ARTICOLE ORIGINALE

TOWARDS HARMONISATION OF CAREGIVER OUTCOME MEASURES


On behalf of the Medical and Scientific Panel of Alzheimer’s Disease International Working Group on Consensus Guidelines for Caregiver Research

WHY DO WE NEED HARMONISATION?

Unlike standard measures of cognition or activities of daily living employed in dementia research, outcome measures used in caregiver intervention studies are far from uniform. Because of the variety in the measurement of intervention effectiveness, comparison between studies is difficult except by statistical manipulation. An international group of caregivers and experts in the field of caregiver research met at the World Alzheimer Congress in Washington D.C. in July 2000. Their aim was to seek agreement on issues of importance in caregiver research, and to achieve harmonisation with regard to which caregiver outcomes should be assessed and how these should be measured. Subsequent to the symposium, a questionnaire was sent to the participants to finalise the minimum data set of recommended caregiver outcome measures. The task, far more complex than initially realised, will require further discussions with caregivers and researchers separately and together. This paper presents the basis of a minimum data set for further refinement and discussion.

Keywords: caregiver, carer, harmonisation, outcome, assessment, measure, dementia.

Rezumat:


Cuvinte cheie: persoană de îngrijire, armonizare, rezultat, evaluare, măsurare, demență.

Abstract:

Unlike standard measures of cognition or activities of daily living employed in dementia research, outcome measures used in caregiver intervention studies are far from uniform. Because of the variety in the measurement of intervention effectiveness, comparison between studies is difficult except by statistical manipulation. An international group of caregivers and experts in the field of caregiver research met at the World Alzheimer Congress in Washington D.C. in July 2000. Their aim was to seek agreement on issues of importance in caregiver research, and to achieve harmonisation with regard to which caregiver outcomes should be assessed and how these should be measured. Subsequent to the symposium, a questionnaire was sent to the participants to finalise the minimum data set of recommended caregiver outcome measures. The task, far more complex than initially realised, will require further discussions with caregivers and researchers separately and together. This paper presents the basis of a minimum data set for further refinement and discussion.

Keywords: caregiver, carer, harmonisation, outcome, assessment, measure, dementia.

* Academic Dept for Old Age Psychiatry, Prince of Wales Hospital, Sydney, Australia
** Institute of Psychiatry, London, UK
*** Silberstein Ageing and Dementia Research Centre, New York University, NY, USA
**** University Centre for Social and Urban Research, University of Pittsburgh, PA, USA
***** Case Western Reserve University, Cleveland, Ohio, USA
****** Alzheimer's Society, London, UK
******* Pennsylvania State University, PA, USA
METHOD

With harmonisation in mind, a group of caregivers and experts in the field of caregiver research, representing different cultures and countries from Africa, Asia, Australia, Canada, Europe, North America, South America, South Africa and the United Kingdom, met at the World Alzheimer Congress in Washington D.C. in July 2000. Their aim was to seek agreement on issues of importance in caregiver research and to achieve agreement as regards to which caregiver outcomes should be assessed and how these should be measured. The concept of harmonisation, as defined by the International Committee on Harmonisation, implies “a group process working toward a mutually satisfying end, not a process wherein some were made to submit to rigid standards imposed by others” (2). This meeting utilised a ‘top down’ (from experts) and ‘bottom up’ (from caregivers) approach in order to reach consensus on important outcome measures in caregiver research. General issues pertinent to the field were discussed, the main points extracted for the purpose of guiding the choice of a minimum data set of essential outcome measures, and issues clarified by subsequent postal questionnaire. This paper aims to summarise the discussion and ensuing recommendations and present issues pertinent to caregiver evaluation research and ultimately achieving harmonization of caregiver outcome measures.

GENERAL ISSUES THAT EMERGED REGARDING OUTCOME MEASUREMENT

Situation-specific factors

Interventions cannot be viewed as ‘one size fits all’. The effectiveness of a particular intervention may be influenced by variables pertaining to the caregiver, the person with dementia and the study itself, such as the relationship between caregiver and person with dementia (spouse, adult-child, other – different roles and responsibilities), stage of dementia (different problems posed at early, middle, or late stage), context of the interventions (home, day care, nursing home), length of the intervention and the duration of follow-up (as there may be a latency before benefits become apparent and some effects may only be transitory). Caregivers may be providing care for more than one person (e.g. for a spouse and parent); this needs to be taken into account when measuring outcomes.

Choosing realistic outcomes

It is important to be realistic about what can and cannot be changed when selecting outcomes. Some outcomes represent the accumulation of a lifetime of exposures, habits or personality styles and are therefore very difficult to change in a brief intervention (e.g. chronic illness, self-esteem). Therefore, when selecting outcomes, it is important to be aware of how close (proximal) or far (distal) the primary outcome is to or from the goal of the intervention. For example, teaching caregivers how to cope with behavioural problems may reduce caregivers’ distress in response to aggression (a proximal measure) but is less likely to influence caregivers’ feelings of self-worth (a distal measure). A related issue is the need to match sufficiently specific and sensitive outcome measures to the expected effects of an intervention. While change in one domain (such as improved coping skills) may spill over into other domains (such as less depression), it is better to specify the most likely effects of treatment and to choose outcome measures accordingly.

Cross-cultural issues

Harmonisation of outcome measures would enable valid comparisons between regions, studies and cultures. One of the difficulties with cross-cultural research is the lack of clarity about whether observed differences are genuine or reflect cultural bias in what one is trying to measure (e.g. due to assumptions of literacy). Care is necessary lest the richness of caregiving in other cultures is lost by forcing its assessment into Western conceptualisations, for example by ignoring spirituality in quality of life measures. The concept of burden varies across cultures; in some regions, such as India, cultural norms may prevent caregivers from admitting to strain.

Caregiver-led research

In order to direct the course of future research, one of the aims of this symposium was to discover which outcome measures were important to caregivers. Contrary to the expectations of researchers, a reduction of psychological effects of caregiving such as depression or subjective burden were not mentioned by caregivers as important, who gave greater weight to practical, information and financial outcomes. Caregivers valued interventions that provided practical assistance to them, and those that led to an improvement in the quality of life of both the caregiver and the person with dementia. Surprisingly, delay of institutionalisation (or nursing home placement) of the person with dementia was not discussed, although often interventions that delay institutionalisation are extolled as positive outcomes (3,4). Attitudes toward institutionalization vary; keeping people at home is not necessarily the best outcome. A delay in institutionalisation might be a negative outcome if people on waiting lists want to be in care and do not want to be a burden. Perhaps the resolution should be to aim for caregiver-led research. In order to achieve harmonisation, we need to survey the experience of a wide variety of people and to canvass many diverse views. An allied approach is Goal Attainment Scales (5), where outcomes are tailored to the individual.
Dissemination of research findings to caregivers and clinicians

Caregivers identified as a priority the bridging of the gap between research and practice, between academics and clinicians, and between health care professionals and people with dementia and their caregivers. Similar arguments apply to information about services and resources. Caregivers stressed the importance of information as a means to empowerment, and expressed concern that without knowledge “sometimes out of the goodness of your own heart you may be doing the wrong thing”. Caregiver knowledge emerged as an important research outcome.

RECOMMENDED MINIMUM DATA SET OF OUTCOME MEASURES FOR CAREGIVER INTERVENTIONS

Basic information

Participants were in agreement that certain basic data should be reported with regard to the characteristics of the intervention used (description of content, number and duration of sessions, time span, size of groups, length of follow-up), the person with dementia (age, sex, marital status, ethnicity, years of education, main occupation, living arrangements, diagnosis criteria and source, severity of cognitive and functional impairment), and the caregiver (age, sex, marital status, ethnicity, years of education, main occupation, relationship to person with dementia, other caregiving responsibilities). Situation-specific information must be taken into account when assessing outcomes.

Specific outcome measures

A model of caregiver burden (6,7) provides a basis for identifying suitable outcomes to measure and facilitates the choice of appropriate proximal measures. This model shows that caregiver strain may fall into four categories – psychological, physical, financial and social – each with distinct outcome measures. In addition, there are a number of protective and exacerbating factors that can be identified and measured. References to and details of specific measures are provided in the text or the accompanying tables but not in both.

Psychological

The characteristics, reliability and validity of a selection of caregiver outcome instruments that measure burden and strain, quality of life, depression and other psychological morbidity, and their ability to address the general issues discussed above, are tabulated (Table 1).

Table 1: Recommended measures of psychological outcomes

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Description</th>
<th>Reliability</th>
<th>Validity</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burden Interview (BI) (19,20) USA</td>
<td>A 22 item inventory that measures the degree to which caregivers perceive their caregiving responsibilities as having an adverse effect on their health, personal and social life, psychological wellbeing, and finances. Each item is rated on a 5-point scale (ranging from ‘never’ to ‘nearly always present’). Total scores range from 0–88, with higher scores signifying higher levels of burden.</td>
<td>Internal reliability: Alpha of 0.91(21). A Japanese version of the BI had α =0.93, &amp; the single global rating of burden (Q22) on the BI and the sum of other BI items were significantly correlated (r=0.65)(22). Test re-test reliability is 0.71(21).</td>
<td>Construct validity: burden was negatively related to morale and positively related to hours spent giving care (23). CES-D and BI total were significantly correlated (r = 0.67).</td>
<td>This scale has been adapted for use in other cultures, e.g. Japan.</td>
</tr>
<tr>
<td>Screen for Caregiver Burden (SCB) (24) USA</td>
<td>A 25-item questionnaire designed to measure objective and subjective burden among spousal caregivers. Items assess various domains, including patient behaviours, disruptions in family and social life, and caregiver affective responses. SCB yields two scores, objective burden (prevalence of caregiving experiences) and subjective burden (ratings from 1 –4 of distress in relation of each experience). Subjective burden scores in excess of 42 considered ‘quite high’.</td>
<td>Internal reliability coefficients are 0.85 (OB) and 0.88 (SB).</td>
<td>Construct validity was supported by relationships of patient behavioural and cognitive functioning with OB and caregiver distress and personality variables with SB. Criterion validity (differences in burden between AD caregivers versus controls) was demonstrated by using age- and sex-matched controls. There were also relationships between increases in SB/OB and deterioration of patient functional/behavioural functioning and SB/OB increases in caregiver distress.</td>
<td>Sensitivity to change was examined by using a sample of individuals with early- to mid-stage AD: over 15–18 months a majority of caregivers changed by a value that was greater than would be expected by measurement error alone. This scale has been used successfully in Germany.</td>
</tr>
</tbody>
</table>
Quality of Life

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Description</th>
<th>Reliability</th>
<th>Validity</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfaction with Life Scale (SWLS) (25) USA</td>
<td>5-item questionnaire that measures global life satisfaction. Caregivers indicated their agreement with each of the 5 questions on a scale from 1 (strongly disagree) to 7 (strongly agree). The possible range of scores is 5 (low satisfaction) to 35 (high satisfaction).</td>
<td>The test-retest correlation coefficient was 0.82, and coefficient alpha was 0.87. Internal consistency was examined by calculating item-total correlations for the 5 items were 0.81, 0.63, 0.61, 0.75 and 0.66.</td>
<td>The correlation between the SWLS and the Cantril Self-Anchoring Striving Scale was found to be 0.62 in college students and 0.66 in an elderly sample.</td>
<td>Has been used in a variety of populations e.g. Iranians living in Norway, Arab Israelis, and Nepalese.</td>
</tr>
</tbody>
</table>

Depression

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Description</th>
<th>Reliability</th>
<th>Validity</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centre for Epidemiologic Studies Depression Scale (CES-D) (26) USA</td>
<td>20-item self-report of depressive symptoms rated according to frequency of occurrence, from 1 (rarely) to 5 (most of the time). Scores range from 0-60, with higher scores indicating more depressive symptoms. A score of 16 is the cut-off for designating subjects likely to be experiencing a significant level of depression.</td>
<td>Has been shown to have a high internal consistency, with coefficient alphas of 0.90 for psychiatric patients and 0.85 for nonpsychiatric individuals</td>
<td>This is a widely used scale. Self-report structure is useful for research purposes. It has been used in a variety of cultural settings – Greece, Spain, Hong Kong, and in Puerto Rican - and African - Americans in USA.</td>
<td></td>
</tr>
<tr>
<td>Beck Depression Inventory (BDI) (27) USA</td>
<td>Self-report questionnaire of depressive symptoms (inc. mood, vegetative, and cognitive aspects of depression) presented in multiple -choice format. The caregiver chooses between four -statements that most appropriate to how they have been feeling over the past week. The maximum score is 63. Scores of 0 -9 indicate the normal range, 10 -15 indicate minimal depression, 16 -19 indicate mild -moderate depression, 20 -29 indicate moderate -severe depression, and 30 -63 indicate severe depression.</td>
<td>Test-retest reliability with 38 patients was above 0.90. Spearman - Brown reliability was 0.93 and internal consistency for test items 0.86 (28).</td>
<td>Concurrent validity coefficient with the Zung SDS is 0.79 in psychiatric patients and 0.54 in college students (29); with the Hamilton Rating Scale it is 0.82 in psychiatric patients. This is a well-known and very widely -used scale. Self-report structure is useful for research purposes. It has been used in countries such as Finland, Saudi Arabia, Malaysia and Taiwan.</td>
<td></td>
</tr>
<tr>
<td>Hamilton Rating Scale for Depression (HRS-D) (30) UK</td>
<td>17-item interviewer checklist with total scores ranging from 0-52, representative of the severity of depression. Scores higher than 14 indicate the presence of depressive symptomatology.</td>
<td>Inter-rater reliability = 0.90. 15 of the individual items had reliability scores which ranged from fair (0.40 to 0.59) to excellent (0.75 –1.00), items which fared poorly included paranoid symptoms, depersonalisation/derealisation, insight, obsessive/compulsive symptoms, hypochondriasis and guilt (31).</td>
<td>Factors: retarded depression, mixed and somatic symptoms and anxiety reaction. While this scale is very widely used (in countries such as Korea, Japan, Italy, Austria, Russia and Spain), it is clinician administered and thus not as amenable for research purposes as other depression scales.</td>
<td></td>
</tr>
</tbody>
</table>

Other Psychological Morbidity

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Description</th>
<th>Reliability</th>
<th>Validity</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Health Questionnaire (GHQ) (32) UK</td>
<td>Self-administered screening instrument to aid in the detection of non -psychotic psychiatric illness. There are several versions of the GHQ. The original scale consists of 60 items, but there are also 36-, 30-, 20-, and 12 -item versions. On the GHQ -30, where scores range from 0 -30, scores &gt; 4 are taken to indicate cases of probable psychiatric illness. Items include questions about psychological distress and</td>
<td>The test-retest reliability for the 60 -item is 0.76, for the 30 items is 0.77, for the 20 -item is 0.73, and for the 12 -item is 0.73.</td>
<td>Construct validity has been examined in both a GP setting and in a medical outpatient population. The coefficients for the GP population were 0.80 (60 item), 0.80 (30 -item), 0.77 (20 item) and 0.77 (12 item). The coefficients for the outpatient population were 0.77 (60 item ), and 0.72 (30-, 20- and 12 -item). Sensitivity in the GP population was 95.7% (60 item), 91.4% (30 item),</td>
<td>The GHQ is widely used in the rating of psychological distress. The shorter versions of the GHQ are commonly used for research purposes. It has been used in Japanese, Nigerian, Polish, Afghan and Greenland Inuit samples. Alternative scoring methods have been reported. The GHQ can be scored by</td>
</tr>
</tbody>
</table>
altered behaviour. For each item, caregivers are asked to compare their recent state with their usual state. Items are rated on a 4-point scale where the essence of the ratings are ‘not at all’, ‘same as usual’, ‘rather more than usual’, or ‘much more than usual’. An item is counted as being present if it is experienced ‘rather’ or ‘much’ ‘more than usual’.

<table>
<thead>
<tr>
<th>Positive and Negative Affect Scale (PANAS) (34) USA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contains ten indicators of positive affect (PA) (e.g. excited, enthusiastic, inspired) and 10 of negative affect (NA) (e.g. distressed, nervous, scared). Respondents rate each of the indicators on a scale that ranges from “1” (very slightly or not at all) to “5” (extreme). A higher score suggests a greater degree of positive or negative affect. Ratings can be gained with several different temporal instructions: how they felt (a) ‘right now (that is, at the present moment)’ (moment instructions); (b) ‘today’ (today); (c) ‘during the past few days’ (past few days); (d) ‘during the past week’ (week); (e) ‘during the past few weeks’ (past few weeks); (f) ‘during the past year’ (year); and (g) ‘in general, that is, on the average (general).</td>
</tr>
</tbody>
</table>

88.2% (20 item) and 93.5% (12 items). Sensitivity in the outpatient population was 80.6% (60 item), 64.5% (30 item & 20 item) and 74.2% (12 item). Specificity for the GP population was 87.8% (60 item), 87% (30 item), 86% (20 item) and 78.5% (12 item). Specificity for the outpatient population was 93.3% (60 item), 91.6% (30 item), 96.7% (20 item) and 95% (12 item).

Convergent validity: Correlations between the PANAS and other brief affect measures range from 0.76 to 0.92. Correlation between the Hopkins Symptom Checklist (HSCL) which measures general distress and dysfunction ranges from 0.65 (today) to 0.74 (past few weeks) with the NA. The correlation between the Beck Depression Inventory and the NA ranges from 0.56 (past few days) to 0.71 (past few weeks).

Discriminant validity: the correlation between the NA and the PA scales ranged from -0.12 (today) to -0.23 (year) over the different time periods. Correlation between the Hopkins Symptom Checklist (HSCL), which measures general distress and dysfunction ranges from -0.19 (past few weeks) to -0.29 (today) with the PA.

This is a flexible scale and is broad-ranging enough to be useful in a range of research situations. It has been used in the Netherlands, Norway and South Africa.

| Internal reliability of the PANAS ranges (over the different time periods) from Cronbach’s coefficient α? 0.86 (year) to 0.90 (today) for the PA and from 0.84 (year) to 0.87 (today, general) for the NA. Test-retest reliability ranged from 0.47 (today, past week) to 0.68 (general) for the PA and from 0.39 (today) to 0.71 (general) for the NA scale. |

Convergent validity: Correlations between the PANAS and other brief affect measures range from 0.76 to 0.92. Correlation between the Hopkins Symptom Checklist (HSCL) which measures general distress and dysfunction ranges from 0.65 (today) to 0.74 (past few weeks) with the NA. The correlation between the Beck Depression Inventory and the NA ranges from 0.56 (past few days) to 0.71 (past few weeks).

Discriminant validity: the correlation between the NA and the PA scales ranged from -0.12 (today) to -0.23 (year) over the different time periods. Correlation between the Hopkins Symptom Checklist (HSCL), which measures general distress and dysfunction ranges from -0.19 (past few weeks) to -0.29 (today) with the PA.

This is a flexible scale and is broad-ranging enough to be useful in a range of research situations. It has been used in the Netherlands, Norway and South Africa.

**Figure 1. Model of Caregiver Burden (after Poulshock & Deimling, 1984)**
Table 2: Recommended measures of social support and isolation

<table>
<thead>
<tr>
<th>Instrument and Expressive Social Support Scale (IESS) (35) USA</th>
<th>Description</th>
<th>Reliability</th>
<th>Validity</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>The IESS is a 28-item scale that measures the perceived adequacy of available support from the social network, as a function of disruptions in relationships. Items are a list of problems covering finances, time and effort demands, lack of adequate companionship, communication, dependency, familial and household problems. Respondents are asked how they have been bothered by each item over the past 6 months and respond on a scale of 1 (most or all the time) to 5 (never), and item scores are summed. Possible scores range from 22 – 130.</td>
<td>Internal consistency: a ranged from 0.89 to 0.93 over two time periods. Confirmatory factor analysis and reliability tests show that excess of responsibilities/demands, lack of money and unsatisfactory intimate relations factors show strong internal consistency and stability over time.</td>
<td>Five factors: excess of responsibilities/demands, lack of money, lack of involvement, unsatisfactory intimate relations and family problems. Scores on the IESS strongly correlates with depression measured with the CES-D (26). But regression analyses show that factors are conceptually different from depression.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Norbeck’s Social Support Questionnaire (NSSQ) (36) USA | Self-administered measure designed to measure multiple dimensions of social support. The NSSQ measures 3 functional properties of social support (affect, affirmation and aid) and two network properties (duration of relationships and frequency of contacts). The subject is asked to list up to 20 significant people in their life, and to specify the relationship of each person. Subjects then rate each person on a 5-point scale (for questions 1 to 6 and 9b, 1 (not at all) to 5 (a great deal); for question 7, 1 (less than 6 months) to 5 (more than 5 years); for question 8, 1 (once a year or less) to 5 (daily) for 9 questions. The sum of ratings for a question gives the respondents score for that question except for question 9. Question 9, losses is scored as a yes, no response, the quantity of losses is the number of categories in which the subject has lost an important relationship and the quality of losses is the rating on the 5-point scale. | Test-retest reliability ranged from 0.85 to 0.92. All items except Q9 are significantly intercorrelated, range from 0.69 to 0.98. | Significant correlations between some items and related subscales on the Social Support Questionnaire.(37) Lack of significant relationships with the Profile of Mood States. | This scale has been used in African-American, Hispanic, Filipino, Mexican and Arab samples in the USA. |
Caregivers’ physical health may be adversely affected by the strain of caregiving. These effects may manifest in a worsening of chronic conditions e.g. hypertension, greater usage of medical services and increased consumption of prescription medication. Number of visits to doctors and volume of medication consumption can be readily measured and used as indices of physical health. Caregiver health may also be measured quantitatively by using omnibus measures of physical health (and well-being), such as the OARS (8) or the widely-used SF-36 (9), self-rated health scales, or more specifically by examining individual functions such as immune functioning (10,11). A different approach is to examine caregivers’ self-perceptions of health, using instruments such as the Subjective Overall Physical Health Questionnaire (12), a brief scale which asks caregivers to rate their own health, and their perception of their health in comparison to others of the same age. The internal reliability of this scale is high (±0.85).
Financial

Caregiving may result in reduced income and/or increased costs for the caregiver. The following are taken into account when assessing financial burden:
- Loss of earnings by the patient or the caregiver
- Increased expenditure on medical consultations and investigations, drugs, personal care and nursing care
- Costs associated with respite or residential care
- Increased time spent caregiving
- Welfare payments

Social

The demands of caregiving may lead to the abandonment by the caregiver of hobbies and social activities, in addition to the social interaction lost by giving up work. Level of social support and perceived isolation may be measured by a number of instruments (Table 2). Measurement of social support is fraught with complexity. Actual support received may differ considerably from perceived support, the rating of which may be highly influenced by the respondent’s psychological state. For some caregivers, practical support is more important than emotional support, for others the reverse holds. The effects of formal support (by paid or professional staff) differ from those of informal supports, i.e. friends and family. Further, simply counting number of visits to the home, e.g. by family members is too crude as such visits may be a source of distress or of succour. Male and female caregivers, spousal carers, is too crude as such visits may be a source of distress or of succour. Male and female caregivers, spousal care recipients, and psychiatric symptoms such as depression, delusions and hallucinations are a major cause of caregiver burden (16). Both the behaviours themselves and the caregivers’ reactions to them are important to measure. If no behaviours occur there can be no reactive dis-

Table 3: Recommended measures of coping

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Description</th>
<th>Reliability</th>
<th>Validity</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health Specific Family Coping Index for Non-Institutional Care (HSFCI) (40) USA</td>
<td>Measures the caregivers’ coping skills, by rating potential and actual health problems in the psychosocial and physical domains of health. Coping is rated in 9 domains, each with scores ranging from 1 -5. Domains include Physical Independence, Therapeutic Competence, Knowledge of Condition, General Hygiene, Health Attitudes, Emotional Competence, Family Living, Physical Enviroment, and Community Resources. A summary score is obtained by summing sub-scores in each domain (min = 9, max = 45)</td>
<td>Inter-rater reliability: á = 0.99</td>
<td>Construct validity: correlation between HSFCI and General Aptitude Family Coping Index is 0.94</td>
<td></td>
</tr>
<tr>
<td>Family Crisis Oriented Personal Evaluation Scale (41) USA</td>
<td>30-item scale that identifies the problem-solving and behavioural strategies utilized in difficult situations. Five subscales: acquiring social support, reframing, seeking spiritual support, mobilising the family to accept/acquire help, and passive appraisal.</td>
<td>Internal reliability has been established at 0.77, with test-retest reliability at 0.71.</td>
<td>Construct validity: determined using a factor analysis procedure.</td>
<td>This scale has been used in the USA and in Austria.</td>
</tr>
<tr>
<td>Dementia Management Strategies Scale (DMSS) (42) USA</td>
<td>Caregivers rate how often during the past month they have used 28 dementia management strategies, on a scale from 1 (never) to 5 (most of the time). Summary scores are generated for three subscales: Use of Criticism, Use of Encouragement, and Active Management.</td>
<td>Reliability of the three subscales ranges from 0.77 to 0.85.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision-making confidence and skill (43) USA</td>
<td>Caregivers respond to 14 items on a scale ranging from 1 (strongly agree) to 5 (strongly disagree). Responses are summed, with possible scores ranging from 14 -70. Higher scores indicate more confidence. Caregivers are then asked to think of a recent important decision and the alternatives they considered. Higher scores indicate better decision-making skill.</td>
<td>Cronbach’s á? has been found to be 0.86 (44).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Exacerbating factors

Factors identified as having the potential to exacerbate existing burden include poor caregiver physical health and social isolation, which have been addressed above, caregivers’ level of knowledge, their coping skills and the premorbid relationship between caregiver and care recipient. Knowledge can be measured quantitatively, for example by the Alzheimer’s Disease Knowledge Test (ADKT) (15). This comprises twenty multiple-choice items concerning facts about AD. This scale has high internal consistency (?= 0.71 to 0.92) and moderate test-retest reliability (?= 0.62). It has been shown to be sensitive to changes in knowledge level after educational interventions. However knowledge has advanced since this was devised and there is a need for an updated scale. A number of instruments have been designed to measure coping style and skills and premorbid relationship, particularly premorbid marital relations (see Tables 3 and 4).

Problem behaviours in the person with dementia, such as aggression, shadowing and constant questioning, and psychiatric symptoms such as depression, delusions and hallucinations are a major cause of caregiver burden (16). Both the behaviours themselves and the caregivers’ reactions to them are important to measure. If no behaviours occur there can be no reactive dis-
tress. If after an intervention for troublesome behaviours caregiver distress subsides, this may be because the behaviours have abated or because the caregiver is reacting with more equanimity to the same behaviours.

Two widely-used instruments that measure caregivers’ reactions to such behaviours are the Revised Memory and Behaviour Problems Checklist (RMBPCL) (17) and the Neuropsychiatric Interview (NPI) (18). The RMBPCL is a 24-item checklist completed by the caregiver that yields a total score as well as three subscores: memory-related, depression, and disruptive behaviours. The NPI comprises 12 behaviours and psychiatric symptoms rated by a clinician after interviewing a caregiver with regard to frequency, severity and caregiver’s distress in reaction.

**Table 4: Recommended measures of caregiver relationships**

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Description</th>
<th>Reliability</th>
<th>Validity</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marital Communication Scale (45) USA</td>
<td>This 10-item scale taps the caregivers’ feelings, beliefs, and attitudes regarding communications with their spouses/partners. Ratings range from 1 (strongly disagree) to 5 (strongly agree).</td>
<td>Internal reliability was 0.68, with test-retest reliability being 0.90.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family Satisfaction Scale (46) USA</td>
<td>This 14-item scale measures family members’ level of family satisfaction on two dimensions: cohesion and adaptability.</td>
<td>Internal reliability for the scale was calculated at 0.92, with test-retest reliability of 0.75 at 5 weeks.</td>
<td></td>
<td>Construct validity was established using factor analysis procedures.</td>
</tr>
<tr>
<td>USA Family Support Scale – modified (47) USA</td>
<td>7-item scale, which measures the caregivers’ perception of family assistance in caring for the patient. Ratings range from 1 (‘none of the time’) to 3 (‘most of the time’).</td>
<td>Internal reliability: Cronbach’s alpha = 0.82.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative impact on elderly-caregiver family relationship (ECR) (7) USA</td>
<td>An 11-item, factor-derived scale, which assesses family members’ negative affect toward a relative, including feelings of anger and resentment.</td>
<td></td>
<td>Has been found to be related to disruptive behaviour of the patient, and of the patient’s impairment in activities of daily living.</td>
<td></td>
</tr>
<tr>
<td>Marital Needs Satisfaction Scale (48) USA</td>
<td>Consists of 24 items designed to measure the extent of marital need satisfaction of older persons. Scores range from 0 to 225 with a higher score indicating greater satisfaction of marital needs. The six dimensions of spousal satisfaction that were assessed included love, respect, communication, personality fulfilment, life meaningfulness, and integration of life experience. Subjects respond on a 5-point scale (very satisfactory to very unsatisfactory).</td>
<td>Internal consistency was measured using the chisquare test and each item was found to be discriminating between subjects scoring in the highest and lowest quartiles at the 0.001 level of significance. Split-half reliability is 0.99, ( \alpha = 0.97 ) (49).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Protective factors

Factors identified as protective against caregiver burden include practical support (Table 2), and adaptive coping mechanisms (Table 3).

**CONCLUSION**

We report here on the outcome of an international discourse on harmonisation of caregiver outcome measures. This paper presents a basis for this discourse and for refinement.

**Acknowledgements**

Many thanks to all the other carers and researchers who came from all round the world to participate in the symposium on the Harmonization of Caregiver Outcomes at the World Alzheimer Congress in July 2000: Henriette Arendse, Olusegun Baiyewu, Beatriz Berg, Kathy Beukes, Carol Binstock, David Coon, Kheng Leng Chew, Jing Jih Chin, Pamela Doty, Sharen Eckert, Della Frazier-Rios, Linda Furlini, Ruth Goldberg, Maria Cristina Gonzalez, Mary Ann
Laubacher, Heidi Lenster, Nancy Emerson Lombardo, Katie aslow, Barbara McSweeney, Elene Misheva, Joanna urry, Martina Nasrun, Myriam Posada-Lemmerman, Kathleen Potts, Marcia Scazufca, K.S. Shaji, Katherine Siggins, Nina Silverstein, Lorraine Singer, Jon Stickey, Magda Tsolaki, Nelly Ulhomic, Matthew Varghese, Don Ward, Bei Wu, Robert Yeoh, Edwin Yu. Particular thanks go to Marcia Or, who facilitated the symposium, to Breda Cullen, who assisted with preparation of the manuscript and Lee-Fay low who provided useful comments.

REFERENCES

29. Kerner SA, Jacobs KW. Correlation between scores on the Beck Depression Inventory and the Zung Self-rating Depression Scale, Psychological Reports 1983; 53:969-970.
38. Stokes JP. Predicting satisfaction with social support from social network structure, American Journal of Community Psychology 1983; 11:141-152.