress (as explicated in Nolan *et al.*, 1998). The authors state that the content validity of CADI was ensured by extensive reference to existing scales and to the empirical and theoretical literature available. The measure includes space for the carer to add, and rate, any additional difficulties not on the scale.	-
(iii) Construct validity	CADI (originally the Carer Perceived Problem Checklist; CPPC) was initially developed to provide data for an empirical test of the transactional model of stress (Nolan *et al.*, 1990). Data obtained from 554 carer responses to care-receiver continence, ADL and mental frailty scales, the CPPC and the Malaise Inventory (Rutter *et al.*, 1970) were subjected to causal path and factor analysis. Factors of the CPPC were found to mediate between factors of the objective caregiving environment and carers' psychological malaise, as measured by the Malaise Inventory.	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	4-point Likert scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-

(v) No. of times used in this review	2	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	7	-
J Feasibility		
(i) Mode of administration	Face-to-face interview, self-report, or 'card format'. For the latter, the carer is given a set of cards, each with a CADI item printed on it. The respondent is asked to sort the cards into 2 piles to separate those statements that apply to them from those that do not. Following this, the carer is asked to look again at the cards that do apply, and to explain in what ways each item is relevant to them.	-
(ii) Number of items	30, plus space for the respondent to add, and rate, other difficulties not on the scale.	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	Not reported	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 60 Carers' Assessment of Managing Index (Nolan et al., 1995)

The Carers' Assessment of Managing Index (CAMI) was developed for use with family caregivers. It can be administered via self-report or interview format. Carers are asked to consider 36 items describing various coping strategies, for example 'establishing a regular routine and sticking to it', to rate whether they use each strategy and, if so, to rate how helpful they find it. The authors recommend using CAMI in conjunction with the Carers' Assessment of Difficulties Index (CADI) and the Carers' Assessment of Satisfactions Index (CASI), as the basis for a comprehensive assessment of carers' needs. CAMI's acceptability to carers is not reported. Reliability and validity testing is limited. This may be explained by the fact that the measure is intended to provide a personal profile of individual carers' coping strategies, rather than an interpretable overall score.

Criterion+indicators	**Data**	**Rating (where applicable)**
Constructs covered: the particular coping strategies that a carer uses, and how helpful they are perceived to be		
A Importance to carers		
(i) Data from interviews with carers and carer representatives	Carer's ability to cope	Met

(ii) Data from conference questionnaire	Carer's ability to cope	Met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Measure designed for family carers for people with various types of disability or illness, including dementia	Met
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	Items were generated following a review of the literature and in-depth interviews conducted with over 100 carers for individuals with a range of dependency needs. Wherever possible, items were constructed using the carer's own words.	Met
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	N/A	N/A
(iii) Response rates	In the initial testing of the instrument, a postal questionnaire was distributed to family carers through the Alzheimer's Disease Society in Wales and Scotland and the Crossroads Care Network in Wales. 266 completed questionnaires were returned. The authors point out the difficulty of calculating a response rate, as the exact number of questionnaires distributed was unknown.	-
(iv) Time to complete	Nolan *et al.* (1998) report that, on a self-administered basis, completing CAMI 'need only take a matter of minutes', but that as part of an assessment interview it can take 10–20 min to complete.	-
(v) Short form available	No	-
vi) Translated measure available	CAMI has been translated into over 15 languages, including most of the major European languages, Japanese, and three languages from the Indian subcontinent. Method of translation not reported. Further information can be obtained from the first author.	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	Yes	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Yes	-
E Reliability		

(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Based on the responses of 266 family carers (see C iii), Cronbach's alpha value was 0.86. Nolan *et al.* (1998) report a Cronbach's alpha value of 0.85.	Met
(iii) Correlation of each item with rest of scale	Not reported	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	No	-
(v) Test-retest reliability	Not reported	-
(vi) Inter-rater reliability	Not reported	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	The instrument is based on the transactional model of stress and coping (as explicated in Nolan *et al.*, 1998). Items were developed through a review of the literature, and interviews with over 100 carers. In addition to the quantitative data on the questionnaire, carers are asked to identify additional coping strategies used that are not included on the original index. In the initial testing of the measure with 266 carers, subjects elaborated further upon certain items on the scale, but no conceptually distinct strategies were added. Together with the manner of development of the instrument, the authors argue that this suggests grounds for confidence in the content validity of CAMI.	-
(iii) Construct validity	The authors state that, as the scale is well grounded theoretically (having been derived explicitly from a transactional model of stress), 'this augers well for good construct validity'. No formal evidence of construct validity is presented, however.	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	4-point Likert scale	-
(ii) Ceiling and floor effects	Not reported	-

(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	No	-
(ii) Norms presented	Nolan *et al.* (1995) provide a list of the 19 most frequently cited coping strategies in their study of 266 carers, plus the percentage of carers who used each strategy and found it either helpful or very helpful.	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in this review	1	-
(vi) Existence of a meaningful overall score	No. Nolan *et al.* (1998) stress that CAMI is intended for use as part of an assessment of individual caregiving circumstances. The instrument provides a detailed profile of the coping strategies that a carer uses, that can be used as a basis for identification of need. CAMI could be used to compare coping profile patterns of different groups of carers, but an overall score is not appropriate.	-
(vii) Number of subscales	0	-
J Feasibility		
(i) Mode of administration	Face-to-face interview, self-report or card format. For the latter, the carer is given a set of cards, each with a CAMI item printed on it. The respondent is asked to sort the cards into 2 piles in order to separate those statements that apply to them from those that do not. Following this, the carer is asked to look again at the cards that do apply, and to explain in what ways each item applies to them.	-
(ii) Number of items	36, plus space for the respondent to add, and rate, other coping strategies that they use.	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	Not reported	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 61 Carers' Assessment of Satisfactions Index (Nolan and Grant, 1992)

The Carers' Assessment of Satisfactions Index (CASI) was developed for use with family caregivers. It can be administered in either self-report or interview format. Carers are asked to consider 30 items describing various aspects of caregiving that may provide a source of satisfaction, for example 'caring provides a challenge'. The carer is asked to indicate whether the statement applies to them and, if so, the degree of satisfaction it provides. The authors recommend using CASI in conjunction with the Carers' Assessment of Managing Index (CAMI) and the Carers' Assessment of Difficulties Index (CADI), as the basis for a comprehensive assessment of carers' needs. Reliability and validity testing of the scale are limited. This may be explained by the fact that the measure is intended to provide a personal profile of individual carers' sources of satisfaction, rather than an interpretable overall score.

Criterion+indicators	**Data**	**Rating (where applicable)**
Constructs covered: aspects of caregiving that the carer finds satisfying, and the degree of satisfaction that these aspects provide		
A Importance to carers		
(i) Data from interviews with carers and carer representatives	Carer's relationship with the care recipient, how the carer feels about caregiving	Met
(ii) Data from conference questionnaire	Carer's relationship with the care recipient, how the carer feels about caregiving	Met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Measure designed for family carers for people with various types of disability or illness, including dementia	Met
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	CASI items were generated from an extensive postal survey of carers and a number of in-depth, semi-structured interviews (Nolan *et al.*, 1996).	Met
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	N/A	N/A

(iii) Response rates	The data from CASI discussed in Nolan *et al.* (1996) was based on 38 interviews with carers, and a postal survey in which 206 questionnaires were returned. The questionnaires were distributed by the Alzheimer's Disease Society Newsletter in Wales and Scotland, and the Crossroads Care Network in Wales. The authors point out the difficulty of calculating a response rate, as the exact number of questionnaires distributed was unknown.	-
(iv) Time to complete	Nolan *et al.* (1998) report that, on a self-administered basis, completing CASI 'need only take a matter of minutes', but that as part of an assessment interview it can take 10–20 min to complete.	-
(v) Short form available	No	-
vi) Translated measure available	CASI has been translated into over 15 languages, including most of the major European languages, Japanese, and three languages from the Indian subcontinent. Method of translation not reported. Further information can be obtained from the first author.	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	Yes	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Yes	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	0.91 (Nolan *et al.*, 1998)	Partially met
(iii) Correlation of each item with rest of scale	Not reported	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	No	-
(v) Test-retest reliability	Not reported	-
(vi) Inter-rater reliability	Not reported	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	CASI items were derived from an extensive postal survey of carers and a number of in-depth semi-structured interviews.	-
(iii) Construct validity	Not reported	-

G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	4-point Likert scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	Not reported	-
(ii) Norms presented	Nolan *et al.* (1998) provide a list of the 9 most frequently cited sources of satisfaction, plus the percentage of carers (*n*=200) who identified each as satisfying. The full results of this study are presented in Nolan *et al.* (1996).	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in this review	2	-
(vi) Existence of a meaningful overall score	No. Nolan *et al.* (1998) stress that CASI is intended primarily for use as part of an assessment of individual caregiving circumstances. The authors point out that a higher score on CASI would not necessarily mean that a carer is more satisfied than someone with a lower score. It is the profile of satisfactions on an individual basis that is the main consideration here.	-
(vii) Number of subscales	0	-
J Feasibility		

(i) Mode of administration	Face-to-face interview, self-report or card format. For the latter, the carer is given a set of cards, each with a CASI item printed on it. The respondent is asked to sort the cards into 2 piles in order to separate those statements that apply to them from those that do not. Following this, the carer is asked to look again at the cards that do apply, and to explain in what ways each item is relevant to them.	-
(ii) Number of items	30, plus space for the respondent to add, and rate, other aspects of caring that they find satisfying.	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	Not reported	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 62 Carers' Checklist (Hodgson et al., 1998)

The Carers' Checklist is a self-report measure designed for use with dementia patients and their carers. The instrument is based on three original measures: the Problem Checklist, the Support Team Assessment Schedule, and the Behaviour and Mood Disturbance Scale (see below). The authors report that acceptability of the measure to carers is good, although they do not state whether carers were involved in item generation. It is reported that the three original measures from which the instrument was devised have been shown to have good reliability and validity, but reliability and validity testing of the Carers' Checklist itself is very limited.

Criterion+indicator	**Data**	**Rating (where applicable)**
Constructs covered: levels of objective and subjective burden arising from symptoms of care-receiver and from unmet needs from services. Physical, financial, emotional, social and overall burden		
A Importance to carers		
(i) Data from interviews with carers and carer representatives	How burdened the carer feels, carer's needs, accessibility and availability of the service, effectiveness of the service, whether the carer has received adequate information about the support services available	Partially met
(ii) Data from conference questionnaire	Carer's needs, carer's satisfaction with the service, how burdened the carer feels	Met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Measure devised for carers for people with dementia.	Met

(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	Not reported. Problem Checklist (Gilleard *et al.*, 1982): not reported. Support Team Assessment Schedule (Higginson and McCarthy, 1993): no. Behaviour and Mood Disturbance Scale (Greene *et al.*, 1982): no.	Unmet/no evidence
C Acceptability		
(i) Acceptability to carers	Acceptability of the Carers' Checklist has been assessed in Hodgson *et al.* (1997). The authors report that carers found the questionnaire useful and relevant to their situation.	Met
(ii) Acceptability to other populations	N/A	N/A
(iii) Response rates	Not reported	-
(iv) Time to complete	15 min	-
(v) Short form available	No	-
vi) Translated measure available	No	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	Not reported	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	The reliability testing reported in E (ii) was carried out on a carer population. The authors state that the three scales from which the measure was devised have been proven to have good reliability and validity.	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	0.93	Partially met
(iii) Correlation of each item with rest of scale	Not reported	-

(iv) Internal consistency assessed by factor-analytic or similar techniques	Not reported. Problem Checklist (Gilleard *et al.*, 1982): carers' responses (*n*=112) to this 25-item scale were subjected to a principal-components analysis, yielding five components with an eigenvalue greater than 1.0: 'dependency', 'disturbance', 'disability', 'demand' and 'wandering'. Behaviour and Mood Disturbance Scale (Greene *et al.*, 1982): items for this scale were initially derived from the literature. The responses of 38 dementia caregivers were then subjected to a principal-factor analysis yielding three factors, the first two referring to overt patient behaviours, and the third relating to patient mood disturbance. The authors report that the factor structure of the Behaviour and Mood Disturbance Scale resembles that of similar scales, and has 'obvious face validity'. Three subscales were created by selecting those items having a loading greater than 0.40 on a particular factor.	-
(v) Test-retest reliability	Not reported. Behaviour and Mood Disturbance Scale (Greene *et al.*, 1982): 18 carers completed the scale again 3 weeks after initial assessment. The correlation coefficient for the total score was 0.84.	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported. Support Team Assessment Schedule (Higginson and McCarthy, 1993): this measure was developed in collaboration with support teams. The input of such teams resulted in significant changes to the scale during piloting.	-

(iii) Construct validity	Not reported. Problem Checklist (Gilleard *et al.*, 1982): this measure was developed in order to provide data for a study of the relationship of particular problems experienced by carers to carer strain, carer mood, and outcomes related to the care recipient. The various components of the checklist identified above (E iv) were found to have varying degrees of impact on carer strain, with 'demand' problems having the strongest relationship. Scores on 'demand' problems, but on none of the other four components, were correlated with negative mood in carers. Lower than average scores on 'demand' problems, but on none of the other four components, were strongly associated with the patient continuing to be cared for in the community at 12-month follow-up. Support Team Assessment Schedule (Higginson and McCarthy, 1993): this measure was devised for use by support teams to provide an indicator of the outcome of palliative care. The authors aimed to establish its validity for this purpose by comparing Support Team Assessment Schedule ratings as completed by support teams, patients and carers. Summed scores of patients and support teams were correlated, Rho=0.66, *P*<0.0001. Support-team ratings were usually closer to those of the patients than to those of the family member. Behaviour and Mood Disturbance Scale (Greene *et al.*, 1982): construct validity was established with correlations between the Behaviour and Mood Disturbance Scale scale and two measures of patient self-care plus an objective measure of patient cognitive functioning.	-
G Responsiveness		
(i) Used to measure change over time?	The authors report that further data are required on the sensitivity of the Carers' Checklist to change over time. None of the three original measures provide data on responsiveness.	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		

(i) Type of response categories	Part One: contains a list of dementia-related problems, plus possible problems relating to services. Column A asks how often the problem applies to the care-receiver or the service, and column B asks how stressful carers find dealing with each problem. Both columns are rated on a 3-point Likert scale. Part Two: contains five scales relating to different types of burden, e.g. emotional or financial. Carers rate how burdensome they find caregiving using visual analogue scales.	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	No	-
(ii) Norms presented	The authors give the score range, mean and standard deviation for each section of the instrument, as found in their study of 120 carers.	-
(iii) Cut off (e.g. for caseness) presented	N/A	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in this review	1	-
(vi) Existence of a meaningful overall score	No. Three scores are computed, relating to: frequency of dementia-related problems, carer burden of dementia-related problems and carer burden scales.	-
(vii) Number of subscales	3 (see above)	-
J Feasibility		
(i) Mode of administration	The instrument is recommended for use as a self-completion questionnaire, but can also form the basis of an interview or act as a focus for discussion.	-
(ii) Number of items	Part one: 30 items, for each of which the carer provides 2 ratings. Part two: 5 items.	-
(iii) Ease of explaining measure to study participants	Easy	-
(iv) Amount of researcher training required	None	-

(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Available from the Mental Health Foundation: www.mentalhealth.org.uk	-
(vii) Cost	£10	-

Table 63 Carers' and Users' Expectations of Services – Carer Version (Lelliott et al., 2003)

Carers' and Users' Expectations of Services – Carer Version (CUES-C) is a self-rated questionnaire, designed to enable carers for people with a severe mental illness to communicate their concerns and needs for help to mental health workers. Each of the 13 CUES-C items is prefaced by a descriptive, normative statement. Part A questions ask the carer to rate the extent to which his/her situation matches that of the statement, or the extent to which there is a problem in that area. Part B questions ask for a rating of the extent to which the carer would like help with the issues described by the normative statement. Part C is an open question about what particular issues, in relation to each domain, the carer would like more help with. This instrument is notable because it is unique in assessing carers' willingness to continue caring and the knowledge of the care recipient's illness, both constructs considered important by carers. Carers gave feedback on the instrument's length and clarity during its development, and its acceptability is good. The reliability and validity testing carried out were performed on a carer population.

Criterion+indicator	**Data**	**Rating (where applicable)**
Constructs covered: carers' satisfaction with various aspects of their experience of caregiving; extent to which more support from services is desired in these areas		
A Importance to carers		
(i) Data from interviews with carers and carer representatives	Carer's mental health, carer's relationship with the care recipient, carer's willingness to continue caring, whether the carer feels recognised, effect of caregiving on the carer's family life, effect of caregiving on the carer's social life, carer's needs, carer's knowledge of the care recipient's illness, accessibility and availability of the service, whether the carer has received adequate information about the support services available, the effect of caregiving on the carer's finances	Met
(ii) Data from conference questionnaire	Carer's relationship with the care recipient, effect of caregiving on the carer's family life, effect of caregiving on the carer's social life, carer's needs, carer's knowledge of the care recipient's illness, carer's satisfaction with the service, carer's depression, carer's anxiety	Met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Measure devised for carers for people with a severe mental illness	Met

(ii) Carer involvement in overall project to develop measure	Partnership research with Rethink, a charity with a large membership of carers. Project team also guided throughout by an advisory group of informal carers.	Met
(iii) Carer involvement in item generation	A literature search was conducted to identify domains relevant to carer's experiences. These were then discussed by a group of carers in a focus group, and by individual carers in five semi-structured interviews.	Partially met
C Acceptability		
(i) Acceptability to carers	75 carers gave structured feedback on limited aspects of the draft instrument during piloting. Two-thirds found the instructions 'clear', and all but one of the rest found them 'usually clear'. 90% thought the length was 'about right'. 48% found CUES-C comprehensive of the issues important to them, but 29% did not and 7% did not know. The remainder did not answer the question. Some modifications were made as a result of feedback from participants in the pilot.	Met
(ii) Acceptability to other populations	N/A	N/A
(iii) Response rates	During piloting, 283 questionnaires were sent to Rethink local group co-ordinators for distribution to carers. The 79 returns received exceeded the stated aims of the authors (Lelliott *et al.*, 1999). In the field trial of the measure with 243 carers, the item concerned with 'choice to care' was most frequently not completed by carers. The authors suggest that this is because of the sensitive and complex nature of the question. Carers were less likely to complete Part B questions than Part A.	-
(iv) Time to complete	About one-third of participants in the pilot took less than 15 min to complete CUES-C, and a further third took between 15 and 30 min.	-
(v) Short form available	No	-
vi) Translated measure available	No	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	Yes	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Yes. The reliability testing reported below was carried out on data from 243 carers.	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence

(ii) Cronbach's alpha	Not reported	Unmet/no evidence
(iii) Correlation of each item with rest of scale	Part A questions: although many correlations between individual Part A responses were significant, the authors report that coefficients (Spearman's Rho) were quite low (below 0.4), with the exception of that between 'your own life' and 'your wellbeing' (Spearman's Rho=0.53). A principal components analysis (see E iv) yielded three factors that accounted for 49% of the variance. Part B questions: the authors report that correlations between Part B responses were generally higher than within Part A, with 20 having coefficients above 0.4. A principal components analysis (see E iv) yielded two factors that accounted for 51% of the variance. Relationship between Part A and Part B questions: the authors report that correlations between Part A and Part B responses were all significant and in the expected direction. That is, lower satisfaction was associated with a greater desire for assistance. There was, however, a difference in the strength of this association for individual items. The correlations for items 1–5 were lower (range 0.19–0.59) than those for items 6–12 (range 0.59–0.72). That is, more carers expressed dissatisfaction with issues relating to their own lives than wanted help for these problems. This was not the case for items relating to direct support for their role as carers.	-

(iv) Internal consistency assessed by factor-analytic or similar techniques	Part A questions: a principal-components analysis of Part A responses yielded three popd (*sic*) factors with an eigenvalue greater than one. Factor 1 (19% of the variance) relates to the impact on carers for the 'burden of caring'. Factor 2 (18% of the variance) relates to quality of help, advice and information given by service providers and the extent to which service providers involve the carer in the planning of treatment and care. Only two items load heavily on to Factor 3 (12% of the variance). Both relate to the extent to which carers feel free either to lead their own lives or to exercise choice about their role as carer. Part B questions: a principal components analysis yielded two factors with eigenvalues greater than one. Factor 1 (30% of the variance) includes all of the items that loaded heavily on to Factors 1 and 3 in the analysis of Part A items. It relates to requesting help for the impact of caring on the carer's own life. The same items load heavily onto Factor 2 (21% of the variance) as loaded onto Factor 2 in the analysis of Part A items. These relate to quality of help, advice and information from service providers, and involvement in planning of care.	-
(v) Test-retest reliability	97 carers completed CUES-C twice, at an interval of between 2 and 14 days. For one of the Part B questions the correlation coefficient was very good (>0.80), for seven Part A and eight Part B questions coefficients were good (0.61–0.80) and for the other 10 they were moderately good (0.41–0.60).	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Lelliott *et al*. (1999) report that the three factors derived from principal components analysis of Part A responses 'had face validity'.	-
(ii) Content validity	A literature review of surveys and other instruments, consultation with carers and piloting were combined to ensure that important domains were included in the instrument. The authors wished to develop a comprehensive but brief instrument, with minimal overlap between the issues addressed by each item. The relatively low correlations between Part A items suggest that this aim has largely been achieved.	-
(iii) Construct validity	Not reported	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet

(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	Part A questions: rated using 3-point Likert scale. Part B questions: rated using 3-point Likert scale. Part C questions: open question.	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	No	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	N/A	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in this review	1	-
(vi) Existence of a meaningful overall score	No. CUES-C is intended for the purpose of enabling carers to communicate their needs for help to the mental health workers who provide a service to the care-receiver. The authors report that the properties of a ‘total’ CUES-C score, or of sub-totals based on the separate factors, have not been tested. Thus it is not known whether the instrument can be used to compare the ‘performance’ of teams or services.	-
(vii) Number of subscales	13 domains; 2 quantitative ‘scores’ relevant to each domain (see above)	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	13 domains; 3 questions in relation to each domain	-
(iii) Ease of explaining measure to study participants	Easy (see C i)	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-

(vi) Freedom of usage	Available to order from the Rethink website: www.rethink.org	-
(vii) Cost	Free	-

Table 64 Finding Meaning Through Caregiving Scale (Farran et al., 1999)

The Finding Meaning Through Caregiving Scale (FMTCS) was designed for use with caregivers of persons with Alzheimer's disease, and can be administered either as a self-rated questionnaire or by interview. The authors wished to design a measure that combined existential theoretical perspectives on the process of finding meaning and its close relationship to suffering and loss, with a more empirically based stress/adaptation approach to the effects of caregiving. As such, two studies were conducted to establish the psychometric properties of the scale. The first focused on establishing validity in relation to criterion measures that had a clear existential base. The second focused on establishing validity with stress/adaptation measures commonly used in caregiver research. Acceptability to carers was not reported.

Criterion+indicator	**Data**	**Rating (where applicable)**
Constructs covered: carers' feelings of loss and powerlessness, and their perceptions of how they find meaning through caring for a person with dementia		
A Importance to carers		
(i) Data from interviews with carers and carer representatives	How the carer feels about caregiving	Partially met
(ii) Data from conference questionnaire	How the carer feels about caregiving	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Measure devised for carers for persons with Alzheimer's disease	Met
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	In a preliminary qualitative study (Farran *et al.*, 1991), 94 family caregivers of persons with dementia answered a series of open-ended questions. Qualitative analysis of their responses identified 6 major themes. Using these 6 themes and specific wording suggested by caregivers, a 135-item FMTCS was developed. This scale was subsequently shortened to the final 43-item FMTCS (see below).	Met
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	N/A	N/A
(iii) Response rates	Not reported	-

(iv) Time to complete	Self-report, 10–15 min; face-to-face interview, 20–25 min	-
(v) Short form available	No	-
vi) Translated measure available	No	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	Not reported	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Yes	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Study 1: in a sample of 46 home-based dementia caregivers, Cronbach's alpha values were:	
	Loss/Powerlessness (LP) subscale, 0.88	Met
	Provisional Meaning (PM) subscale, 0.92	Partially met
	Ultimate Meaning (UM) subscale, 0.95	Partially met
	Entire measure, 0.91	Partially met
	Study 2: in a sample of 208 spouse caregivers, Cronbach's alpha values were:	
	Loss/Powerlessness, 0.89	Met
	Provisional Meaning, 0.88	Met
	Ultimate Meaning, 0.91	Partially met
	Entire measure, 0.91	Partially met

(iii) Correlation of each item with rest of scale	Preliminary analyses revealed that the Pearson's r bivariate correlation coefficients and subscale-to-total correlations were consistent across both studies. Expected inverse correlations were noted between LP and PM, UM, and Total Meaning (overall score), r=-0.20 to -0.81. Expected positive correlations were noted between PM and UM and Total Meaning, r=0.44–0.81. Correlations between PM and UM suggested some conceptual overlap between these 2 subscales, r=0.44 and 0.58. The authors state that this conceptual overlap is theoretically supported by existential perspectives that suggest that provisional meaning is based upon the values that one holds. If these values are spiritual/religious in nature, a close relationship between PM and UM would be expected. Individual item correlations are not reported.	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	The authors used confirmatory factor analysis because they believed that the original subscales had a strong theoretical base. The scores of 208 subjects of Study 2 were used to estimate the model of 3 underlying factors: X, Y and Z. Fit of the model was adequate with a goodness of fit index of 0.763, χ^2=1444, df=832, and a coefficient of determination of 0.998. Factor 1 (X) was estimated to be correlated 0.29 and 0.14, respectively with Factor 2 (Y) and Factor 3 (Z). The Factor 2/3 (Y–Z) correlation was estimated to be 0.55. Confirmatory factor analysis supported the construct validity of the 3 subscales. 16 of the 19 LP subscale items loaded highest on Factor 1. All 19 PM subscale items had highest loadings on Factor 2, and all 5 UM subscale items loaded highest on Factor 3. Only 3 LP items had somewhat higher loadings on Factor 2 (PM). These were not moved to Factor 2 because their weights were still positively associated with Factor 1 and their content was more consistent with Factor 1. The authors state that the conceptual overlap between the LP and PM subscales suggested by factor analysis may reflect the simultaneous nature of feelings of loss/powerlessness and the process of finding meaning.	-
(v) Test-retest reliability	The 46 subjects of Study 1 completed the FMTCS again 1 month after initial completion. Spearman correlations ($P<0.0000$) were as follows: 0.85 for LP, 0.85 for PM, 0.89 for UM and 0.80 for the entire measure.	-
(vi) Inter-rater reliability	Not reported	-

F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	The content of the measure was derived from interviews with caregivers (see B iii).	-

(iii) Construct validity	Study 1: study subjects were 46 home-based dementia caregivers. Loss/Powerlessness: convergent validity was hypothesised to be established by positive associations between LP and the subscales of the Non-Death Grief Experience Inventory-Form B (GEI-B; Sanders *et al.*, 1985), and between LP and the 'existential vacuum' and 'goal-seeking' subscales of the Life Attitude Profile-Revised (LAP-R; Reker *et al.*, 1987). These positive correlations were found, with strongest relationships between LP and selected GEI-B subscales (despair, social isolation, loss of vigor, physical symptoms, depersonalization), r=0.37–0.51, P=0.01; and the LAP-R 'existential vacuum' and 'goal-seeking' subscales, r=0.50 and 0.51, P=0.01. The authors hypothesised that discriminant validity would be established by inverse correlations between PM, UM and Total Meaning and these same GEI-B and LAP-R subscales. This hypothesis was supported by the inverse relationships between these subscales, r=-0.36 to -0.61, P=0.01. Provisional Meaning: convergent validity was hypothesised to be demonstrated by positive relationships between PM and the remaining LAP-R subscales and total scores (purpose, coherence, life control, personal meaning index, life attitude balance index) and the Ladder of Life Index (LOLI; Reker, 1992). Positive relationships were found between PM and the selected LAP-R subscales, r=0.48–0.59, P=0.01. PM had the strongest relationship with present meaning on the LOLI, r=0.45, P=0.01, suggesting that present PM may be more relevant than either past or future meaning. Discriminant validity was hypothesised to be established by inverse or insignificant relationships between LP and these same LAP-R and LOLI subscales, and lower positive or insignificant relationships between UM and these subscales. These inverse relationships were found between LP and the LAP-R subscales, r=-0.34 to -0.43, P=0.01. Discriminant validity between PM and UM was also supported, with weaker correlations between UM and 4 out of the 5 LAP-R scales (purpose, life control, personal meaning index, life attitude balance), and insignificant relationships for all LOLI subscales. Ultimate Meaning: convergent validity would be established by positive relationships between UM and Public and Private Religiosity measures (Idler and Kasl, 1991). The hypothesised positive relationships were found, r=0.35–0.77, P=0.01. UM had the strongest relationship with private religiosity, r=0.77, P=0.01. Discriminant validity would be established by inverse relationships between LP and these same subscales, and lower positive or insignificant relationships between PM and the religiosity subscales. The hypothesised but non-	-

G Responsiveness		
(i) Used to measure change over time?	Yes	Met
(ii) Evidence that change over time correlates with other measures?	Horton-Deutsch *et al*. (2002) included the FMTCS in a battery of measures designed to test the effect of the 'PLUS nursing intervention' on caregivers of elderly persons with depression. They report trends toward improvements in caregiver satisfaction (P=0.10) and higher meaning (FMTCS) (P=0.14). Caregivers also reported less depression (P=0.05) and fewer physical complaints (P=0.05) post-intervention.	-
(iii) Are the changes over time found significant to carers?	Not reported	-
H Precision		
(i) Type of response categories	5-point Likert scale (changed from 7-point Likert scale in Study 1, due to response patterns and to reduce interview time).	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	No	-
(ii) Norms presented	Means and standard deviations are given for all of the subscales in studies 1 and 2, and for the Total Meaning scores in Study 2.	-
(iii) Cut off (e.g. for caseness) presented	N/A	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in this review	1	-
(vi) Existence of a meaningful overall score	Yes. The Total Meaning score is derived by reverse scoring Loss/Powerlessness items and then summing the 3 subscales.	-
(vii) Number of subscales	3	-
J Feasibility		
(i) Mode of administration	Self-report/face-to-face interview	-
(ii) Number of items	43 (LP subscale, 19; PM subscale, 19; UM subscale, 5)	-

(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 65 Gratifications of Caregiving Questionnaire (Schwartz and Gidron, 2002)

The Gratifications of Caregiving Questionnaire is a self-report measure designed for use with parents of mentally ill adult children living at home. Parents are asked to rate their level of agreement with eight statements about things they might find rewarding or gratifying about the experience of caregiving. Schwartz and Gidron attribute the instrument to Greenberg and colleagues (1994). No reference to the questionnaire was found in this article, however, or in searches of Medline and PsycINFO. We have therefore extracted the available data from Schwartz and Gidron's article. As such, there is very little evidence available on the psychometric properties of the scale.

Criterion+indicator	**Data**	**Rating (where applicable)**
Constructs covered: the extent to which parents of mentally ill adult children perceive their caregiving experience as psychologically and emotionally rewarding		
A Importance to carers		
(i) Data from interviews with carers and carer representatives	How the carer feels about caregiving	Partially met
(ii) Data from conference questionnaire	How the carer feels about caregiving	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Measured devised for parents of mentally ill adult children living at home	Met
(ii) Carer involvement in overall project to develop measure	Not reported	Unmet/no evidence
(iii) Carer involvement in item generation	Not reported	Unmet/no evidence
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	N/A	N/A
(iii) Response rates	Not reported	-

(iv) Time to complete	Not reported	-
(v) Short form available	No	-
vi) Translated measure available	No	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	Not reported	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Yes (see E ii)	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	0.86	Met
(iii) Correlation of each item with rest of scale	Not reported	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	Not reported	-
(v) Test-retest reliability	Not reported	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	Not reported	-
G Responsiveness		
(i) Used to measure change over time?	No	-
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	5-point Likert scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		

(i) Minimal difference considered important by people the outcome measure will be used on is presented	No	-
(ii) Norms presented	In Schwartz and Gidron's study of 93 parents caring for an adult child with a mental illness at home, mean scores for each item ranged from 2.08 for 'Becoming more self-confident' (potential score range=0–4; SD=1.40) to 3.18 for 'Sense of satisfaction from fulfilling parental duties' (potential score range=0–4; SD=0.91). Mean overall score was 2.73 (potential score range=0–4; SD=0.81).	-
(iii) Cut off (e.g. for caseness) presented	N/A	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in this review	1	-
(vi) Existence of a meaningful overall score	Yes. The mean of all 8 items is calculated to provide an overall index score on gratifications.	-
(vii) Number of subscales	0	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	8	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 66 Picot Caregiver Rewards Scale (Revised) (Picot et al., 1997)

The Picot Caregiver Rewards Scale (PCRS) is designed to measure carers' perceptions of positive feelings and outcomes related to caregiving. It was developed for use with carers for older adults with dementia or any other type of illness/disability. The revised scale consists of 16 items, derived from interviews with caregivers and caregiving literature. Evidence of its acceptability to carers is not presented, however. Carers rate their agreement with statements such as 'I feel I have a closer relationship with my [elder]' and 'I feel more important'. Reliability and validity were tested in two separate studies and are good.

Criterion+indicator	**Data**	**Rating (where applicable)**
Constructs covered: perceived rewards of caregiving		

A Importance to carers		
(i) Data from interviews with carers and carer representatives	Carer's relationship with the care recipient, how the carer feels about caregiving	Met
(ii) Data from conference questionnaire	Carer's relationship with the care recipient, how the carer feels about caregiving	Met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Measure devised for adult caregivers of older adults with either dementia or any other type of illness/disability.	Met
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	Items generated from caregiver interviews and caregiving literature.	Met
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	N/A	N/A
(iii) Response rates	While all subjects involved in the two psychometric testings responded to all other items in the questionnaire, 48% of subjects in the first testing did not respond to one particular item. This item was subsequently removed from the revised questionnaire.	-
(iv) Time to complete	Not reported	-
(v) Short form available	No	-
vi) Translated measure available	No	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	Yes	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Yes	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	In the first psychometric testing, Cronbach's alpha was 0.83. In the second psychometric testing, in which the scale was reduced from 25 items to the final 16, Cronbach's alpha was 0.88.	Met

(iii) Correlation of each item with rest of scale	During piloting of the instrument, 3 items were deleted because they had correlations of above 0.80 or below 0.30. In the first psychometric testing, correlations ranged from -0.45 to 0.72, with a mean of 0.17. The authors argued that this reflects the inclusion of the extreme negative correlations and that, with one exception; the Cronbach's alpha would decrease with the deletion of each item. In the second psychometric testing, correlations of the 16 items ranged from 0.05 to 0.61, with a mean of 0.31. Cronbach's alpha was decreased by the deletion of each item. Individual item correlations not reported.	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	Although the researchers believed that the PCRS was bidimensional, it was administered and scored as a unidimensional measure during the initial piloting and first psychometric testing, due to the inability to satisfy minimum sample requirements for a factor analysis. During the second psychometric testing, the assumed 2-factor structure of Internal and External Rewards was subjected to confirmatory factor analysis. Fit of the model was assessed using the adjusted goodness of fit index, χ^2 and the root mean-square residual. The fit of the model was poor. The 2 latent concepts were highly correlated, raising the question of whether a 2-factor or a 1-factor structure would best describe the data. The fit statistics for the 1-factor structure were the same as those for the 2-factor structure, and a 1-factor structure was finally chosen for parsimony.	-
(v) Test-retest reliability	The PCRS was re-administered to a subsample of 20 caregivers 2–4 weeks after initial administration in the second psychometric testing. Correlation of the two scores resulted in a coefficient of stability of 0.75.	-
(vi) Inter-rater reliability	Not reported	-
F Validity		
(i) Face validity	Not reported	-

(ii) Content validity	The authors report that the *a priori* content validity of the scale is evident based on its manner of construction, since items were based on the literature and caregiver interviews. It was also sent to two content experts, who rated the relevance of the items, using the written theoretical definition of caregiver rewards and the scale's objectives. All items were considered at least somewhat relevant, and 13 were considered quite relevant and very relevant.	-
(iii) Construct validity	First psychometric testing: as hypothesized based on the literature, perceived caregiver rewards (as measured by the PCRS) were positively associated (r=0.22, P=0.04) with caregiving demands, as measured by the TRIMS Behavioural Problem Checklist (Niederehe, 1988). As hypothesised, perceived caregiver rewards were significantly correlated (r=0.26, P=0.02) with palliative coping, as measured by a subscale of the Jalowiec Coping Scale (Jalowiec, 1988). However, the hypothesized negative relationship between perceived caregiver rewards and perceived caregiver costs as measured by the Cost of Care Index (Kosberg, Cairl, and Keller, 1990) was not supported (r=0.07, P=0.05). Second psychometric testing: the authors hypothesised that the PCRS would be negatively correlated with measures of depression and caregiver burden. As expected, higher levels of rewards were correlated with lower levels of depression (r=-0.30 P=0.0001), as measured by the Center for Epidemiological Studies Depression (CES-D) Scale (Radloff, 1977). Higher levels of rewards were also correlated with lower levels of caregiver burden (r=-0.35, P=0.0001), as measured by the Zarit Burden Interview (Zarit and Zarit, 1986).	-
G Responsiveness		
(i) Used to measure change over time?	Yes. The passage of 1 year's time resulted in a significant decline in perceived caregiver rewards, as measured by the PCRS.	Met
(ii) Evidence that change over time correlates with other measures?	Not reported	-
(iii) Are the changes over time found significant to carers?	Not reported	-
H Precision		
(i) Type of response categories	5-point Likert scale	-
(ii) Ceiling and floor effects	Not reported	-

(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Confirmatory factor analysis of the original 25 items resulted in the deletion of 9 items. The authors suggest that some eliminated items may not have functioned well because of the potential for social desirability bias.	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	No	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	N/A	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in review	1	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	None	-
J Feasibility		
(i) Mode of administration	Scale administered by face-to-face interview in both psychometric testings	-
(ii) Number of items	16	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	The 25-item scale used in the second psychometric testing is available on request from the first author. This includes the 16 items of the revised scale.	
(vii) Cost	Free	-

Table 67 Professional Support Questionnaire (Reinhard, 1994)

The Professional Support Questionnaire (PSQ) was devised primarily for use within a study of the effects of professional support and personal control on caregiver burden (Reinhard, 1994). It is a self-report questionnaire, asking respondents to rate the extent of professional support provided in each of seven ‘instrumental’ areas and four ‘affective’ areas. Reliability and validity are reasonably good but the acceptability of the scale to carers, and its responsiveness to change over time, are not presented.

Criterion+indicator	**Data**	**Rating (where applicable)**

Constructs covered: instrumental and affective support received from mental health professionals		
A Importance to carers		
(i) Data from interviews with carers and carer representatives	Whether the service provides emotional support, whether the carer feels recognised, attitudes of health-care providers, accessibility and availability of the service, effectiveness of the service, whether the carer has received adequate information about the support services available	Met
(ii) Data from conference questionnaire	Carer's satisfaction with the service	Met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Measure devised for caregivers of mentally ill relatives	Met
(ii) Carer involvement in overall project to develop measure	Not reported	Unmet/no evidence
(iii) Carer involvement in item generation	The authors report that an 'advisory group of family caregivers' reviewed all items of the scale, providing evidence of content validity.	Partially met
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	N/A	N/A
(iii) Response rates	Not reported	-
(iv) Time to complete	Not reported	-
(v) Short form available	No	-
vi) Translated measure available	No	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	Not reported	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Yes	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	0.74 in original study, 0.94 in a study of 126 family caregivers of young adults with severe mental illness (Doornbos, 2002)	(Partially) met
(iii) Correlation of each item with rest of scale	Not reported	-

(iv) Internal consistency assessed by factor-analytic or similar techniques	No	-
(v) Test-retest reliability	Not reported	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Items were derived from the literature documenting the expressed needs of families. The authors report that review of all items by the family caregiver advisory group and psychiatric nurses provided evidence of content validity.	-
(iii) Construct validity	Concurrent validity of the total (index) score of the PSQ was supported by a correlation of 0.60 ($P<0.001$) with a separate item measuring respondents' satisfaction with the 'amount of information, practical advice and emotional support mental health professionals have given'.	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	5-point Likert scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	No	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	N/A	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in this review	1	-

(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	2: 'Instrumental support' and 'Affective support'	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	11	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 68 Revised Scale for Caregiving Self-Efficacy (Steffen et al., 2002)

The Caregiver Self-Efficacy Scales (Zeiss *et al.*, 1999) were revised and extended by Steffen and colleagues (2002) to produce the Revised Scale for Caregiving Self-Efficacy. The measure is administered by face-to-face interview. It was devised for use with carers for persons with dementia and measures the caregiver's level of perceived self-efficacy in relation to three domains: obtaining respite, responding to disruptive patient behaviours and controlling upsetting thoughts. Respondents rate their level of confidence, from 0 to 100%, that they can carry out 15 tasks, such as asking a friend or family member to do errands for them. The subscales are scored separately, with no overall score being calculated. Reliability and validity of the measure were established in two consecutive studies, and are good.

Criterion+indicator	**Data**	**Rating (where applicable)**
Constructs covered: carer's level of confidence in their ability to carry out specific tasks related to coping with caregiving		
A Importance to carers		
(i) Data from interviews with carers and carer representatives	Carer's confidence in providing care	Partially met
(ii) Data from conference questionnaire	Carer's confidence in providing care	Met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Measure devised for family caregivers of persons with dementia.	Met
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet

C Acceptability		
(i) Acceptability to carers	During pre-testing of the 14 items of the initial Caregiver Self-Efficacy Scales (Zeiss *et al.*, 1999), proposed formats were tested on 10 carers. The authors found that many items needed to be made more difficult, as caregivers were doing more already than the initial item content addressed. They also found that interview format was clearer and easier for older adults. These findings influenced further item development for the Revised Scale for Caregiving Self-Efficacy.	Partially met
(ii) Acceptability to other populations	N/A	N/A
(iii) Response rates	The authors report that all of the Study 2 participants who were asked to participate in a second testing to establish test-retest reliability (*n*=100) agreed to do so, and were re-assessed 2 weeks later.	-
(iv) Time to complete	Not reported	-
(v) Short form available	No	-
vi) Translated measure available	No	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	Yes	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Yes	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	Study 1 (169 women caring for a family member with Alzheimer's disease or other dementing disorder): 0.88 for 'Obtaining Respite' subscale; 0.84 for 'Responding to Disruptive Patient Behaviours' subscale; 0.86 for 'Controlling Upsetting Thoughts' subscale. Study 2 (145 men and women caring for a relative or close friend with Alzheimer's disease or other dementing disorder): 0.85 for 'Obtaining Respite' subscale; 0.82 for 'Responding to Disruptive Patient Behaviours' subscale'; 0.85 for 'Controlling Upsetting Thoughts' subscale.	Met

(iii) Correlation of each item with rest of scale	Individual item correlations not reported. Using Study 1 data, the 'Obtaining Respite' subscale shows low correlations with the other two subscales (*r*=0.09 and 0.15). The 'Responding to Disruptive Patient Behaviours' subscale is, however, significantly correlated with the 'Controlling Upsetting Thoughts' subscale (*r*=0.52). The authors state that these relationships are not surprising, because both the latter subscales assess an ability to regulate emotional responses, whereas the first does not.	-

(iv) Internal consistency assessed by factor-analytic or similar techniques	Study 1: in order to reduce the initial, 51-item scale (including the 14 items of the original Caregiver Self-Efficacy Scales), the authors first eliminated all items with high skewness, kurtosis or very low SDs. This resulted in a 34-item scale. Separate internal-consistency analyses were then conducted for items in each of the 3 hypothesised subscales, to eliminate any items that reduced the Cronbach's alpha for that subscale. This resulted in the removal of 1 item. An oblique factor analysis (principal axis factoring; promax) was conducted on the remaining 33 items. Items that failed to load 0.3 or higher on one factor, or that loaded greater than 0.3 on two or more factors, were eliminated; this resulted in 4 items being removed. A second factor analysis (same extraction and rotation methods) was run on the remaining 29 items, resulting in a 9-factor solution. 15 of the items loaded significantly on to one of the first 3 factors, whereas the last 6 factors each had 2 or 3 items with loadings >0.3. The 14 items that loaded on to these last 6 factors were eliminated to create a cleaner and more interpretable factor structure. The third and final factor analysis resulted in a solution consisting of 15 items loading on 3 factors, accounting for 62% of the variance. Study 2: the same oblique factor analysis as used in Study 1 was applied to the Study 2 data. This resulted in a very similar factor structure. A structural equations approach to confirmatory factor analysis was also used to examine the relative fit of the simple 3-factor solution to Study 2 data. The model that was tested specified 3 factors, with each item loading on 1 factor only. A value of less than 3.0 for a χ^2 divided by its degrees of freedom was used as an indication of adequate fit. In the current model, this index was $\chi^2(138.62)/df(87)=1.59$, which was clearly in the acceptable range of values. A Bentler Comparative Fit Index (CFI) of 0.90 or greater is viewed as supportive of an adequate fit; the CFI for Study 2 data was 0.93. Finally, the *t* values for the factor loadings were examined; these values were all significant, with coefficients ranging from 3.7 to 13.3. The mean *t* value was 9.0. In total, these results support the 3-factor solution ('Obtaining Respite', 'Controlling Upsetting Thoughts about Caregiving' and 'Responding to Disruptive Patient Behaviours') for the 15 self-efficacy items.	-

(v) Test-retest reliability	100 Study 2 participants were reassessed with the Scale for Caregiving Self-Efficacy 2 weeks after initial assessment. Results were as follows: *r*=0.76 for 'Obtaining Respite' subscale; *r*=0.70 for 'Responding to Disruptive Patient Behaviours' subscale; *r*=0.76 for 'Controlling Upsetting Thoughts' subscale. The authors report that these coefficients are in the acceptable range, especially given their view that self-efficacy is not a global, trait-like construct but is sensitive to context.	-
(vi) Inter-rater reliability	Not reported	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-

(iii) Construct validity	Study 1: as hypothesised, depression, as measured by the Beck Depression Inventory (Beck *et al.*, 1961), was significantly negatively correlated with all three subscales of the Revised Scale for Caregiving Self-Efficacy ($r=-0.16$, $P<0.05$ for 'Obtaining Respite'; $r=-0.31$, $P<0.001$ for 'Responding to Disruptive Behaviours'; $r=-0.44$, $P<0.001$ for 'Controlling Upsetting Thoughts'). As hypothesised, trait anger, as measured by the Trait Anger Expression Inventory (Spielberger, 1988), was significantly negatively related to 'Responding to Disruptive Behaviours' ($r=-0.41$, $P<0.001$), but not to the other subscales. As hypothesised, trait anxiety, as measured by the Trait Anxiety Inventory (Spielberger, 1983), was significantly negatively associated with 'Controlling Upsetting Thoughts' ($r=-0.62$, $P<0.001$). Contrary to predictions, trait anxiety also showed a strong negative relationship to 'Responding to Disruptive Behaviours' ($r=0.53$, $P<0.001$). As hypothesised, perceived available social network size, as measured by the Arizona Social Support Interview Schedule (Barrera, 1980), was significantly related to 'Obtaining Respite', but this correlation was very modest ($r=0.16$, $P<0.05$). Study 2: the Beck Depression Inventory-Short Form (Beck and Beck, 1972), the Trait Anger Expression Inventory(Spielberger, 1988), the brief version of the Multiple Affect Adjective Checklist (MAACL)-Anxiety Subscale (Zuckerman and Lubin, 1965) and the Perceived Social Support-Family Scale (Procidano and Heller, 1983) were used in Study 2 to establish the same hypothesised relationships between the Self-Efficacy subscales, depression, trait anger, trait anxiety and perceived social support. The magnitudes of the obtained correlations correspond to the hypothesised patterns. The authors also wanted to use a multimethod approach to demonstrate construct validity. They hypothesised that caregivers placed in the 'Critical/Borderline Critical' subgroup on the basis of their Expressed Emotion ratings on Five Minute Speech Samples (Magana *et al.*, 1986) would have lower scores for 'Responding to Disruptive Patient Behaviours' and 'Controlling Upsetting Thoughts' than would caregivers in the 'Low Critical' subgroup. Using independent sample t tests, caregivers in these 2 subgroups were compared. As predicted, 'Critical/Borderline Critical' caregivers were significantly lower on 'Responding to Disruptive Behaviours', $t(90)=3.61$, $P<0.05$, and on 'Controlling Upsetting Thoughts', $t(90)=2.14$, $P<0.05$, than were caregivers in the low-criticism group. As anticipated, the 2 groups did not differ on 'Obtaining Respite' scores. Responses from participants in the 2 studies were compared with each other as a final method of documenting construct validity. Participants in Study 1 were responding to an initial assessment before participating in an	-

G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	The respondent is asked to rate their level of confidence that they can do each activity, from 0 to 100%.	-
(ii) Ceiling and floor effects	The initial Caregiver Self-Efficacy Scales (Zeiss *et al.*, 1999) suffered from some ceiling effects. All of the subscales of the Revised Scale for Caregiving Self-Efficacy, however, were normally distributed and had acceptable levels of skewness. Responses covered the possible range of scores.	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	No	-
(ii) Norms presented	Steffen *et al*. (2002) give the average confidence level and SD for each item of the scale, based on the mean response for the pooled sample of Study 1 and 2 participants (*n*=314).	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in this review	1	-
(vi) Existence of a meaningful overall score	No. The authors report that use of a total score reflecting the sum of the 15 items is contrary to the view of self-efficacy as domain-specific, and can mask significant relationships between subscales and other constructs. For these reasons, they strongly advocate using scores for the 3 subscales rather than a total score.	-
(vii) Number of subscales	3	-
J Feasibility		

(i) Mode of administration	Face-to-face interview. The authors strongly believe that the scale should be administered by an interviewer because of the complexity of the task and the need to evaluate carers' understanding of the concepts involved.	-
(ii) Number of items	15	-
(iii) Ease of explaining measure to study participants	The authors report that, in their experience, some caregivers require clarification (see J i).	-
(iv) Amount of researcher training required	Not reported	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Available on request from the first author	-
(vii) Cost	Free	-

Table 69 Short Sense of Competence Questionnaire (Vernooij-Dassen et al., 1999)

The Short Sense of Competence Questionnaire (SSCQ) is a seven-item self-report measure designed to assess the problems of carers for people with dementia. The instrument was developed from the 27-item Sense of Competence Questionnaire (SCQ; Vernooij-Dassen *et al.*, 1996), and most items were initially derived from Zarit's Burden Questionnaire (Zarit *et al.*, 1980). Respondents rate their level of agreement with statements such as 'I feel that my [care recipient] behaves the way s/he does to annoy me', and 'I feel strained in my interaction with my [care recipient]'. Validity of the scale is good, although there are only limited data regarding reliability and data on responsiveness to change over time and acceptability to carers are not presented.

Criterion+indicator	**Data**	**Rating (where applicable)**
Constructs covered: carer's sense of competence in dealing with the burden of caring for a person with dementia		
A Importance to carers		
(i) Data from interviews with carers and carer representatives	Carer's relationship with the care recipient, effect of caregiving on the carers' family life, effect of caregiving on the carer's social life, carer's confidence in providing care	Met
(ii) Data from conference questionnaire	Carer's relationship with the care recipient, effect of caregiving on the carer's family life, effect of caregiving on the carer's social life, carer's confidence in providing care	Met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	Measure devised for carers for people with dementia	Met

(ii) Carer involvement in overall project to develop measure	Not reported	Unmet/No Evidence
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/No Evidence
(ii) Acceptability to other populations	N/A	N/A
(iii) Response rates	In a study of the French version of the SSCQ (Vernooij-Dassen *et al.*, 2003), the standardised index of missing values was deemed acceptable.	-
(iv) Time to complete	Less than 5 min	-
(v) Short form available	No. The SSCQ is itself a short form of the 27-item SCQ (Vernooij-Dassen *et al.*, 1996).	-
vi) Translated measure available	The original version of the SSCQ is Dutch. There is also a French version. Translation method for both versions is unknown.	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	Not reported	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Yes	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	0.76 in a study of 141 caregivers of community-living dementia patients (Vernooij-Dassen *et al.*, 1996); 0.68 in a study of 90 caregivers of dementia patients awaiting nursing-home admission (Furst, 1996); 0.76 in a study of 166 partners of non-institutionalised stroke patients (Scholte op Reimer *et al.*, 1998); 0.87 in a study of 84 daughters caring for disabled older parents (Dautzenberg *et al.*, 1996); 'exceeded the criterion of 0.70' in study of the French version (Vernooij-Dassen *et al.*, 2003).	Met in all but one study
(iii) Correlation of each item with rest of scale	All items had correlations of at least 0.2; range, 0.36–0.58.	-

(iv) Internal consistency assessed by factor-analytic or similar techniques	Factor analysis revealed three constructs measured by the original SCQ: Satisfaction with the demented person as a recipient of care; Satisfaction with one's own performance as a caregiver; Consequences of involvement in care for the personal life of the caregiver (Vernooij-Dassen, 1993; Vernooij-Dassen *et al.*, 1996). The SSCQ was designed to include the most relevant items of the original SCQ. To this end, the items with the highest factor loadings within each original SCQ domain were selected for inclusion in the SSCQ.	-
(v) Test-retest reliability	Not reported	-
(vi) Inter-rater reliability	N/A	N/A
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	The original, 27-item SCQ was based on the family-crisis model (Bengtson and Kuypers, 1985). Most issues were included in Zarit's Burden Scale (Zarit *et al.*, 1980). Content validity of the SSCQ was addressed by including questions from the three component domains of the SCQ (see E iv) and was assessed by a panel of 39 experts (Gerritsen and Van der Ende, 1994).	-
(iii) Construct validity	The authors aimed at a high Pearson correlation (>0.80) between the SCQ and the SSCQ and similar Pearson correlations between the original SCQ and predicting characteristics to those between the SSCQ and predicting characteristics. The construct validity of the SSCQ was supported by the high Pearson correlation between the two scales (0.88). Correlation coefficients and significance thresholds between the SCQ and predicting characteristics, and between the SSCQ and predicting characteristics, were similar.	-
G Responsiveness		
(i) Used to measure change over time?	No	Unmet
(ii) Evidence that change over time correlates with other measures?	N/A	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	5-point Likert scale	-

(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	No	-
(ii) Norms presented	In a sample of 141 caregivers of community-living dementia patients (Vernooij-Dassen *et al.*, 1996), results were as follows: mean score ±SD=4.44±1.96; range=0–7.	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in this review	1	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	None	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	7	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

5.3.2 Published measures developed for non-carer populations

Table 70 Giessen Symptom List, Short Version (Brahler and Scheer, 1995)

The Giessen Symptom List, Short Version (GSL-24) is a self-report measure of perceived physical health. Respondents use a 5-point Likert scale, ranging from 'not at all' to 'very', to respond to questions asking how burdened they feel by 24 symptoms relating to the four subscales of exhaustion, stomach complaints, aching limbs and heart complaints. Both the GSL-24 and the original, 57-item Giessen Symptom List (GSL) from which it was derived were developed and piloted in German. We were only able to assess data on the scale that was given in English, and as such there may be additional psychometric data available that are not reported here. The reliability and validity of the scale are good, although not tested specifically for carers.

Criterion+indicator	Data	Rating (where applicable)
Constructs covered: subjective appraisal of physical health		
A Importance to carers		
(i) Data from interviews with carers and carer representatives	Carer's physical health	Partially met
(ii) Data from conference questionnaire	Carer's physical health	Met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/no evidence
(ii) Acceptability to other populations	Not reported	Unmet/no evidence
(iii) Response rates	Grasel's (2002) study of carers' and former carers' physical health had a 94.6% response rate at 12-month follow-up. Grasel's (1995) study of somatic symptoms and caregiving strain had a response rate of 59%.	-
(iv) Time to complete	Not reported	-
(v) Short form available	No. The GSL-24 is itself a shortened version of the 57-item GSL (Brahler and Scheer, 1995).	-
vi) Translated measure available	The original version of the GSL-24 is German. Method of translation for the English version is unknown.	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	No	-
E Reliability		
(i) Split-half reliability	0.84 in the standardization sample (representative sample of the adult normal population, *n*=1601)	Met

(ii) Cronbach's alpha	0.91 in the standardization sample	Partially met
(iii) Correlation of each item with rest of scale	Not reported	-
(iv) Internal consistency assessed by factor-analytic or similar techniques	On the basis of principal components factor analysis (with varimax rotation) of the original GSL (Brahler and Scheer, 1995), and by stipulating that only discriminating items be used in scoring, the standardised scoring relates to only 4 symptom complexes (subscales) consisting of 6 items each. These 24 items form the GSL-24. Subscales are as follows: Exhaustion, Stomach complaints, Aching limbs, Heart complaints.	-
(v) Test-retest reliability	Grasel (1995, 2002) reports that the average value of the test-retest reliability for different groups of patients is 0.78 in obtained sensitivity to change for disease-related symptoms.	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Not reported	-
(iii) Construct validity	Grasel (1995) reports that specific somatic dysfunctions have been shown to give a specific picture of symptoms on the GSL-24. Psychosomatic patients also reveal a greater number and wider range of somatic symptoms on the GSL-24 than patients with specific dysfunctions.	-
G Responsiveness		
(i) Used to measure change over time?	Yes	Met
(ii) Evidence that change over time correlates with other measures?	Grasel (2002) used the GSL-24 to investigate whether change in carers' perceived physical health over a 1-year period correlates with whether or not they are still actively caring at follow-up. Change in perceived physical health over a 1-year period was measured in 2 groups of carers: those still caring for a dementia patient at home 12 months after initial assessment, and those who had ceased to provide care for at least 6 months to 1 year after initial assessment. The somatic symptoms of the active caregivers remained stable at a high level, but decreased significantly (24%) in the group of former caregivers.	-
(iii) Are the changes over time found significant to carers?	Not reported	-
H Precision		

(i) Type of response categories	5-point Likert scale	-
(ii) Ceiling and floor effects	Grasel (2002) reports that the occurrence of illnesses in active caregivers increased after 12 months to a moderate but significant degree, whereas no significant changes in the extent of somatic complaints (as measured by the GSL-24) were observed. The author notes that after several years of caregiving, carers manifest an average degree of physical complaints well above that for the normal population, rated according to age and gender. There is thus every indication that with these high values, the GSL-24 is not sufficiently sensitive to indicate any further deterioration in physical health.	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Grasel (2002) points out that self-assessment errors, for example the tendency to adjust answers to social norms and expectations, cannot be excluded when asked about health aspects. However, in longitudinal data such as those of the 2002 study, these errors have no significant influence on the results, because it can be assumed that they do not occur to varying degrees depending on the time of assessment or the caregiver groups.	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	No	-
(ii) Norms presented	Brahler and Scheer (1995) provide standardized values based on a large representative sample (*n*=1601) of adults. These values are gender-dependent and given according to age groups in decades for the subscales as well as for overall somatic symptoms. The raw data were converted into percentile norms (PN); PN>50 corresponded to an above-average manifestation of symptoms.	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in this review	1	-
(vi) Existence of a meaningful overall score	Yes (derived by summing scores on the 4 subscales)	-
(vii) Number of subscales	4	-
J Feasibility		

(i) Mode of administration	Self-report	-
(ii) Number of items	24	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 71 Older Adult Health and Mood Questionnaire (Kemp and Adams, 1995)

The Older Adult Health and Mood Questionnaire (OAHMQ) is a 22-item self-report instrument designed to assess depressive symptoms in elderly people. DSM-III-R criteria for diagnosing a major depressive episode state that the person must have cognitive, behavioural, and/or physiologic symptoms of depression, plus a pervasive change in mood. This measure therefore incorporates two classes of item: those that reflect dysphoric mood (odd items), and those that reflect behavioural, cognitive, or physiologic symptoms (even items). Its acceptability to carers is not reported. Reliability and validity are good.

Criterion+indicator	**Data**	**Rating (where applicable)**
Constructs covered: depressive symptoms in elderly people		
A Importance to carers		
(i) Data from interviews with carers and carer representatives	Carer’s mental health	Met
(ii) Data from conference questionnaire	Carer’s depression	Partially met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet
(ii) Acceptability to other populations	The input of 15 older adults at the design stage resulted in the elimination of 12 items from the measure, due to lack of clarity.	Partially met
(iii) Response rates	Not reported	-

(iv) Time to complete	Not reported	-
(v) Short form available	No	-
vi) Translated measure available	Spanish. Forward translation and blind English backward translation were used. The two forms were compared for equivalency and found to be highly similar.	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	Yes	-
E Reliability		
(i) Split-half reliability	Spearman-Brown split-half correlation for the total scale was 0.93, 0.75 for the even items (corrected for unequal length) and 0.91 for the odd items (also corrected).	(Partially) met
(ii) Cronbach's alpha	Standardised item alpha for the full scale was 0.93, 0.80 for the even items and 0.92 for the odd items.	(Partially) met
	In a study of depressive symptoms among Latino caregivers (Ranney and Aranda, 2001), Cronbach's alpha was 0.91 for the English version and 0.86 for the Spanish version.	(Partially) met
(iii) Correlation of each item with rest of scale	In an initial validation study, item analysis revealed that 4 items did not contribute sufficiently to the total score ($r<0.15$, $P>0.05$). The authors report that these 4 items were subsequently revised.	-

(iv) Internal consistency assessed by factor-analytic or similar techniques	A principal components analysis was conducted on data from 91 elderly subjects, plus 148 spousal caregivers of dementia patients. Initial statistics suggested one large eigenvalue for both samples (9.02 and 6.57, respectively), and several factors with eigenvalues between 1.67 and 1.00 (5 and 6, respectively). Examination of the rotated pattern matrices suggested that 4 factors were represented strongly by more than one item. Principal-components analyses were then conducted on the samples forcing them into 4-factor models. These 4-factor models were found to be both meaningful and similar. The factors accounted for 60% of the variance for sample 1, and 52% for sample 2. An unrestricted principal-components analysis (oblique rotation) was then performed on the pooled sample. 4 factors were extracted using the default criteria of eigenvalues exceeding 1.0. Factor 1 contains 10 items reflecting themes of loss of meaning, hope, and pleasure (anhedonia); factor 2 contains 4 items pertaining to low energy (lethargy); factor 3 contains 2 items reflecting diurnal variations in energy (diurnal); and factor 4 contains 6 items referring to changes in mood (dysphoria). The authors point out that factors 2 and 3 (lethargy and diurnal) are comprised solely of even items, and factor 4 (dysphoria) is comprised solely of odd items. This supports the construction of the scale in that the odd items were created to reflect mood-related aspects of depression, and the even items were created to reflect cognitive, behavioural and somatic aspects. Although factor 1 (anhedonia) is comprised of both odd and even items, the theme of loss of meaning, hope and pleasure is clear.	-
(v) Test-retest reliability	The measure was re-administered to 37 subjects 2 weeks after initial administration. This produced a test-retest coefficient of 0.87 (*P*<0.001), indicating that the instrument is stable over short periods of time.	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	The content of the instrument was based upon DSM-III-R criteria, and on the known differences in the display of depression that occur among older persons.	-

(iii) Construct validity	It was hypothesised that in the low score range, there would be a higher percentage of non-mood items endorsed than mood items, because many of the former reflect normal age changes in older people. As total scores increase, it was hypothesised that the percentage of odd items would be higher, because the depressive mood symptoms would become more prominent than those demonstrating effects of normal aging. This was found to be the case. At the lower end of the scale (total scores below 11), significantly more even items were endorsed (2.83 compared with 1.86; $t(34)=-3.05$, $P<0.01$). Significantly more odd items were endorsed at the high end (9.13 compared with 7.57; $t(56)=5.01$, $P<0.001$). Women scored significantly higher than did men on the scale (9.1 compared with 7.7, $P<0.05$), and divorced individuals scored higher than did married individuals (9.6 compared with 8.8, $P<0.05$), in keeping with the literature. Scores on the OAHMQ were also compared to clinical ratings to help establish criterion validity. High scores on both odd and even items were expected to reflect a major depressive episode (MDE), and low scores on both were expected to be normal. A pattern of high scores on the odd items but lower scores on the even items was expected to characterise individuals with clinically significant symptomatology. These hypotheses were supported. The mean was 14.3 for persons clinically diagnosed with MDE ($n=30$), 12.0 for persons diagnosed with depressive symptomatology ($n=49$), and 4.9 for persons judged non-depressed. Persons diagnosed with MDE scored equally high on both odd and even items ($P>0.05$). Those with clinically significant symptoms scored higher on odd items than on even items ($t=6.4$, $P<0.01$). Non-depressed persons scored equally on even and odd items ($P>0.05$). A random sample of 37 subjects from a caregiver study completed the OAHMQ, the Geriatric Depression Scale (Yesavage *et al.*, 1983) and the SCL-90-R (Derogatis, 1980). Pearson correlations of 0.70 ($P<0.001$) were obtained between the OAHMQ and both the Geriatric Depression Scale and the depression subscale of the SCL-90-R, indicating that the OAHMQ shares about 50% of the variance ($R^2=0.49$) with widely used measures of depression in older adults. The authors suggest that the OAHMQ can be seen to be similar enough to well-established scales to bolster validity, yet not so similar as to be only duplicative.	-

G Responsiveness		
(i) Used to measure change over time?	Yes. The authors report that the OAHMQ 'has shown itself sensitive to therapeutic changes in at least one study' (Kemp *et al.*, 1992).	Met
(ii) Evidence that change over time correlates with other measures?	Not reported	-
(iii) Are the changes over time found significant to carers?	Not reported	-
H Precision		
(i) Type of response categories	Binary (T or F)	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	To minimize the number of 'false-positives' based on too high a proportion of physiologic symptoms, only 4 such items were included in the scale.	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	No	-
(ii) Norms presented	In a study with a sample of 112 older adults who visited a Geriatric Evaluation Clinic, the average item was endorsed by 45% of the group. The odd (mood) items were endorsed significantly ($P<0.05$) less often than the even items. The inter-quartile ranges were as follows: 25% endorsed 4 or fewer items; 50% endorsed 9 or fewer items; 75% endorsed 16 or fewer items; 25% endorsed 17 or more items.	-
(iii) Cut off (e.g. for caseness) presented	Yes. A number of analyses were conducted to find the cut-off scores that best maximised 'hits' and minimised 'false-positives'. These were found to be 0–3=No Depression/Normal; 4–10=Clinically Significant Depressive Symptomatology; 11 or above=Probable MDE. When screening for MDE only, the sensitivity of the scale was 0.80 and the specificity was 0.87. When screening for any depressive disorder, sensitivity was 0.93 and specificity was 0.87, given a score above 11. The scale was less sensitive and specific for identifying 'clinically significant symptomatology' (0.76 and 0.54, respectively).	-
(iv) Calibrated against other life events or experiences	No	-

(v) No. of times used in this review	1	-
(vi) Existence of a meaningful overall score	Yes	-
(vii) Number of subscales	None. But total scores on odd and even items can be calculated separately, in order to determine the effect of different interventions on each class of symptoms.	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	22	-
(iii) Ease of explaining measure to study participants	The items are written in short, simple sentences and printed in large type to enhance reading and comprehension.	-
(iv) Amount of researcher training required	None	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	Freely available	-
(vii) Cost	Free	-

Table 72 Schedule for the Evaluation of Individual Quality of Life – Direct Weighting (SEIQoL-DW) (O'Boyle et al., 1996)

The Schedule for the Evaluation of Individual Quality of Life – Direct Weighting (SEIQoL-DW) comprises a three-stage interview process. The first stage asks respondents to nominate the five areas of their lives that most significantly determine their overall quality of life. The second and third stages elicit the respondents' ratings of their current status in each of these five areas, and the relative importance of each area to overall quality of life. The measure is designed to allow the respondent to describe his/her individual quality of life, without being restricted to predetermined areas. No evidence of the measure's acceptability to carers was found. Evidence on the reliability and validity of the instrument was often conflicting.

Criterion+indicator	**Data**	**Rating (where applicable)**
Constructs covered: aspects of life considered by the individual to be crucial to his/her quality of life, current level of satisfaction with each aspect, relative importance of each aspect to overall quality of life		
A Importance to carers		
(i) Data from interviews with carers and carer representatives	Carer's quality of life	Partially met
(ii) Data from conference questionnaire	Carer's quality of life, carer's satisfaction with life	Met
B Carer involvement in outcome measure development		

(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet/No Evidence
(ii) Acceptability to other populations	In a study of individuals with HIV/AIDS (Hickey *et al.*, 1996), the authors reported that the measure was found to be 'acceptable and practicable'.	Met
(iii) Response rates	Moons *et al*. (2004) found that 8% of their study population of congenital heart disease patients were unable to provide valid responses to the Dutch version of the SEIQoL-DW, due to difficulty understanding the procedure. Seven patients (5%) also declined to participate in the follow-up study. In a study investigating individual quality of life in advanced cancer patients (Waldron *et al.*, 1999), all patients successfully completed the SEIQoL-DW.	-
(iv) Time to complete	About 5 or 10 min (Hickey *et al.*, 1996); median of 15 min (Waldron *et al.*, 1999)	-
(v) Short form available	No. The SEIQoL-DW is itself a short form of the SEIQoL (O'Boyle *et al.*, 1993).	-
vi) Translated measure available	Joyce *et al*. (2003) state that, as cues and weights are evaluated by each individual in his own preferred language, no translation is needed for the respondent, and that validity of the SEIQoL-DW is therefore not culture-bound.	-
D Appropriateness		
(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	No	-
E Reliability		
(i) Split-half reliability	Not reported	N/A
(ii) Cronbach's alpha	Not reported	N/A
(iii) Correlation of each item with rest of scale	Not reported	-

(iv) Internal consistency assessed by factor-analytic or similar techniques	Moons *et al*. (2004) state that with respect to the SEIQoL-DW, 'assessment of internal consistency is irrelevant because the items nominated by the patients are not intended to be interrelated'.	-
(v) Test-retest reliability	Moons *et al*. (2004) performed a test-retest on 98 patients for whom no medical or psychosocial changes occurred during an interval of 1 year after the first application. A paired t test showed no difference in scoring between the two applications (t=0.59, P=0.56).	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-
(ii) Content validity	Hickey *et al*. (1996) argue that, because individual patients nominate the items that are important for their quality of life, the content of the SEIQoL-DW is by definition valid. In terms of the internal structure of the measure, Moons *et al*. (2004) suggest that if respondents do not understand the distinction between the actual status (step 2) and the relative importance (step 3) of the five cues, a high correlation between the two scores can be expected. They found a low to moderate correlation of r=0.26 (P<0.001), providing confirmatory evidence for internal structure.	-

(iii) Construct validity	As participants select the areas they wish to rate, this may be seen as enhancing construct validity. However Moons *et al*. (2004) have questioned the construct validity of this measure, arguing that quality of life should be considered in terms of satisfaction with life. They therefore suggest that it is necessary to re-evaluate whether the SEIQoL-DW does indeed measure quality of life. They hypothesise that the instrument assesses determinants that contribute to individual's quality of life. Moons *et al*. hypothesised that patient's scores on a Linear Analogue Scale (LAS) ranging from 0 to 100, and measuring perceived health status, would be highly correlated with actual status on the SEIQoL-DW in patients who reported health as important for their quality of life. They found a correlation of r=0.69 (P<0.0001) between the two variables. They also hypothesised that the rating of actual status on financial means would be lower in patients who were unemployed than in other patients. This relationship was supported, as the score of unemployed patients was significantly lower (t=7.9; P<0.0001). They predicted that patients who are not able to work due to disabilities would score lower on actual level of health than their non-disabled counterparts. This hypothesis was substantiated by the significant difference between the two groups (t=2.76; P=0.006). Because Moons and his colleagues assumed that the SEIQoL-DW does not measure quality of life, but rather determinants of quality of life, a low to moderate correlation between the SEIQoL-DW index score and the score on a Linear Analogue Scale of quality of life was predicted. Indeed, a correlation coefficient of r=0.48 (P<0.001) was found. The authors state that this confirms the hypothesis that the SEIQoL-DW is not an indicator of quality of life. In a study of 40 healthy subjects, Browne *et al*. (1997) found that cue weights derived from the original SEIQoL and the SEIQoL-DW differed enough to suggest that the procedures are not interchangeable. They suggest that this is because the judgment analysis used in the SEIQoL is designed to access unconscious thought, whereas the direct-weighting process of the SEIQoL-DW accesses only conscious thought.	-
G Responsiveness		
(i) Used to measure change over time?	Yes	Met

(ii) Evidence that change over time correlates with other measures?	Moons *et al*. (2004) used the data from their test-retest 1 year after initial administration to examine responsiveness of the SEIQoL-DW. They hypothesised that changes in scores on the LAS for perceived health would be highly correlated with changes in scores on actual status in patients who reported health as important for their quality of life. However, they found that these scored were not correlated (r=0.16, P=0.23). Health is only one determinant of quality of life. Moons *et al*. therefore hypothesised that a deterioration of health status would not necessarily result in a decreased SEIQoL-DW index score. 22 patients experienced complications between test and retest. This change in health status corresponded with a decrease in perceived health on the LAS, whereas quality of life on the LAS remained stable. The score on the SEIQoL-DW index increased from 74.1 to 80.8 (t=2.74, P=0.012). The authors state that this demonstrates that SEIQoL-DW index scores are independent of changes in health.	-
(iii) Are the changes over time found significant to carers?	N/A	-
H Precision		
(i) Type of response categories	Part 1: open question, asking respondent to nominate 5 areas of life ('cues') that are important for their overall quality of life. Part 2: 100-mm vertical visual analogue scale, labelled at the upper and lower extremities by the terms 'as good as could possibly be', and 'as bad as could possibly be', respectively. The respondent is asked to rate his/her actual status on each cue. Part 3: direct-weighting procedure. The respondent adjusts 5 interlocking disks, mounted on a backing disk with a 100-point scale around the circumference, to form a pie chart in which the size of each segment corresponds to the relative importance of each cue in overall quality of life.	-
(ii) Ceiling and floor effects	In Moons *et al*. (2004), floor and ceiling effects were evaluated by calculating the percentage of patients with the lowest (0) and highest possible score (100) on the SEIQoL-DW index. Low floor and ceiling scores were observed, with 0% (0/579) having the lowest possible score and 1% (6/579) having the highest possible score.	-
(iii) Precision of scales	Not reported	-

(iv) Sources of potential bias in scoring	The interviewer is asked to estimate how well the respondent understood the method, plus his/her level of fatigue/boredom.	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	No	-
(ii) Norms presented	O'Boyle *et al*. (2000) report that normative data for a UK sample have recently been collected using a similar method to that of the SEIQoL-DW, and that national population norms are available on pertinent domains of quality of life, as are data on the relative importance of these domains to people with long-standing illness (Bowling, 1996).	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	No	-
(v) No. of times used in this review	1	-
(vi) Existence of a meaningful overall score	Yes. The individual's overall quality-of-life score can be calculated by multiplying the level of satisfaction by the weight (relative importance) for each cue, summing the products, and dividing by 100 (score range 0–100). A higher score indicates better quality of life.	-
(vii) Number of subscales	3-stage interview; status on 5 'cues' rated	-
J Feasibility		
(i) Mode of administration	Face-to-face interview	-
(ii) Number of items	3 (see above)	-
(iii) Ease of explaining measure to study participants	Variable; see C(ii) and (iii)	-
(iv) Amount of researcher training required	Not reported	-
(v) Weighting used in scoring	Yes (see above)	-
(vi) Freedom of usage	Available from the Department of Psychology, Royal College of Surgeons in Ireland	-
(vii) Cost	€38.00	-

Table 73 World Health Organization Quality of Life Measure – Brief Version (The WHOQOL Group, 1996)

The World Health Organization Quality of Life Measure – Brief Version (WHOQOL-BREF) is a self-report measure of 26 items, designed for use in a variety of populations. The instrument measures respondents' levels of satisfaction within four domains related to quality of life: physical, psychological, social and environmental.

Respondents use a five-point Likert scale to answer questions such as ‘to what extent do you find your life to be meaningful?’ and ‘how well are you able to concentrate?’ The measure also includes two separate questions measuring perceptions of overall quality of life and health. The WHOQOL-BREF is a shortened version of the WHOQOL-100, which was developed in 15 international field centres simultaneously. Reliability, validity and responsiveness to change over time are good, but the acceptability of the WHOQOL-BREF to carers has not been studied.

Criterion+indicator	**Data**	**Rating (where applicable)**
Constructs covered: perceived quality of life in various domains including the physical, psychological, social and environmental		
A Importance to carers		
(i) Data from interviews with carers and carer representatives	Carer’s quality of life	Partially met
(ii) Data from conference questionnaire	Carer’s quality of life, carer’s satisfaction with life	Met
B Carer involvement in outcome measure development		
(i) Outcome measure devised for use with carers specifically	No	Unmet
(ii) Carer involvement in overall project to develop measure	No	Unmet
(iii) Carer involvement in item generation	No	Unmet
C Acceptability		
(i) Acceptability to carers	Not reported	Unmet
(ii) Acceptability to other populations	Not reported	Unmet
(iii) Response rates	Not reported	-
(iv) Time to complete	Not reported	-
(v) Short form available	No. The WHOQOL-BREF is itself a short form of the 100-item WHOQOL-100 (The WHOQOL Group, 1994).	-
vi) Translated measure available	The WHOQOL-BREF was derived from the original WHOQOL-100, which was developed in 15 international field centres simultaneously. Data from field centres in 18 countries were used to select items for inclusion in the WHOQOL-BREF. The WHOQOL-BREF was available in 19 different languages at the time of publication (The WHOQOL Group, 1996). The methodology for developing further language versions of the WHOQOL-BREF includes forward and backward translation.	-
D Appropriateness		

(i) Whether carers were included in the piloting of the measure	No	-
(ii) Whether the reliability or validity of the measure has been established on a carer population	No	-
E Reliability		
(i) Split-half reliability	Not reported	Unmet/no evidence
(ii) Cronbach's alpha	3 data-sets were used to select items for inclusion in the WHOQOL-BREF (The WHOQOL Group, 1998). The first included data from 15 international field centres, the second from 13 field centres, and the third from 5 field centres. All of these centres had been involved in the development of the WHOQOL or WHOQOL-100, and study samples included healthy and ill adults of both sexes. Based on these 3 data-sets, Cronbach's alpha for each of the 4 domains of the WHOQOL-BREF were as follows:	
	Physical health: 0.80–0.84	Met
	Psychological: 0.75–0.77	Met
	Social relationships: 0.66–0.69	Partially met
	Environment: 0.80	Met
	The authors state that Cronbach's alpha values for the 'Social relationships' domain should be read with caution, as there are only 3 items in this domain, rather than the minimum 4 generally recommended for assessing internal reliability.	Met
	Fang *et al*. (2002) report that Cronbach's alpha ranged from 0.74 to 0.85 across the domains of the Taiwanese version of the WHOQOL-BREF, as tested in a sample of 136 patients with HIV.	
(iii) Correlation of each item with rest of scale	Individual correlations were not reported. The 26-item WHOQOL-BREF initially contained the item from each facet of the WHOQOL-100 that correlated most highly with the total score, calculated as the mean of all facets. It also contained 2 items from the Overall Quality of Life and General Health facets. After examination by a panel, however, 3 items from the environmental domain were substituted because they were highly correlated with the psychological domain. A further 3 items were substituted because it was felt that other items within the facet could better explain the concept.	-

(iv) Internal consistency assessed by factor-analytic or similar techniques	Confirmatory factor analysis of the WHOQOL-100, at facet level, suggested a 4-domain solution (physical health, psychological, social relationships, environment). The 4 domains then all load on to a second-order factor, representing global quality of life. This 4-domain structure was tested using items in the WHOQOL-BREF. In the first 2 data-sets (see E ii), an acceptable-fit index (one measure of which is a Comparative Fit Index (CFI) of 0.90 or higher) was achieved when the data was applied to the 4 domain structure. In the third data-set, the initial CFI was 0.87, suggesting that alterations to the model were necessary. When 3 pairs of error variances were allowed to co-vary, and 2 items were allowed to cross-load on other domains, the CFI increased to 0.901. Multi-sample analysis was then undertaken to assess whether parameter estimates were similar across all 3 data-sets. All parameter estimates were constrained to be equal across datasets, with the exception of 2 of the 24 items, as these were known to cross-load on other domains in the case of the third data-set. In the multivariate model, the CFI reached 0.900, suggesting that the parameter estimates assessed were equivalent across all data-sets.	-
(v) Test-retest reliability	Data used to assess test-retest reliability (The WHOQOL Group, 1998) included a majority (87%) of well subjects from 4 centres participating in the field trial of the WHOQOL-100 (total *N*=391). The interval between test and retest ranged from 2 to 8 weeks. Correlations between items at time points 1 and 2 were generally high, ranging from 0.56 for item 8 (How safe do you feel in your daily life?) to 0.84 for item 12 (Have you enough money to meet your needs?). The test-retest reliabilities for domains were 0.66 for physical health, 0.72 for psychological, 0.76 for social relationships and 0.87 for environment. Fang *et al*. (2002) report that test-retest reliability ranged from 0.64 to 0.79 across domains at average 4-week retest interval, in a study of the Taiwanese version of the WHOQOL-BREF with 136 HIV patients.	-
(vi) Inter-rater reliability	N/A	-
F Validity		
(i) Face validity	Not reported	-

(ii) Content validity	Each item selected for inclusion in the WHOQOL-BREF correlated highly with one of the 24 facets of the WHOQOL-100 (see E iii). The development of the WHOQOL-100 included exploration of the quality-of-life construct in focus groups within 15 culturally diverse field centres.	-
(iii) Construct validity	Domain scores calculated using the WHOQOL-100 and the WHOQOL-BREF were very similar (The WHOQOL Group, 1998). Differences between domain scores based on either assessment ranged from 0 to 0.27 (mean difference=0.11). Correlations between domain scores based on the WHOQOL-100 and those calculated using items included in the WHOQOL-BREF ranged from 0.89 (domain 3) to 0.95 (domain 1). The authors report that the WHOQOL-100 has demonstrated criterion validity. The WHOQOL Group (1998) report that the WHOQOL-100 has previously been shown to have excellent ability in discriminating between ill and well respondents. The WHOQOL-BREF was shown to be comparable to the WHOQOL-100 in discriminating between these subject groups, with similar values and significant differences between ill and well subjects apparent in all domains. Fang *et al*. (2002) report that the scores of 136 patients with HIV and 213 healthy persons on the 4 domains of the Taiwan version of the WHOQOL-BREF correlated positively with self-evaluated health status and happiness, and negatively with number and severity of symptoms. Scores on all domains except the environmental domain discriminated between healthy subjects and HIV-infected patients.	-
G Responsiveness		
(i) Used to measure change over time?	Yes	Met

(ii) Evidence that change over time correlates with other measures?	Fung and Chien (2002) used the Chinese version of the WHOQOL-BREF to measure the effectiveness of a mutual support group for 60 family caregivers of a relative with dementia. The WHOQOL-BREF and the Neuropsychiatric Inventory Caregiver Distress Scale (NPI-D; Kaufer *et al.*, 1998) were administered before and after the 12-week intervention. Caregivers who participated in the support group showed significantly greater improvement in perceived quality of life on the psychological, social, and overall quality-of-life domains of the WHOQOL-BREF than did the control group. Caregivers in the experimental group also showed significant reduction in distress levels when managing client's symptoms post-test, as measured by the NPI-D. O'Carroll *et al*. (2000) tested the responsiveness to change over time of the WHOQOL-100 and the WHOQOL-BREF before and 3 months after liver transplantation in 50 patients, and also in 21 non-transplanted control patients. They report that domains on both measures were highly correlated, were sensitive to change following transplant, and remained stable on repeat assessment in non-transplanted control patients. However, the sensitivity to change was significantly reduced for the Social domain in the WHOQOL-BREF.	-
(iii) Are the changes over time found significant to carers?	Not reported	-
H Precision		
(i) Type of response categories	5-point Likert scale	-
(ii) Ceiling and floor effects	Not reported	-
(iii) Precision of scales	Not reported	-
(iv) Sources of potential bias in scoring	Not reported	-
I Interpretability		
(i) Minimal difference considered important by people the outcome measure will be used on is presented	No	-
(ii) Norms presented	No	-
(iii) Cut off (e.g. for caseness) presented	No	-
(iv) Calibrated against other life events or experiences	Respondents are asked whether they are currently ill, and if so, what they think is wrong with their health.	-
(v) No. of times used in this review	1	-

(vi) Existence of a meaningful overall score	No. The mean score of items within each domain is used to calculate the 4 domain scores. There are also 2 items that are examined separately, assessing an individual's overall perception of quality of life and their overall perception of their health.	-
(vii) Number of subscales	4 domains, plus 2 separate items (see above)	-
J Feasibility		
(i) Mode of administration	Self-report	-
(ii) Number of items	26	-
(iii) Ease of explaining measure to study participants	Not reported	-
(iv) Amount of researcher training required	Not reported	-
(v) Weighting used in scoring	No	-
(vi) Freedom of usage	The appropriate-language version, and permission for using it, can be obtained from The WHOQOL Group, Programme on Mental Health, WHO, CH-1211, Geneva 27, Switzerland.	-
(vii) Cost	Unknown	-

5.4 Results: unpublished measures

Carer organisations provided five unpublished measures that they had used, and one further measure was found on a carer organisation website. None had had its psychometric properties formally tested; for this reason, they are described here but not reviewed in detail. They are not included in the summary of results that follows. Names of the organisations have been removed.

5.4.1 Questionnaire 1

This two-page questionnaire to evaluate carers' groups comprises nine questions, most open-ended. It ascertains what respondents like and dislike about the carers' group, why they attend and what improvements might be made.

5.4.2 Questionnaire 2

This five-page questionnaire for carers for people with Alzheimer's disease comprises 13 questions, ascertaining demographics, whether the carer has been contacted by the care recipient's named worker and how much support he or she has received from services.

5.4.3 Questionnaire 3

This questionnaire comprises items on demographics and the usefulness of services, in particular individual and group support, help coping with transitions into care, workshops and outings.

5.4.4 Questionnaire 4

This six-item questionnaire, designed for a survey of a day hospital, comprises questions on carers' satisfaction with the support and care provided by staff, along with carers' appraisal of care recipients' quality of life and mental health.

5.4.5 Questionnaire 5

This evaluation form for day trips comprises eight questions on the carer's ratings of the day trips undertaken and their impact on the family.

5.4.6 Questionnaire 6

This 25-item questionnaire, which takes 5–10 min to complete, is a survey of carers' activities and health. It covers the amount of time spent caring, tasks performed and the carer's own physical health and psychological well-being and worries.

5.5 Summary of results: measures used three or more times

The following summary does not consider data presented for Criterion A (Importance to Carers), which are discussed in the Synthesis below (Section 7).

5.5.1 Measures developed for use with carers for people with mental health problems or dementia

Fourteen of the 49 published measures were originally developed for use with carers for people with mental health problems or dementia (Table 74). Eleven of these measured different aspects of caregiving: five assessed burden (Family Burden Interview Schedule, Pearlin Caregiver Measures for Carers for People with Alzheimer's, Screen for Caregiver Burden, Caregiver Burden Interview, Social Behaviour and Assessment Schedule); two appraisal of caregiving (Caregiver Reaction Assessment, Experience of Caregiving Inventory); one stress (Relatives' Stress Scale); one strain (Strain Scale); one the day-to-day experience of caregiving (Caregiving Hassles Scale); and one several caregiving domains (Involvement Evaluation Questionnaire). The remaining three measures developed specifically for use with carers assessed social participation and satisfaction (Social Satisfaction Scale); psychological distress (Neuropsychiatric Inventory Caregiver Distress Scale); and behavioural problems and their effects (Revised Memory and Behaviour Problems Checklist).

Table 74 Measures developed for use with carers

Construct and instrument	Number of uses[a]	Care recipient's diagnosis
Caregiving (11 instruments)		
Caregiver Burden Interview	93	Dementia
Caregiver Reaction Assessment	3	Dementia
Caregiving Hassles Scale	3	Dementia
Experience of Caregiving Inventory	6	Mental health problems

Family Burden Interview Schedule	5	Mental health problems
Involvement Evaluation Questionnaire	8	Mental health problems
Pearlin Caregiving Scales	19	Dementia
Relatives' Stress Scale	14	Dementia
Screen for Caregiver Burden	10	Dementia
Social Behaviour Assessment Schedule	14	Mental health problems
Strain Scale	9	Dementia
Social functioning and participation (1 instrument)		
Social Satisfaction Scale	3	Dementia
Carer's psychological distress (1 instrument)		
Neuropsychiatric Inventory Caregiver Distress Scale	3	Dementia
Behavioural problems and effects (1 instrument)		
Revised Memory and Behaviour Problems Checklist	40	Dementia

[a]Number of studies found in this review that used the measure, not including any by the original author.

Of the measures developed for use with carers, 10 were developed for carers for people with dementia and four for carers for people with mental health problems. The number of times our review indicated they had been used ranged from three to 93.

Papers on 11 of the measures developed specifically for carers presented no or little evidence of reliability and validity. The measures with good reliability and validity were of behavioural problems and their effects (Revised Memory Behaviour Problems Checklist); appraisal of caregiving (Caregiver Reaction Assessment and Experience of Caregiving Inventory); the day-to-day experience of caregiving (Caregiving Hassles Scale) and the Involvement Evaluation Questionnaire, which measured several caregiving domains. Only two measures had involved carers in the development at all: one had generated the items through carer interviews and taken feedback to ascertain its acceptability to them (Experience of Caregiving Inventory); the other had used carers to generate all the items but did not report its acceptability to them (Screen for Caregiver Burden). Neither of these had involved a carer in the project team. None of the remainder had used carers to generate items, and only one (the Involvement Evaluation Questionnaire) had presented evidence of its acceptability to them.

5.5.2 Measures developed for use with non-carers

The remaining 35 measures were not originally developed for carers (Table 75). Seven were for assessing depression, anxiety or both (Beck Depression Inventory, CES Depression Scale, Geriatric Hospital Depression Scale, Hospital Anxiety and Depression Scale, Hamilton Rating Scale for Depression, Self-Rating Depression Scale, State-Trait Anxiety Inventory); four for social support (Social Support Questionnaire, Inventory of Socially Supportive Behaviours, Interpersonal Support Evaluation List, Social Support Appraisals Scale); four for some aspect of the family

(Family Assessment Measure, Family Satisfaction Scale, Family Environment Scale, Family Adaptability and Cohesion Scale II); two for personality (NEO Personality Inventory, Life Orientation Test); five for psychological well-being or affect (Affect Balance Scale, General Health Questionnaire, Positive and Negative Affect Schedule, Profile of Mood States, Brief Symptom Inventory); two for health (Multilevel Assessment Instrument for physical health alone, Medical Outcomes Study SF-36 for both physical and mental health); two for stress (Perceived Stress Scale, Daily Hassles and Uplifts Scale); four for coping (Coping Responses Inventory, Mastery Scale, Ways of Coping Checklist, Jalowiec Coping Scale); one for life satisfaction (Life Satisfaction Index) and one for sleep quality (Pittsburg Sleep Quality Index). Three measures had been developed for people who were caring for care recipients who did not have mental health problems: one assessed strain for carers for people recovering from hip and heart surgery (Caregiver Strain Index) and two assessed several constructs for carers for elderly people (Caregiver Burden Scale, Family Caregiving Inventory). None of the three had been validated for carers for people with mental health problems, so they are included here as instruments developed for non-carers.

Twenty of the 35 had been developed for use with the general population and eight for use in psychiatric or medical populations; three were for use with elderly people and did not specify the setting (psychiatric or general). This includes one for use both with the general population and in psychiatric or medical settings. Three further measures were developed for carers for other groups. Information describing for which population the remaining measure was developed was not reported. The number of times each had been used with carers for people with mental health problems ranged from three to 91, with a tendency for the depression scales to have been used more often.

Table 75 Measures developed for use with non-carer populations

Concept and instrument	Number of uses[a]	Population for which developed
Depression/anxiety (7 instruments)		
Beck Depression Inventory	34	Psychiatric and other
CES Depression Scale	91	General
Geriatric Depression Scale	16	Elderly
Hospital Anxiety and Depression Scale	5	Medical outpatients
Hamilton Rating Scale for Depression	12	Psychiatric (depressed)
Self-Rating Depression Scale	9	Psychiatric (depressed)
State-Trait Anxiety Inventory	7	General
Stress (2 instruments)		
Perceived Stress Scale	14	General
Daily Hassles and Uplifts Scale	3	General
Coping (4 instruments)		
Coping Responses Inventory	13	General/psychiatric/medical
Mastery Scale	9	General

Ways of Coping Checklist	22	General
Jalowiec Coping Scale	3	Medical (hypertensive and A&E)
Personality (2 instruments)		
NEO Personality Inventory	5	General
Life Orientation Test	3	General
Social support (4 instruments)		
Social Support Questionnaire	7	General
Inventory of Socially Supportive Behaviours	5	General
Interpersonal Support Evaluation List	6	General
Social Support Appraisals Scale	5	General
Family (4 instruments)		
Family Assessment Measure	3	General
Family Satisfaction Scale	3	General
Family Environment Scale	3	Not reported
Family Adaptability and Cohesion Scale II	4	General
Health (2 instruments)		
Multilevel Assessment Instrument	28	Elderly
Medical Outcomes Study SF-36	14	Medical
Psychological well-being/symptoms (5 instruments)		
Affect Balance Scale	17	General
General Health Questionnaire	54	General
Positive and Negative Affect Schedule	6	General
Profile of Mood States	3	General
Brief Symptom Inventory	40	Psychiatric and medical
Life satisfaction (1 instrument)		
Life Satisfaction Index	13	Rural ageing
Sleep quality (1 instrument)		
Pittsburgh Sleep Quality Index	4	Patients with sleep disorders
Caregiving (3 instruments)		
Caregiver Burden Scale	9	Carers for frail elderly
Family Caregiving Inventory	4	Carers for frail elderly
Caregiver Strain Index	6	Carers for people recovering from hip and heart surgery

[a] *Number of studies found in this review that used the measure, not including any by the original author.*

None of these measures had had their validity and reliability tested specifically in relation to carers for people with mental health problems or had their acceptability to carers assessed, as far as the literature obtained for this review indicates. The degree of psychometric testing for the original population, however, was much better, overall, than for the measures developed for carers. Three of the depression scales had good or fairly good psychometric properties (Beck Depression Inventory, CES Depression Scale, Geriatric Depression Scale), while for another three only minimal psychometric data had been reported (Hamilton Rating Scale for Depression, Hospital Anxiety and Depression Rating Scale, Self-Rating Depression Scale). Evidence for the reliability and validity of the State-Trait Anxiety Inventory was weak. The two health scales had fairly good psychometric properties; as did three of the instruments assessing coping (Coping Responses Inventory, Jalowiec Coping Scale, Ways of Coping Checklist). For the other coping scale (the Mastery Scale), little evidence with which to assess its psychometric properties was reported. Four of the psychological well-being scales had fairly good psychometric properties (Positive and Negative Affect Schedule, General Health Questionnaire, Brief Symptom Inventory, Profile of Moods State), but only limited psychometric data were available to assess the Affect Balance Scale. The personality measures had good psychometric properties, as did the social support scales and family systems or environment scales. The Life Satisfaction scale had limited evidence with which to assess its psychometric properties. The Sleep Quality Index had reasonably good psychometric properties. Of the three measures developed for other carer groups, two had reasonably good psychometric properties (Family Caregiving Inventory, Caregiver Strain Index) but little evidence regarding psychometric properties is reported for the third (Caregiver Burden Scale).

5.6 Summary of results: measures from 1994 onwards used fewer than three times

The following summary does not consider data presented for Criterion A (Importance to Carers), which are discussed in the Synthesis (Section 7).

5.6.1 Measures developed for use with carers for people with mental health problems or dementia

Eleven of the 15 published measures were originally developed for use with carers for people with mental health problems or dementia (Table 76). These measured different aspects of caregiving: three assessed perceived rewards of caregiving (Picot Caregiver Rewards Scale, Gratifications of Caregiving Questionnaire, Carers' Assessment of Satisfactions Index), two carers' sense of confidence in their ability to cope with caregiving (Revised Scale for Caregiving Self-Efficacy, Short Sense of Competence Questionnaire), one burden (Carers' Checklist), one finding meaning (Finding Meaning Through Caregiving Scale), one professional support (Professional Support Questionnaire), one difficulties and associated stress (Carers' Assessment of Difficulties Index), one coping strategies (Carers' Assessment of Managing Index) and one several caregiving domains (Carers' and Users' Expectations of Services – Carer Version).

Table 76 Measures developed for use with carers

Construct and instrument	Number of uses[a]	Care recipient's diagnosis

Caregiving (11 instruments)		
Carers' Assessment of Difficulties Index	2	Dementia
Carers' Assessment of Managing Index	1	Dementia
Carers' Assessment of Satisfactions Index	2	Dementia
Carers' Checklist	1	Dementia
Carers' and Users' Expectations of Services – Carer Version	1	Mental health problems
Finding Meaning Through Caregiving Scale	1	Dementia
Gratifications of Caregiving Questionnaire	1	Mental health problems
Picot Caregiver Rewards Scale	1	Dementia
Professional Support Questionnaire	1	Mental health problems
Revised Scale for Caregiving Self-Efficacy	1	Dementia
Short Sense of Competence Questionnaire	1	Dementia

[a] *Number of studies found in this review that used the measure.*

Of the measures developed for use with carers, eight were developed for carers for people with dementia and three for carers for people with mental health problems. The number of times our review indicated they had been used ranged from one to two.

Papers on three of the measures developed specifically for carers presented no or little evidence of reliability and validity. The measures with good reliability and validity were of finding meaning (Finding Meaning Through Caregiving Scale), and carers' sense of confidence in their ability to cope with caregiving (Revised Scale for Caregiving Self-Efficacy). The Professional Support Questionnaire, Picot Caregiver Rewards Scale, Carers' Assessment of Managing Index, Short Sense of Competence Questionnaire, Carers' and Users' Expectations of Services – Carer Version and Carers' Assessment of Difficulties Index had moderate reliability and validity.

The following six measures had involved carers in item development: Carers' Assessment of Managing Index, Carers' Assessment of Satisfactions Index, Professional Support Questionnaire, Picot Caregiver Rewards Scale, Finding Meaning Through Caregiving Scale and Carers' and Users' Expectations of Services – Carer Version. Only the project team of Carers' and Users' Expectations of Services – Carer Version were guided throughout by an advisory group of carers. This measure was also developed in partnership with Rethink, a charity with a large carer membership. Only the Carers' Checklist and Carers' and Users' Expectations of Services – Carer Version provided evidence of good acceptability to carers, although carers only reported on limited aspects of acceptability.

5.6.2 Measures developed for use with non-carers

The remaining four measures were not originally developed for carers (Table 77). Two were for assessing quality of life, one for physical symptoms and one for depression in older adults. Three of the four had been developed for use with the general population, and one for screening older adults for depression in a medical setting. Each had been used once with carers for people with mental health problems.

Table 77 Measures developed for use with non-carer populations

Construct and instrument	Number of uses[a]	Care recipient's diagnosis
Quality of life (2 instruments)		
Schedule for the Evaluation of Individual Quality of Life – Direct Weighting	1	General/psychiatric/medical
World Health Organization Quality of Life Measure – Brief Version	1	General/psychiatric/medical
Physical symptoms (1 instrument)		
Giessen Symptom List	1	General
Depression (1 instrument)		
Older Adult Health and Mood Questionnaire	1	Elderly

[a] *Number of studies found in this review that used the measure.*

None of these measures had had their acceptability to carers assessed, as far as the literature obtained for this review indicates. A moderate amount of reliability and validity testing had been carried out on the Older Adult Health and Mood Questionnaire specifically in relation to carers. None of the other measures had had their reliability or validity tested in relation to a carer population. The degree of psychometric testing for the original population in the case of the World Health Organization Quality of Life Measure – Brief Version and the Older Adult Health and Mood Questionnaire were good, however. Psychometric testing of the Giessen Symptom List was moderately good. A moderate amount of psychometric data were presented for the Schedule for the Evaluation of Individual Quality of Life – Direct Weighting, although there was some conflicting evidence concerning construct validity.

Section 6 Discussion

6.1 Limitations

6.1.1 Time frame

The time frame of the study limited its scope in terms of the exhaustiveness of the literature review. In particular, we limited our search of electronic databases to two, although these were unarguably the most important sources of this kind of literature; and the search of PsycINFO was limited to the last 5 years. A limited number of instruments were not retrievable. As mentioned above, we report psychometric data predominately found in the original validation paper.

6.1.2 Response to conference questionnaire

Distributing a questionnaire to delegates at the Rethink conference was an opportunity to obtain information about important outcomes from an additional sample of carers and carer representatives. There were around 150 delegates at the Rethink conference, 28 of whom were carers or carer representatives who completed our questionnaire. It is not possible to determine a response rate to the questionnaire because the denominator is imprecise – we do not know how many carers and carer representatives attended the conference. Nevertheless, it is clear that we obtained data from only a small proportion of delegates. For this reason, we have treated it with caution, and kept it distinct from data obtained in the interviews.

6.1.3 The concept of an ‘outcome’

We had anticipated asking participants in the consultation exercise for their views of each individual measure. Not only did the time frame of the study preclude doing so, but early attempts demonstrated that it was not a helpful approach. We were keen that consultation-exercise participants should think about outcomes in as broad a way as possible. Showing them individual measures would have resulted in them focusing on the detail of the measure rather than the broad outcome it assessed. Instead, the researcher asked for participants’ views of a list of outcomes that had been assessed previously. The researcher was careful only to show participants this list after they had generated their own outcomes, although for several participants, especially those who were carers, it was only with the aid of the list that the researcher was able to convey what was meant by an outcome. Indeed, the main difficulty in the consultation exercise was communicating the concept of an outcome to participants. It was not a familiar concept and, even once explained, many carers, who comprised the majority of participants, did not consider it very relevant to them. In part, this was because carers who participated in the consultation exercise found it extremely difficult to focus on themselves and their caregiving role, perhaps because they considered it selfish to do so. It was also considered irrelevant because generally carers considered that the best approach to evaluating services was a carer-driven one, in which each carer was asked about their experience of caregiving in an open-ended way. This is illustrated by the following excerpts:

> *Interviewer: If I wanted to find out how things were going, what should I ask you?*
>
> *Participant: Doesn't matter what you ask, as long as you ask something…just say 'are you OK?' and conversation will follow on.*
>
> (JF: carer of a daughter with mental illness)
>
> *Interviewer: What would be the most effective way of you feeding back information to make sure that you could improve the service in the way that you think is key? Does that make sense?*
>
> *Participant: Well I would want to write something of…what I think is valuable. I might miss out quite a number of things. In fact, it might be quite useful to have this as a basis [refers list of outcome measures] and select from that what you want to add.*
>
> (SH: carer of a daughter with mental illness)

Despite these difficulties, we were able to obtain rich information about the aspects of caregiving that carers considered important, and with probing determine the outcomes they considered key.

6.1.4 Representativeness of participants in the consultation exercise

An explicit aim of this study was to seek the views of a broad range of carers. We were successful in recruiting participants from a variety of ethnic minority groups, although clearly not all were represented. We found it particularly difficult to recruit young carers, and gay and lesbian carers. In attempt to address this difficulty, we purposively recruited a representative from a young carers' organisation, and a representative from a gay and lesbian carers' organisation. Despite this, our study is undoubtedly limited by their under-representation. For reasons of time, it was necessary to rely on carer organisations for recruitment. Consequently, almost our entire sample comprised members of carer organisations. Carers who belong to carer organisations are likely to differ from carers who do not, but it is unclear how this difference has affected our findings.

A further bias in our sample was towards carers who prefer to contribute to research and evaluation through a semi-structured interview approach rather than a questionnaire approach. When we recruited carers, we advised them that they would be invited to participate in either a one-to-one interview or a focus group. That participants then expressed a preference for semi-structured interviews, rather than questionnaire studies, may be a consequence of self-selection. Nevetheless, although the sample may have been biased towards carers who prefer a semi-structured interview approach, all had previous experience of receiving postal questionnaires and so were able to compare the different methods.

6.2 Consultation exercise

Purposive sampling ensured that a broad range of views was represented in the consultation exercise. In particular, the views of carers from several minority groups were represented, including carers from ethnic minority groups, young carers, gay and lesbian carers and carers caring from a distance. Generally, carers and representatives of carers from minority groups tended to contribute additional information to the discussion, such as the importance of addressing language issues when designing questionnaires, rather than different views about which outcomes were important. There were no consistent differences between the views expressed by carer representatives and carers for people with mental health problems on the one hand, and carer representatives and carers for people with dementia on the other.

There was relatively little consensus among participants regarding the outcomes considered most important. As previously discussed, this may have been partly because participants did not perceive the question to be particularly relevant, or alternatively because they considered it important that a broad range of outcomes be assessed. Most participants were happy to endorse the importance of most outcomes and few thought just one outcome would be adequate. This view is likely to reflect the fact that participants perceived caregiving to affect all aspects of carers' lives, and therefore any instrument assessing their experience needs to be similarly broad.

Although participants considered numerous outcomes important, those identified can be grouped into broad categories, as follows.

- Outcomes that related to the carer himself or herself, for example the carer's health or quality of life. These are outcomes that may be considered important by all individuals, regardless of whether or not they are carers.
- Outcomes that related to the carer's experience of caregiving, such as the effect of caregiving on the carer's life or their knowledge of the care recipient's illness. Clearly, this group of outcomes would only be relevant to individuals caring for a relative or friend, although it would not be restricted to carers for people with mental health problems or dementia.
- Outcomes that related to the service that the carer received, for example its accessibility or effectiveness. This group comprises outcomes that would likely be relevant to any individual receiving a service.

In addition to commenting on which outcomes should be evaluated, participants were also asked for their views about how evaluations should be conducted. On this subject there was greater consensus. Most participants disliked the use of questionnaires, particularly those comprising only closed questions. This was primarily because participants considered that such questionnaires do not adequately capture carers' experiences. They also expressed a preference for face-to-face administration over self-completion questionnaires, a finding consistent with previous research, such as that by Weinberger and colleagues (1996). Participants suggested that less-structured methods of obtaining information such as one-to-one or group interviews are much preferred by carers. The practical difficulties in obtaining information in this way were noted by providers, however. One possibility, proposed by several participants, was the use of a combination of methods. For example, a questionnaire administered face-to-face would be preferred to one administered by post. Or a questionnaire combined with an unstructured interview would enable the carer to communicate their experiences more accurately. A key concern expressed by most participants was that carers receive prompt feedback following their involvement in any evaluation.

6.3 Literature review

As anticipated, our review of the literature found a very large number of papers using outcome measures in studies of carers, yielding a total of 64 published measures eligible for review and six unpublished measures provided by carer organisations or the internet. As noted, the comprehensiveness of the instrument assessment is determined by that of the literature search, and it is possible that instruments for which we found no psychometric data might have had such data presented in articles that we failed to find. Nevertheless, the magnitude of the search findings attests to the comprehensiveness of the search.

6.4 Instrument assessment

The majority of measures had been developed for non-carers. Among the measures developed specifically for carers, most were for carers for people with dementia rather than carers for people with mental health problems. Overall they presented less psychometric data, although there were notable exceptions. Overall, the standard of psychometric testing was highly variable, with some alarming misconceptions evident, particularly about how to determine test-retest reliability. For all measures, psychometric data were more common on reliability and validity than on responsiveness, precision or interpretability, which were rarely tested. This might be due to the fact that these types of psychometric properties have only more recently been emphasised.

The importance of the constructs measured is discussed elsewhere, but it is worth noting that the degree of involvement of carers in the development of the measures designed to be used with them was in general limited, as was the degree of ascertaining a measure’s acceptability to them. However, carer involvement in item generation was more common in the instruments developed in the last 10 years, with over one-third (6/15) having some involvement. For only one measure, the Carers’ and Users’ Expectations of Services – Carer Version, were carers involved in the project team. This is perhaps not surprising as it is only recently that consumer involvement has become more common.

Defining an outcome measure for carers was sometimes contentious. We excluded any measure that on close inspection was found to assess constructs relating to the care recipient by collecting data from the carer, and also excluded measures that collected data on carers for reasons other than to assess the carers' own outcomes, such as measures of 'expressed emotion'. We included measures of personality, however, which might not be seen as an outcome per se.

It might seem intuitive that measures originally developed for carers would be better for use with that group than those developed for other populations, but this cannot be assumed. In many cases, measures developed for other groups had been much better validated. The obvious drawback, however, is that there is rarely evidence of their validity for carers specifically, or their acceptability to them. It might also be speculated that measures developed for use with the general population would be more appropriate to carers than those developed for psychiatric or medical populations, but again this cannot be assumed. The discussion that follows will consider the findings of the psychometric review in conjunction with those of the consultation exercise, to determine the extent to which such judgements can be made.

Section 7 Synthesis

7.1 Which outcome to assess

It can be seen from the range of constructs identified during the consultation exercise, and the array of different instruments developed, that the concept of an outcome is a broad one. This indicates that researchers and service providers have been interested in the effect of providing care on many different aspects of carers' well-being. In this review, we have not recommended one particular outcome that should be assessed by researchers and service providers. Rather, the outcomes selected for evaluation should be those that researchers and/or service providers consider most relevant. For example, if a service appoints a carer-support officer with the intention of providing carers with greater practical and emotional support, then it might be appropriate for this enhancement to be assessed using instruments that measure carers' psychological well-being and caregiving activities. By contrast, if a service is keen to routinely assess the level of health needs among carers, they may wish to use instruments that measure carers' physical and mental health. However, although we suggest that the choice of which outcome to assess should remain with researchers and service providers, we do recommend that they select an outcome that is considered important by carers.

7.2 Importance of the instruments' constructs to carers

In this study, emphasis was placed on whether or not each instrument measured a construct considered important by carers. Of the 64 published instruments assessed, 54 met or partially met this criterion. The remaining 10 instruments did not.

Of the ten instruments that assessed constructs not considered important by carers in the consultation exercise, one had been devised for use with carers (Revised Memory and Behaviour Problems Checklist) but the remaining nine had not. Four assessed social support (Social Support Appraisals Scale, Social Support Questionnaire, Inventory of Socially Supportive Behaviours, Interpersonal Support Evaluation List). Two assessed personality (Life Orientation Test, NEO Personality Inventory), two assessed stress (Daily Hassles and Uplifts Scale, Perceived Stress Scale) and one assessed sleep (Pittsburgh Sleep Quality Index). Given the orientation of this study, these instruments will not be considered further. This is not to say that these constructs are unimportant to researchers; indeed, there is substantial evidence to indicate that social support, for example, is an important predictor of carers' experience (Harvey, 2000). Nor does it imply that no further exploration of the role of these and other constructs in carers' experience is required. Rather, it indicates that these outcomes are not considered important to carers, and so if an intervention aimed at supporting carers is to be evaluated, outcomes other than these should be assessed.

The constructs important to carers measured by the 54 instruments were:

- depression and/or anxiety (the Neuropsychiatric Inventory Caregiver Distress Scale, Beck Depression Inventory, CES Depression Scale, Geriatric Depression

Scale, Hamilton Rating Scale for Depression, Self-Rating Depression Scale, Hospital Anxiety and Depression Scale, State-Trait Anxiety Index, Older Adult Health and Mood Questionnaire);

- mood or psychological well-being (Positive and Negative Affect Schedule, Affect Balance Scale, Brief Symptom Inventory, General Health Questionnaire, Profile of Mood States Inventory);
- physical and mental health (Medical Outcomes Study SF-36, Multilevel Assessment Instrument);
- physical health (Giessen Symptom List);
- how the carer feels about caregiving (Caregiver Reaction Assessment, Caregiving Hassles Scale, Experience of Caregiving Inventory, Relatives' Stress Scale, Strain Scale, Caregiver Strain Index, Family Caregiving Inventory, Finding Meaning Through Caregiving Scale, Gratifications of Caregiving Questionnaire, Picot Caregiver Rewards Scale);
- satisfaction with caring (Carers' Assessment of Satisfactions Index);
- burden (Caregiver Burden Interview, Family Burden Interview Schedule, Pearlin Caregiver Measures, Screen for Caregiver Burden, Social Behaviour Assessment Schedule, Caregiver Burden Scale, Carers' Checklist);
- effect of caregiving on the carer's family (Family Adaptability and Cohesion Scale II, Family Assessment Measure, Family Environment Scale, Family Satisfaction Scale);
- effect of caregiving on the carer's social life (Social Satisfaction Scale);
- confidence in providing care (Revised Scale for Caregiving Self-Efficacy, Short Sense of Competence Questionnaire);
- professional support the carer receives (Professional Support Questionnaire);
- coping (Coping Responses Inventory, Jalowiec Coping Scale, Mastery Scale, Ways of Coping Checklist, Carers' Assessment of Managing Index);
- satisfaction with life (Life Satisfaction Index);
- quality of life (Schedule for the Evaluation of Individual Quality of Life, WHO Quality of Life Measure)

While a few of the instruments so far described assessed more than one construct important to carers, typically they were restricted to one domain, for example two or three aspects of caregiving.

By contrast, three instruments were notably broad-ranging and incorporated several constructs important to carers.

- The Involvement Evaluation Questionnaire examined the carer's mental health, carer's physical health, effect of caregiving on the carer's family life, effect of caregiving on the carer's social life, amount of caregiving activities the carer undertakes, how burdened the carer feels, how the carer feels about caregiving and the effect of caregiving on the carer's finances.
- The Carers' Assessment of Difficulties Index measured the carer's physical health, effect of caregiving on the carer's family life, effect of caregiving on the carer's social and daily life, the type of caregiving activities the carer undertakes, how the carer feels about caregiving, carer's relationship with the care recipient, how burdened the carer feels, carer's confidence in providing

care, effect of caregiving on the carer's finances, attitudes of health-care providers, and effectiveness of and satisfaction with the service.

- The Carers' and Users' Expectations of Services – Carer Version measured the carer's depression and/or anxiety, effect of caregiving on the carer's family life, effect of caregiving on the carer's social life, carer's relationship with the care recipient, effect of caregiving on the carer's finances, carer's willingness to continue caring, whether the carer feels recognised, carer's needs, carer's knowledge of the care recipient's illness, accessibility and availability of the service, whether the carer has received adequate information about the support services available and the carer's satisfaction with the service.

To synthesise the two threads of the study, we have no longer distinguished instruments that met Criterion A (measure a construct important to carers) from instruments that partially met Criterion A. Likewise we have no longer distinguished instruments that met outcomes generated in the interviews (Ai) and instruments that met outcomes determined important from the conference questionnaire (Aii). This is because all measured a construct of at least some importance to carers. Similarly, instruments used three or more times are no longer distinguished from those used less often.

7.3 Other carer-focused criteria (involvement, acceptability and appropriateness)

Of the 54 instruments that measured a construct important to carers, 25 were devised specifically for use with carers and so could have possibly involved carers in their development. However, as noted in the previous section, the degree of involvement of carers in the development of the instruments devised for use with them was in general limited.

Only one measure, the Carers' and Users' Expectations of Services – Carer Version, involved carers in the project team (Criterion Bii). For nine instruments, carers were involved in the generation of items for the instrument (Caregiver Strain Index, Experience of Caregiving Inventory, Screen for Caregiver Burden, Carers' Assessment of Managing Index, Carers' Assessment of Satisfactions Index, Carers' and Users' Expectations of Services – Carer Version, Finding Meaning Through Caregiving Scale, Picot Caregiver Rewards Scale, Professional Support Questionnaire; Criterion Biii). For only four instruments devised specifically for use with carers did the authors assess its acceptability to them (Criterion Ci), namely the Experience of Caregiving Inventory, the Involvement Evaluation Questionnaire, the Carers' and Users' Expectations of Services – Carer Version and the Revised Scale for Caregiving Self-Efficacy. The remaining 12 instruments developed for use with carers did not involve carers in any aspect of their generation.

The appropriateness of the instrument for a carer population (Criteria Di and Dii) had been established for 14 of the 25 instruments developed for use with carers (Caregiver Reaction Assessment, Caregiving Hassles Scale, Experience of Caregiving Inventory, Involvement Evaluation Questionnaire, Pearlin Caregiving Measures for Carers for People with Alzheimer's Disease, Relatives' Stress Scale, Revised Memory and Behaviour Problems Checklist, Social Satisfaction Scale, Carers' Assessment of Difficulties Index, Carers' Assessment of Managing Index, Carers' Assessment of Satisfactions Index, Carers' and Users' Expectations of Services – Carer Version, Picot Caregiver Rewards Scale, Revised Scale for

Caregiving Self-Efficacy). For a further nine, some degree of appropriateness had been established, with either the instrument having been piloted on carers or the reliability and validity having been established on a carer population, but not both (Caregiver Burden Interview, Screen for Caregiver Burden, Strain Scale, Carers' Checklist, Finding Meaning Through Caregiving Scale, Gratifications of Caregiving Questionnaire, Neuropsychiatric Inventory Caregiver Distress Scale, Professional Support Questionnaire, Short Sense of Competence Questionnaire). Appropriateness of the Social Behaviour and Assessment Schedule had not been established and data were not available to determine the appropriateness of the Family Burden Interview Schedule. For two further instruments, which had not originally been devised for use with carers, the Ways of Coping Checklist and the Older Adult Health and Mood Questionnaire, appropriateness for carers had been established.

Understandably, carers had not been involved in either the project team or the generation of items for the remaining 29 instruments that assessed constructs important to carers but were not devised for carers originally. Nor had their acceptability or appropriateness to carers been established, although if they continue to be used with a carer population this may be an important next step. That only three instruments (the CES Depression Scale, the Older Adult Health and Mood Questionnaire and the Schedule for the Evaluation of Individual Quality of Life) had evidence indicating acceptability to the population for whom they were designed demonstrates how acceptability has been generally overlooked by those who develop and use instruments.

The remaining psychometric properties for the 54 instruments measuring constructs important to carers were variable, both in terms of whether or not they had been assessed, and in terms of their quality.

7.4 Reliability

The reliability of 17 of the 54 instruments was good (Caregiver Reaction Assessment, Caregiving Hassles, Involvement Evaluation Questionnaire, CES Depression Scale, General Health Questionnaire, Beck Depression Inventory, Geriatric Depression Scale, Family Adaptability and Cohesion Scale II, Family Assessment Measure, Profile of Mood States Inventory, Carers' Assessment of Difficulties Index, Finding Meaning Through Caregiving Scale, Giessen Symptom List, Older Adult Health and Mood Questionnaire, Picot Caregiver Rewards Scale, Revised Scale for Caregiving Self-Efficacy, World Health Organization Quality of Life Measure). The reliability for a further 17 was moderate (Experience of Caregiving Inventory, Screen for Caregiver Burden, Social Satisfaction Scale, Brief Symptom Inventory, Caregiver Strain Index, Coping Responses Inventory, Family Caregiving Inventory, Family Environment Scale, Family Satisfaction Scale, Jalowiec Coping Scale, Life Satisfaction Index, Medical Outcomes Study SF-36, Multilevel Assessment Instrument, Positive and Negative Affect Schedule, Ways of Coping Checklist, Carers' Assessment of Managing Index, Short Sense of Competence Questionnaire). Little, or no, evidence was available to assess the reliability of the remaining 20 instruments.

7.5 Validity

Validity was demonstrated to be good for 21 of the 54 instruments (Caregiver Reaction Assessment, Caregiving Hassles, Experience of Caregiving Inventory, Involvement Evaluation Questionnaire, Neuropsychiatric Inventory Caregiver Distress Scale, Relatives' Stress Scale, Beck Depression Inventory, CES Depression Scale, Coping Responses Inventory, Family Adaptability and Cohesion Scale II, Family Assessment Measure, Family Environment Scale, General Health Questionnaire, Hospital Anxiety and Depression Scale, Jalowiec Coping Scale, Profile of Mood States Inventory, Finding Meaning Through Caregiving Scale, Older Adult Health and Mood Questionnaire, Picot Caregiver Rewards Scale, Revised Scale for Caregiving Self-Efficacy, World Health Organization Quality of Life Measure). For a further 18 instruments it was moderate (Screen for Caregiver Burden, Affect Balance Scale, Brief Symptom Inventory, Caregiver Strain Index, Geriatric Depression Scale, Life Satisfaction Index, Mastery Scale, Medical Outcomes Study SF-36, Multilevel Assessment Instrument, Positive and Negative Affect Schedule, Self-Rating Depression Scale, State-Trait Anxiety Inventory, Ways of Coping Checklist, Carers' Assessment of Difficulties Index, Carers' and Users' Expectations of Services – Carer Version, Giessen Symptom List, Professional Support Questionnaire, Short Sense of Competence Questionnaire). For a further 13 instruments, no evidence was available to assess validity. For one of the remaining instrument, validity was demonstrated as poor (Family Caregiving Inventory), and for the other, the Schedule for the Evaluation of Individual Quality of Life, the study in which most validity evidence was presented concluded that the instrument did not in fact measure quality of life.

There were 13 instruments for which both reliability and validity could be considered good. Two of these assessed depression (CES Depression Scale, Beck Depression Inventory), three assessed psychological symptoms (General Health Questionnaire, Profile of Mood States Inventory, Older Adult Health and Mood Questionnaire), four assessed aspects of caregiving (Caregiver Reaction Assessment, Caregiving Hassles, Finding Meaning Through Caregiving Scale, Revised Scale for Caregiving Self-Efficacy), two assessed aspects of family functioning (Family Adaptability and Cohesion Scale II, Family Assessment Measure), and one assessed quality of life (World Health Organization Quality of Life Measure). The thirteenth, the Involvement Evaluation Questionnaire, assessed several domains.

7.6 Responsiveness

Data to determine responsiveness were available for only 14 instruments. An instrument's ability to detect change over time reliably is of vital importance if it is to be used to evaluate services, and this is an area where further work is clearly needed. Where responsiveness could be assessed, it was generally good (Caregiver Reaction Assessment, Involvement Evaluation Questionnaire, Screen for Caregiver Burden, Beck Depression Inventory, Family Assessment Measure, Positive and Negative Affect Schedule, Profile of Mood States Inventory, Self-Rating Depression Scale, Finding Meaning Through Caregiving Scale, Giessen Symptom List, Older Adult Health and Mood Questionnaire, Picot Caregiver Rewards Scale, Schedule for the Evaluation of Individual Quality of Life, World Health Organization Quality of Life Measure).

7.7 Precision

As with responsiveness, evidence regarding instruments' precision was rarely available. Eleven instruments examined potential sources of bias (Social Behaviour Assessment Schedule, Affect Balance Scale, Beck Depression Inventory, CES Depression Scale, Family Adaptability and Cohesion Scale II, Interpersonal Support Evaluation List, NEO Personality Inventory, Social Support Questionnaire, Older Adult Health and Mood Questionnaire, Picot Caregiver Rewards Scale, Schedule for the Evaluation of Individual Quality of Life); whereas four (Medical Outcomes Study SF-36, Giessen Symptom List, Revised Scale for Caregiving Self-Efficacy, Schedule for the Evaluation of Individual Quality of Life) examined ceiling and floor effects in addition to potential sources of bias. Clearly, this is another area where further work is required.

7.8 Interpretability

Interpretability was judged as good if a meaningful overall score could be calculated and either norms or cut-offs were reported. Twenty instruments met these criteria (Caregiving Hassles, Neuropsychiatric Inventory, Beck Depression Inventory, CES Depression Scale, Family Adaptability and Cohesion Scale II, Family Assessment Measure, Family Satisfaction Scale, General Health Questionnaire, Geriatric Depression Scale, Hamilton Rating Scale for Depression, Hospital Anxiety and Depression Scale, Profile of Mood States, Screen for Caregiver Burden, State-Trait Anxiety Inventory, Finding Meaning Through Caregiving Scale, Giessen Symptom List, Gratifications of Caregiving Questionnaire, Older Adult Health and Mood Questionnaire, Schedule for the Evaluation of Individual Quality of Life – Direct Weighting, Short Sense of Competence Questionnaire). A further 20 met some of the criteria (Caregiver Reaction Assessment, Caregiver Burden Interview, Caregiver Burden Scale, Experience of Caregiving Inventory, Involvement Evaluation Questionnaire, Relatives' Stress Scale, Strain Scale, Affect Balance Scale, Brief Symptom Inventory, Caregiver Strain Index, Jalowiec Coping Scale, Life Satisfaction Index, Mastery Scale, Medical Outcomes Study SF-36, Positive and Negative Affect Schedule, Self-Rating Depression Scale, Social Satisfaction Scale, Carers' Assessment of Difficulties Index, Picot Caregiver Rewards Scale, Professional Support Questionnaire). The interpretability of the 14 remaining was judged as poor, or information was unavailable.

7.9 Feasibility

The majority of instruments were self-report. Only nine required a structured interview (Neuropsychiatric Inventory, Social Behaviour Assessment Schedule, Family Caregiving Inventory, Hamilton Rating Scale for Depression, Multilevel Assessment Instrument, Pittsburgh Sleep Quality Index, Picot Caregiver Rewards Scale, Revised Scale for Caregiving Self-Efficacy, Schedule for the Evaluation of Individual Quality of Life). A further seven could be administered as either a structured interview or a self-report rating scale (Coping Responses Inventory, NEO Personality Inventory, Carers' Assessment of Difficulties Index, Carers' Assessment of Managing Index, Carers' Assessment of Satisfactions Index, Carers' Checklist, Finding Meaning Through Caregiving Scale). In addition, the Carers' Assessment of Difficulties Index, Carers' Assessment of Managing Index and Carers' Assessment of

Satisfactions Index could be administered in 'card format', which is the only example of a semi-structured measure that would yield essentially qualitative data.

The two criteria considered key when judging feasibility were the length of time the instrument took to complete and the cost of using it. The first of these was judged as particularly important because of evidence obtained in the consultation exercise regarding carers' willingness to complete lengthy assessments. The feasibility of using four instruments, the Family Caregiving Inventory, Social Behaviour Assessment Schedule, Profile of Mood States and the Multilevel Assessment Instrument, was compromised by their completion time, 1–2 h. Moreover, the latter appeared to be unobtainable. The feasibility of using nine further instruments was compromised by their cost (Coping Responses Inventory, Family Assessment Measure, Family Environment Scale, General Health Questionnaire, Hospital Anxiety and Depression Scale, Profile of Mood States, State-Trait Anxiety Inventory, Carers' Checklist, Schedule for the Evaluation of Individual Quality of Life). Otherwise the remaining 41 were relatively brief and free to use.

7.10 Range of constructs assessed

A key finding of the consultation exercise was that a broad range of outcomes should be assessed, because caregiving has an impact on several aspects of carers' lives. The findings from the instrument-assessment exercise have identified several reliable and valid instruments, but most measured only one construct. By contrast, three instruments identified in the review attempt to assess a broad range of constructs all of which are important to carers. These instruments are the Involvement Evaluation Questionnaire, the Carers' Assessment of Difficulties Index and the Carers' and Users Expectations of Services. The first of these, the Involvement Evaluation Questionnaire, comprises a core module that assesses the consequences of caregiving, which is typically used in conjunction with other modules assessing socio-demographic status, financial consequences of caregiving, use of professional help and psychological distress via the 12-item General Health Questionnaire. It has good reliability and validity and some evidence to demonstrate responsiveness.

The second instrument assessing multiple constructs, the Carers' Assessment of Difficulties Index, similarly assesses aspects of caregiving, the financial consequences of caregiving, attitudes of health-care providers and effectiveness of and satisfaction with the service, along with the carer's physical health. Its authors recommend its use in conjunction with the Carers' Assessment of Management Index and the Carers' Assessment of Satisfactions Index, which would broaden it further. Evidence for the instrument's reliability is good, although validity testing is limited. This may be explained by the fact that the measure is intended to provide a personal profile of individual carers' difficulties, rather than an interpretable overall score.

The Carers' and Users' Expectations of Services – Carer Version is the third instrument to measure multiple constructs. As with the previous two instruments, it assesses aspects of caregiving, the financial consequences of caregiving, aspects of the service carers have received and carers' mental health. Uniquely, it also assesses carers' willingness to continue caring and whether the carer feels recognised. Evidence regarding the instrument's reliability and validity is limited, however.

Given the further work required to establish the full psychometric properties of these instruments, it remains to be seen whether one instrument can assess several constructs reliably and validly, or whether it is better to use a battery that comprises several instruments, each measuring a single construct well. The advantages of using a battery of reliable and valid instruments may, however, be outweighed by the disadvantage that, according to our consultation exercise findings, carers object to lengthy assessments.

7.11 Outcomes important to carers for which no instruments have been developed

For two outcomes important to carers, namely their morale and their receipt of emotional support, it appears that no measure has yet been developed. These outcomes may not require the development of new instruments, but could be assessed by incorporating a few questions into an instrument currently in use.

7.12 Carers from minority groups

As previously mentioned, no outcomes specific to carers from minority groups were identified in the consultation exercise. It was emphasised, however, that conventional methods of assessment are not appropriate for young carers, for whom alternative methods need to be developed.

7.13 Methods of assessment

It is clear from the consultation exercise that carers prefer face-to-face assessments rather than postal questionnaires. The most-preferred method of assessment is by semi-structured interview. Carers' preferences for semi-structured interviews constrast with the needs of researchers and service providers to obtain quantitative data. If questionnaires are employed then, in addition to closed questions, space should be incorporated for carers' comments. The language used in questionnaires should be acceptable to carers from all ethnic minority groups. This means translated versions may be required, although generating reliable and valid translations of each instrument in numerous languages may be unfeasible. Questionnaires should be easy for carers to understand and complete.

7.14 Summary

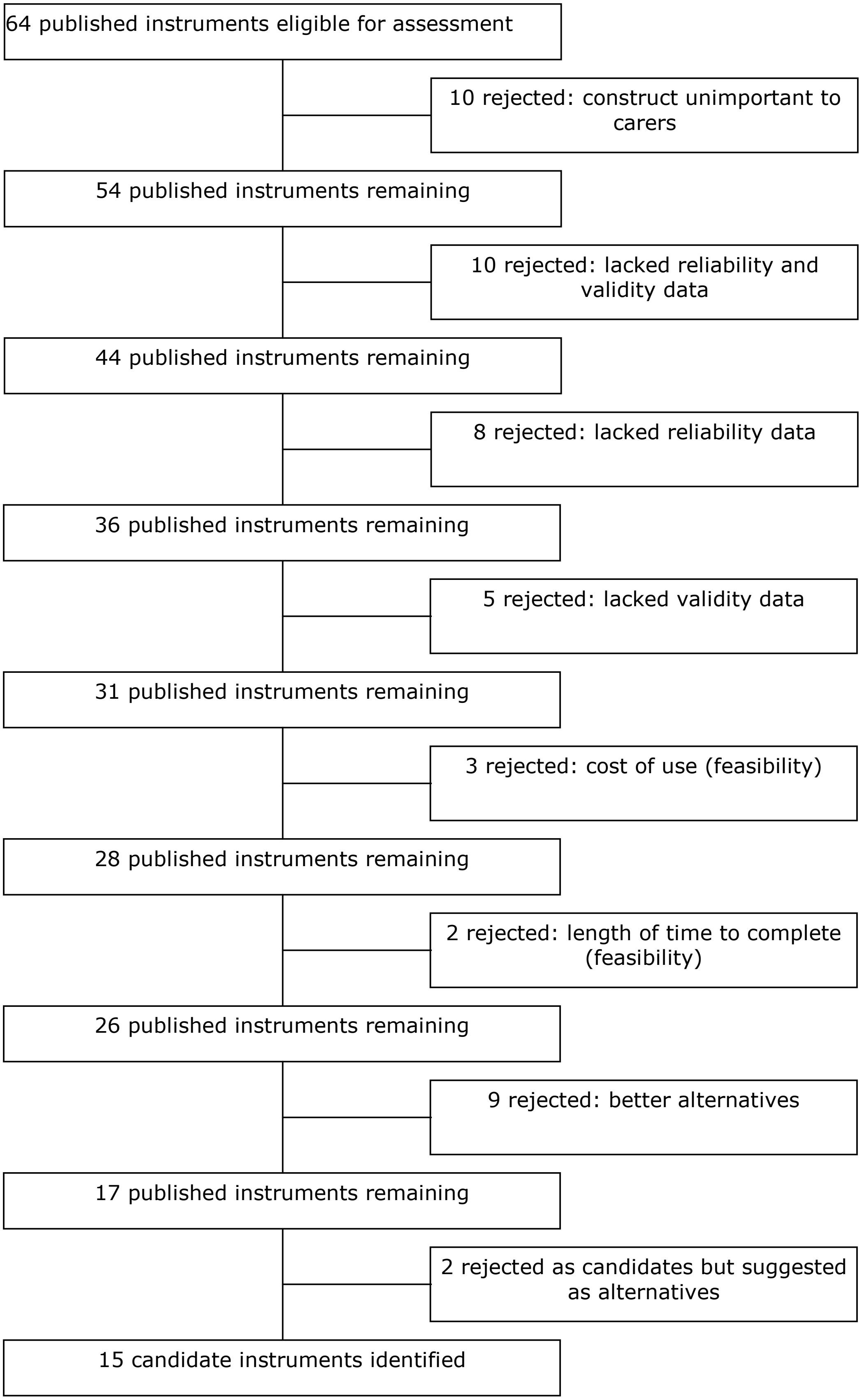
When selecting candidate measures for recommendation, our primary focus was on whether the outcome assessed by the instrument was considered important by carers (Criterion A), and whether the measure was reliable (Criterion E), valid (Criterion F), interpretable (Criterion I) and feasible to use (Criterion J). Benchmarks were applied when assessing the importance to carers of the construct assessed and some aspects of reliability (see Table 9). Benchmarks were not appropriate when assessing validity, interpretability and feasibility. Instead researchers made a judgement about the quality of each instrument in relation to these criteria. Instruments had to assess outcomes considered important by carers and be reliable, valid, interpretable and feasible to be included as candidate measures. Further criteria were also used to assess instruments, namely carers' involvement in the development of the instrument (Criterion B), the acceptability of and appropriateness for carers of the instrument (Criteria C and D), and the instrument's responsiveness (Criterion G) and precision (Criterion H). Lack of these data did not preclude instruments from our candidate list but instead is noted as an area requiring further research. This approach was adopted because so few instruments had data pertaining to responsiveness and precision, and instruments developed for the general population were not expected to have involved carers in their development or assessed their appropriateness for and acceptability to carers.

The process of identifying candidate instruments is shown in Figure 2. Our assessment of the psychometric properties of the 64 instruments indicates that 26 assess outcomes considered important by carers, and are reliable, valid, interpretable and feasible to use.

- Four assess depression, the first two of which have the added advantage of being responsive: Older Adult Health and Mood Questionnaire, Beck Depression Inventory, CES Depression Scale and Geriatric Depression Scale.
- Two assess psychological symptoms: Brief Symptom Inventory and General Health Questionnaire, although the latter is not free to use.
- One assesses mood and also has evidence of responsiveness: Positive and Negative Affect Schedule.
- One assesses physical health and also has evidence of responsiveness: Geissen Symptom List Short Version.
- One assesses physical and mental health: Medical Outcomes Study SF-36*.*
- One assesses the family environment: Family Adaptability and Cohesion Scale II.
- Two assess coping: Jalowiec Coping Scale, Ways of Coping Checklist.
- Twelve assess aspects of caregiving:
  - ? three of these assess how the carer feels about caregiving (Experience of Caregiving Inventory, Caregivers' Reaction Assessment and Caregiver Strain Index); the first has evidence of responsiveness;
  - ? one assesses burden and has evidence of responsiveness (Screen for Caregiver Burden);
  - ? two assess the positive aspects of caregiving, both of which have evidence of responsiveness (Finding Meaning Through Caregiving Scale, Picot Caregiver Rewards Scale);
  - ? two assess carers' sense of competence in caregiving (Short Sense of Competence Questionnaire, Revised Scale for Caregiving Self-Efficacy).
  - ? one assesses the day-to-day experience of providing care (Caregiving Hassles Scale);
  - ? one assesses carers' perception of the professional support they have received (Professional Support Questionnaire);
  - ? two assess several caregiving domains (Involvement Evaluation Questionnaire, Carers' Assessment of Difficulties Index); the latter has evidence of responsiveness.
- One assesses quality of life and has evidence of responsiveness (World Health Organisation Quality of Life Measure).
- One assesses satisfaction with life (Life Satisfaction Index), although it should be noted that most consultation-exercise participants did not consider this a very important construct.

All but four of these instruments are self-report rating scales. The Picot Caregiver Rewards Scale and the Revised Scale for Caregiving Self-Efficacy both require a structured interview. The Finding Meaning Through Caregiving Scale and the Carers' Assessment of Difficulties Index can be administered as either a structured interview or a self-report questionnaire. Nine have been translated into numerous other languages (Carers' Assessment of Difficulites Index, Beck Depression Inventory, CES Depression Scale, Geriatric Depression Scale, Brief Symptom Inventory, General Health Questionnaire, Medical Outcomes Study SF-36, Positive and Negative Affect Schedule, World Health Organsiation Quality of Life Measure); eight have been translated into between one and four other languages (Involvement Evaluation Questionnaire, Family Adaptability and Cohesion Scale II, Jalowiec Coping Scale, Caregiver Strain Index, Screen for Caregiver Burden, Giessen Symptom List Short Version, Older Adult Health and Mood Questionnaire, Short Sense of Competence Questionnaire), whereas those remaining appear to be available only in English. Two outcomes important to carers, namely their morale and their receipt of emotional support, are not covered by any of the instruments.

Figure 2 Process of selection of candidate instruments



Section 8 Review of findings

8.1 Approaches to assessment

Having consulted a diverse range of carers, representatives of carer organisations and service providers, and reviewed the plethora of instruments used to assess carer outcomes, we can make the following recommendations.

To assess the effectiveness of services aimed at improving carers' experiences the following needs to be done.

- Information should preferably be gathered using semi-structured interviews.
- If semi-structured interviews are not feasible, assessments should at least be conducted face-to-face. If questionnaires are used, it is preferable that they are given to carers in a face-to-face setting, rather than by post.
- If questionnaires are used, they should not only comprise closed questions but should also incorporate space for carers to add additional comments. Questionnaires should be easy for carers to understand and complete.
- Carers invited to evaluate services should be given prompt feedback regarding findings. They should also be informed about the outcome of the evaluation.
- Conventional approaches to assessment should not be used for young carers. Alternatives could include group work and games.
- Those evaluating services should assess a broad range of carer outcomes. We therefore recommend the Involvement Evaluation Questionnaire, which measures a broad range of constructs, is acceptable to and appropriate for carers, and has good psychometric properties. However, it is acceptable to carers that, where a service intends to achieve change, the outcome measures used focus on assessing these potential changes.

Findings from our consultation exercise, and from previous research, demonstrate that caregiving affects many aspects of carers' lives and this is reflected in the range of constructs assessed by the instruments we have recommended. It would, however, be impractical for researchers or service providers to attempt to assess all of these aspects in any one study. Instead, they will need to identify a handful of domains that are pertinent to the purpose of their evaluation.

8.2 Candidate instruments for use or further development

Of the 64 published instruments assessed, 26 assessed an outcome considered important by carers, and were reliable, valid, interpretable and feasible to use. We acknowledge that these instruments are, contrary to carers' preferences, mostly self-report questionnaires and are not designed to be administered in a semi-structured way. Four are designed to be conducted as structured or semi-structured interviews, rather than as self-report questionnaires. Two of these assess the positive aspects of caregiving (Finding Meaning Through Caregiving Scale, Picot Caregiver Rewards Scale), one carers' sense of competence (Revised Scale for

Caregiving Self-Efficacy) and the fourth how the carer feels about caregiving (Carers' Assessment of Difficulties Index). We have persisted with recommending self-report questionnaires, despite carers' objections, in recognition of the need of researchers and service providers to assess outcome quantitatively, in order to evaluate services systematically or conduct research that is generalisable.

The outcomes considered important by carers can be categorised into those that assess carers' well-being, those that assess the experience of caregiving, and those that assess carers' needs for professional support. For several outcomes considered important by carers, more than one instrument has been identified as reliable, valid, interpretable and feasible to use. The relative merits and demerits of each instrument will now be considered to identify candidate measures for use and further development. This process is also shown in Figure 2.

8.2.1 Instruments that assess carers' well-being

Physical and mental health

One candidate instrument:

- *Medical Outcomes Study SF-36*: further work will be required to determine its responsiveness and precision, as well as its acceptability to and appropriateness for carers.

One alternative instrument:

- *Giessen Symptom List Short Version*: an alternative for those seeking to assess physical health alone that has already been shown to be responsive to change, although its acceptability to and appropriateness for carers would need to be established.

Psychological well-being

Two candidate instruments:

- *Brief Symptom Inventory, General Health Questionnaire*: the acceptability to and appropriateness for carers and the responsiveness of both instruments will need to be determined by further work. A disadvantage of the General Health Questionnaire is that it is not free to use. We considered, however, that its merits in terms of reliability and validity outweighed its relatively low cost.

Five rejected instruments:

- *Beck Depression Inventory, CES Depression Scale, Geriatric Depression Scale, Postive and Negative Affect Scale, Older Adult Health and Mood Questionnaire*: we have not recommended these instruments because, although reliable and valid, they assess carers' mental health in its narrowest sense. They focus on depression and/or anxiety, both of which are constructs covered by the Medical Outcomes Study SF-36 and Multilevel Assessment Instrument recommended above, rather than the broad construct of psychological well-being that is important to carers.

Coping

Two candidate instruments:

- *Jalowiec Coping Scale, Ways of Coping Checklist*: data indicate that while the validity of the Jalowiec Coping Scale is superior, the Ways of Coping Checklist has the advantage that its appropriateness for carers has been established. Further work will be required to determine the appropriateness of the Jalowiec Coping Scale to carers. Both will require further work to determine their responsiveness and their acceptability to carers.

Carers' quality of life

One candidate instrument:

- *World Health Organisation Quality of Life Measure*: this instrument is recommended because it is the only reliable and valid quality of life measure identified in this review. However, the costs of using the instrument are unclear, and further work will be required to determine its acceptability to and appropriateness for carers.

One rejected instrument:

- *Life Satisfaction Index*: this instrument measures life satisfaction, a similar construct to quality of life. It is not recommended, however, because the reliability and validity of the candidate measure are better.

8.2.2 Instruments that assess the experience of caregiving

Caregiving activities

One candidate instrument:

- *Involvement Evaluation Questionnaire*: this was the only reliable and valid instrument identified by our review that assessed the type and amount of caregiving activities that carers undertake, as well as the financial consequences of caregiving. Further work will be required to establish its precision.

Effect of caregiving on the family

One candidate instrument:

- *Family Adaptability and Cohesion Scale II*: further work will be required to determine its acceptability to and appropriateness for carers and its responsiveness.

One rejected instrument:

- *Caregiving Hassles Scale*: this instrument is not recommended because, although reliable and valid, its responsiveness and acceptability to carers have yet to be established and the constructs it assesses are contained within the Involvement Evaluation Questionnaire.

How the carer feels about caregiving

Three candidate instruments:

- *Experience of Caregiving Inventory, Caregiver Strain Index, Carers' Assessment of Difficulties Index*: the responsiveness of the Experience of Caregiving Inventory and the Carers' Assessment of Difficulties Index will need to be determined from further work. Further work will also be required to determine the acceptability to carers and responsiveness of the Caregiver Strain Index. It is notable that the Carers' Assessment of Difficulties Index is the only candidate measure that can be administered in a semi-structured way that provides essentially qualitative data, as well as a self-report measure or structured interview.

One rejected instrument:

- *Caregiver Reaction Assessment*: this instrument measures a similar construct to the Experience of Caregiving Inventory. It is not recommended because, although reliable and valid, the Experience of Caregiving Inventory has better data regarding its appropriateness of and acceptability to carers.

Burden

One candidate instrument:

- *Screen for Caregiver Burden*: the acceptability of this instrument to carers remains to be determined. This is particularly important given that participants in the consultation exercise were divided about the importance and acceptability to carers of the burden construct.

Carers' sense of competence

One candidate instrument:

- *Revised Scale for Caregiving Self-Efficacy*: this instrument has the disadvantage that it does not provide a meaningful overall score. Moreover, further work will be required to determine its responsiveness.

One rejected instrument:

- *Short Sense of Competence Questionnaire*: although this instrument does have evidence of its appropriateness for carers, reliability and validity, that of the candidate instrument is better. Moreover, the candidate instrument has data regarding precision, which this instrument lacks.

Positive aspects of caregiving

Two candidate instruments:

- *Finding Meaning Through Caregiving Scale, Picot Caregiving Rewards Scale*: both of these instruments have good reliability and validity and have evidence of responsiveness; however, further work will be required to determine the acceptability to carers of both, and the appropriateness to carers of the latter.

One alternative instrument:

- *Experience of Caregiving Inventory*: this instrument has good reliability and validity, acceptability to and appropriateness for carers and includes a subscale assessing postive aspects of caregiving, although its responsiveness remains to be established.

8.2.3 Instruments that assess carers' needs for professional support

Carers' perception of the professional support they have received

One candidate instrument:

- *Experience of Caregiving Inventory*: the Experience of Caregiving Inventory includes a subscale assessing carers' appraisal of services, and has good reliability, validity, acceptability and appropriateness. We therefore recommend use of this instrument in studies where carers' perception of the professional support they have received is to be assessed.

One alternative instrument:

- *Professional Support Questionnaire*: we acknowledge that in some circumstances an instrument that only measures carers' perception of the professional support they have received may be required. In these circumstances, the Professional Support Questionnaire could be used; however, further work will be required to fully establish its reliability and validity, acceptability to and appropriateness for carers, and its responsiveness.

It should be noted that there may be instruments designed to assess physical and mental health, psychological well-being and coping that have superior psychometric properties to those recommended here, but that were not identified in our literature review because they have never been used to assess carer outcomes.

Section 9 Recommendations

1 Our finding that carers preferred open-ended or semi-structured interviews to questionnaires is based on a sample drawn from carer groups. These are, inevitably, a self-selected group and it is unclear how representative they are of carers overall. *We recommend a replication of the consultation exercise with a more representative sample of carers not enrolled in carer organisations*.

2 In the restricted time span of this review we were unable to ensure that views of specific carer groups (gay and lesbian carers, young carers, ethnic minorities, etc.) were adequately reported. Our carers did, however, have a range of backgrounds and we remain sceptical of the value of specific recommendations or policies for these groups at a national level. Locally, however, it was clear that the special needs of young carers are different and the area is poorly researched. *We recommend that local needs-assessments are conducted where there are high concentrations of 'untypical' carers such as in areas with high concentrations of specific ethnic minorities. We also recommend urgent study of the extent of young carers and an identification of their needs.*

3 The carer preference for more unstructured interviews rather than questionnaires expressed their concern that they had so often been neglected by services and a desire to have their experience recognised rather than a strongly held belief that this was the most effective way to mould services. We believe that the recently introduced right to a carer assessment as part of the Care Programme Approach will, when comprehensively implemented, meet this entirely legitimate aspiration. *We recommend that the routine assessment of carer needs and experience for service planning purposes be conducted by the use of validated and tested questionnaires, preferably self-report.*

4 Our review of the literature demonstrates three relatively discrete aims in carer assessment:
4.1 the identification of carers whose own health is at risk;
4.2 a better understanding of the carer experience;
4.3 the identification of specific carer needs for professional support.

We recommend that in local decision-making the purpose of conducting carer assessments should be carefully considered and the instruments selected to reflect this purpose.

5 We have outlined above the instruments which have been used for each of these three aims and have indicated which we consider have the best-established psychometric properties. *We recommend that service planners use these selected instruments, unless there are overwhelming local reasons for choosing others, so that benchmarking across services can be established.*

6 The Involvement Evaluation Questionnaire is the most reliable and valid of the comprehensive assessments of the experience of caregiving and has the additional benefit of including the 12-item General Health Questionnaire so that it simultaneously assesses carer-wellbeing. It also has the advantage that its appropriateness for and acceptability to carers has been established, as has its responsiveness, interpretability and feasibility. *We recommend it as currently the best all-round assessment of carers for routine use.*

7 While all of the instruments we have included in our final list measure factors considered important to carers and all have acceptable data on reliability and validity, all lack some data regarding other psychometric properties. *We recommend that studies are conducted to establish these properties, namely appropriateness for, and acceptability to carers, precision, responsivity and interpretability.*

8 *We recommend that any work into commissioning new scales should currently be restricted to neglected areas of the carer experience such as morale and emotional support.*

9 The Carers' and Users' Expectation of Services – Carers Version has no published data on reliability, thereby precluding it from our final list of recommended measures. However, it is the only instrument we reviewed which specifically addresses carers' needs and has rapidly attracted approval from clinicians and established a reputation for ease of use (feasibility). *We recommend that studies to establish the psychometric properties of Carers' and Users' Expectation of Services – Carers Version are conducted as a matter of urgency.*

10 Our study used two thresholds to identify potential instruments. The first requirement was that they had to have at least three published studies by groups beyond their developers. However, all instruments introduced in the last 10 years were included if they had been published once. This was to identify recent changes in practice in the development and focus of newer scales. Our findings indicated a recent emphasis on positive aspects of the carer experience. However, there was no evidence of significantly improved methodology or quality in these scales (in particular, no evidence of increased carer involvement in their development); nor was there any evidence of a significant surge in new scale development. *We recommend that the choice of scales to measure carer outcomes should continue to emphasise the traditional process of replication and establishment of psychometric properties.*

References 1 Cited articles not pertaining to instruments reviewed in the instrument assessment exercise

Arksey, H., O'Malley, L., Baldwin, S., Harris, J., Newbronner, E., Hare, P. and Mason, A. 2002. *Services to support carers of people with mental health problems scoping exercise*. London: NCCSDO

Burns, A., Lawlor, B. and Craig, S. (eds) 1999. *Assessment scales in old age psychiatry*. London: Martin Dunitz

Department of Health. 1998. *A matter of chance for carers? Inspection of Local Authority support for carers*. London: HMSO

Department of Health. 1999a. *Caring about carers. A national strategy for carers*. London: Department of Health

Department of Health. 1999b. *A National Service Framework for mental health.* London: Department of Health

Department of Health. 2001. *A National Service Framework for older people.* London: Department of Health

Fitzpatrick, R., Davey, C., Buxton, M.J. and Jones, D.R. 1998. Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment* **2(14)**

Harvey, K. 2000. *Relatives caring for a psychotic family member: factors that influence contact and experience*. University of London: PhD Thesis

Maurin, J. and Barmann Boyd, C. 1990. Burden of mental illness on the family: a critical review. *Archives of Psychiatric Nursing* **IV**: 99–107

Platt, S. 1985. Measuring the burden of psychiatric illness on the family: an evaluation of some rating scales. *Psychological Medicine* **15**: 383–93

Schene, A., Tessler, R.C. and Gamache, G. 1994. Instruments measuring family or caregiver burden in severe mental illness. *Social Psychiatry and Psychiatric Epidemiology* **29**: 228–40

Weinberger, M., Oddone, O.Z., Samsa, G.P. and Landsman, P.B. 1996. Are health-related quality-of-life measures affected by the mode of administration? *Journal of Clinical Epidemiology* **49**: 135–40

References 2 Original articles pertaining to the 49 instruments used three or more times

Note: this list includes articles that describe a revised version.

Affect Balance Scale

Bradburn, N.M. 1969. Two dimensions of psychological well-being: positive and negative affect. In *The structure of psychological well-being*: 53–89. Chicago: Aldine Publishing Company

Beck Depression Inventory

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J. and Erbaugh, J. 1961. An inventory for measuring depression. *Archives of General Psychiatry* **4**: 53–63

Brief Symptom Inventory

Derogatis, L. and Melisaratos, N. 1983. The Brief Symptom Inventory: an introductory report. *Psychological Medicine* **13**: 595–605

Caregiver Burden Interview

Zarit, S.H., Reever, K.E. and Bach-Peterson, J. 1980. Relatives of the impaired elderly: correlates of feeling of burden. *The Gerontologist* **20**: 649–55

Caregiver Burden Scale

Montgomery, R.J.V., Stull, D.E. and Borgatta, E.F. 1985. Measurement and the analysis of burden. *Research on Aging* **7(1)**: 137–52

Caregiver Reaction Assessment

Given, C.W., Given, B., Stommel, M., Collins, C., King, S. and Franklin, S. 1992. The caregiver reaction assessment (CRA) for caregivers to persons with chronic physical and mental impairments. *Research in Nursing and Health* **15(4)**: 271–83

Caregiver Strain Index

Robinson, B.C. 1983. Validation of a caregiver strain index. *Journal of Gerontology* **38**: 344–8

Caregiving Hassles Scale

Kinney, J.M. and Stephens, M.A. 1989. Caregiving Hassles Scale: assessing the daily hassles of caring for a family member with dementia. *Gerontologist* **29(3)**: 328–32

Centre for Epidemiological Studies Depression Scale

Radloff, L. 1977. The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement* **1**: 385–401

Coping Responses Inventory

Moos, R.H. 1988. Coping Responses Inventory: a measure of approach and avoidance coping skills. In Zalaquett, C.P. and Wood, R.J. (eds), *Evaluating stress: a book of resources*: 51–65. Lanham: Scarecrow

Daily Hassles and Uplifts Scale

Kanner, A.D., Coyne, J.C. and Schaefer, C. 1981. Comparisons of two modes of stress measurements: Daily hassles and uplifts vs. major life events. *Journal of Behavioural Medicine* **4**: 1–39

Experience of Caregiving Inventory

Szmukler, G., Burgess, P., Herrman, H., Benson, S., Colusa, S. and Bloch, S. 1996. Caring for relatives with a serious mental illness: the development of the 'Experience of Caregiving Inventory'. *Social Psychiatry and Psychiatric Epidemiology* **31**: 137–48

Family Adaptability and Cohesion Scale II

Olsen, D.H., McCubbin, H.I., Barnes, H., Larsen, A., Muxen, M. and Wilson, M. 1982. Family Adaptability and Cohesion Scale. In *Family Inventories*: 1–20. Minneapolis: Life Innovations

Family Assessment Measure version III

Skinner, H.A., Steinhauer, P.D. and Santa-Barbara, J. 1995. *Family Assessment Measure Version III: Technical Manual*. Multi-Health Systems

Family Burden Interview Schedule

Tessler, R.C., Fisher, G.A. and Gamache, G.M. 1992. *The Family Burden Interview Schedule: Manual.* Amherst, MA: Social and Demographic Research Institute, University of Massachusetts

Family Caregiving Inventory

Archbold, P.G. and Stewart, B.J. 1986. *Family Caregiving Inventory.* Portland, OR: Department of Family Nursing, School of Nursing, Oregon Health Sciences University

Family Environment Scale

Moos, R.H. and Moos, B.S. 1981. *Family Environment Scale.* Palo Alto, CA: Consulting Psychologists Press

Family Satisfaction Scale

Olson, D.H. and Wilson, M. 1982. Family Satisfaction Scale. In Olsen, D.H., McCubbin, H.I., Barnes, H., Larsen, A., Muxen, M. and Wilson, M. (eds), *Family Inventories*: 21–28. Minneapolis: Life Innovations

General Health Questionnaire

Goldberg, D. 1978. *The General Health Questionnaire.* Windsor: NFER-Nelson

Geriatric Depression Scale

Yesavage, J.A., Brink, T.L., Rose, T.L., Lum, O., Huang, V., Adey, M. and Leirer, O. 1983. Development and validation of a geriatric depression screening scale: a preliminary report. *Journal of Psychiatric Research* **17**: 37–49

Hamilton Rating Scale for Depression

Hamilton, M. 1960. A rating scale for depression. *Journal of Neurological Psychiatry* **26**: 56–62

Hospital Anxiety and Depression Scale

Zigmond, A.S. and Snaith, R.P. 1983. The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica* **67**: 361–70

Interpersonal Support Evaluation List

Cohen, S., Mermelstein, R., Karmarck, T. and Hoberman, H.M. 1984. Measuring the functional components of social support. In Sarason, I.G. and Sarason, B.R. (eds), *Social support: theory, research, and applications*: 73–94. The Hague: Martinus Nijhoff

Inventory of Socially Supportive Behaviours

Barrera, M., Sandler, I.N. and Ramsay, T.B. 1981. Preliminary development of a social of social support: studies on college students. *American Journal of Community Psychology* **9**: 435–47

Involvement Evaluation Questionnaire

Schene, A. and van Wijngaarden, B. 1992. *The Involvement Evaluation Questionnaire.* Amsterdam: Department of Psychiatry, University of Amsterdam

van Wijngaarden, B. 2003. *Consequences for caregivers of patients with severe mental illness.* University of Amsterdam: PhD Thesis

Jalowiec Coping Scale

Jalowiec, A. 1979. *Stress and coping in hypertensive and emergency room patients*. University of Illinois: Master's Thesis

Life Orientation Test

Scheier, M. and Carver, C.S. 1985. Optimism, coping and health: assessment and implications of generalised outcomes expectancies. *Health Psychology* **4**: 219–47

Scheier, M., Carver, C.S. and Bridges, M.W. 1994. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology* **67**: 1063–78

Life Satisfaction Index

Wood, V., Wylie, M.L. and Sheafor, B. 1969. An analysis of a short self-report measure of life satisfaction: correlation with rater judgements. *Journal of Gerontology* **24**: 465–9

Mastery Scale

Pearlin, L.I. and Schooler, C. 1978. The structure of coping. *Journal of Health and Social Behaviour* **19**: 2–21

Medical Outcomes Study SF-36

McHorney, C.A., Ware, J.E. and Raczek, A.E. 1993. The MOS 36-item Short-Form Health Survey (SF-36). II Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Medical Care* **31**: 247–63

McHorney, C.A., Ware, J.E., Lu, R. and Sherbourne, C.D. 1994. The MOS 36-item Short-Form Health Survey (SF-36). III Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Medical Care* **32**: 40–66

Ware, J.E. and Sherbourne, C.D. 1992. The MOS 36-item Short-Form Health Survey (SF-36). I Conceptual framwork and item selection. *Medical Care* **30**: 473–83

Multilevel Assessment Instrument

Lawton, M.P., Moss, M., Fulcomer, M. and Morton, H.K. 1982. A research and service oriented multilevel assessment instrument. *Journal of Gerontology* **37**: 91–9

NEO Personality Inventory

Costa, P.T. and McCrae, R.R. 1985. *The NEO Personality Inventory manual.* Odessa, FL: Psychological Assessment Resources

Costa, P.T., McCrae, R.R. and Dye, D.A. 1991. Facet scales for agreeableness and conscientiousness: a revision of the NEO Personality Inventory. *Individual Differences* **12**: 887–98

Neuropsychiatric Inventory Caregiver Distress Scale

Kaufer, D.I., Cummings, J.L., Christine, D., Bray, T., Castellon, S., Masterman, D. *et al.* 1998. Assessing the impact of neuropsychiatric symptoms in Alzheimer's disease: the Neuropsychiatric Inventory Caregiver Distress Scale. *Journal of the American Geriatrics Society* **46(2)**: 210–5

Pearlin Caregiving Measures

Pearlin, L.I., Mullan, J.T., Semple, S.J. and Skaff, M.M. 1990. Caregiving and the Stress Process: An Overview of Concepts and Their Measures. *Gerontologist* **30(5)**: 583–94

Skaff, M.M. and Pearlin, L.I. 1992. Caregiving: role engulfment and the loss of self. *Gerontologist* **32(5)**: 656–64

Perceived Stress Scale

Cohen, S., Karmarck, T. and Mermelstein, R. 1983. A global measure of perceived stress. *Journal of Health and Social Behaviour* **24**: 385–96

Pittsburgh Sleep Quality Index

Buysse, D.J., Reynolds, C.F., Monk, T.H., Berman, S.R. and Kupfer, D.J. 1989. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Research* **28**: 193–213

Positive and Negative Affect Schedule

Watson, D. and Clark, L.A. 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology* **54**: 1063–70

Profile of Mood States

McNair, D.M., Lorr, M. and Droppleman, L.F. 2003. *POMS Profile of Mood States Manual*. Multi-Health Systems

Relatives' Stress Scale

Greene, J.G., Smith, R., Gardiner, M. and Timbury, G.C. 1982. Measuring behavioural disturbance of elderly demented patients in the community and its effects on relatives: a factor analytic study. *Age and Ageing* **11**: 121–6

Revised Memory and Behaviour Problems Checklist

Teri, L., Truax, P., Logsdon, R., Uomoto, J., Zarit, S. and Vitaliano, P.P. 1992. Assessment of behavioural problems in dementia: the Revised Memory and Behaviour Problems Checklist. *Psychology and Aging* **7**: 622–31

Screen for Caregiver Burden

Vitaliano, P.P., Russo, J., Young, H.M., Becker, J. and Maiuro, R.D. 1991. The Screen for Caregiver Burden. *The Gerontologist* **31**: 76–83

Self-Rating Depression Scale

Zung, W.W.K. 1965. A self-rating depression scale. *Archives of General Psychiatry* **12**: 63–70

Social Behaviour Assessment Schedule

Platt, S., Weyman, A., Hirsch, S. and Hewett, S. 1980. *Social Behaviour Assessment Schedule* (SBAS), 3rd edn. Windsor: NFER-Nelson

Social Satisfaction Scale

George, L.K. and Gwyther, L.P. 1986. Caregiver well-being: a multidimensional examination of family caregivers of demented adults. *The Gerontologist* **26**: 253–9

Social Support Appraisals Scale

Vaux, A., Phillips, J., Holly, L., Thomson, B., Williams, D. and Stewart, D. 1986. The Social Support Appraisals (SS-A) Scale: studies of reliability and validity. *American Journal of Community Psychology* **14**: 195–219

Social Support Questionnaire

Sarason, I.G., Levine, H.M., Basham, R.B. and Sarason, B.R. 1983. Assessing social support: the Social Support Questionnaire. *Journal of Personality and Social Psychology* **44**: 127–39

State-Trait Anxiety Inventory

Spielberger, C.D., Gorsuch, R.L. and Lushene, R.E. 1974. *Manual for the Stait-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press (out of print); now available from www.mindgarden.com

Strain Scale

Gilleard, C.J. 1984. *Living with dementia: community care of the elderly mentally infirm*. Beckenham: Croom Helm

Ways of Coping Checklist

Folkman, S. and Lazarus, R.S. 1985. If it changes it must be a process: study of emotion and coping during three stages of college examination. *Journal of Personality and Social Psychology* **48**: 150–70

Vitaliano, P.P., Russo, J., Carr, J.E., Maiuro, R.D. and Becker, J. 1985. The Ways of Coping Checklist: revision and psychometric properties. *Multivariate Behavioural Research* **20**: 3–26

References 3 Additional articles pertaining to the 49 instruments used three or more times

Aaronson, N.K., Acquadro, C., Alonso, J., Apolone, G., Bucquet, D., Bullinger, M. *et al.* 1992. International Quality of Life Assessment (IQOLA) Project. Netherlands Cancer Institute, Amsterdam. *Quality of Life Research* **1**: 349–51

Andresen, E.M., Malmgren, J.A., Carter, W.B. and Patrick, D.L. 1994. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiological Studies Depression Scale). *American Journal of Preventative Medicine* **10**: 77–84

Arai, Y., Kudo, K., Hosokawa, T., Washio, M., Miura, H. and Hisamichi, S. 1997. Reliability and validity of the Japanese version of the Zarit Caregiver Burden interview. *Psychiatry and Clinical Neurosciences* **51**: 281–7

Archbold, P.G., Stewart, B.J., Greenlick, M.R. and Harvath, T.A. 1990. Mutuality and preparedness as predictors of caregiver role strain. *Research in Nursing Health* **13**: 375–84

Aroian, K.J., Patsdaughter, C.A., Levin, A. and Gianan, M.E. 1995. Use of the Brief Symptom Inventory to assess psychological distress in three immigrant groups. *International Journal of Social Psychiatry* **41**: 31–4

Aylard, P.R., Gooding, J.H., McKenna, P.S. and Snaith, R.P. 1987. A validation study of three anxiety and depression self-assessment scales. *Journal of Psychosomatic Research* **31**: 261–8

Barczak, P., Kane, N., Andrews, S., Congdon, A.M., Clay, J.C. and Betts, T. 1988. Patterns of psychiatric morbidity in a genito-urinary clinic: a validation of the Hospital Anxiety and Depression Scale. *British Journal of Psychiatry* **152**: 698–700

Barraca, J., López Yarto, L. and Olea, J. 2000. Psychometric properties of a New Family Life Satisfaction Scale Universidad Autónoma de Madrid. *Spain European Journal of Psychological Assessment* **16**: 98–106

Barrera, M., Sandler, I.N. and Ramsay, T.B. 1981. Preliminary development of a social of social support: studies on college students. *American Journal of Community Psychology* **9**, 435–47

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J. and Erbaugh, J. 1961. An inventory for measuring depression. *Archives of General Psychiatry* **4**: 561–71

Beck, A.T., Rial, W.Y. and Rickets, K. 1974. Short form of Depression Inventory: cross-validation. *Psychological Reports* **34**: 1184–6

Beck, A.T., Steer, R.A. and Brown, G.K. 1996. *Manual for the Beck Depression Inventory*, 2nd edn. San Antonio, TX: The Psychological Corporation

Blackman, M., Pitcher, S. and Rauch F. 1986. A preliminary outcome study of a community group treatment programme for emotionally disturbed adolescents. *Canadian Journal of Psychiatry* **31(2)**: 112–239

Bobes, J., Bulbena, A., Luque, A., Dal-Re, R., Ballesteros, J. and Ibarra, N. 2003. A comparative psychometric study of the Spanish versions with 6, 17, and 21 items of the Hamilton Depression Rating Scale. *Medicina Clinica (Barcelona)* **120(18)**: 693–700

Bloomquist, M.L. and Harris WG. 1984. Measuring family functioning with the MMPI: A reliability and concurrent validity study of three MMPI scales. *Journal of Clinical Psychology* **40**: 1209–14

Bradburn, N. 1969. *The Structure of Well-Being*. Chicago, IL: Aldine

Bramley, P.N., Easton, A.M., Morley, S. and Snaith, R.P. 1988. The differentiation of anxiety and depression by rating scales. *Acta Psychiatrica Scandinavica* **77**: 136–9

Brislin, R.W. 1980. Translation and content analysis of oral and written materials. In Triandis, H.C. and Berry, J.W. (eds), *Handbook of cross-cultural psychology-methodology*: 389–444. Boston: Allen and Bacon

Burns, A., Lawlor, B. and Craig, S. 1999. *Assessment scales in old age psychiatry*. London: Martin Dunitz

Cantril, A.H. and Roll, C.W. 1971. *Hopes and Fears of the American People*. New York: Universe Books

Carpenter, S. and Andrykowski, M.A. 1998. Psychometric evaluation of the Pittsburgh sleep quality index. *Journal of Psychosomatic Research* **45**: 5–13

Ceolim, M.F. and Menna-Barreto, L. 2000. Sleep/wake cycle and physical activity in healthy elder people. *Sleep Research Online* **3**: 87–95

Chen, M.L. and Hu, L.C. 2002. The generalizability of Caregiver Strain Index in family caregivers of cancer patients. *International Journal of Nursing Studies* **39**: 823–9

Chou, K.-R., Chu, H., Tseng, C.-L. and Lu, R.-B. 2003. The measurement of caregiver burden. *Journal of Medicine Scienc*e **23(2)**: 73–82

Costa, P.T. and McCrae, R.R. 1991. Facet scales for agreeableness and conscientiousness: a revision of the Neo Personality Inventory. *Personality and Individual Differences* **12(9)**: 887–98

Crowne, D.P. and Marlowe, D.A. 1960. A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology* **24**: 349–54

Deimling, G.T. and Bass, D.M. 1984. The strengths and resources of families caring for impaired elders. a report to the Retirement Research Foundation. Cleveland, OH: Benjamin Rose Institute

Dekker, J.J.M.D. 1996. *Het Amsterdams Sociaal Psychiatrisch Dienstencentrum: een vergelijkend onderzoek naar een multi-functionele eenheid en de reguliere 24-uurs zorg*. University of Amsterdam: PhD Thesis

Derogatis, L.R. 1983. *SCL-90-R administration, scoring, and procedures manual*, vol. II. Baltimore, MD: Clinical Psychometric Research

Derogatis, L.R., Lipman, R.S., Covi, L., Rickels, K. and Uhlenhuth, E.H. 1970. Dimensions of outpatient neurotic pathology: Comparison of a clinical versus an empirical assessment. *Journal of Consulting Psychology* **34**: 164–71

Derogatis, L.R., Lipman, R.S., Covi, L. and Rickels K. 1971. Neurotic symptom dimensions: As perceived by psychiatrists and patients of various social classes. *Archives of General Psychiattry* **24**: 454–64

Derogatis, L.R., Lipman, R.S., Rickels, K., Uhlenhuth, E.H. and Covi, L. 1974. The Hopkins Symptom Checklist (HSCL): A Self-Report Symptom Inventory. *Behavioural Science* **19**: 1–15

Engfer, A., Schneewind, K.A. and Hinderer, J. 1977. *Die familien-klima-salen (FKS). Ein fragebogen zur erhebung perzipierter familienumwelten nach, R.H. Moos.* Munich: University of Munich

Epstein, N.B., Baldwin, L.M. and Bishop, D. 1982. The McMaster family assessment device. *Journal of Marital and Family Therapy* **9(2)**: 171–80

Eysenck, H.J. 1958. A short questionnaire for the measurement of two dimensions of personality. *Journal of Applied Psychology* **42**: 14–17

Eysenck, H.J. and Eysenck, S.B. 1968. *Nanual: Eysenck Personality Inventory*. San Diego, CA: Educational Testing Service

Ferrell, B.R., Grant, M., Chan, J. and Ferrell, C. 1995. The impact of cancer pain education on family caregivers of elderly patients. *Oncological Nursing Forum* **22**: 1211–18

Folstein, M.F., Folstein, S.E. and McHugh, P.R. 1975. "Mini-Mental State". A practical method for grading the cognitive sate of patients for the clinician. *Journal of Psychiatric Research* **12**: 189–98

Gamache, G. 1996. The Family Burden Interview Schedule - Short Form (FBIS/SF). In Sederer, L. and Dickey, B. (eds), *Outcomes Assessment in Clinical Practice*: 110–12. Baltimore, MD: Williams and Wilkins

Garfinkel, P.E., Garner, D.M., Rose, J., Darby, P.L., Brandes, J.S., O'Hanlon, J. and Walsh, N. 1983. A comparison of characteristics in the families of patients with anorexia nervosa and normal controls. *Psychological Medicine* **13**: 821–8

Gaudreau, P. 2000. *Vers une version francaise du PANAS: Analyses en composantes principales avant, pendant, et apres une competition sportive.* Montreal: University of Montreal

Goldberg, D. and Hillier, V.F. 1979. A scaled version of the General Health Questionnaire. *Psychological Medicine* **9**: 139–45

Greene, J.G., Smith, R., Gardiner, M. and Timbury, G.C. 1982. Measuring behavioural disturbance of elderly demented patients in the community and its effects on relatives: a factor analytic study. *Age and Ageing* **11**: 121–6

Grizenko, N. and Sayegh, L. 1990. Evaluation of the effectiveness of a psychodynamically oriented day treatment program for children with behaviour problems: a pilot study. *Canadian Journal of Psychiatry* **35**: 519–25

Groth-Marnat, G. 1990. *The handbook of psychological assessment*, 2nd edn. New York: John Wiley and Sons

Goh, D.S., King, D.W. and King, L.A. 1982. Psychometric evaluation of the Eysenck Personality Questionnaire. *Educational and Psychological Management* **42**: 297–309

Guilford, J.S., Zimmerman, W.S. and Guilford, J.P. 1976. *The Guilford-Zimmerman Temperament Survey handbook: Twenty-five years of research and application*. San Diego, CA: EDITS

Hamilton, M. 1960. A rating scale for depression. *Journal of Neurological and Neurosurgical Psychiatry* **23**: 56–62

Hamilton, M. 1967. Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology* **6**: 278–96

Haskell, D.S., Pugatch, D. and McNair, D.M. 1969. Time-limited psychotherapy for whom? *Archives of General Psychiatry* **21**: 546–52

Hersch, E.L., Kral, V.A. and Palmer, B. 1978. Clinical value of the London Psychogeriatric Rating Scale. *Journal of the American Geriatrics Society* **26**: 348- 354

Holstein, R.M. 1970. *Patient and therapist initial expectancies for psychotherapeutic change as they relate to treatment outcomes.* Unpublished dissertation

Huizink, A.J., Robles de Medina, P.G., Mulder, E.J.H., Visser, G.H.A. and Buitelaar, J.K. 2001. *Does coping mediate the effects of stress in pregnancy?*, chapter 6. Thesis, University of Utrecht

Ibbotson, T., Maguire, P. and Selby, T. 1989. *Validating self-rating questionnaires.* Paper presented to the Annual Conference of the British Psychosocial Oncology Group

Jacob, T. 1991. *Comparison of alcoholic, depressed, and control families using the FAM-III.* Unpublished manuscript

Jacob, T. 1995. The role of the time frame in the assessment of family functioning. *Journal of Marital and Family Therapy* **21(3)**: 281–6

Joyce, J., Leese, M. and Szmukler, G. 2000. The Experience of Caregiving Inventory: further evidence. *Social Psychiatry and Psychiatric Epidemiology* **35**: 185–9

Katerndahl, D.A. 2002. Factors influencing care seeking for a self-defined worst panic attack. *Psychiatric Services* **53**: 464–70

Katz, S., Downs, T., Cash, H. and Grotz, R. 1970. Progress in development of the index of ADL. *The Gerontologist* **10**: 20–31

Kawada, T., Suzuki, S., Kubota, F., Ohnishi, N. and Satoh, K. 1999. Content and Cross Validity of the Todai Health Index Depression Scale in Relation to the Center for Epidemiological Studies Depression Scale and the Zung Self-rating Depression Scale. *Journal of Occupational Health* **41**: 154–9

Kim, H., Garvin, B.J. and Moser, D.K. 1999. Stress during mechanical ventilation: benefit of having concrete objective information before cardiac surgery. *American Journal of Critical Care* **8**: 118–26

Kim, J.-S. and Lee, E.-H. 2003. Cultural and noncultural predictors of health outcomes in Korean daughter and daughter-in-law caregivers. *Public Health Nursing* **20(2)**: 111–19

Kramer, J.J.A.M. 1998. *Families under pressure. The development and psychometric evaluation of the interview for the Burden on the Family (IBF).* [in Dutch]. University of Groningen: PhD Thesis

Lai, J.C.L., Cheung, H., Lee, W.-M. and Yu, H. 1988. The utility of the Revised Life Orientation Test to Measure Optimism among Hong Kong Chinese. *International Journal of Psychology* **33**: 45–56

Langner, S. 1983. *Reliability of the Jalowiec Coping Scale.* Chicago, IL: University of Illinois

Lazarus, R.S. and Launier, R. 1978. Stress-related transactions between the person and the environment. In Pervin, L.A. and Lewis, M. (eds), *Perspective in interactional psychology*: 287–327. New York: Plenum

Levene, J.E. 1991. *The influence of expressed emotion and family interaction on the course of illness in Schizophrenia.* Doctoral Thesis

Lorr, M., McNair, D.M., Weinstein, G.J., Michaux, W.W. and Raskin, A. 1961. Meprobamate and chlorpromazine in psychotherapy. *Archives of General Psychiatry* **4**: 381–9

Lorr, M., McNair, D.M. and Weinstein, G.J. 1964. Early effects of chlordiazepoxide (Librium) used with psychotherapy. *Journal of Psychiatric Research* **1**: 257–70

Lucas, R.E., Diener, E., Grob, A., Suh, E. and Shao, L. 2000. Cross-cultural evidence for the fundamental features of extraversion. *Journal of Personality and Social Psychology* **79**: 452–68

MacCarthy, B. and Brown, R. 1989. Psychosocial factors in Parkinson's disease. *British Journal of Clinical Psychology* **28**: 41–52

MacKinnon, A., Jorm, A.F., Christensen, H., Korten, A.E., Jacomb, P.A. and Rodgers, B. 1999. A short form of the Positive and Negative Affect Schedule: evaluation of factorial validity and invariance across demographic variables in a community sample. *Personality and Individual Differences* **27**: 405–16

Magne-Ingvar, U. and Öjehagen, A. 2002. *Significant others of persons with mental health problems: the testing of a questionnaire on the burden of significant others*. Unpublished

Marusic, I. and Bratko, D. 1998. Relations of masculinity and femininity with personality dimensions of the five-factor model. *Sex Roles: A Journal of Research* **38(1–2)**: 29–44

McCrae, R., Costa, P., Zonderman, A. and Siegler, I. 1996. *Stress, coping and personality in aging men and women*. Baltimore, MD: National Institute on Aging, Gerontology Research Center

McCrae, R.R. and Allik, J. (eds) 2002. *The Five-Factor Model of Personality Across Cultures (International and Cultural Psychology)*. New York: Kluwer Academic/Plenum Publishers

Mermelstein, R., Lichtenstein, E. and McIntyre, K. 1983. Partner support and relapse in smoking-cessation programs. *Journal of Consulting and Clinical Psychology* **51(3)**: 465–6

Montazeri, A., Vahdaninia, M., Ebrahimi, M. and Jarvandi, S. 2003. The Hospital Anxiety and Depression Scale (HADS): translation and validation study of the Iranian version. *Health Quality of Life Outcome*s **1(1)**: 14

Moorey, S., Greer, S., Watson, M., Gorman, C., Rowien, A., Tunmore, R., Robertson, B. and Bliss, J. 1991. The factor structure and factor stability of the HAD Scale in Patients with Cancer. *British Journal of Psychiatry* **158**: 255–9

Moos, R.H. 1974. *Family environment scales.* Palo Alto, CA: Consulting Psychologists Press

Moos, R.H. and Moos, B.H. 1981. *Family Environment Scale Manual*. Palo Alto, CA: Consulting Psychogists Press

Mumford, D.B., Tareen, I.A.K. and Bajwa, M.A.Z. 1991. The translation and evaluation of an Urdu version of the Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica* **83**: 81–5

Murphy, S.P. 1982. *Factors influencing adjustment and quality of life of haemodialysis patients: a multivariate approach.* University of Illinois: PhD Thesis

Mykletun, A., Storfal, E. and Dahl, A. 2001. Hospital Anxiety and Depression (HAD) scale: factor structure, item analyses and internal consistency in a large population. *British Journal of Psychiatry* **179**: 540–4

Neugarten, B.L., Havighurst, R.J. and Tobin, S.S. 1961. The measurement of life satisfaction. *Journal of Gerontology* **16**, 134–43

Nilchaikovit, T., Lortrakul, M. and Phisansuthidet, U. 1996. Development of Thai version of Hospital Anxiety and Depression Scale in cancer patients. *Journal of Psychiatric Association of Thailand* **41(1)**: 18–30

Olson, D.H., Russell, C.S. and Sprenkle, D.H. 1983. Circumplex model of marital and family systems, 6: Theoretical update. *Family Process* **22**: 69–83

Olson, D.H., Portner, J. and Lavee, Y. 1987. Family adaptability and cohesion evaluation scales [FACES III]. In Corcoran, K. and Fischer, J. (eds), *Measures for clinical practice: a sourcebook*: 428–31. New York: Free Press

O'Rourke, N. and Tuokko, H.A. 2003. Psychometric properties of an abridged version of The Zarit Burden Interview within a representative Canadian caregiver sample. *Gerontologist* **43**: 121–7

Pearlin, L.I. and Schooler, C. 1978. The structure of coping. *Journal of Health and Social Behaviour* **19**: 2–21

Perczek, R., Carver, C.S., Price, A.A. and Pozo-Kaderman, C. 2000. Coping, mood, and aspects of personality in Spanish translation and evidence of convergence with English versions. *Journal of Personality Assessment* **74**: 63–87

Perez, C.M. 1989. Maternal depression, parenting stress, social support, and coping mechanisms: a comparison on Puerto Rican mothers of clinic and non-clinic latency-age boys. Doctoral dissertation. *Dissertation Abstracts International* **50**: 3172

Pfeiffer, E. 1975. A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. *Journal of the American Geriatrics Society* **23**: 433–8

Pino, C., Simons, N. and Slowinowski, M.J. 1984. The Children's Family Environment Scale (CFES). *Family Therapy* **11(1)**: 85–6

Poulshock, S.W. and Deimling, G.T. 1984. Families caring for elders in residence: Issues in the measurement of burden. *Journal of Gerontology* **39**: 230–9

Procidano, M. and Heller, K. 1983. Measures of perceived social support from friends and from family: three validational studies. *American Journal of Community Psychology* **11**: 1–24

Radloff, L.S. 1977. The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement* **1**: 385–401

Raskin, A., Schulterbrandt, J., Reatig, N. and McKeon, J. 1969. Replication of factors and psychopathology in interview, ward behavior, and self-report ratings of hospitalized depressives. *Journal of Nervous and Mental Disease* **148**: 87–96

Reynolds, W.M. and Kobak, K.A. 1995. Reliability and validity of the Hamilton Depression Inventory: a paper-and-pencil version of the Hamilton Depression Rating Scale Clinical Interview. *Psychological Assessment* **7**: 472–83

Rosenberg, M. 1965. *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press

Royuela, A., Macias, J.A., Gil-Verona, J.A. *et al.* 2002. Sleep in schizophrenia: a preliminary study using the Pittsburgh Sleep Quality Index. *Neurobiology of Sleep-Wakefulness Cycle* **2**: 37–9

Salvador-Carulla, L., Garcia-Mellado, M.J., Velazquez, R., Romero, C. and Alonso, F. 1998. A reliability study of the Spanish version of the Social Behaviour Schedule (SBS) in a population of adults with learning disabilities. *Journal of Intellectual Disability Research* **42**: 22–8

Sarason, I.G., Sarason, B.R., Shearin, E.N. and Pierce, G.R. 1987. A brief measure of social support: practical and theoretical implications. *Journal of Social and Personal Relationships* **4**: 497–510

Schene, A.H. and van Wijngaarden, B. 1993. Family members of psychotic patients. An inquiry among Ypsilon members [in Dutch]. *Maandblad Geestelijke Volksgezondheid* **9**: 899–914

Shacham, S., Reinhardt, L.C., Raubertas, R.F. and Cleeland, C.S. 1983. Emotional states and pain: interindividual and intraindividual measures of association. *Journal of Behavioural Medicine* **6**: 405–18

Shao, L. 1997. *Extraversion and positive affect.* University of Illinois at Urbana: PhD Thesis

Shaul, S.L. 1981. *Loneliness: a comparison of two group counseling strategies with adults.* University of Washington: PhD Thesis

Skinner, H.A., Steinhauer, P.D. and Santa-Barbara J. 1983. The Family Assessment Measure. *Canadian Journal of Community Mental Health* **2(2)**: 91–105

Smith, G.T., McCarthy, D.M. and Anderson, K.G. 2000. On the sins of short-form development. *Psychological Assessment* **12**: 102–11

Speilberger, C.D., Gorsuch, R.L. and Lushene, R.E. 1974. *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press

Stam, H. and Cuipers, P. 2001. Effects on family interventions on burden of relatives of psychiatric patients in the Netherlands: a pilot study. *Community Mental Health Journal* **37**: 179–87

Steer, R.A., Rissmiller, D.J. and Beck, A.T. 2000. Use of the Beck Depression Inventory with depressed geriatric patients. *Behaviour Research and Therapy* **38**: 311–18

Sultan, S. and Bureau, B. 1999. Which optimism in health psychology? *European Review of Applied Psychology* **49**: 43–51

Szmukler, G.I., Burgess, P., Herrman, H., Benson, A., Colusa, S. and Bloch, S. 1996. Caring for relatives with serious mental illness: the development of the Experience of Caregiving Inventory. *Social Psychiatry and Psychiatric Epidemiology* **31**: 137–48

Takada, K. 2002. Study on Quality of Life (QOL) and Mental Health among Japanese living in Australia. Sampling and identification in the general population of the Japanese community overseas: use of electronic residential telephone directory and surname statistics. *Japanese Bulletin of Social Psychiatry* **11**: 1–10

Tatsuki, S. 1999. *Theoretical and empirical study of family system: validation of Olson's Circumplex model of marital and family systems* [in Japanese]. Tokyo: Kawashima Shoten

Tennakoon, L., Fannon, D., Doku, V., O'Ceallaigh, S., Soni, W., Santamaria, M. *et al.* 2000. Experience of caregiving: relatives of people experiencing first episode psychosis. *British Journal of Psychiatry* **177**: 529–33

Turner, J.A. and Romano, J.M. 1984. Self-report screening measures for depression in chronic pain patients. *Journal of Clinical Psychology* 40, 909–13

van Wijk, A.S. and Haveman, M.J. 1992. *Transmurale zorg in de psychiatrie: een experimentele verkenning naar bet sociaal-psychiatrisch centrum te Maastricht*. Maastricht: RU Limburg en Riagg/Vijverdal Combinatie

van Wijngaarden, B., Schene, A.H. and Koeter MWJ 1996. *The consequences of depressive disorders for those involved with the patient: a study on the psychometric qualities of the Involvement Evaluation Questionnaire* [in Dutch]. Amsterdam: Department of Psychiatry, University of Amsterdam

van Wijngaarden, B., Schene, A.H., Koeter, M. *et al.* and the EPSILON study group. 2000. Caregiving in Schizophrenia: development, internal consistency and reliability of the European version of the Involvement Evaluation Questionnaire – EU version (IEQ-EU). *British Journal of Psychiatry* **177** (suppl. 39): s21–7

Vaughn, C. and Leff, J. 1976. The measurement of expressed emotion in the families of psychiatric patients. *British Journal of Clinical Psychology* **15**: 157–65

Voyer, M. and Vezina, J. no date. *Contribution of hassles, appraisals and coping to psychological distress amongst elderly widows.* Quebec: Laval University

Wan, W., Luk, C.-L. and Lai, J. 1999. Personality correlates of loving styles amongst Chinese students in Hong Kong. *Personality and Individual Differences* **29**: 169–75

Watson, D. and Clark, L.A. 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology* **54**: 1063–70

Watson, 1988. Intraindividual and interindividual analyses of Positive and Negative Affect: their relation to health complaints, perceived stress and daily activities. *Journal of Personality and Social Psychology* **54**: 1020–30

Wolthaus, J.E.D., Dingemans, P.M.A.J., Schene, A.H., Linszen, D.H., Wiersma, D., van den Bosch, R.J. *et al.* 2002. Caregiver burden in recent-onset schizophrenia and spectrum disorders: the influence of symptoms and personality traits. *Journal of Nervous and Mental Disorders* **190**: 421–7

Yam, B.M.C. and Shiu, T.Y. 2003. Research in Brief Perceived stress and sense of coherence among critical care nurses in Hong Kong: a pilot study*. Journal of Clinical Nursing* **12**: 144–6

References 4 Original articles pertaining to the 15 instruments developed in the last 10 years and used fewer than three times

Note: this list includes articles that describe a revised version.

Carers' Assessment of Difficulties Index

Nolan, M. and Grant, G. 1992. *Regular respite. An evaluation of a hospital rota bed scheme for elderly people*. London: Age Concern Institute of Gerontology

Carers' Assessment of Managing Index

Nolan, M., Keady, J. and Grant, G. 1995. CAMI: a basis for assessment and support with family carers. *British Journal of Nursing* **4(14)**: 822–6

Carers' Assessment of Satisfactions Index

Nolan, M. and Grant, G. 1992. *Regular respite. An evaluation of a hospital rota bed scheme for elderly people*. London: Age Concern Institute of Gerontology

Carers' Checklist

Hodgson, C., Higginson, I. and Jefferys, P. 1998. *Carers' Checklist: an outcome measure for people with dementia and their carers*. London: The Mental Health Foundation

Carers' and Users' Expectations of Services – Carer Version

Lelliott, P., Beevor, A., Hogman, G., Hyslop, J., Lathlean, J. and Ward, M. 2003. Carers' and users' expectations of services – carer version (CUES-C): a new instrument to support the assessment of carers of people with a severe mental illness. *Journal of Mental Health* **12(2)**: 143–52

Finding Meaning Through Caregiving Scale

Farran, C.J., Miller, B.H., Kaufman, J.E., Donner, E. and Fogg, L. 1999. Finding meaning through caregiving: development of an instrument for family caregivers of persons with Alzheimer's disease. *Journal of Clinical Psychology* **55(9)**: 1107–25

Gratifications of Caregiving Questionnaire

Schwartz, C. and Gidron, R. 2002. Parents of mentally ill adult children living at home: rewards of caregiving. *Health and Social Work* **27(2)**: 145–54

Picot Caregiver Rewards Scale

Picot, S.J., Youngblut, J. and Zeller, R. 1997. Development and testing of a measure of perceived caregiver rewards in adults. *Journal of Nursing Measurement* **5(1)**: 33–52

Professional Support Questionnaire

Reinhard, S.C. 1994. Living with mental illness: effects of professional support and personal control on caregiver burden. *Research in Nursing and Health* **17(2)**: 79–88

Revised Scale for Caregiving Self-Efficacy

Steffen, A.M., McKibbin, C., Zeiss, A.M., Gallagher-Thompson, D. and Bandura, A. 2002. The Revised Scale for Caregiving Self-Efficacy: Reliability and validity studies. *Journal of Gerontology* **57B(1)**: 74–86

Zeiss, A.M., Gallagher-Thompson, D., Lovett, S., Rose, J. and McKibbin, C. 1999. Self-efficacy as a mediator of caregiver coping: development and testing of an assessment model. *Journal of Clinical Geropsychology* **5(3)**: 221–30

Short Sense of Competence Questionnaire

Vernooij-Dassen, M.J., Felling, A.J., Brummelkamp, E., Dauzenberg, M.G., van den Bos, G.A. and Grol, R. 1999. Assessment of caregiver's competence in dealing with the burden of caregiving for a dementia patient: a Short Sense of Competence Questionnaire (SSCQ) suitable for clinical practice. *Journal of the American Geriatrics Society* **47(2)**: 256–7

Giessen Symptom List, Short Version

Brahler, E. and Scheer, J.W. 1995. Giessen Symptom List in German]. Seattle, WA: Huber

Older Adult Health and Mood Questionnaire

Kemp, B.J. and Adams, B.M. 1995. The Older Adult Health and Mood Questionnaire: a measure of geriatric depressive disorder. *Journal of Geriatric Psychiatry and Neurology* **8(3)**: 162–7

Schedule for the Evaluation of Individual Quality of Life – Direct Weighting

O'Boyle, C.A., Browne, J., Hickey, A., McGee, H. and Joyce, C.R.B. 1996. *Manual for the SEIQoL-DW*. Dublin: Department of Psychology, Royal College of Surgeons in Ireland

World Health Organization Quality of Life Measure – Brief Version

The WHOQOL Group. 1996. *WHOQOL-BREF: Introduction, Administration, Scoring and Generic Version of the Assessment*. *Field Trial Version*. Geneva: World Health Organization

---

References 5 Additional articles pertaining to the 15 instruments developed in the last 10 years and used fewer than three times

Archbold, P., Stewart, B., Greenlick, M.R. and Horwath, T. 1990. Mutuality and preparedness as predictors of caregiver role strain. *Research in Nursing and Health* **13**: 375–84

Barrera, M. 1980. A method for the assessment of social support in community survey research. *Connections* **3**: 8–13

Beck, A.T. and Beck, R.W. 1972. Screening depressed patients in family practice: a rapid technique. *Postgraduate Medicine* **52**: 81–5

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J. and Erbaugh, J. 1961. An inventory for measuring depression. *Archives of General Psychiatry* **4**: 561–71

Bengtson, V.L. and Kuypers, J. 1985. The family support cycle: psychosocial issues in the aging family. In Munnichs, J.M.A., Olbrich, E., Mussen, P. and Coleman, P.G. (eds), *Life-span and change in a gerontological perspective*: 257–73. New York: Academic Press

Bengtson, V., Rosenthal, C. and Burton, L. 1990. Families and aging: diversity and heterogeneity. In Binstock, R.H. and George, L.K. (eds), *Handbook of aging and the social sciences*: 263–387. Boston, MA: Academic Press

Bowling, A. 1996. The effects of illness on quality of life: findings from a survey of households in Great Britain. *Journal of Epidemiology and Community Health* **50**: 149–55

Browne, J.P., O'Boyle, C.A., McGee, H.M., McDonald, N.J. and Joyce, C.R.B. 1997. Development of a direct weighting procedure for quality of life domains. *Quality of Life Research* **6**: 301–9

Dautzenberg, M.G.H., Diederiks, J.P.M., Philipsen, H. and Stevens, F.C.J. 1996. Women of a middle generation and informal care for the elderly. *Tijdschrift voor gerontologie en geriatrie* **27**: 141–8

Derogatis, J. 1980. *The Symptom Checklist - 90*. Baltimore, MD: John Hopkins University Press

Doornbos, M. 2002. Predicting family health in families of young adults with severe mental illness. *Journal of Family Nursing* **8**: 241–63

Fang, C.T., Hsiung, P.C., Yu, C.F., Chen, M.Y. and Wang, J.D. 2002. Validation of the World Health Organization quality of life instrument in patients with HIV infection. *Quality of Life Research* **11**: 753–62

Farran, C.J., Keane-Hagerty, E., Salloway, S., Kupferer, S. and Wilken, C.S. 1991. Finding meaning: an alternate paradigm for Alzheimer's disease family caregivers. *Gerontologist* **31**: 483–9

Fung, W. and Chien, W. 2002. The effectiveness of a mutual support group for family caregivers of a relative with dementia. *Archives of Psychiatric Nursing* **16**: 134–44

Furst, M.E.G. 1996. *Indication for Admission into a Psychogeriatric Nursing Home*. Nijmegen: University of Nijmegen

Gerritsen, J.C. and Van der Ende, P.C. 1994. The development of a care-giving burden scale. *Age and Aging* **23**: 483–91

Gilleard, C.J., Boyd, W.D. and Watt, G. 1982. Problems in caring for the elderly mentally infirm at home. *Archives of Gerontology and Geriatrics* **1**: 151–8

Grasel, E. 1995. Somatic symptoms and caregiving strain among family caregivers of older patients with progressive nursing needs. *Archives of Gerontology and Geriatrics* **21**: 253–66

Grasel, E. 2002. When home care ends – changes in the physical health of informal caregivers caring for dementia patients: a longitudinal study. *Journal of the American Geriatrics Society* **50**: 843–9

Greenberg, J.S., Greenley, J.R. and Benedict, P. 1994. Contributions of persons with serious mental illness to their families. *Hospital and Community Psychiatry* **45**: 475–80

Greene, J.G., Smith, R., Gardiner, M. and Timbury, G.C. 1982. Measuring behavioural disturbance of elderly demented patients in the community and its effects on relatives: a factor analytic study. *Age and Aging* **11**: 121–6

Hickey, A.M., Bury, G., O'Boyle, C.A., Bradley, F., O'Kelly, F.D. and Shannon, W. 1996. A new short form individual quality of life measure (SEIQoL-DW):

application in a cohort of individuals with HIV/AIDS. *British Medical Journal* **313**: 29–33

Higginson, I.J. and McCarthy, M. 1993. Validity of the support team assessment schedule: do staffs' ratings reflect those made by patients or their families? *Palliative Medicine* **7**: 219–28

Hodgson, C., Higginson, I. and Jefferys, P. 1997. *Developing outcome measures in community services for people with dementia and their carers*. Report to the Mental Health Foundation

Horton-Deutsch, S.L., Farran, C.J., Choi, E.E. and Fogg, L. 2002. The PLUS Intervention: a pilot test with caregivers of depressed older adults. *Archives of Psychiatric Nursing* **16**: 61–71

Idler, E.L. and Kasl, S.V. 1991. Health perceptions and survival: do global evaluations of health status really predict mortality? *Journal of Gerontology: Social Sciences* **46**: 555–65

Jalowiec, A. 1988. Confirmatory factor analysis of the Jalowiec Coping Scale. In Waltz, C. and Strickland, O. (eds), *Measurement of nursing outcomes: client outcomes*: 287–308. New York: Springer Publishing Company

Joyce, C.R.B., Hickey, A., McGee, H.M. and O'Boyle, C.A. 2003. A theory-based method for the evaluation of individual quality of life: The SEIQoL. *Quality of Life Research* **12**: 275–80

Kaufer, D.I., Cummings, J.L., Christine, D., Bray, T., Castellon, S., Maasterman, D. *et al.* 1998. Assessment of the impact of neuropsychiatric symptoms in Alzheimer's disease: Neuropsychiatric Inventory Caregiver Distress Scale. *Journal of the American Geriatrics Society* **46**: 210–15

Kemp, B.J., Corgiat, M. and Gill, C. 1992. Effects of brief cognitive-behavioural group psychotherapy on older persons with and without disabling illness. *Behaviour, Health and Aging* **2**: 21–8

Kosberg, J.I., Cairl, R.E. and Keller, D.M. 1990. Components of burden: interventive implications. *Gerontologist* **30**: 236–42

Lawton, M.P., Kleban, M.H., Moss, M., Rovine, M. and Glicksman, A. 1989. Measuring caregiving appraisal. *Journal of Gerontology: Psychological Sciences* **44**: 61–71

Lelliott, P., Beevor, A., Hogman, G., Hyslop, J., Lathlean, J. and Ward, M. 1999. *The CUES Project: Carers' and users' expectations of services. A part of the Department of Health's OSCA initiative (Outcomes of Social Care for Adults)*. Final Report to the Department of Health

Magana, A.B., Goldstein, M.J., Karno, M., Miklowitz, D.J., Jenkins, J. and Falloon, I. 1986. A brief method for assessing expressed emotion in relatives of psychiatric patients. *Psychiatry Research* **17**: 203–12

Moons, P., Marquet, K., Budts, W. and De Geest, S. 2004. Validity, reliability and responsiveness of the 'Schedule for the Evaluation of Individual Quality of Life - Direct Weighting' (SEIQoL-DW) in congenital heart disease. *Health and Quality of Life Outcomes* **2**: 27–34

Niederehe, G. 1988. TRIMS Behavioral Problem Checklist (BPC). *Psychopharmacological Bulletin* **24**: 771–8

Niederehe, G. and Fruge, E. 1984. Dementia and family dynamics: clinical research issues. *Journal of Geriatric Psychiatry* **17**: 21–60

Nolan, M., Grant, G. and Keady, J. 1996. *Understanding Family Care: a Multidimensional Model of Caring and Coping*. Buckingham: Open University Press

Nolan, M., Grant, G. and Keady, J. 1998. *Assessing the needs of family carers: a guide for practitioners*. Bangor: University of Wales, Bangor

Nolan, M.R., Grant, G. and Ellis, N.C. 1990. Stress is in the eye of the beholder: reconceptualizing the measurement of carer burden. *Journal of Advanced Nursing* **15**: 544–55

O'Boyle, C.A., McGee, H.M., Hickey, A., Joyce, C.R.B., Brown, J. and O'Malley, K. 1993. *The schedule for the evaluation of individual quality of life. User manual*. Dublin: Department of Psychology, Royal College of Surgeons in Ireland

O'Boyle, C.A., McGee, H.M. and Browne, J.P. 2000. Measuring response shift using the Schedule for Evaluation of Individual Quality of Life. In Schwartz, C.E. and Sprangers, M.A.G. (eds), *Adaptation to changing health. Response shift in quality-of-life research*: 123–36. Washington DC: American Psychological Association

O'Carroll, R.E., Smith, K., Couston, M., Cossar, J.A. and Hayes, P.C. 2000. A comparison of the WHOQOL-100 and the WHOQOL-BREF in detecting change in quality of life following liver transplantation. *Quality of Life Research* **9**: 121–4

Procidano, M.E. and Heller, K. 1983. Measures of perceived social support from friends and from family: three validation studies. *American Journal of Community Psychology* **11**: 1–24

Radloff, L.S. 1977. The CES-D Scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement* **1**: 385–401

Ranney, M.J. and Aranda, M.P. 2001. Factors associated with depressive symptoms among Latino family dementia caregivers. *Journal of Ethnic and Cultural Diversity in Social Work* **10**: 1–21

Reker, G.T. 1992. *The Life Attitude Profile (LAP) and the Life Attitude Profile-Revised (LAP-R)*. Unpublished manuscript: Trent University

Reker, G.T., Peacock, E.J. and Wong PTP 1987. Meaning and purpose in life and well-being: A life-span perspective. *Journal of Gerontology: Psychological Sciences* **42**: 44–9

Rutter, M., Graham, P. and Yule, W. 1970. *A Neuropsychiatric Study in Childhood*. London: Heinemann

Sanders, C.M., Mauger, P.A. and Strong, P.N. 1985. *A manual for the Grief Experience Inventory*. Blowing Rock, NC: The Center for the Study of Separation and Loss

Scholte op Reimer, W.J.M., De Haan, R.J. and Pijnenborg, J.M.A. 1998. Assessment of burden in partners of stroke patients with the Sense of Competence Questionnaire. *Stroke* **29**: 373–9

Spielberger, C.D. 1983. *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press

Spielberger, C.D. 1988. *State-Trait Anger Expression Inventory: Research Edition*. Odessa, FL: Psychological Assessment Resources

Taylor, R.J. and Chatters, L.M. 1986. Church-based informal support among elderly blacks. *Gerontologist* **26**: 637–42

The WHOQOL Group 1994. Development of the WHOQOL: rationale and current status. *International Journal of Mental Health* **23**: 24–56

The WHOQOL Group 1998. Development of the World Health Organization WHOQOL-BREF Quality of Life Assessment. *Psychological Medicine* **28**: 551–8

Vernooij-Dassen, M. 1993. *Dementia and Home Care: Determinants of the Sense of Competence of Primary Caregivers and the Effect of Professionally Guided Caregiver Support*. Lisse: Swets and Zeitlinger

Vernooij-Dassen, M., Felling, A. and Persoon, J. 1996. Predictors of sense of competence in primary caregivers of demented persons. *Social Science Medicine* **43**: 41–9

Vernooij-Dassen, M., Kurz, X., Scuvee-Moreau, J. and Dresse, A. 2003. The measurement of sense of competence in caregivers of patients with dementia. *Revue d'épidémiologie et de santé publique* **51**: 227–35

Waldron, D., O'Boyle, C.A., Kearney, M., Moriarty, M. and Carney, D. 1999. Quality-of-life measurement in advanced cancer: assessing the individual. *Journal of Clinical Oncology* **17**: 3603–11

Yesavage, J.A., Brink, T.L., Rose, T.L., Lum, O., Huang, V., Adey, M. and Leirer, O. 1983. Development and validation of a geriatric depression screening scale: a preliminary report. *Journal of Psychiatric Research* **17**: 37–49

Zarit, S.H., Reever, K.E. and Bach-Peterson, J. 1980. Relatives of the impaired elderly: Correlates of feelings of burden. *Gerontologist* **20**: 649–55

Zarit, S.H. and Zarit, J.M. 1986. *The Memory and Behavior Problem Checklist 1987R and the Burden Interview: Instruments and instructions manual*. University Park, PA: Pennsylvania State University

Zeiss, A.M., Gallagher-Thompson, D., Lovett, S., Rose, J. and McKibbin, C. 1999. Self-efficacy as a mediator of caregiver coping: development and testing of an assessment model. *Journal of Clinical Geropsychology* **5**: 221–30

Zuckerman, M. and Lubin, B. (1965) *Manual for the Multiple Affect Adjective Checklist*. San Diego, CA: Edits

References 6 Articles pertaining to ineligible instruments

Barrowclough, C., Lobban, F., Hatton, C. and Quinn, J. 2001. An investigation of models of illness in carers of schizophrenia patients using the Illness Perception Questionnaire. *British Journal of Clinical Psychology* **40(4)**: 371–85

Baumgarten, M., Battista, R.N., Infante-Rivard, C., Hanley, J.A., Becker, R. and Gauthier, S. 1992. The psychological and physical health of family members caring for an elderly person with dementia. *Journal of Clinical Epidemiology* **45**: 61–70

Bowers, S.P. 2000. Gender role identity and the caregiving experience of widowed men. *Sex Roles* **41(9)**: 655

Brennan, P.F., Moore, S.M. and Smyth, K.A. 1995. The effects of a special computer network on caregivers of persons with Alzheimer's disease. *Nursing Research* **44**: 166–72

Elmståhl, S., Malmberg, B. and Annerstedt, L. 1996. Caregiver's burden of patients 3 years after stroke assessed by a novel caregiver burden scale. *Archives of Physical Medicine and Rehabilitation* **77**: 177–82

Fuller-Jonap, F. and Haley, W.E. 1995. Mental and physical health of male caregivers of a spouse with Alzheimer's disease. *Journal of Aging and Health* **7**: 99–118

Goode, K.T., Haley, W.E., Roth, D.L. and Ford, G.R. 1998. Predicting longitudinal changes in caregiver physical and mental health: a stress process model. *Health Psychology* **17**: 190–8

Goodman, C.C. and Pynoos, J. 1990. A model telephone information and support program for caregivers of Alzheimer's patients. *Gerontologist* **30**: 399–404

Haley, W.E., Levine, E.G., Brown, S.L., Berry, J.W. and Hughes, G.H. 1987. Psychological, social, and health consequences of caring for a relative with senile dementia. *Journal of the American Geriatrics Society* **35**: 405–11

Haley, W.E., Roth, D.L., Coleton, M.I., Ford, G.R., West, C.A. and Collins, R.P. 1996. Appraisal, coping, and social support as mediators of well-being in black and white family caregivers of patients with Alzheimer's disease. *Journal of Consulting and Clinical Psychology* **64**: 121–9

Hobbs, T.R. 1997. Depression in the caregiving mothers of adult schizophrenics: a test of the resource deterioration model. *Community Mental Health Journal* **33**: 387–99

Kaye, L.W. and Applegate, J.S. 1990. Men as elder caregivers: a response to changing families. *American Journal of Orthopsychiatry* **60**: 86–95

McMillan, S.C. and Mahom, M. 1994. The impact of hospice services on the quality of life of primary caregivers. *ONF* **21(7)**: 1189–95

Morrissey, E., Becker, J. and Rubert, M.P. 1990. Coping resources and depression in the caregiving spouses of Alzheimer patients. *British Journal of Medical Psychology* **63**: 161–71

Owen, J.E. 2000. Association of life events and psychological distress in family caregivers of dementia patients. *Aging and Mental Health* **6(1)**: 62–71

Provencher, H.L. and Mueser, K.T. 1997. Positive and negative symptom behaviours and caregiver burden in the relatives of persons with schizophrenia. *Schizophrenia Research* **26**: 71–80

Rauktis, M.E., Koeske, G.F. and Tereshko, O. 1995. Negative social interactions, distress, and depression among those caring for a seriously and persistently ill relative. *American Journal of Community Psychology* **23(2)**: 279–99

Roth, D.L., Haley, W.E., Owen, J.E., Clay, O.J. and Goode, K.T. 2001. Latent growth models of the longitudinal effects of dementia caregiving: a comparison of African American and White family caregivers. *Psychology and Aging* **16**: 427–36

Stephens, M.A., Norris, V.K., Kinney, J.M., Ritchie, S.W. and Grotz, R.C. 1988. Stressful situations in caregiving: relations between caregiver coping and well-being. *Psychology and Aging* **3**: 208–9

Struening, E.L., Stueve, A., Vine, P., Kreisman, D.E., Link, B.G. and Herman, D.B. 1995. Factors associated with grief and depressive symptoms in caregivers of people with serious mental illness. *Research in Community and Mental Health* **8**: 91–124

Appendices

Appendix 1 Topic guide used in interviews

Topic guide for use with in one-to-one interviews and focus groups with carers

A	***Tell me about your caregiving experience*** Who do you care for? How long have you been caring for? What sorts of things do you do? Do you share responsibility for caring?
B	***How do your caregiving responsibilities affect your life?*** Work, social life, physical health, emotional/mental health, finance/income, household routine, family/household interaction, schooling/education of children, children generally, interaction with others outside the household/family.
C	***What support from services does the care recipient receive?*** Who does he/she see from the CMHT (psychiatrist, CPN, psychologist, occupational therapist, social worker)? Does he/she go to a day centre or a drop-in centre? ***What support do you receive from services?*** Do you see anyone from the CMHT? A carer-support worker? Do you have contact with an organisation like Rethink? ***What should support services for carers be offering/provide? What do you want out of a support service?***
D	***What outcomes should support services be evaluating?*** Considering the things you mentioned earlier about the way in which your caregiving responsibilities affect your life, what sorts of things should be asked about when evaluating a service? Prompt using list of outcomes. ***What would be the most important outcome?***

E	***Have you ever evaluated a support service?*** Have you ever been asked by professionals/services providers about your experiences of support services? Have you been asked how satisfied you are with a support service? Have you been asked to fill in any forms, or has someone come to talk to you about services? ***If yes, what were you asked?*** Were you asked to evaluate the right sorts of things? ***Did you receive any feedback from the evaluation?*** ***When services for carers are introduced, is it important that they are evaluated?***
F	***How should services be evaluated?*** By questionnaire/semi-structured interview/unstructured interview/ face-to-face/telephone/post? ***If by questionnaire, how long should it be? How should it be worded/structured?*** ***Who should evaluate services?*** A researcher? Service providers? ***When should services be evaluated?***

Topic guide for use with representatives from carer organisations

A	***Tell me about your experiences with carers both from your personal experience and through your role in the carer organisation.*** How much contact do you have with carers? What sorts of things to the carers do? How do the carers care for? How long have they been caring? What sorts of things do you do?
B	***How do you think caregiving responsibilities affect carers' lives?*** Work, social life, physical health, emotional/mental health, finance/income, household routine, family/household interaction, schooling/education of children, children generally, interaction with others outside the household/family.
C	***In your experience what are the most common types of support that care recipients receive from services?*** Seeing someone from the CMHT (psychiatrist, CPN, psychologist, occupational therapist, social worker)? Do they go to day centres or a drop-in centres? ***In your experience what types of support do carers receive from services?*** Anyone from the CMHT? A carer-support worker? An organisation like Rethink? ***What should support services for carers be offering/provide? What do carers want out of a support service?***

D	***What outcomes should support services be evaluating?*** Considering the things you mentioned earlier about the way in which caregiving responsibilities affect carers, what sorts of things should be asked about when evaluating a service? Prompt using list of outcomes. ***What would be the most important outcome?***
E	***Are you aware of any evaluations of services for carers?*** Are you aware of service providers asking carers about their experiences with a support service? Have these involved asking carers how satisfied they are? Have they involved asking carers to fill in any forms, or someone talking to carers to get feedback on the services they have received? ***If yes, what was asked?*** Were these the right sorts of things? ***Did the carers receive any feedback from the evaluation?*** ***When services for carers are introduced, is it important that they are evaluated?***
F	***How should services be evaluated?*** By questionnaire/semi-structured interview/unstructured interview/ face-to-face/telephone/post? ***If by questionnaire, how long should it be? How should it be worded/structured?*** ***How should participants be recruited?*** Randomly? Support groups? ***Who should evaluate services?*** A researcher? Service providers? ***When should services be evaluated?***

Topic guide for use with service providers

A	***Tell me about your experiences with carers both from your personal experience and through your role as a service provider.*** How much contact do you have with carers? What sorts of things to the carers do? How do the carers care for? How long have they been caring? What sorts of things do you do?

B	***In your experience what are the most common types of support that care recipients receive from services?*** Seeing someone from the CMHT (psychiatrist, CPN, psychologist, occupational therapist, social worker)? Do they go to day centres or a drop-in centres? ***In your experience what types of support do carers receive from services?*** Anyone from the CMHT? A carer-support worker? An organisation like Rethink? What should support services for carers be offering/provide? What do carers want out of a support service?
C	***What outcomes should support services be evaluating?*** Prompt using list of outcomes. ***What would be the most important outcome?*** As a service provider, do you think your view about the outcomes that should be evaluated will differ from carers' views?
D	***Are you aware of any evaluations of services for carers?*** Are you aware of service providers asking carers about their experiences with a support service? Have these involved asking carers how satisfied they are? Have they involved asking carers to fill in any forms, or someone talking to carers to get feedback on the services they have received? ***If yes, what was asked?*** Were these the right sorts of things? ***Did the carers receive any feedback from the evaluation?*** ***When services for carers are introduced, is it important that they are evaluated?***
E	***How should services be evaluated?*** By questionnaire/semi-structured interview/unstructured interview/ face-to-face/telephone/post? ***If by questionnaire, how long should it be? How should it be worded/structured?*** ***How should participants be recruited?*** Randomly? Support groups? ***Who should evaluate services?*** A researcher? Service providers? ***When should services be evaluated?***
F	***How do you think services providers can evaluate their services for carers?*** ***What are the challenges of evaluating a service?***

Appendix 2 List of outcomes used in interviews as a prompt

The effect caregiving has on your family life	The effect caregiving has on your daily life
The sorts of caregiving activities that you do	The effect caregiving has on your social life
Your well-being	Your mental health (e.g. whether you are depressed or anxious)
Your ability to cope	Your physical health
Your quality of life	How you feel about caregiving
Amount of time spent providing care	How burdened you feel by caregiving
Your satisfaction with the service	Your confidence in providing care
Your relationship with the person you care for	Your knowledge of the person's illness

Appendix 3 Questionnaire used in interviews

Thank you for agreeing to participate in this study. In addition to the information we collect during the interview, we would like to ask you some additional questions. This is so that we know something about the sorts of people who have taken part in the study.

The information you give us will be kept confidential. Please feel free to ask the researcher any questions you have.

Some questions about you

1 What is your date of birth?

day		month		year			
				1	9		

2 What sex are you? (please tick one box)

male ☐ 0 female ☐ 1

3 What is your ethnicity? (please tick one box)

White ☐ 1
Black-Caribbean ☐ 2
Black-African ☐ 3
Black-Other ☐ 4
Indian ☐ 5
Pakistani ☐ 6
Bangladeshi ☐ 7
Chinese ☐ 8
something else ☐ 9

4 Are you a <u>member</u> of a carer support group e.g. Rethink, Alzheimer's Disease Society? (please tick one box)

no ☐ 0 yes ☐ 1

5 Do you work (paid or voluntary) for a carer support group? (please tick one box)

no ☐ 0 yes ☐ 1

6 Do you care for a relative or friend with a mental illness? (please tick one box)

no ☐ 0 yes ☐ 1

If yes, please answer the questions overleaf.

If no, please finish here and thank you for completing the questionnaire.

Some questions about the person you care for

7 What relationship are you to the person with mental illness? (please tick one box)

I am the person's…

mother/father	☐	1	girlfriend/boyfriend	☐	5
daughter/son	☐	2	friend	☐	6
sister/brother	☐	3	neighbour	☐	7
wife/husband/partner	☐	4	something else	☐	9

8 Does the person with mental illness usually live with you? (please tick one box)

no ☐ 0 yes ☐ 1

9 During the last four weeks, how often have you <u>seen</u> the person with mental illness? (please tick one box)

every day in the last four weeks	☐	6
nearly every day in the last four weeks	☐	5
more than 4 times (but not every day) in the last four weeks	☐	4
4 times in the last four weeks (about once a week)	☐	3
2 or 3 times in the last four weeks	☐	2
once in the last four weeks	☐	1
I haven’t seen the person in the last four weeks	☐	0

10 During the last four weeks, how much <u>telephone contact</u> have you had with the person with mental illness? (please tick one box)

every day in the last four weeks	☐	6
nearly every day in the last four weeks	☐	5
more than 4 times (but not every day) in the last four weeks	☐	4
4 times in the last four weeks (about once a week)	☐	3
2 or 3 times in the last four weeks	☐	2
once in the last four weeks	☐	1
none in the last four weeks	☐	0

11 Are you the person’s <u>main</u> carer? (please tick one box)

no ☐ 0 yes ☐ 1

Thank you for completing this questionnaire.

Appendix 4 Conference questionnaire

Which carer outcomes should professionals assess?

Family and friends often make a great contribution to the care of people with mental health problems, and the Government has recognised that they need more support. Services for carers are being developed, but to ensure that carers are offered the right sort of support it is important that services are evaluated.

To evaluate services for carers, we need to know what **outcomes** to assess. Outcomes typically assessed include carers' physical or mental health, the types of caregiving tasks carers undertake, or carers' satisfaction with services. The problem is that until now, carers themselves have not been asked what outcomes they think are most important.

The purpose of this study is to find out from carers and representatives from support organisations which outcomes we should assess when evaluating services. By completing this questionnaire you will help us do that. The questionnaire should only take you a few minutes to complete, and it is anonymous. Once you have completed the questionnaire please return it to conference staff along with your evaluation form of the Rethink Conference. If you would prefer to complete this questionnaire at a later date, please post it to Amy Langman at St. George's Hospital Medical School, Department of Psychiatry, Jenner Wing, Cranmer Terrace, London SW17 0RE.

If you have any questions about this project please telephone Amy Langman on 020 8725 3522.

Some questions about your caregiving role

1 Are you a <u>member</u> of a carer support group e.g. Rethink, Alzheimer's Disease Society, Carers National Association? (please tick one box)

no ☐ yes ☐

2 Do you <u>work</u> (paid or voluntary) for a carer support group? (please tick one box)

no ☐ yes ☐

3 Do you care for a relative or friend with a mental illness? (please tick one box)

no ☐ yes ☐

If yes, please continue with question 4.

If no, please go to question 8.

Some questions about the person you care for

4 What relationship are you to the person with mental illness? (please tick one box)

I am the person's ……

mother/father	☐	girlfriend / boyfriend	☐
daughter / son	☐	friend	☐
sister/brother	☐	neighbour	☐
wife/husband/partner	☐	other	☐

5 What mental illness does the person you care for have (for example Alzheimer's disease, schizophrenia, depression, dementia)? (please write below)

6 Does the person with mental illness usually live with you? (please tick one box)

no ☐ yes ☐

7 How many years have you been caring for the person with mental illness? (please write your answer in the boxes)

☐☐ years

What outcomes do you think should be assessed?

8 Imagine a new service aimed at supporting carers has been introduced in your area. You have been receiving this service for three months. The service providers are keen to evaluate how good (or bad) their new service is at supporting you. What sorts of things should they ask you about?

Please read the whole list (they are not in any order) and tick all the things that you think it would be important for the service-providers to assess.

☐ The sorts of caregiving activities that you do
☐ Your satisfaction with the service
☐ Your quality of life
☐ The amount of time that you spend providing care
☐ The effect caregiving has on your daily life
☐ The effect caregiving has on your social life
☐ Your satisfaction with life
☐ The effect caregiving has on your family life
☐ Whether you are depressed
☐ How you feel about caregiving
☐ Your relationship with the person you care for
☐ How burdened you feel by caregiving
☐ Your physical health
☐ Your confidence in providing care
☐ Your knowledge of the person's illness
☐ Whether you are anxious
☐ Your morale
☐ Your needs
☐ Your ability to cope

9 Now read the list again, and tick the <u>one</u> thing you think it is <u>most</u> important for the service providers to assess.

☐ The sorts of caregiving activities that you do
☐ Your satisfaction with the service
☐ Your quality of life
☐ The amount of time that you spend providing care
☐ The effect caregiving has on your daily life
☐ The effect caregiving has on your social life
☐ Your satisfaction with life
☐ The effect caregiving has on your family life
☐ Whether you are depressed
☐ How you feel about caregiving
☐ Your relationship with the person you care for
☐ How burdened you feel by caregiving
☐ Your physical health
☐ Your confidence in providing care
☐ Your knowledge of the person’s illness
☐ Whether you are anxious
☐ Your morale
☐ Your needs
☐ Your ability to cope

10 What things not on the list do you think service providers should assess? (please write them below)

____

____

____

____

11 What is the best way for service-providers to evaluate the services they are offering carers? (please tick one box)

By interviewing the carer face-to-face ☐

By interviewing carers in groups ☐

By asking carers to complete a questionnaire ☐

A few more questions about you

The following questions will help us know whether a variety of people have completed the questionnaire.

12 What is your date of birth?

day		month		year			
				1	9		

13 What sex are you? (please tick one box)

male ☐ female ☐

14 What is your ethnicity? (please tick one box)

White ☐

Black-Caribbean ☐

Black-African ☐

Black-Other ☐

Indian ☐

Pakistani ☐

Bangladeshi ☐

Chinese ☐

Other ☐

15 Please write anything else you think is relevant below.

....

....

....

....

....

....

Thank you for completing this questionnaire.

Appendix 5 Literature review search terms

Medline search terms

1 caregivers.mp. or CAREGIVERS/

2 (care$ adj giv$).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]

3 carer$.mp.

4 (informal$ adj care$).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]

5 befriend$.mp.

6 (home$ adj2 care$).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]

7 ((famil$ or neighbour$ or relative$) adj2 (care$ or caring or support)).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]

8 ((famil$ or neighbour$ or relative$) adj2 (care$ or caring or support)).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]

9 ((famil$ or neighbour$ or relative$) adj2 (care$ or caring or support)).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]

10 ((famil$ or neighbour$ or relative$) adj2 (care or caring or support)).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]

11 ((famil$ or neighbour$ or relative$) adj (care or caring or support)).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]

12 ((famil$ or neighbour$ or relative$) adj (care or caring or support)).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]

13 ((famil$ or neighbour$ or relative$) adj (care or caring or support)).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]

14 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13

15 Mental Disorders/

16 exp Mental Health/

17 (psych$ adj2 (disorder$ or problem$)).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]

18 Alzheimer Disease/ or Dementia/ or dement$.mp.

19 Schizophrenia/ or schizophren$.mp.

20 alzheimer$.mp.

21 (mental$ adj2 disease).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]

22 (psych$ adj2 ill$).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]

23 (mental$ adj2 impair$).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]

24 (eating adj2 disorder$).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]

25 (cognitive adj2 disorder$).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]

26 (mood adj2 disorder$).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]

27 (anxiety adj2 disorder$).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]

28 personality disorder$.mp. or Personality Disorders/

29 anorexi$.mp. [mp=title, abstract, cas registry/ec number word, mesh subjectheading]

30 bulimi$.mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]

31 exp substance related disorders/

32 hysteri$.mp.

33 hysteri$.mp.

34 neuro$.mp.

35 psycho$.mp.

36 (bipolar adj2 disorder$).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading]

37 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36

38 14 and 37

PsycINFO Search Terms

1 caregiver burden/ or caregivers/

2 limit 1 to english language

3 caregiver.mp. [mp=title, abstract, heading word, table of contents, key concepts]

4 limit 3 to english language

5 (care$ adj giv$).mp. [mp=title, abstract, heading word, table of contents, key concepts]

6 5

7 limit 6 to english language

8 carer$.mp.

9 8

10 limit 9 to english language

11 informal$ care$.mp.

12 limit 11 to abstracts

13 11

14 limit 13 to english language

15 befriend$.mp.

16 limit 15 to english language

17 (home adj2 care).mp. [mp=title, abstract, heading word, table of contents, key concepts]

18 limit 17 to english language

19 ((famil$ or neighbour$ or relative$) adj2 (care or caring or support)).mp. [mp=title, abstract, heading word, table of contents, key concepts]

20 limit 19 to english language

21 2 or 4 or 7 or 10 or 14 or 16 or 18 or 20

22 mental disorders.mp. or exp Mental Disorders/

23 22

24 limit 23 to english language

25 mental health.mp. or exp Mental Health/

26 limit 25 to english language

27 (psych$ adj2 (disorder$ or problem$)).mp. [mp=title, abstract, heading word, table of contents, key concepts]

28 limit 27 to english language

29 exp Senile Dementia/ or exp "Depression (Emotion)"/ or exp Alzheimers Disease/ or dement$.mp.

30 limit 29 to english language

31 schizophren$.mp. [mp=title, abstract, heading word, table of contents, key concepts]

32 limit 31 to english language

33 alzheimer$.mp. [mp=title, abstract, heading word, table of contents, key concepts]

34 limit 33 to english language

35 (mental$ adj2 disease).mp. [mp=title, abstract, heading word, table of contents, key concepts]

36 limit 35 to english language

37 (psych$ adj2 ill$).mp. [mp=title, abstract, heading word, table of contents, key concepts]

38 limit 37 to english language

39 (mental$ adj2 impair$).mp. [mp=title, abstract, heading word, table of contents, key concepts]

40 limit 39 to english language

41 (eating adj2 disorder$).mp. [mp=title, abstract, heading word, table of contents, key concepts]

42 limit 41 to english language

43 exp Mental Disorders/ or exp Schizophrenia/ or cognitive disorder$.mp.

44 limit 43 to english language

45 exp Schizophrenia/ or cognitive disorder.mp.

46 limit 45 to english language

47 (mood adj2 disorder$).mp. [mp=title, abstract, heading word, table of contents, key concepts]

48 limit 47 to english language

49 (anxiety adj2 disorder$).mp. [mp=title, abstract, heading word, table of contents, key concepts]
50 limit 49 to english language
51 (personality adj2 disorder$).mp. [mp=title, abstract, heading word, table of contents, key concepts]
52 limit 51 to english language
53 anorexi$.mp. or exp Anorexia Nervosa/
54 limit 53 to english language
55 exp Bulimia/ or bulimi$.mp.
56 limit 55 to english language
57 exp Drug Rehabilitation/ or exp Drug Abuse/ or exp Affective Disorders/ or substance related disorders.mp.
58 limit 57 to english language
59 hysteri$.mp.
60 limit 59 to english language
61 exp Addiction/ or neuro$.mp.
62 limit 61 to english language
63 psycho$.mp.
64 limit 63 to english language
65 (bipolar adj2 disorder$).mp. [mp=title, abstract, heading word, table of contents, key concepts]
66 limit 65 to english language
67 24 or 26 or 28 or 30 or 32 or 34 or 36 or 38 or 40 or 42 or 46 or 48 or 50 or 52 or 54 or 58 or 60 or 62 or 64 or 66
68 cancer.mp. or exp Neoplasms/
69 limit 68 to english language
70 (21 and 67) not 69
71 limit 70 to yr=1998–2003
72 limit 71 to journal articles

Appendix 6 Ineligible instruments

The first set of parentheses gives the original author (not referenced) and the second set of parentheses gives the paper(s) using the measure that were found in this review (see References 6).

Bem Sex Role Inventory (Bem) (Bowers, 2000; Fuller-Jonap and Haley, 1995; Kaye and Applegate, 1990) N.B. on inspection, not an outcome measure.

Caregiving Impact Scale (Poulshock and Deimling) (Brennan *et al.*, 1995; Goodman and Pynoos, 1990; Haley *et al.*, 1987; Stephens *et al.*, 1988) N.B. Authors requested this measure not be included in this review.

Cornell Medical Index (Brodman *et al*.) (Fuller-Jonap and Haley, 1995; Goode *et al.*, 1998; Haley *et al.*, 1996; Morrissey *et al.*, 1990; Roth *et al.*, 2001) N.B. Measure no longer available.

Louisville Older Person Events Scale (Murrell) (Baumgarten *et al.*, 1992; Hobbs, 1997; Owen, 2000) N.B. On inspection, not an outcome measure.

Disclaimer: