Final Report: Dementia Outcomes Measurement Suite Project

APPENDICES

Centre for Health Service Development

September, 2007
Jan Sansoni

Nick Marosszeky

Yun-Hee Jeon

Lynn Chenoweth

Graeme Hawthorne

Madeleine King

Marc Budge

Siggi Zapart

Emily Sansoni

Kate Senior

Patsy Kenny

Lee-Fay Low

Suggested citation

Author Affiliations
1. Director, Australian Health Outcomes Collaboration and Principal Research Fellow, Centre for Health Service Development, University of Wollongong
2. Research Fellow, Centre for Health Service Development, University of Wollongong
3. Fellow, Australian Primary Health Care Research Institute, ANU
4. Professor, Health and Ageing Research Unit, University of Technology Sydney and South Eastern Sydney and Illawarra Area Health Service
5. Associate Professor, Department of Psychiatry, University of Melbourne
6. Senior Lecturer, Centre for Health Economics Research and Evaluation, University of Technology Sydney
7. Head, Department of Geriatric Medicine, Australian National University Medical School
8. Research Officer, Centre for Health Economics Research and Evaluation, University of Technology Sydney
9. Research Assistant, Centre for Health Service Development, University of Wollongong
10. Division Head, Education and Training, Menzies School of Health Research
11. Senior Research Officer, Centre for Health Economics Research and Evaluation, University of Technology Sydney
12. Postdoctoral Research Fellow, Dementia Collaborative Research Centre for Assessment and Better Care Outcomes, University of New South Wales

Acknowledgements
The project team would like to thank all members of the National Expert Panel for their assistance and in particular the contributions of Dianne Wikstrom, Prof. Henry Brodaty and Dr. Rod McKay to the deliberations of the Expert Measurement Group. The project team would also like to thank Dr Marion Haas, Deputy Director, Centre for Health Economics Research and Evaluation for her feedback; and Elizabeth Cuthbert, Darcy Morris and Erin Gleeson from the Centre for Health Service Development for their administrative, editing and research support.

Peer Review Statement
This report has been subject to peer review by an international expert in dementia, in order to assess its scientific merit and its contribution to the literature and clinical practice. In doing so, this report was evaluated in terms of its clinical utility, practicality and validity. The international expert was independently consulted by the Australian Government Department of Health and Ageing to conduct the review, as part of the Department’s quality and evaluation processes. The review made a number of suggestions which were used to improve the final version of this report. In conclusion, the international expert found that “this report is thorough, up to date, well written, and gives a comprehensive overview of the most useful and well validated, assessment tools for the field of dementia diagnosis, cognitive and functional assessment and care.”
# Table of Contents

**APPENDIX 1: NATIONAL EXPERT PANEL**

Appendix 1.1: National Expert Panel: Terms of Reference ......................................................... 1
Appendix 1.2: Membership of the National Expert Panel ................................................................. 4

**APPENDIX 2: THE EXPERT MEASUREMENT GROUP**

Appendix 2.1: Terms of Reference and the Membership of the Expert Measurement Group ............. 5

**APPENDIX 3: REVISED AHOC REVIEW SHEET**

Appendix 4.1: Footnotes to Section 3 - The Standardization of Clinical Terminology ................. 12

**APPENDIX 4A: The NINCDS-ADRDA criteria for the clinical diagnosis of Alzheimer’s disease (McKahn, 1984) ................................................................. 13

**APPENDIX 4B: The State of California Alzheimer’s Disease Diagnostic and Treatment Centres (ADDTC) Criteria for Ischemic Vascular Dementia (Chui, et al. 1992) .................................................. 14

**APPENDIX 4C: Diagnostic criteria for the Clinical Diagnosis of Vascular Dementia (Roman, et al. 1993), from NINDS-AIREN ................................................................. 15

**APPENDIX 4D: The Consensus Criteria for the Clinical Diagnosis of PROBABLE and POSSIBLE dementia of the Lewy body type (DLB) (McKeith, et al. 1996) .................................................. 16

**APPENDIX 4E: Clinical and Pathological Criteria for Fronto-Temporal Dementia, Manchester-Lund Criteria (The Lund and Manchester Groups, 1994) .................................................. 17

**APPENDIX 4F: Definitions, Classifications and Diagnostic Guidelines for Four Major Types of Dementia based on the ICD-10 (WHO, 1992) and the ICD-10-AM Mental Health Manual (2002) .................................................. 19

**APPENDIX 4G: ICF components and domains, with examples of contents ......................................... 21

**APPENDIX 4H: The Hachinski Ischaemic Scale (Hachinski, et al. 1975) ........................................ 22

**APPENDIX 4I: Short form of the Informant Questionnaire on Cognitive Decline in the Elderly (Short IQCODE) ................................................................. 23

**APPENDIX 4J: Different Stages of the EADC MCI Diagnostic Procedure ........................................ 24

**APPENDIX 4K: Clinical Dementia Rating Scale ................................................................. 25

**APPENDIX 4L: The Global Deterioration Scale for Assessment of Primary Degenerative Dementia (Reisberg, 1982; 2000) ................................................................. 26

**APPENDIX 5: REVIEWS OF DEMENTIA STAGING AND DESCRIPTIVE INSTRUMENTS**

Appendix 5.1: Global Deterioration Scale ......................................................................................... 28
Appendix 5.2: Clinical Dementia Rating ......................................................................................... 37
Appendix 5.3: Dementia Severity Rating Scale ............................................................................... 49
Appendix 5.4: Blessed Dementia Scale ......................................................................................... 56
Appendix 5.5: Sandoz Clinical Assessment – Geriatric ...................................................................... 65

**APPENDIX 6: REVIEWS OF DEMENTIA SPECIFIC HEALTH RELATED QUALITY OF LIFE INSTRUMENTS**

Appendix 6.1: Quality of Life in Alzheimer’s Disease ....................................................................... 73
### APPENDIX 7: REVIEWS OF COGNITIVE ASSESSMENT INSTRUMENTS

<table>
<thead>
<tr>
<th>Appendix 7.1</th>
<th>Modified Mini-Mental State Exam</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix 7.2</td>
<td>Alzheimer’s Disease Assessment Scale – Cognition</td>
</tr>
<tr>
<td>Appendix 7.3</td>
<td>The General Practitioner Assessment of Cognition</td>
</tr>
<tr>
<td>Appendix 7.4</td>
<td>Rowland Universal Dementia Assessment Scale</td>
</tr>
<tr>
<td>Appendix 7.5</td>
<td>Minimum Data Set Cognition Scale</td>
</tr>
</tbody>
</table>

### APPENDIX 8: REVIEWS OF MULTI-ATTRIBUTE UTILITY MEASURES

<table>
<thead>
<tr>
<th>Appendix 8.1</th>
<th>EQ-5D Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix 8.2</td>
<td>Assessment of Quality of Life</td>
</tr>
</tbody>
</table>

### APPENDIX 9: REVIEWS OF SOCIAL ISOLATION MEASURES

<table>
<thead>
<tr>
<th>Appendix 9.1</th>
<th>Loneliness Scale</th>
</tr>
</thead>
</table>

### APPENDIX 10: REVIEWS OF ASSOCIATED SYMPTOM MEASURES

<table>
<thead>
<tr>
<th>Appendix 10.1</th>
<th>Global Assessments of Behavioural and Psychological Symptoms of Dementia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix 10.1.1</td>
<td>Neuropsychiatric Inventory</td>
</tr>
<tr>
<td>Appendix 10.1.2</td>
<td>Behavioural Pathology in Alzheimer’s Disease Rating Scale</td>
</tr>
<tr>
<td>Appendix 10.1.3</td>
<td>Dementia Behaviour Disturbance Scale</td>
</tr>
<tr>
<td>Appendix 10.1.4</td>
<td>The Neurobehavioural Rating Scale</td>
</tr>
<tr>
<td>Appendix 10.1.5</td>
<td>Consortium to Establish a Registry for Alzheimer’s Disease – Behavioral Rating Scale for Dementia</td>
</tr>
<tr>
<td>Appendix 10.2</td>
<td>Measures to Assess Delirium</td>
</tr>
<tr>
<td>Appendix 10.2.1</td>
<td>Confusion Assessment Method</td>
</tr>
<tr>
<td>Appendix 10.2.2</td>
<td>Delirium Rating Scale-Revised-98</td>
</tr>
<tr>
<td>Appendix 10.3</td>
<td>Individual Symptom Measures for Associated Symptoms</td>
</tr>
<tr>
<td>Appendix 10.3.1</td>
<td>Rating Scale for Aggressive Behaviour in the Elderly</td>
</tr>
<tr>
<td>Appendix 10.3.2</td>
<td>Cohen Mansfield Agitation Inventory – Long Form</td>
</tr>
<tr>
<td>Appendix 10.3.3</td>
<td>Pittsburgh Agitation Scale</td>
</tr>
<tr>
<td>Appendix 10.3.4</td>
<td>Rating Anxiety in Dementia</td>
</tr>
<tr>
<td>Appendix 10.3.5</td>
<td>Apathy Evaluation Scale</td>
</tr>
<tr>
<td>Appendix 10.3.6</td>
<td>Geriatric Depression Scale</td>
</tr>
<tr>
<td>Appendix 10.3.7</td>
<td>Cornell Scale for Depression in Dementia</td>
</tr>
</tbody>
</table>

### APPENDIX 11: REVIEWS OF MEASURES OF FUNCTION

<table>
<thead>
<tr>
<th>Appendix 11.1</th>
<th>Alzheimer’s Disease Co-operative Study – Activities of Daily Living Inventory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix 11.2</td>
<td>Disability Assessment for Dementia</td>
</tr>
</tbody>
</table>

---

**Centre for Health Service Development**

APPENDIX 6.2: DEMQOL

APPENDIX 6.3: Quality of Life in Late-Stage Dementia

APPENDIX 7: REVIEWS OF COGNITIVE ASSESSMENT INSTRUMENTS

APPENDIX 8: REVIEWS OF MULTI-ATTRIBUTE UTILITY MEASURES

APPENDIX 9: REVIEWS OF SOCIAL ISOLATION MEASURES

APPENDIX 10: REVIEWS OF ASSOCIATED SYMPTOM MEASURES

APPENDIX 11: REVIEWS OF MEASURES OF FUNCTION
<table>
<thead>
<tr>
<th>Appendix 11.3</th>
<th>Cleveland Scale for Activities of Daily Living</th>
<th>367</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>APPENDIX 12:</strong></td>
<td>REVIEWS OF PATIENT AND CARER SATISFACTION MEASURES</td>
<td>375</td>
</tr>
<tr>
<td>Appendix 12.1</td>
<td>Short Assessment of Patient Satisfaction</td>
<td>376</td>
</tr>
<tr>
<td>Appendix 12.2</td>
<td>The Consultation Satisfaction Questionnaire</td>
<td>381</td>
</tr>
<tr>
<td>Appendix 12.3</td>
<td>Satisfaction with Care at the End of Life in Dementia Scale</td>
<td>387</td>
</tr>
<tr>
<td><strong>APPENDIX 13:</strong></td>
<td>SOME RESOURCES RELEVANT FOR THE ASSESSMENT OF CALD POPULATIONS</td>
<td>392</td>
</tr>
<tr>
<td><strong>APPENDIX 14:</strong></td>
<td>APPROPRIATENESS OF RECOMMENDED ASSOCIATED SYMPTOMS INSTRUMENTS FOR INDIGENOUS PEOPLE</td>
<td>393</td>
</tr>
<tr>
<td><strong>APPENDIX 15:</strong></td>
<td>KIMBERLEY INDIGENOUS COGNITIVE ASSESSMENT TOOL</td>
<td>406</td>
</tr>
</tbody>
</table>
Appendix 1: National Expert Panel

Appendix 1.1: National Expert Panel: Terms of Reference

Australian Government Department of Health and Ageing
and
Centre for Health Services Development
University of Wollongong

Development of a Dementia Outcomes Measurement Suite (DOMS)
DOMS National Expert Panel (DOMS-NEP)

Preamble:
The purpose of this project is to develop a set of recommended measures/tools for the assessment, diagnosis, screening and outcomes monitoring of dementia and associated conditions that are applicable for the Australian health care context. By developing a set of recommended measures it is hoped to standardise assessment and evaluation procedures used in this field to enhance comparability of findings across research and practice settings.

Goals:
1. Reach agreement on recommendations for the standardisation of clinical terminology relating to dementia, as informed by the Consultation Paper prepared by the research team.
2. Reach agreement on recommendations for the best instruments for use in the Australian Health Care context, as informed by the recommendations of the Expert Measurement Group members of the research team.
3. Advise on issues concerning the project scope.
4. Reach agreement on the conclusions and recommendations in the final report to the Australian Government Department of Health and Ageing.

Objectives:
1. Review, provide feedback, and discuss:
   1.1. Recommendations for the standardisation of clinical terminology relating to dementia as identified in a literature review prepared by the research team;
   1.2. The most commonly used measurement instruments for the assessment, diagnosis and outcome evaluation in formal and informal contexts for people with dementia;
   1.3. The inclusion and exclusion criteria for instrument selection;
   1.4. Recommendations for suitable domains to capture each stage of the disease and in each care context;
   1.5. General directions and issues for implementation and appropriate use of measures;
   1.6. The degree of consensus about the above issues.
2. Participate in regular consultation with the research team over the course of the study.

Membership:
Representatives from all states and territories within Australia, including:
1. Representatives of key community/consumer groups and aged care groups:
   Alzheimer’s Australia
   Carers Australia
   Council on the Ageing
   ACSA (The Aged and Community Services Australia)
2. Clinical practitioner representatives from different health care contexts:
   GPs
   Psycho-geriatricians / other Consultant Psychiatrists
   Geriatricians / other Consultant Physicians
   Dementia Clinical Nurse Consultants and Specialists
Specialist Mental Health Nurses  
Occupational Therapist specialising in dementia  
Clinical Psychologists specialising in dementia  
Social Worker specialising in dementia and aged care  
Aged Care Assessment Team manager

3. Diversity/multicultural health representatives:  
   Multicultural Mental Health Services  
   Remote area health services  
   National Cross-Cultural Dementia Network

4. Dementia and aged care researchers:  
   Academic appointments  
   Clinically-based researchers

5. Representative of Expert Measurement Group
6. Representative of National Evaluator for the Dementia Health Priority
7. Australian Institute of Health & Welfare Representative
8. Australian Government Department of Health & Ageing Representative
9. Minister’s Dementia Health Priority Taskforce

Quorum:  
The required number of members that will constitute a quorum of the membership for meetings is 10.

Chairman:  
Associate Professor Marc Budge  
Head, Department of Geriatric Medicine  
Australian National University Medical School  
Medical Director, Aged Care and Rehabilitation Services  
ACT Health.

Secretaries:  
Dr Lynn Chenoweth  
Professor of Aged & Extended Care Nursing, University of Technology Sydney, and South Eastern Sydney-Illawarra Area Health Service.

Dr Yun-Hee Jeon  
Fellow, Australian Primary Health Care Research Institute, College of Medicine and Health Sciences, Australian National University.

Responsibilities:  
Members:  
1. Attend /contribute via teleconference to three meetings between May and February 2007, as outlined in the Objectives 1, 2, and 3 listed (above).  
2. Participate in regular consultation between meetings with the research team.

Chairman:  
1. Convene and chair the DOMS-NEP meetings three times during the research project.  
2. Refer all work undertaken in the project to the DOMS-NEP for ensuring the quality and performance of the project.  

Secretaries:  
1. Set the dates for the three meetings, in consultation with members and the Chairman.  
2. Provide members and the Chairman with an agenda one week prior to the meetings.  
3. Arrange and advise members of teleconferencing arrangements for meetings.  
4. Prepare Minutes of the meetings, have signed off by the Chairman and send copies to all members within two weeks of meetings.
5. Provide advice on re-imbursement for travel costs and make teleconference arrangements for the meetings.

**Conduct of the Meetings:**
Frequency: Three meetings between May 2006 and September 2007

**Dates:**
Meeting 1 – May 3, 2006
Meeting 2 – August 4, 2006
Meeting 3 – August 17, 2007
There was NEP representation (Prof. H. Brodaty, Dr. R. McKay) at the Special EMG meeting of 8/6/2007 to discuss the measures used to assess the associated symptoms of dementia.

**Times:**
1.00-5.00pm

**Venue:**
Conference Room 1
Primrose House
South East Sydney-llawarra Area Health Service executive offices
Cnr. Russell & Malua Streets
Dolls Point (Sydney) NSW
(reception) 02-9947 9832

**Report to:**
1. DOMS Expert Measurement Group
2. Australia government Department of Health and Ageing
3. Centre for Health Services Development, University of Wollongong
4. Research Partners
### Appendix 1.2: Membership of the National Expert Panel

<table>
<thead>
<tr>
<th>Composition</th>
<th>Membership</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minister's Dementia Health Priority Taskforce</td>
<td>Professor Henry Brodaty (Old Age Psychiatry, Prince of Wales Hospital, NSW)</td>
</tr>
<tr>
<td>Key community groups</td>
<td></td>
</tr>
<tr>
<td>Alzheimers NSW</td>
<td>Dr Robert Yeoh (GP, Fairfield, NSW)</td>
</tr>
<tr>
<td>Carer's NSW</td>
<td>Ms Toni Payne (Manager, Training and Policy, NSW)</td>
</tr>
<tr>
<td>Council on the Ageing</td>
<td>Ms Lisa Langley (Policy officer, NSW)</td>
</tr>
<tr>
<td>National Cross-Cultural Dementia Network</td>
<td>Mr René Grypma (Senior clinical psychologist, CNAHS-MH-Aged Care, SA)</td>
</tr>
<tr>
<td>Key community groups</td>
<td></td>
</tr>
<tr>
<td>Composition</td>
<td>Membership</td>
</tr>
<tr>
<td>Clinical practitioner/service representatives</td>
<td></td>
</tr>
<tr>
<td>Geriatricians</td>
<td>Associate Professor Marc Budge (The Canberra Hospital, ACT)</td>
</tr>
<tr>
<td>Psychologists</td>
<td>Dr Mike Bird (Greater Southern Area Health Service, NSW)</td>
</tr>
<tr>
<td>Psycho-geriatricians</td>
<td>Dr Rod McKay (Aged Care Psychiatry, Braeside Hospital, NSW)</td>
</tr>
<tr>
<td>Clinical Nurse Consultants in Dementia/ Psychogeriatrics</td>
<td>Mr Bryan McMinn (Hunter and New England Area Health Service, NSW) Ms Colleen McKinnon (South Eastern Sydney and Illawarra Area Health Service, NSW)</td>
</tr>
<tr>
<td>Dementia Researchers/Academics</td>
<td>Dr Kaarin Anstey (ANU, ACT) Professor Helen Edwards (Head of the School of Nursing at Queensland University of Technology, QLD) Professor Lynn Chenoweth (Aged and extended care, UTS, NSW) Dr Yun-Hee Jeon (Australian Primary Health Care Research Institute, ANU, ACT) Mr Richard Fleming (Hammond Care Group, NSW)</td>
</tr>
<tr>
<td>GP/Geriatrician</td>
<td>Dr George Stathers (NSW)</td>
</tr>
<tr>
<td>Aged Care Assessment Team</td>
<td>Ms Colleen O'Neill (Social Work Manager, War Memorial Hospital, NSW)</td>
</tr>
<tr>
<td>Diversity/Multicultural Health Rural and Remote</td>
<td>Ms Clarissa Mulas, on behalf of Mr Abd Malak (Director, Diversity Health Institute, NSW) Ms Judy Ratajec (Coordinator, Northern Territory Older Persons Support Service, NT)</td>
</tr>
<tr>
<td>Community/Residential Care</td>
<td>Dr Meredith Gresham (NSW)</td>
</tr>
<tr>
<td>Australian Institute of Health and Welfare's (AIHW)</td>
<td>The Australian Dementia Data Analysis and Development Project Ms Ann Peut (Head, Ageing and Aged Care Unit, AIHW, ACT)</td>
</tr>
<tr>
<td>Department of Health and Ageing</td>
<td>Mr Kevin Vassarotti (Director, Population and Ageing Research Section, Office for an Ageing Australia)</td>
</tr>
<tr>
<td>National Evaluator for the Dementia Health Priority</td>
<td>Ms Lynne Pezzullo (Associate Director, Access Economics, ACT)</td>
</tr>
<tr>
<td>Representative of Expert Measurement Group</td>
<td>Ms Jan Sansoni (Director, Australian Health Outcomes Collaboration, ACT)</td>
</tr>
</tbody>
</table>
Appendix 2: The Expert Measurement Group

Appendix 2.1: Terms of Reference and the Membership of the Expert Measurement Group

Australian Government Department of Health and Ageing
and
Centre for Health Services Development
University of Wollongong

Development of a Dementia Outcomes Measurement Suite (DOMS)
DOMS Expert Measurement Group (DOMS-EMG)

Goals:
1. Make recommendations concerning search strategies for the instrument reviews.
2. Make recommendations concerning exclusion/inclusion criteria to determine the leading instruments in each instrument category to be systematically reviewed.
3. Make recommendations concerning the review and elaboration of AHOC instrument review protocol.
4. Make recommendations concerning the weighting of instrument review criteria.
5. Review the systematic reviews of selected instruments from each of the instrument review teams.
6. Liaise with and make recommendations to the NEP concerning recommended instruments and measurement issues.

Objectives:
1. Review, provide feedback, and discuss:
   1.1. Recommendations for instrument search strategies.
   1.2. Feedback concerning the most commonly used measurement instruments for the assessment, diagnosis and outcome evaluation in the field for people with dementia as advised by the NEP.
   1.3. Recommendations concerning inclusion and exclusion criteria for instrument selection.
2. Provide recommendations concerning the revision and elaboration of the AHOC instrument review protocol.
3. Provide recommendations concerning the weighting of instrument review criteria.
4. Review the systematic reviews of recommended instruments from each of the instrument reviewers.
5. Liaise with and make recommendations to the NEP concerning recommended instruments and measurement issues.
6. Participate in regular consultation with the research team over the course of the study.

Membership:
Assoc Prof Marc Budge
Ms Jan Sansoni
Dr Madeleine King
Assoc Prof Graeme Hawthorne
Ms Diane Wikstrom (Clinical Nurse Consultant)
Mr Nick Marosszeky (will provide research and secretariat support to the EMG)

Quorum:
The required number of members that will constitute a quorum of the membership for meetings is 4.
Chairman:
Jan Sansoni
Director
Australian Health Outcomes Collaboration and
Principal Research Fellow, Centre for Health Service Development
University of Wollongong

Secretary:
Nick Marosszeky
Research Fellow
Centre for Health Service Development
University of Wollongong

Responsibilities:
Members:
1. Attend and/or contribute via teleconference to three meetings between May and September 2007, as outlined in the Objectives 1, 2, and 3 listed (above).
2. Participate in regular consultation between meetings with the research team.

Chairman:
1. Convene and chair the DOMS-EMG meetings three times during the research project.
2. Refer all work undertaken in the project to the DOMS-NEP for ensuring the quality and performance of the project.
3. Report the consensus recommendations of the DOMS-EMG to the Commonwealth Department of Health & Ageing and to the DOMS-NEP.

Secretary:
1. Set the dates for the three meetings, in consultation with members and the Chairman.
2. Provide members and the Chairman with an agenda one week prior to the meetings.
3. Arrange and advise members of teleconferencing arrangements for meetings.
4. Prepare minutes of the meetings, have signed off by the Chairman and send copies to all members within two weeks of meetings.
5. Provide advice on re-imbursement for travel costs and make teleconference arrangements for the meetings.

Conduct of the Meetings:
Frequency:
Five meetings were held between May 2006 and September 2007.

Meeting Dates:
The first meeting was held on 8th June 2006. A working group meeting was held on August 8th at the Australian Health Outcomes Conference. Meetings were also held on the 14th November 2006, 7th June (including NEP representation) and the 17th August 2007.

Report to:
1. DOMS National Expert Panel
2. Australian Government Department of Health and Ageing
3. Centre for Health Services Development, University of Wollongong
4. Research Partners
Appendix 3: Revised AHOC Review Sheet

AHOC INSTRUMENT REVIEW SHEET

Title:

Abbreviations:

Author(s) Name:

Author(s) Address:

Supplied by:

Cost:

Training requirements:

Purpose:

Administration time:

Instrument Type:

Structure:

Scoring:

Developed for:

Normative Data:

Clinical Data:

Applications:

Carer and/or Patient Use of Instrument:
**Psychometric Criteria**

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td>□ Alpha &gt;0.70 □ Marginal or inadequate internal consistency (&lt;0.70) □ No information found on internal consistency</td>
<td>□ ICC &gt;0.70 Time intervals and confidence intervals reported □ Marginal or inadequate test-retest reliability ICC&lt;0.70 □ No information found on test-retest reliability</td>
<td>□ Agreement reported and adequate □ Inadequate inter-rater agreement □ No information provided</td>
</tr>
<tr>
<td>Test – retest</td>
<td>□ ICC &gt;0.70 Time intervals and confidence intervals reported □ Marginal or inadequate test-retest reliability ICC&lt;0.70 □ No information found on test-retest reliability</td>
<td>□ ICC &gt;0.70 Time intervals and confidence intervals reported □ Marginal or inadequate test-retest reliability ICC&lt;0.70 □ No information found on test-retest reliability</td>
<td>□ Agreement reported and adequate □ Inadequate inter-rater agreement □ No information provided</td>
</tr>
<tr>
<td>Inter – rater</td>
<td>□ Agreement reported and adequate □ Inadequate inter-rater agreement □ No information provided</td>
<td>□ Agreement reported and adequate □ Inadequate inter-rater agreement □ No information provided</td>
<td>□ Agreement reported and adequate □ Inadequate inter-rater agreement □ No information provided</td>
</tr>
</tbody>
</table>

Internal consistency
The extent to which items in a (sub) scale are inter-correlated; a measure of the homogeneity of a (sub)scale
Cronbach's alpha should be between 0.70 and 0.90 for every dimension / sub-scale

Test – retest
The extent to which the same results are obtained on repeated administrations of the same questionnaire when no change in physical functioning has occurred
Calculation of an intraclass correlation coefficient (ICC); and an ICC > 0.70 is desired
Preferred if time interval and confidence intervals were presented
<table>
<thead>
<tr>
<th>VALIDITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td></td>
<td>□ Patients and experts were involved during item selection and/or item reduction</td>
<td>□ Patients were consulted for reading and comprehension □ No patient involvement □ No information found on content validity □ There is an adequate coverage of relevant domains □ There is limited coverage of relevant domains</td>
</tr>
<tr>
<td>Construct</td>
<td></td>
<td>□ Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used</td>
<td>□ Limited/inadequate construct validity reported □ No information provided</td>
</tr>
<tr>
<td>Construct: Internal Structure</td>
<td></td>
<td>□ No evidence provided/failed a test of dimensionality □ Some evidence provided to support internal structure □ Substantial evidence provided to support internal structure</td>
<td></td>
</tr>
<tr>
<td>Construct: Correlation with other measures</td>
<td></td>
<td>□ Correlations with other measures are reported □ Correlations not reported</td>
<td></td>
</tr>
<tr>
<td>Construct: Discriminant Validity</td>
<td></td>
<td>□ Scale differentiates between relevant categories of respondents □ No information on discriminant validity</td>
<td></td>
</tr>
<tr>
<td>Criterion</td>
<td></td>
<td>□ Comparison made to criterion measures □ No comparison with criterion measures provided</td>
<td></td>
</tr>
</tbody>
</table>
## Interpretability

The degree to which one can assign qualitative meaning to quantitative scores

Do authors provide the following:

- Presentation of means and SD of scores before and after treatment
- Comparative data on the distribution of scores in relevant subgroups
- Information on the relationship of scores to well-known functional measures or clinical diagnosis
- Information on the association between changes in scores and patients' global ratings of the magnitude of change they have experienced

<table>
<thead>
<tr>
<th>Authors provide 2 or more types of information on interpretability</th>
<th>Authors provide limited information to assist with interpretability</th>
<th>No information provided</th>
</tr>
</thead>
</table>

## RESPONSIVENESS

### Floor and ceiling effects

The questionnaire fails to demonstrate a worse score in patients who clinically deteriorated and an improved score in patients who clinically improved.

Authors should provide descriptive statistics of the distribution of scores.

- Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected
- Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score
- No information provided on floor and ceiling effects

### Sensitivity to change

The ability to detect important change over time in the concept being measured.

- Hypotheses were formulated and results were in agreement
- An adequate metric was used (ES, SRM, comparison with external standard)
- No information on sensitivity to change was provided
- MCID - Information was provided about the magnitude of score differences which would be clinically meaningful
- MCID – No information was provided
Cultural Applicability
and Cultural Adaptations:

Gender Appropriateness:

Age Appropriateness:

Summary:

Reporter:

Date of report:

References

Adequacy checks were modified from Bot et al. 2004 and represent world’s best practice for the selection of health measurement instruments (see Mokkink et al. 2006).


Appendix 4: Footnotes to Section 3 - The Standardization of Clinical Terminology
APPENDIX 4A: The NINCDS-ADRDA criteria for the clinical diagnosis of Alzheimer’s disease. (McKahn, 1984)

I. Criteria for the clinical diagnosis of PROBABLE Alzheimer’s disease include:
   • dementia established by clinical examination and documented by the Mini-Mental Test, Blessed Dementia Scale, or some similar examination, and confirmed by neuropsychological tests;
   • deficits in two or more areas of cognition;
   • progressive worsening of memory and other cognitive functions;
   • no disturbance of consciousness;
   • onset between ages 40 and 90, most often after age 6; and
   • absence of systemic disorders or other brain diseases that in and of themselves could account for the progressive deficits in memory and cognition.

II. The diagnosis of PROBABLE Alzheimer’s disease is supported by:
   • progressive deterioration of specific cognitive functions such as language (aphasia), motor skills (apraxia), and perceptions (agnosia);
   • impaired activities of daily living and altered patterns of behaviour;
   • family history of similar disorders, particularly if confirmed neuropathologically; and
   • laboratory results of:
     o normal lumbar puncture as evaluated by standard techniques,
     o normal pattern or non-specific changes in EEG, such as increased slow-wave activity, and evidence of cerebral atrophy on CT with progression documented by serial observation.

III. Other clinical features consistent with the diagnosis of PROBABLE Alzheimer’s disease, after exclusion of causes of dementia other than Alzheimer’s disease, include:
   • plateaus in the course of progression of the illness;
   • associated symptoms of depression, insomnia, incontinence, delusions, illusions, hallucinations, catastrophic verbal, emotional, or physical outbursts, sexual disorders, and weight loss;
   • other neurologic abnormalities in some patients, especially with more advanced disease and including motor signs such as increased muscle tone, myoclonus, or gait disorder;
   • seizures in advanced disease; and
   • CT normal for age.

IV. Features that make the diagnosis of PROBABLE Alzheimer’s disease uncertain or unlikely include:
   • sudden, apoplectic onset;
   • focal neurologic findings such as hemiparesis, sensory loss, visual field deficits, and incoordination early in the course of the illness; and
   • seizures or gait disturbances at the onset or very early in the course of the illness.

V. Clinical diagnosis of POSSIBLE Alzheimer’s disease:
   • May be made on the basis of the dementia syndrome, in the absence of other neurologic, psychiatric, or systemic disorders sufficient to cause dementia, and in the presence of variations in the onset, in the presentation, or in the clinical course.

VI. Criteria for diagnosis of DEFINITE Alzheimer’s disease are:
   • the clinical criteria for probable Alzheimer’s disease and histopathologic evidence obtained from a biopsy or autopsy

VII. Classification of Alzheimer’s disease for research purposes should specify features that may differentiate subtypes of the disorder, such as:
   • familial occurrence;
   • onset before age of 65;
   • presence of trisomy-21; and
   • coexistence of other relevant conditions such as Parkinson’s disease.
### APPENDIX 4B: The State of California Alzheimer’s Disease Diagnostic and Treatment Centres (ADDTC) Criteria for Ischemic Vascular Dementia (Chui, et al. 1992)

<table>
<thead>
<tr>
<th><strong>I. Dementia</strong></th>
<th><strong>III. Possible IVD</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dementia is a deterioration from a known or estimated prior level of intellectual function sufficient to interfere broadly with the conduct of the patient’s customary affairs of life, which is not isolated to a single narrow category of intellectual performance, and which is independent of level of consciousness. This deterioration should be supported by historical evidence and documented by either bedside mental status testing or ideally by more detailed neuropsychological examination, using tests that are quantifiable, reproducible, and for which normative data are available.</td>
<td>A clinical diagnosis of POSSIBLE IVD may be made when there is</td>
</tr>
<tr>
<td>1. Dementia; 2. Evidence of two or more ischemic strokes by history, neurologic signs, and/or neuroimaging studies (CT or T₁-weighted MRI); or Occurrence of a single stroke with a clearly documented temporal relationship to the onset of dementia; 3. Evidence of at least one infarct outside the cerebellum by CT or T₁-weighted MRI.</td>
<td>1. Dementia; and one or more of the following:</td>
</tr>
<tr>
<td>A. The criteria for the clinical diagnosis of PROBABLE IVD include ALL of the following:</td>
<td>2a. A history or evidence of a single stroke (but not multiple strokes) without a clearly documented temporal relationship to the onset of dementia; or</td>
</tr>
<tr>
<td>1. Dementia; 2. Evidence of two or more ischemic strokes by history, neurologic signs, and/or neuroimaging studies (CT or T₁-weighted MRI); or Occurrence of a single stroke with a clearly documented temporal relationship to the onset of dementia; 3. Evidence of at least one infarct outside the cerebellum by CT or T₁-weighted MRI.</td>
<td>2b. Binswanger’s syndrome (without multiple strokes) that includes all of the following:</td>
</tr>
<tr>
<td>B. The diagnosis of PROBABLE IVD is supported by</td>
<td>i. Early-onset urinary incontinence not explained by urologic disease, or gait disturbance (e.g. parkinsonian, magnetic, apraxic, or “senile” gait) not explained by peripheral cause, ii. Vascular risk factors, and iii. Extensive white matter changes on neuroimaging.</td>
</tr>
<tr>
<td>1. Evidence of multiple infarcts in brain regions known to affect cognition; 2. A history of multiple transient ischemic attacks; 3. History of vascular risk factors (e.g. hypertension, heart disease, diabetes mellitus); 4. Elevated Hachinski Ischemia Scale (original or modified version).</td>
<td>IV. Definite IVD</td>
</tr>
<tr>
<td>C. Clinical features that are thought to be associated with IVD, but await further research, include</td>
<td>A diagnosis of DEFINITE IVD requires histopathologic examination of the brain, as well as</td>
</tr>
<tr>
<td>1. Relatively early appearance of gait disturbance and urinary incontinence; 2. Periventricular and deep white matter changes on T₁-weighted MRI that are excessive for age; 3. Focal changes in electrophysiologic studies (e.g. EEG, evoked potentials) or physiologic neuroimaging studies (e.g. SPECT, PET, NMR spectroscopy).</td>
<td>A. Clinical evidence of dementia;</td>
</tr>
<tr>
<td>D. Other clinical features that do not constitute strong evidence either for or against a diagnosis of PROBABLE IVD include</td>
<td>B. Pathologic confirmation of multiple infarcts, some outside of the cerebellum.</td>
</tr>
<tr>
<td>1. Periods of slowly progressive symptoms; 2. Illusions, psychosis, hallucinations, delusions; 3. Seizures.</td>
<td>Note: If there is evidence of Alzheimer’s disease or some other pathologic disorder that is thought to have contributed to the dementia, a diagnosis of MIXED dementia should be made.</td>
</tr>
<tr>
<td>E. Clinical features that cast doubt on a diagnosis of PROBABLE IVD include</td>
<td>V. Mixed dementia</td>
</tr>
<tr>
<td>1. Transcortical sensory aphasia in the absence of corresponding focal lesions on neuroimaging studies; 2. Absence of central neurologic symptoms/signs, other than cognitive disturbance.</td>
<td>A diagnosis of MIXED dementia should be made in the presence of one or more other systemic or brain disorders that are thought to be causally related to the dementia. The degree of confidence in the diagnosis of IVD should be specified as possible, probable, or definite, and the other disorder(s) contributing to the dementia should be listed. For example: mixed dementia due to probable IVD and possible Alzheimer’s disease or mixed dementia due to definite IVD and hypothyroidism.</td>
</tr>
</tbody>
</table>

| **VI. Research classification** |
| Classification of IVD for RESEARCH purposes should specify features of the infarcts that may differentiate subtypes of the disorder, such as |
| Location: cortical, white matter, periventricular, basal ganglia, thalamus |
| Size: volume Distribution: large, small, or microvessel |
| Severity: chronic ischemia versus infarction |
| Aetiology: embolism, atherosclerosis, arteriosclerosis, cerebral amyloid angiopathy, hypoperfusion. |
APPENDIX 4C: Diagnostic criteria for the Clinical Diagnosis of Vascular Dementia (Roman, et al. 1993), from NINDS-AIREN

I. Criteria for the clinical diagnosis of PROBABLE vascular dementia include all of the following:

1. **Dementia** defined by cognitive decline from a previously higher level of functioning and manifested by impairment of memory and of two or more cognitive domains (orientation, attention, language, visuospatial functions, executive functions, motor control, and praxis), preferably established by clinical examination and documented by neuropsychological testing; deficits should be severe enough to interfere with activities of daily living not due to physical effects of stroke alone.

   **Exclusion criteria:** cases with disturbance of consciousness, delirium, psychosis, severe aphasia, or major sensorimotor impairment precluding neuropsychological testing. Also excluded are systemic disorders or other brain diseases (such as AD) that in and of themselves could account for deficits in memory and cognition.

2. **Cerebrovascular disease**, defined by the presence of focal signs on neurologic examination, such as hemiparesis, lower facial weakness, Babinski sign, sensory deficit, hemianopia, and dysarthria consistent with stroke (with or without history of stroke), and evidence of relevant CVD by brain imaging (CT or MRI) including multiple large-vessel infarcts or a single strategically placed infarct (angular gyrus, thalamus, basal forebrain, or PCA or ACA territories), as well as multiple basal ganglia and white matter lacunes or extensive periventricular white matter lesions, or combinations thereof.

3. A **relationship between the above two disorders**, manifested or inferred by the presence of one or more of the following: (a) onset of dementia within 3 months following a recognized stroke; (b) abrupt deterioration in cognitive functions; or fluctuating, stepwise progression of cognitive deficits.

II. Clinical features consistent with the diagnosis of PROBABLE vascular dementia include the following:

(a) early presence of a gait disturbance (small-step gait or marche à petits pas, or magnetic, apraxicataxic or parkinsonian gait)

(b) history of unsteadiness and frequent, unprovoked falls

(c) early urinary frequency, urgency, and other urinary symptoms not explained by urologic disease

(d) pseudobulbar palsy

(e) personality and mood changes, abulia, depression, emotional incontinence, or other subcortical deficits including psychomotor retardation and abnormal executive function.

III. Features that make the diagnosis of vascular dementia uncertain or unlikely include

(a) early onset of memory deficit and progressive worsening of memory and other cognitive functions such as language (transcortical sensory aphasia), motor skills (apraxia), and perception (agnosia), in the absence of corresponding focal lesions on brain imaging

(b) absence of focal neurologic signs, other than cognitive disturbance

(c) absence of cerebrovascular lesions on brain CT or MRI.

IV. Clinical diagnosis of POSSIBLE vascular dementia may be made in the presence of dementia (section I-1) with focal neurologic signs in patients in whom brain imaging studies to confirm definite CVD are missing; or in the absence of clear temporal relationship between dementia and stroke; or in patients with subtle onset and variable course (plateau or improvement) of cognitive deficits and evidence of relevant CVD.

V. Criteria for diagnosis of DEFINITE vascular dementia are (a) clinical criteria for PROBABLE vascular dementia; (b) histopathologic evidence of CVD obtained from biopsy or autopsy; (c) absence of neurofibrillary tangles and neuritic plaques exceeding those expected for age; and (d) absence of other clinical or pathologic disorder capable of producing dementia.

VI. Classification of vascular dementia for research purposes may be made on the basis of clinical, radiologic and neuropathologic features, for subcategories or defined conditions such as cortical vascular dementia, subcortical vascular dementia, BD and thalamic dementia.

The term “AD with CVD” should be reserved to classify patients fulfilling the clinical criteria for possible AD and who also present clinical or brain imaging evidence of relevant CVD. Traditionally, these patients have been included with VaD in epidemiological studies. The term mixed dementia hitherto, should be avoided.

1. The central feature required for a diagnosis of DLB is progressive cognitive decline of sufficient magnitude to interfere with normal social or occupational function. Prominent or persistent memory impairment may not necessarily occur in the early stages but is usually evident with progression. Deficits on tests of attention and of frontal sub-cortical skills and visuospatial ability may be especially prominent.

2. Two of the following core features are essential for a diagnosis of probable DLB, and one is essential for possible DLB:
   a. fluctuating cognition with pronounced variations in attention and alertness
   b. recurrent visual hallucinations that are typically well formed and detailed
   c. spontaneous motor features of parkinsonism.

3. Features supportive of the diagnosis are:
   a. repeated falls
   b. syncope
   c. transient loss of consciousness
   d. neuroleptic sensitivity
   e. systematised delusions
   f. hallucinations in other modalities.

4. A diagnosis of DLB is less likely in the presence of:
   a. stroke disease, evident as focal neurologic signs or on brain imaging
   b. evidence on physical examination and investigation of any physical illness or other brain disorder sufficient to account for the clinical picture.

Revised criteria for the clinical diagnosis of dementia with Lewy bodies (DLB) (McKeith, et al. 2005).

1. Central feature (essential for a diagnosis of possible or probable DLB)
   Dementia defined as progressive cognitive decline of sufficient magnitude to interfere with normal social or occupational function.
   Prominent or persistent memory impairment may not necessarily occur in the early stages but is usually evident with progression.
   Deficits on tests of attention, executive function, and visuospatial ability may be especially prominent.

2. Core features (two core features are sufficient for a diagnosis of probable DLB, one for possible DLB)
   Fluctuating cognition with pronounced variations in attention and alertness
   Recurrent visual hallucinations that are typically well formed and detailed
   Spontaneous features of parkinsonism

3. Suggestive features (If one or more of these is present in the presence of one or more core features, a diagnosis of probable DLB can be made. In the absence of any core features, one or more suggestive features is sufficient for possible DLB. Probable DLB should not be diagnosed on the basis of suggestive features alone.)
   REM sleep behaviour disorder
   Severe neuroleptic sensitivity
   Low dopamine transporter uptake in basal ganglia demonstrated by SPECT or PET imaging

4. Supportive features (commonly present but not proven to have diagnostic specificity)
   Repeated falls and syncope
   Transient, unexplained loss of consciousness
   Severe autonomic dysfunction, e.g., orthostatic hypotension, urinary incontinence
   Hallucinations in other modalities
   Systematized delusions
   Depression
   Relative preservation of medial temporal lobe structures on CT/MRI scan
   Generalized low uptake on SPECT/PET perfusion scan with reduced occipital activity
   Abnormal (low uptake) MIBG myocardial scintigraphy
   Prominent slow wave activity on EEG with temporal lobe transient sharp waves

5. A diagnosis of DLB is less likely:
   In the presence of cerebrovascular disease evident as focal neurologic signs or on brain imaging
   In the presence of any other physical illness or brain disorder sufficient to account in part or in total for the clinical picture
   If parkinsonism only appears for the first time at a stage of severe dementia

6. Temporal sequence of symptoms
   DLB should be diagnosed when dementia occurs before or concurrently with parkinsonism (if it is present). The term Parkinson disease dementia (PDD) should be used to describe dementia that occurs in the context of well-established Parkinson disease. In a practice setting the term that is most appropriate to the clinical situation should be used and generic terms such as LB disease are often helpful. In research studies in which distinction needs to be made between DLB and PDD, the existing 1-year rule between the onset of dementia and parkinsonism DLB continues to be recommended. Adoption of other time periods will simply confound data pooling or comparison between studies. In other research settings that may include clinicopathologic studies and clinical trials, both clinical phenotypes may be considered collectively under categories such as LB disease or alpha-synucleinopathy.
APPENDIX 4E: Clinical and Pathological Criteria for Fronto-Temporal Dementia, Manchester-Lund Criteria (The Lund and Manchester Groups, 1994)

Core diagnostics include:

1. **Behavioural disorder**
   - insidious onset and slow progression
   - early loss of personal awareness (neglect of personal hygiene and grooming)
   - early signs of disinhibition (such as unrestrained sexuality, violent behaviour, inappropriate jocularity, restless pacing)
   - mental rigidity and inflexibility
   - hyperorality (oral/dietary changes, over eating, food fads, excessive smoking and alcohol consumption, oral exploration of objects)
   - stereotyped and preservative behaviour (wandering, mannerisms such as clapping, singing, dancing, ritualistic preoccupations such as hoarding, toileting and dressing)
   - utilisation behaviour (unrestrained exploration of objects in the environment)
   - distractibility, impulsivity and impersistence
   - early loss of insight into the fact that the altered condition is due to a pathological change of own mental state.

2. **Affective Symptoms**
   - depression, anxiety, excessive sentimentality, suicidal and fixed ideation, delusion (early and evanescent)
   - hypochondriasis, bizarre somatic preoccupation (early and evanescent)
   - emotional unconcern (emotional indifference and remoteness, lack of empathy and sympathy, apathy)
   - amimia (inertia, aspontaneity).

3. **Speech Disorder**
   - progressive reduction of speech (aspontaneity and economy of utterance)
   - stereotypy of speech (repetition of limited repertoire of words, phrases or themes)
   - echolalia and preservation
   - late mutism.

4. **Spatial orientation and praxis preserved** (intact abilities to negotiate the environment)

5. **Physical Signs**
   - early primitive reflexes
   - early incontinence
   - late akinesia, rigidity, tremor
   - low and labile blood pressure.

6. **Investigation**
   - normal EEG despite clinically evident dementia
   - brain imaging (structural or functional, or both) predominant frontal or anterior temporal abnormality or both
   - neuropsychology (profound failure on frontal lobe tests in the absence or severe amnesia, aphasia or perceptual spatial disorder).

Supportive diagnostic features include:

1. onset before 65
2. positive family history of similar disorder in a first degree relative
3. bulbar palsy, muscular weakness and wasting, fasciculations (motor neuron disease).

Diagnostic exclusion features include:

1. abrupt onset with ictal events
2. head trauma related to onset
3. early severe amnesia
4. early spatial disorientation, lost in surroundings, defective localization of objects
5. early severe apraxia
6. logoclonic speech with rapid loss of train of thoughts
7. myoclonus
8. cortical bulbar and spinal deficits
9. cerebellar ataxia
10. choreo-athetosis
11. early, severe, pathological EEG
12. brain imaging (predominant post-central structural or functional deficit, multifocal cerebral lesions on CT or MRI)
13. laboratory tests indicating brain involvement or inflammatory disorder (such as multiple sclerosis, syphilis, AIDS and herpes simplex encephalitis).

**Relative diagnostic exclusion features include:**

1. typical history of chronic alcoholism
2. sustained hypertension
3. history of vascular disease (such as angina, claudication).
### APPENDIX 4F: Definitions, Classifications and Diagnostic Guidelines for Four Major Types of Dementia based on the ICD-10 (WHO, 1992) and the ICD-10-AM Mental Health Manual (2002)

<table>
<thead>
<tr>
<th>Definitions and Classifications</th>
<th>Diagnostic Guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dementia in Alzheimer's disease</strong>&lt;br&gt;Alzheimer's disease is a primary degenerative cerebral disease of unknown etiology with characteristic neuropathological and neurochemical features. The disorder is usually insidious in onset and develops slowly but steadily over a period of several years. <strong>Dementia in Alzheimer's disease with early onset</strong>&lt;br&gt;Dementia in Alzheimer's disease with onset before the age of 65, with a relatively rapid deteriorating course and with marked multiple disorders of the higher cortical functions. <strong>Alzheimer's disease, type 2</strong>&lt;br&gt;Presenile dementia, Alzheimer's type</td>
<td><strong>Dementia in Alzheimer's disease</strong>&lt;br&gt;The essential features for a definite diagnosis of AD include:&lt;br&gt;(a) Presence of a dementia as described above (Table in Appendix 4B);&lt;br&gt;(b) Insidious onset with slow deterioration. While the onset usually seems difficult to pinpoint in time, realisation by others that the defects exist may come suddenly. An apparent plateau may occur in the progression;&lt;br&gt;(c) Absence of clinical evidence, or findings from special investigations, to suggest that the mental state may be due to other systemic or brain disease which can induce a dementia (e.g. hypothyroidism, hypercalcemia, vitamin B12 deficiency, niacin deficiency, neurosyphilis, normal pressure hydrocephalus, or subdural haematoma);&lt;br&gt;(d) Absence of a sudden, apoplectic onset, or of neurological signs of focal damage such as hemiparesis, sensory loss, visual field defects, and incoordination occurring early in the illness (although these phenomena may be superimposed later).&lt;br&gt;In a certain proportion of cases, the features of Alzheimer's disease and vascular dementia may both be present. In such cases, double diagnosis (and coding) should be made. When the vascular dementia precedes the Alzheimer's disease, it may be impossible to diagnose the latter on clinical grounds.&lt;br&gt;Includes: primary degenerative dementia of the Alzheimer's type <strong>Dementia in Alzheimer's disease with late onset</strong>&lt;br&gt;Dementia in Alzheimer's disease with onset after the age of 65, usually in the late 70s or thereafter, with a slow progression, and with memory impairment as the principal feature.&lt;br&gt;<strong>Alzheimer's disease, type 1</strong>&lt;br&gt;Primary degenerative dementia of the Alzheimer's type, senile onset</td>
</tr>
<tr>
<td><strong>Dementia in Alzheimer's disease, unspecified</strong></td>
<td><strong>Dementia in Alzheimer's disease</strong>&lt;br&gt;The essential features for a definite diagnosis of AD include:&lt;br&gt;(a) Presence of a dementia as described above (Table in Appendix 4B);&lt;br&gt;(b) Insidious onset with slow deterioration. While the onset usually seems difficult to pinpoint in time, realisation by others that the defects exist may come suddenly. An apparent plateau may occur in the progression;&lt;br&gt;(c) Absence of clinical evidence, or findings from special investigations, to suggest that the mental state may be due to other systemic or brain disease which can induce a dementia (e.g. hypothyroidism, hypercalcemia, vitamin B12 deficiency, niacin deficiency, neurosyphilis, normal pressure hydrocephalus, or subdural haematoma);&lt;br&gt;(d) Absence of a sudden, apoplectic onset, or of neurological signs of focal damage such as hemiparesis, sensory loss, visual field defects, and incoordination occurring early in the illness (although these phenomena may be superimposed later).&lt;br&gt;In a certain proportion of cases, the features of Alzheimer's disease and vascular dementia may both be present. In such cases, double diagnosis (and coding) should be made. When the vascular dementia precedes the Alzheimer's disease, it may be impossible to diagnose the latter on clinical grounds.&lt;br&gt;Includes: primary degenerative dementia of the Alzheimer's type <strong>Dementia in Alzheimer's disease with late onset</strong>&lt;br&gt;Dementia in Alzheimer's disease with onset after the age of 65, usually in the late 70s or thereafter, with a slow progression, and with memory impairment as the principal feature.&lt;br&gt;<strong>Alzheimer's disease, type 1</strong>&lt;br&gt;Primary degenerative dementia of the Alzheimer's type, senile onset</td>
</tr>
</tbody>
</table>
| **Vascular dementia**<br>Vascular dementia is the result of infarction of the brain due to vascular disease, including hypertensive cerebrovascular disease. The infarcts are usually small but cumulative in their effect. Onset is usually in later life. | **Vascular dementia**<br>The diagnosis presupposes the presence of a dementia as described above. Impairment of cognitive function is commonly uneven, so that there may be memory loss, intellectual impairment, and focal neurological signs. Insight and judgement may be relatively well preserved. An abrupt onset or a stepwise deterioration, as well as the presence of focal...
Vascular dementia of acute onset
Usually develops rapidly after a succession of strokes from cerebrovascular thrombosis, embolism or haemorrhage. In rare cases, a single large infarction may be the cause.

Multi-infarct dementia
Gradual in onset, following a number of transient ischaemic episodes which produce an accumulation of infarcts in the cerebral parenchyma. Predominantly cortical dementia

Subcortical vascular dementia
Includes cases with a history of hypertension and foci of ischaemic destruction in the deep white matter of the cerebral hemispheres. The cerebral cortex is usually preserved and this contrasts with the clinical picture which may closely resemble that of dementia in Alzheimer's disease.

Mixed cortical and subcortical vascular dementia

Other vascular dementia

Vascular dementia, unspecified

Dementia in Pick's disease
The following features are required for a definite diagnosis:
(a) A progressive dementia;
(b) A predominance of frontal lobe features with euphoria, emotional blunting, and coarsening of social behaviours, disinhibition, and either apathy or restlessness;
(c) Behavioural manifestations, which commonly precede frank memory impairment.

Dementia in Parkinson's disease
Dementia developing in an individual with advanced, usually severe, Parkinson's disease.

Dementia in:  
- paralysis agitans
- parkinsonism

neurological signs and symptoms, increases the probability of the diagnosis; in some cases, confirmation can be provided only by computerised axial tomography or, ultimately, neuropathological examination. Associated features are: hypertension, carotid bruit, emotional lability with transient depressive mood, weeping or explosive laughter, and transient episodes of clouded consciousness or delirium, often provoked by further infarction. Personality is believed to be relatively well preserved, but personality changes may be evident in a proportion of cases with apathy, disinhibition, or accentuation of previous traits such as egocentricity, paranoid attitudes, or irritability.

Includes: arteriosclerotic dementia
### APPENDIX 4G: ICF components and domains, with examples of contents

<table>
<thead>
<tr>
<th>Component</th>
<th>Domains / Chapter headings</th>
</tr>
</thead>
</table>
| **Body Functions: eight chapters** | • Mental functions e.g. memory function, intellectual functions  
• Sensory functions and pain e.g. hearing function, smell function  
• Voice and speech functions e.g. articulation functions  
• Functions of the cardiovascular, haematological, immunological and respiratory systems e.g. blood pressure functions, respiratory muscle functions  
• Functions of the digestive, metabolic and endocrine systems e.g. ingestion functions, endocrine gland functions  
• Genitourinary and reproductive functions e.g. menstruation functions  
• Neuromusculoskeletal and movement-related functions e.g. mobility of joint functions  
• Functions of the skin and related structures e.g. repair functions of the skin                                                                                                                                                                                                                                                                                                                                 |
| **Body Structures: eight chapters** | • Structures of the nervous system e.g. spinal cord and related structures  
• The eye, ear and related structures e.g. structure of eyeball, structure of inner ear  
• Structures involved in voice and speech e.g. structure of mouth  
• Structures of the cardiovascular, immunological and respiratory systems  
• Structures related to the digestive, metabolic and endocrine systems e.g. structure of intestine, structure of gall bladder and ducts  
• Structures related to the genitourinary and reproductive systems e.g. structure of the urinary system, structure of pelvic floor  
• Structures related to movement e.g. structure of head and neck region  
• Skin and related structures e.g. structure of skin glands                                                                                                                                                                                                                                                                                                                                 |
| **Activities & Participation: nine chapters** | • Learning and applying knowledge e.g. learning to read, solving problems  
• General tasks and demands e.g. carrying out daily routine  
• Communication e.g. speaking, conversation  
• Mobility e.g. getting around inside or outside home  
• Self-care e.g. washing oneself, dressing  
• Domestic life e.g. preparing meals, acquiring a place to live  
• Interpersonal interactions and relationships e.g. relating with strangers, formal relationships, family relationships  
• Major life areas e.g. work and employment, remunerative employment  
• Community, social and civic life e.g. recreation and leisure, religion and spirituality                                                                                                                                                                                                                                                                                                                                 |
| **Environmental Factors: five chapters** | • Products and technology e.g. products and technology for communication; design, construction and building products and technology of buildings for public use  
• Natural environment and human-made changes to environment e.g. physical geography, light, sound, air quality  
• Support and relationships e.g. immediate family, health professionals  
• Attitudes e.g. individual attitude of friends, individual attitude of health professionals  
• Services, systems and policies e.g. social security services, systems and policies                                                                                                                                                                                                                                                                                                                                 |

APPENDIX 4H: The Hachinski Ischaemic Scale (Hachinski, et al. 1975)

These criteria for vascular dementia are based upon the Hachinski ischaemic score, originally derived on the basis of cerebral blood flow patterns in people with dementia. On the weighted scale a score of 7 or more is taken to indicate vascular dementia while a score of 4 or less suggests that this is an unlikely diagnosis.

**Hachinski ischaemic score**
- Abrupt onset 2
- Stepwise progression 1
- Fluctuating course 2
- Nocturnal confusion 1
- Relative preservation of personality 1
- Depression 1
- Somatic complaints 1
- Emotional incontinence 1
- History of hypertension 1
- History of strokes 2
- Evidence of associated atherosclerosis 1
- Focal neurological symptoms 2
- Focal neurological signs 2
### APPENDIX 4I: Short form of the Informant Questionnaire on Cognitive Decline in the Elderly (Short IQCODE)

Created by F. Jorm, Centre for Mental Health Research, The Australian National University, Canberra, Australia [www.anu.edu.au/iqcode].

In research studies the IQCODE is preceded by questions on the subject’s sociodemographic characteristics and physical health. There is no copyright on the Short IQCODE, but please keep the author informed of research projects which make use of it.

Now we want you to remember what your friend or relative was like 10 years ago and to compare it with what he/she is like now. 10 years ago was in 19__. Below are situations where this person has to use his/her memory or intelligence and we want you to indicate whether this has improved, stayed the same or got worse in that situation over the past 10 years. Note the importance of comparing his/her present performance with 10 years ago. So if 10 years ago this person always forgot where he/she had left things, and he/she still does, then this would be considered ‘Hasn’t changed much’. Please indicate the changes you have observed by circling the appropriate answer.

#### Compared with 10 years ago how is this person at:

<table>
<thead>
<tr>
<th>1. Remembering things about family and friends eg occupations, birthdays, addresses</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much improved</td>
<td>A bit improved</td>
<td>Not much change</td>
<td>A bit worse</td>
<td>Much worse</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2. Remembering things that have happened recently</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much improved</td>
<td>A bit improved</td>
<td>Not much change</td>
<td>A bit worse</td>
<td>Much worse</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3. Recalling conversations a few days later</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much improved</td>
<td>A bit improved</td>
<td>Not much change</td>
<td>A bit worse</td>
<td>Much worse</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4. Remembering his/her address and telephone number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much improved</td>
<td>A bit improved</td>
<td>Not much change</td>
<td>A bit worse</td>
<td>Much worse</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5. Remembering what day and month it is</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much improved</td>
<td>A bit improved</td>
<td>Not much change</td>
<td>A bit worse</td>
<td>Much worse</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6. Remembering where things are usually kept</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much improved</td>
<td>A bit improved</td>
<td>Not much change</td>
<td>A bit worse</td>
<td>Much worse</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>7. Remembering where to find things which have been put in a different place from usual</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much improved</td>
<td>A bit improved</td>
<td>Not much change</td>
<td>A bit worse</td>
<td>Much worse</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>8. Knowing how to work familiar machines around the house</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much improved</td>
<td>A bit improved</td>
<td>Not much change</td>
<td>A bit worse</td>
<td>Much worse</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>9. Learning to use a new gadget or machine around the house</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much improved</td>
<td>A bit improved</td>
<td>Not much change</td>
<td>A bit worse</td>
<td>Much worse</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>10. Learning new things in general</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much improved</td>
<td>A bit improved</td>
<td>Not much change</td>
<td>A bit worse</td>
<td>Much worse</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>11. Following a story in a book or on TV</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much improved</td>
<td>A bit improved</td>
<td>Not much change</td>
<td>A bit worse</td>
<td>Much worse</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>12. Making decisions on everyday matters</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much improved</td>
<td>A bit improved</td>
<td>Not much change</td>
<td>A bit worse</td>
<td>Much worse</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>13. Handling money for shopping</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much improved</td>
<td>A bit improved</td>
<td>Not much change</td>
<td>A bit worse</td>
<td>Much worse</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>14. Handling financial matters eg the pension, dealing with the bank</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much improved</td>
<td>A bit improved</td>
<td>Not much change</td>
<td>A bit worse</td>
<td>Much worse</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>15. Handling other everyday arithmetic problems eg knowing how much food to buy, knowing how long between visits from family or friends</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much improved</td>
<td>A bit improved</td>
<td>Not much change</td>
<td>A bit worse</td>
<td>Much worse</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>16. Using his/her intelligence to understand what’s going on and to reason things through</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much improved</td>
<td>A bit improved</td>
<td>Not much change</td>
<td>A bit worse</td>
<td>Much worse</td>
<td></td>
</tr>
</tbody>
</table>
# APPENDIX 4J: Different Stages of the EADC MCI Diagnostic Procedure

<table>
<thead>
<tr>
<th>First Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Medical Practitioners</td>
</tr>
<tr>
<td>Neurologist, Geriatricians</td>
</tr>
</tbody>
</table>

“MCI syndrome”?

1) cognitive complaint emanating from the patient and/or his/her family,
2) the subject and/or informant report a decline in cognitive functioning relative to previous abilities during the past year,
3) cognitive disorders evidenced by clinical evaluation: impairment in memory and/or another cognitive domain,
4) cognitive impairment does not have major repercussions on daily life. However, the subject may report difficulties concerning complex day-to-day activities,
5) no dementia

<table>
<thead>
<tr>
<th>Second Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurologist, Geriatricians</td>
</tr>
<tr>
<td>Memory Clinics</td>
</tr>
</tbody>
</table>

Which **MCI syndrome subtype**?

<table>
<thead>
<tr>
<th>Third Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory Clinics</td>
</tr>
</tbody>
</table>

Which underlying **aetiopathogenic** sub-type?
## APPENDIX 4K: Clinical Dementia Rating Scale

<table>
<thead>
<tr>
<th>Clinical Dementia Rating (COR)</th>
<th>0</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Memory</strong></td>
<td>None: No memory loss or slight inconsistent forgetfulness</td>
<td>Questionable: Consistent slight forgetfulness; partial recall of events; “fogging” forgetfulness</td>
<td>Moderate: Moderate memory loss; more marked for recent events; deficit interferes with everyday activities</td>
<td>Severe: Severe memory loss; only highly learned material retained; new material rapidly lost</td>
<td>Severe: Severe memory loss; only fragments remain</td>
</tr>
<tr>
<td>Orientation</td>
<td>Fully oriented</td>
<td>Fully oriented except for slight difficulty with time relationships</td>
<td>Moderate difficulty with time relationships; oriented for place of examination; may have geographic disorientation elsewhere</td>
<td>Severe difficulty with time relationships; usually disoriented to time, often to place</td>
<td>Oriented to person only</td>
</tr>
<tr>
<td>Judgement &amp; Problem Solving</td>
<td>Solves everyday problems and handles business and financial affairs well; judgement good in relation to past performance</td>
<td>Slight impairment in solving problems, similarities and differences</td>
<td>Moderate difficulty in handling problems, similarities and differences; social judgement usually maintained</td>
<td>Severely impaired in handling problems, similarities and differences; social judgement usually impaired</td>
<td>Unable to make judgements or solve problems</td>
</tr>
<tr>
<td>Community Affairs</td>
<td>Independent function at usual level in job, shopping, volunteer and social groups</td>
<td>Slight impairment in these activities</td>
<td>Unable to function independently at these activities although may still be engaged in some; appears normal to casual inspection</td>
<td>No pretence of independent function outside home</td>
<td>Appears too ill to be taken to functions outside a family home</td>
</tr>
<tr>
<td>Home and Hobbies</td>
<td>Life at home; hobbies and intellectual interests well maintained</td>
<td>Life at home, hobbies and intellectual interests slightly impaired</td>
<td>Mild but definite impairment of function at home; more difficult chores abandoned; more complicated hobbies and interests abandoned</td>
<td>Only simple chores preserved; very restricted interests, poorly maintained</td>
<td>No significant function in home</td>
</tr>
<tr>
<td>Personal Care</td>
<td>Fully capable of self-care</td>
<td>Needs prompting</td>
<td>Requires assistance in dressing, hygiene, keeping of personal effects</td>
<td>Requires much help with personal care; frequent incontinence</td>
<td></td>
</tr>
</tbody>
</table>

Score only as decline from previous usual level due to cognitive loss, not impairment due to other factors.

©Copyright by Washington University, St. Louis, MO.
## APPENDIX 4L: The Global Deterioration Scale for Assessment of Primary Degenerative Dementia (Reisberg, 1982; 2000)

<table>
<thead>
<tr>
<th>Level</th>
<th>Clinical Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No cognitive decline</td>
</tr>
<tr>
<td></td>
<td>No subjective complaints of memory deficit. No memory deficit evident on clinical interview.</td>
</tr>
<tr>
<td>2</td>
<td>Very mild cognitive decline (Age Associated Memory Impairment)</td>
</tr>
<tr>
<td></td>
<td>Subjective complaints of memory deficit, most frequently in following areas: (a) forgetting where one has placed familiar objects; (b) forgetting names one formerly knew well. No objective evidence of memory deficit on clinical interview. No objective deficits in employment or social situations. Appropriate concern with respect to symptomatology.</td>
</tr>
<tr>
<td>3</td>
<td>Mild Cognitive decline (Mild Cognitive Impairment)</td>
</tr>
<tr>
<td></td>
<td>Earliest clear-cut deficits. Manifestations in more than one of the following areas: (a) patient may have gotten lost when travelling to an unfamiliar location; (b) co-workers become aware of patient's relatively poor performance; (c) word and name finding deficit becomes evident to intimates; (d) patient may read a passage or a book and retain relatively little material; (e) patient may demonstrate decreased facility in remembering names upon introduction to new people; (f) patient may have lost or misplaced an object of value; (g) concentration deficit may be evident on clinical testing. Objective evidence of memory deficit obtained only with an intensive interview. Decreased performance in demanding employment and social settings. Denial begins to become manifest in-patient. Mild to moderate anxiety accompanies symptoms.</td>
</tr>
<tr>
<td>4</td>
<td>Moderate cognitive decline (Mild Dementia)</td>
</tr>
<tr>
<td></td>
<td>Clear-cut deficit on careful clinical interview. Deficit manifest in following areas: (a) decreased knowledge of current and recent events; (b) may exhibit some deficit in memory of one's personal history; (c) concentration deficit elicited on serial subtractions; (d) decreased ability to travel, handle finances, etc. Frequently no deficit in following areas: (a) orientation to time and place; (b) recognition of familiar persons and faces; (c) ability to travel to familiar locations. Inability to perform complex tasks. Denial is dominant defence mechanism. Flattening of affect and withdrawal from challenging situations frequently occur.</td>
</tr>
<tr>
<td>5</td>
<td>Moderately severe cognitive decline (Moderate Dementia)</td>
</tr>
<tr>
<td></td>
<td>Patient can no longer survive without some assistance. Patient is unable during interview to recall a major relevant aspect of their current lives, e.g., an address or telephone number of many years, the names of close family members (such as grandchildren), the name of the high school or college from which they graduated. Frequently some disorientation to time (date, day of week, season, etc.) or to place. An educated person may have difficulty counting back from 40 by 4s or from 20 by 2s. Persons at this stage retain knowledge of many major facts regarding themselves and others. They invariably know their own names and generally know their spouse's and children's names. They require no assistance with toileting and eating, but may have some difficulty choosing the proper clothing to wear.</td>
</tr>
<tr>
<td>6</td>
<td>Severe cognitive decline (Moderately Severe Dementia)</td>
</tr>
<tr>
<td></td>
<td>May occasionally forget the name of the spouse upon whom they are entirely dependent for survival. Will be largely unaware of all recent events and experiences in their lives. Retain some knowledge of their past lives but this is very sketchy. Generally unaware of their surroundings, the year, the season, etc. May have difficulty counting from 10, both backward and, sometimes, forward. Will require some assistance with activities of daily living, e.g., may become incontinent, will require travel assistance but occasionally will be able to travel to familiar locations. Diurnal rhythm frequently disturbed. Almost always recall their own name. Frequently continue to be able to distinguish familiar from unfamiliar persons in their environment. Personality and emotional changes occur. These are quite variable and include: (a) delusional behaviour, e.g., patients may accuse their spouse of being an impostor, may talk to imaginary figures in the environment, or to their own reflection in the mirror; (b) obsessive symptoms, e.g., person may continually repeat simple cleaning activities; (c) anxiety symptoms, agitation, and even previously nonexistent violent behaviour may occur; (d) cognitive abulia, i.e., loss of willpower because an individual cannot carry a thought long enough to determine a purposeful course of action.</td>
</tr>
<tr>
<td>7</td>
<td>Very severe cognitive decline (Severe Dementia)</td>
</tr>
<tr>
<td></td>
<td>All verbal abilities are lost over the course of this stage. Frequently there is no speech at all -only unintelligible utterances and rare emergence of seemingly forgotten words and phrases. Incontinent of urine, requires assistance toileting and feeding. Basic psychomotor skills, e.g., ability to walk, are lost with the progression of this stage. The brain appears to no longer be able to tell the body what to do. Generalized rigidity and developmental neurologic reflexes are frequently present.</td>
</tr>
</tbody>
</table>

©2000 Barry Reisberg, MD. Reproduced with permission.
Appendix 5: Reviews of Dementia Staging and Descriptive Instruments
Appendix 5.1: Global Deterioration Scale

AHOC INSTRUMENT REVIEW SHEET

Title: Global Deterioration Scale.

Abbreviations: GDS.

Author(s) Name: Dr Barry Reisberg, Steven H. Ferris, Mony J. De Leon and Thomas Crook.

Author(s) Address: Dr Barry Reisberg (1982):
Department of Psychiatry,
New York University Medical Center
550 First Ave., New York, NY 10016

Supplied by: www.geriatric-resources.com/html/gds.html

Cost: Nil.

Training requirements: Clinician/caregiver training is usual, but not complicated.

Purpose: To identify the presence and severity/degree of impairment in dementia.

Administration time: 5-10 minutes.

Instrument Type: Clinical rating scale for dementia, global severity/level of impairment scale.

Structure: The GDS is the main part of a clinical rating system called the Global Deterioration Scale Staging System. Three independent measures are included in the Staging System: the GDS, the Brief Cognitive Rating Scale (BCRS) and the Functional Assessment Staging System (FAST). These provide caregivers with an overview of the stages of cognitive decline in those suffering from a primary degenerative dementia such as Alzheimer’s disease. The GDS is broken down into 7 different stages based on severity. Each stage is numbered (1-7) and given a short title (e.g. No Cognitive Decline, Mild Cognitive Decline etc.) followed by a description of the characteristics for that stage.

Scoring: The rating is based on semi-structured interviews with the patient and the caregiver. The interview covers increasingly worse cognitive deficits and the stage which reflects the cognitive capacity of the patient then becomes the rating/ score (McKeith, et al. 1999). Stages 1-3 are the pre-dementia stages in which the person is able to function quite well in their daily lives, or as usual. Stages 4-7 are the stages which reveal loss of cognitive and other functions that are needed for successful living. Once at stage 5, an individual can no longer live without assistance. For a more detailed description of each stage see Reisberg, et al. (1982) or www.geriatricresources.com/html/gds.html.

Developed for: The GDS allows caregivers and clinical staff to rate the patient’s dementia severity by observing their behavioural characteristics. For more specific assessments, use of the accompanying Brief Cognitive Rating Scale (BCRS) and the Functional Assessment Staging (FAST) measures is recommended.

Normative Data: Choi, et al. (2003) administered the GDS to 34 non-demented control subjects. Heun, et al. (1998) obtained normative data from 287 subjects from a general population sample aged from 60 to 100 years.
Clinical Data:

Applications:
Bakker, et al. (2004) identified the GDS as a prognostic measure for psychiatric function disorders, including paranoia and somatic co-morbidity. It is widely employed by health clinicians and available for use by caregivers with different levels of expertise in Europe (Ramirez Diaz, et al. 2005) and in Australia. It is used to assess cognition, behavioural disturbance and activities of daily living, and to identify the presence and severity of dementia according to descriptors of 7 stages ranging from normal cognition to advanced dementia. The items listed and the possible responses are easy to interpret and score.

Carer and/or Patient Use of Instrument:
Assessment and scoring of person with dementia by clinician/caregiver.

Psychometric Criteria

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td></td>
<td>□ Alpha &gt;0.7 □ Marginal or inadequate internal consistency (&lt;0.70) X No information found on internal consistency</td>
<td>No information on internal consistency or Cronbach’s alpha found. The instrument’s design is such that internal consistency is not able to be measured.</td>
</tr>
<tr>
<td>Test – retest</td>
<td>Reisberg, et al. (1996)</td>
<td>X ICC &gt;.70 Time intervals and confidence intervals reported □ Marginal or inadequate test-retest reliability ICC&lt;.70 □ No information found on test-retest reliability</td>
<td>Test-retest reliability has been demonstrated regularly at over 0.90. Two independent inter-rater reliability studies were conducted: Study 1 in a</td>
</tr>
<tr>
<td>Inter – rater</td>
<td>Reisberg, et al. (1996)</td>
<td>X Agreement reported and adequate □ Inadequate inter-rater agreement □ No information provided</td>
<td>Inter-rater reliability has regularly been demonstrated to be over 0.90. Two independent inter-rater reliability studies were conducted: Study 1 in a</td>
</tr>
</tbody>
</table>
Two different raters used the Global Deterioration Scale and the Mini-Mental State Examination to assess dementia severity. Study 1 (N=22) resulted in an intraclass correlation coefficient (ICC) of .99 (p< .01) against the M-OSPD total score. Study 2 (N= 19) resulted in an ICC of .96 (p= .01) against the M-OSPD total score.

Between the four different raters, kappa values ranged from .93 to 1.00, revealing high inter-rater reliability.

Testing inter-rater reliability of the GDS in a sample of 43 subjects with probable Alzheimer Disease resulted in an ICC of .82.

### VALIDITY

<table>
<thead>
<tr>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall, et al. (1990)</td>
<td>X Patients and experts were involved during item selection and/or item reduction</td>
<td>Content validity was established by constructing a 30-item questionnaire which, following a principal components analysis, the items sorted themselves into seven groups corresponding with the GDS stages.</td>
</tr>
<tr>
<td>Bakker, et al. (2004)</td>
<td>□ Patients were consulted for reading and comprehension</td>
<td>Caregivers and clinicians were involved in development and testing, and the GDS rated highly against other measures for content of items.</td>
</tr>
<tr>
<td>Reisberg, et al. (1982)</td>
<td>□ No information found on content validity</td>
<td>The GDS was designed to, and is able to accurately outline all the stages of primary degenerative dementia from the earliest to the most severe signs of cognitive decline.</td>
</tr>
<tr>
<td>Auer &amp; Reisberg (1996)</td>
<td>X Results were acceptable in accordance with the hypotheses and an adequate comparison</td>
<td>Differences in scores on the Alzheimer’s Disease Assessment Scale-Cognitive Function, Clinician’s Interview-Based impression of Change</td>
</tr>
</tbody>
</table>

Content

- The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire

- Overall, et al. (1990)


- Solomon, et al. (1999)

- Reisberg, et al. (1982)

Construct

- The extent to which scores on the questionnaire relate to other measures in a manner

- Auer & Reisberg (1996)
that is consistent with theoretically derived hypothesis concerning the domains that are measured. A factor analysis of a 30-item questionnaire found the items to fall into 7 clusters which corresponded with the stages in the GDS.

<table>
<thead>
<tr>
<th>Construct: Internal Structure</th>
<th>Overall, et al. (1990)</th>
<th>□ No evidence provided/failed a test of dimensionality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information provided on factor structure</td>
<td>□ Some evidence provided to support internal structure</td>
<td></td>
</tr>
<tr>
<td></td>
<td>□ Substantial evidence provided to support internal structure</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>that is consistent with theoretically derived hypothesis concerning the domains that are measured.</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ Limited/inadequate construct validity reported</td>
</tr>
<tr>
<td>□ No information provided</td>
</tr>
</tbody>
</table>

Auer, et al. (1996)


measure was used (with both clinical and caregiver information considered), Progressive Deterioration Scale, Mini-Mental State Examination (MMSE) and the Global Deterioration Scale were assessed. The panel determined that statistically significant differences in scores on all scales except the MMSE were likely associated with functional or cognitive differences that were clinically relevant for patients, reflecting stabilization that would have beneficial consequences for caregivers and health care resource use.

The study population consisted of cognitively normal individuals and dementia patients. The GDS was evaluated highly against the Empirical Behavioural Pathology in Alzheimer’s Disease Rating Scale (E-BEHAVE-AD) (employing caregiver-based behavioural assessment and a mental status assessment (Mini-Mental State Examination). The study aimed to estimate life expectancy of psychogeriatric patients and to identify prognostic characteristics – on admission – for survival after discharge. The functional assessments consisted of the Global Deterioration Scale (GDS), the Help Index (HI) and the Activities of Daily Life (ADL). The diagnostic characteristics were assessed by two experts (a psychogeriatrician and a clinical psychologist). These experts completed a standardized Functional Assessment List (FAL) based on the DSM-IV and ICD-9. The GDS was of prognostic significance (HR=1.58: 95%CI: 1.11-2.23) (HR = hazard ratio).
### Construct: Correlation with other measures

Comparisons made to other measures

<table>
<thead>
<tr>
<th>Authors</th>
<th>Correlations with other measures</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solomon, et al. (1999)</td>
<td>X Correlations with other measures are reported</td>
<td>The relationships between the test scores of 100 successive admissions of patients who were screened with each of five dementia instruments as part of an initial evaluation (MMSE, BIMC, ADAS-Cog, ADL, and GDS) were calculated. Correlations between the MMSE, BIMC, ADAS-Cog and GDS were all high and statistically significant ((p&lt;0.001)). The range of correlation coefficients for the GDS was (r = 0.81-0.87). These high intercorrelations suggest that each of listed instruments is measuring similar cognitive functions.</td>
</tr>
<tr>
<td>Reisberg, et al. (1982)</td>
<td>□ Correlations not reported</td>
<td></td>
</tr>
<tr>
<td>McDowell (2006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Choi, et al. (2003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Korner, et al. (1996)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Construct: Discriminant Validity

The scale differentiates between relevant categories of respondent e.g. sick vs. well, varying degrees of severity

<table>
<thead>
<tr>
<th>Authors</th>
<th>Scale differentiates between relevant categories of respondents</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bakker, et al. (2004)</td>
<td>X</td>
<td>The average score on the GDS was 4.2 (SD=1.3). The average score on the Activities of Daily Living scale (ADL) was 2.4 (SD=1.7) Patients with dementia who scored higher in the GDS had a lower probability of survival, than patients who scored lower on the GDS. After adjusting for other patient characteristics such as age, gender and type of discharge.</td>
</tr>
<tr>
<td>Heun, et al. (1998)</td>
<td>□ No information on discriminant validity</td>
<td>Using Receiver Operating Characteristics (ROC) analysis a threshold score of ≤3 was</td>
</tr>
</tbody>
</table>
The GDS was found to be the best discriminator between dementia patients and non-demented controls, with 53% sensitivity and 97% specificity. The threshold for optimal sensitivity was a score of ≤2, with 94% sensitivity but only 49% specificity. The authors interpreted this as adequate validity to detect dementia in the general population.

### Criterion

<table>
<thead>
<tr>
<th>Information on the relationship of scores to gold standard measures or clinical diagnosis is provided</th>
<th>Solomon, et al. (1999)</th>
<th>X Comparison made to criterion measures □ No comparison with criterion measures provided</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reisberg et al. (1982)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Inter-correlations between the GDS and other standard measures, including the Mini-Mental State Examination (MMSE), Alzheimer’s Disease Assessment Scale (ADAS), and the Blessed Information-Memory-Concentration Scale (BIMC), were high and allowed for conversion between test scores to help clinicians stage, diagnose and measure the rate of progression in dementia.

The GDS was significantly correlated with CAT scan measures of ventricular dilation ($r = .62$), and CAT scan cortical assessments of sulcal enlargement ($r = .53$). There was also significant relationships between PET scans indicating lower rates of metabolism corresponded with cognitive decline as indicated by the GDS ($r = .69-.83$).

### Interpretability

<table>
<thead>
<tr>
<th>The degree to which one can assign qualitative meaning to quantitative scores. Do authors provide the following:</th>
<th>Reisberg, et al. (1988)</th>
<th>□ Authors provide 2 or more types of information on interpretability X Authors provide limited information to assist with interpretability □ No information provided</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presentation of means and SD of scores before and after treatment Comparative data on the distribution of scores in relevant subgroups Information on the relationship of scores to well-known functional measures or clinical diagnosis Information on the association between changes in scores and patients’ global ratings of the magnitude of change they have experienced.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The GDS is made up of detailed clinical descriptions of seven major stages in dementia, ranging from normal cognition and functioning in daily living to very severe dementia and little or no ability to function in daily life without assistance. These descriptors are scored from 1 to 7, depending on demonstrated and perceived ability and functioning.
<table>
<thead>
<tr>
<th>RESPONSIVENESS</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floor and ceiling effects</td>
<td>The questionnaire fails to demonstrate a worse score in patients who clinically deteriorated and an improved score in patients who clinically improved&lt;br&gt;Authors should provide descriptive statistics of the distribution of scores</td>
<td>□ Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected&lt;br&gt;□ Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score&lt;br&gt;X No information provided on floor and ceiling effects</td>
<td>No information found on floor or ceiling effects.</td>
</tr>
<tr>
<td>Sensitivity to change</td>
<td>The ability to detect important change over time in the concept being measured</td>
<td>Solomon, et al. (1999)&lt;br&gt;Auer, et al. (1996)&lt;br&gt;Eisdorfer, et al. (1992)</td>
<td>□ Hypotheses were formulated and results were in agreement&lt;br&gt;□ An adequate metric was used (ES, SRM, comparison with external standard)&lt;br&gt;□ No information on sensitivity to change was provided&lt;br&gt;□ MCID - Information was provided about the magnitude of score differences which would be clinically meaningful&lt;br&gt;□ MCID – No information was provided.</td>
</tr>
</tbody>
</table>
Cultural Applicability and Cultural Adaptations:
A German study did not report any difficulties with translating the GDS for use in the clinical setting and showed no differences in scoring or interpretation by translation. Ramirez Diaz (2005) states that the GDS is used widely in memory clinics throughout the European Union (38% of all measures used to screen for and determine dementia severity). It has been reported in the international literature, however, that it is unknown if the measure is used in culturally diverse health populations.

Gender Appropriateness: Appropriate for use with both genders.

Age Appropriateness: The GDS has been mainly used with those 60 years and over and is appropriate for use with elderly people.

Summary:
While the face validity of the GDS is high, guidelines for assigning patients ratings has not been explained in great depth and so empirical validity is not so well-documented (Kane & Kane, 2004). Eisdorfer, et al. (1992) also pointed out that in the development of the GDS, there was no explicit discussion of how Reisberg, et al. (1982) related the stages described in the scale to the progression of dementia, instead basing the scale on descriptions of observations of dementia patients. However, the cited authors and several others attest to the clinical usefulness of the GDS. It has high score correlations for individual items and the total score against standard measures, like the Mini-Mental State Examination, in identifying the presence of dementia and the level of severity in activities of daily living and cognitive functioning. It is easy to understand and score by caregivers and clinical staff at different levels of expertise.

Burns, et al. (2004) indicates measures such as the GDS and the CDR are widely used as staging measures in descriptive and intervention studies. It is noted that specialist clinicians are less likely to use these global staging instruments than other clinical or research personal. Such instruments may not be particularly useful for fine differentiation at an early stage of dementia. However, global functional scales like the GDS and the CDR have their place in broadly describing people with dementia; particularly for research purposes and in residential care and community care settings.

Reporter: Lynn Chenoweth and Emily Sansoni

Date of report: 18/01/07

References


Appendix 5.2: Clinical Dementia Rating

AHOC INSTRUMENT REVIEW SHEET

Title: Clinical Dementia Rating.

Abbreviations: CDR.

Author(s) Name: John C Morris

Author(s) Address: J Morris
Memory and Aging project
Washington University School of Medicine
Department of Neurology
4488 Forest Park Ave., Suite 130
St. Louis, MO 63108
USA.

Supplied by: With the author’s written permission.

Cost: Free, with author’s written permission (Contact Tom Heuser, Alzheimer’s Disease Research Centre, address as above).

Training requirements: Clinical skills to elicit appropriate information and judge its relevance are required; physicians and non-physician health professionals can administer the CDR after appropriate training.

Free training is available online, training videotapes can be purchased, and on-site training (mini-fellowship) is also offered.

Purpose: To clinically stage the severity of cognitive-functional impairment in dementia by standardized and reliable means.

A version suitable for use in chronic care facility settings is also available (Marin, Flynn et al. 2001)

Administration time: 40-75 minutes.

Instrument Type: Clinical rating scale. Interviewer administered semi-structured interview of the patient and a reliable informant or collateral source (usually a close family member) to produce a clinical rating.

Structure: Impairment levels are determined in 6 cognitive-functional domains:

- Memory (15 items for informant and 10 items for patients)
- Orientation (8 items for informant and 8 items for patients)
- Judgment and Problem Solving (6 items for informant and 9 items for patients)
- Community affairs (10 items for informant)
- Home and hobbies (5 items for informant)
- Personal care (4 items for informant)

Scoring: The CDR Table provides descriptive anchors that guide the clinician in making appropriate ratings based on interview data and clinical judgment. In addition to ratings on a 5-point scale for each domain (except Personal Care, which is rated on a 4-point scale) an overall CDR score is derived by a standard algorithm. This score is useful for globally staging the level of impairment*: 0 = no impairment (normal), 0.5 = very mild/questionable, 1 = mild, 2 = moderate, 3 = severe impairment/dementia. This global score is achieved by first assessing each domain separately using the same levels (in the CDR “Table”). Individual items are not weighted.
“Notes:
- Scored as decline from previous level due to cognitive loss, not impairment due to other factors or causes.
- The clinician uses his/her judgement in assigning patients to levels, based on the standard set of information collected in the CDR “Worksheets”.

Developed for:
The CDR is used in both research and clinical settings to characterize the level of cognitive and functional performance in patients at risk for, or suspected of having, Alzheimer’s disease or other dementing disorder. It is used to clinically stage the severity of cognitive-functional impairment in dementia.

Normative Data:
Normative data for rating scales is difficult to obtain. The CDR was however used as part of the Consortium to Establish a Registry for Alzheimer’s Disease (CERAD) assessment battery in a study that obtained normative data for the rural elderly in America (Ganguli, Ratcliffe & Huff 1991).

Clinical Data:


The CDR has also been used in studies investigating the association between memory complaints, dementia stage and depression (Berg, McKeel, Miller, Baty, et al. 1993), and the association between cognition and functioning and behavioral symptoms (Tractenberg, Weiner, Cummings, Patterson, et al. 2005).

Applications:
Common applications include patient evaluation in memory assessment clinics, research studies of normal elderly and those with dementia, and clinical trials of therapeutic agents that might influence dementia progression.

Carer and/or Patient Use of Instrument:
Patient and carers provide information to the clinician.
### Psychometric Criteria

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal consistency</strong>&lt;br&gt;The extent to which items in a (sub) scale are intercorrelated; a measure of the homogeneity of a (sub)scale.&lt;br&gt;Cronbach's alpha should be between 0.70 and 0.90 for every dimension / sub-scale.</td>
<td>Marin, Flynn, et al. (2001)</td>
<td>□ Alpha &gt;0.70 □ Marginal or inadequate internal consistency (&lt;0.70) X No information found on internal consistency</td>
<td>No papers, including the original paper describing the development of the instrument, cite internal consistency evidence. The instrument’s design is such that internal consistency is not able to be measured.</td>
</tr>
<tr>
<td><strong>Test – retest</strong>&lt;br&gt;The extent to which the same results are obtained on repeated administrations of the same questionnaire when no change in physical functioning has occurred&lt;br&gt;Calculation of an intraclass correlation coefficient (ICC); and an ICC &gt; 0.70 is desired&lt;br&gt;Preferred if time interval and confidence intervals were presented</td>
<td>Marin, Flynn, et al. (2001)</td>
<td>X ICC &gt;.70 Time intervals and confidence intervals reported □ Marginal or inadequate test-retest reliability ICC&lt;.70 □ No information found on test-retest reliability</td>
<td>Good test re-test validity with ICC’s for the domains ranging from 0.86 to 0.93. Global ICC was 0.92. Testing was conducted at one month interval.</td>
</tr>
<tr>
<td>VALIDITY</td>
<td>Studies Reported &amp; References</td>
<td>Adequacy Checks</td>
<td>Comment</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------------------</td>
<td>-----------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Content       | Hughes, Berg, et al. (1982)  | □ Patients and experts were involved during item selection and/or item reduction  
□ Patients were consulted for reading and comprehension  
X No patient involvement  
□ No information found on content validity  
X There is an adequate coverage of relevant domains  
□ There is limited coverage of relevant domains | The instrument was developed by a team of physicians experienced in the field. Patients were only involved as subjects in the validation stage. |
| Construct     | Blessed, Tomlinson, et al. (1968)  
Folstein, Folstein, et al. (1975)  
Hughes, Berg, et al. (1982)  
Katzman, Brown, et al. (1983)  
Morris, Rubin, et al. (1987)  
Berg, Miller, et al. (1988)  
Faber-Langendoen, Morris, et al. (1988)  
Morris, McKeel, et al. (1988)  
Morris, Drazier, et al. (1989)  
Rubin, Morris, et al. (1989)  
Romanelli, Morris, et al. (1990)  
Welsh, Butters, et al. (1992)  
Rubin, Kinscherf, et al. (1993)  
Juva, Sulkava, et al. (1994)  
Dooneief, Marder, et al. (1996)  
Fillenbaum, Peterson, et al. (1996)  
Haroutunian, Perl, et al. (1998)  
Sano, Albert, et al. (1999)  
Marin, Flynn, et al. (2001)  
Shah, Carr, et al. | X Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used  
□ Limited/inadequate construct validity reported  
□ No information provided | Shows expected correlations with the following measures of cognitive functioning:  
Face -Hands Test (FHT) \( r = 0.57 \);  
Mini-Mental State Exam (MMSE);  
Abbreviated Mental Test (AMT);  
Short Portable mental Status Questionnaire (SPMSQ) \( r = 0.84 \);  
Short Blessed test (SBT);  
Elderly Cognitive Assessment Battery (ECAB);  
Short version of Blessed Information, Memory and Concentration Test (sBIMC);  
Dementia Scale – Cognitive (DS-C).  
Also shows expected correlations with the;  
Physical Performance test (PPT);  
Aphasia Battery (AB);  
Alzheimer's Disease Cooperative and Schwab & England (ADL);  
Visual Analogue scale (VAS);  
as well as items measuring general cognitive and physical functioning, and neuropsychological and psychopathology symptoms. |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>□ No evidence provided/failed a test of dimensionality</td>
<td>X Some evidence provided to support internal structure</td>
<td>□ Substantial evidence provided to support internal structure</td>
<td></td>
<td></td>
<td></td>
<td>The instrument has been shown to correlate with the Blessed Dementia Scale (BDS).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>X Correlations with other measures are reported</td>
<td>□ Correlations not reported</td>
<td></td>
<td>The instrument has been shown to correlate with the Blessed Dementia Scale (BDS).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>X Scale differentiates between relevant categories of respondents</td>
<td>□ No information on discriminant validity</td>
<td>The instrument has been shown to be a good predictor of dementia progression, nursing home placement, and survival.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-mortem examinations also indication CDR stage is significantly associated with hippocampal volume loss and shape, density of senile plaques, neurons in the entohortal cortex, Alzheimer’s Disease legions (both gross and microscopic) and neurofibrillary tangles.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Criterion</td>
<td>Interpretability</td>
<td>Studies report agreement with other measures (ratings determined by expert raters) ranged from 53 to 95%.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Fillenbaum, Peterson, et al. (1996)  
Morris, Ernesto, et al. (1997)  
Folstein, Folstein, et al. (1975)  
Hughes, Berg, et al. (1982)  
Katzman, Brown, et al. (1983)  
Botwinick, Storandt, et al. (1986)  
Morris, Rubin, et al. (1987)  
Botwinick, Storandt, et al. (1988)  
Burke, Miller, et al. (1988)  
Berg, Miller, et al. (1988)  
Faber-Langendoen, Morris, et al. (1988)  
Morris, McKeel, et al. (1988)  
Rubin, Morris, et al. (1989)  
Morris, Drazner, et al. (1989)  
Morris, Heyman, et al. (1989)  
McCulla, Coats, et al. (1989)  
Davis, Morris, et al. (1990)  
Romanelli, Morris, et al. (1990)  
Summers, DeBoynton, et al. (1990)  
Morris, McKeel, et al. (1991) | X Authors provide 2 or more types of information on interpretability  
□ Authors provide limited information to assist with interpretability  
□ No information provided | Studies provide considerable information about interpretability. Most provide at least 2 or more types of information. |
<p>| Welsh, Butters, et al. (1992) |
| Berg, McKeel, et al. (1993) |
| Rubin, Kinscherf, et al. (1993) |
| Juva, Sulkava, et al. (1994) |
| Dooneief, Marder, et al. (1996) |
| Fillenbaum, Peterson, et al. (1996) |
| O'Connor, Blessed, et al. (1996) |
| Morris, Storandt, et al. (1996) |
| Thal, Carta, et al. (1996) |
| Morris, Ernesto, et al. (1997) |
| Berg, McKeel, et al. (1998) |
| Haroutunian, Perl et al. (1998) |
| Rogers, Farlow, et al. (1998) |
| Sano, Albert, et al. (1999) |
| Waite, Grayson, et al. (1999) |
| Rockwood, Strang, et al. (2000) |
| Tractenberg, Schafer, et al. (2001) |
| Marin, Flynn, et al. (2001) |
| Morris, Storandt, et al. (2001) |
| Lim, Chin, et al. (2005) |
| Perneczky, Wagenpfeil, et al. (2006) |</p>
<table>
<thead>
<tr>
<th>RESPONSIVENESS</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
</table>

**Cultural Applicability and Cultural Adaptations:**

The instrument is available in Chinese for Taiwan, Czech, Dutch, Dutch for Belgium, English for Australia, English for the UK, Finnish, French, French for Belgium, French for Canada, German, German for Austria, Hebrew, Polish, Spanish, Spanish for Argentina, Swedish.

**Gender Appropriateness:**

It is appropriate for use with both genders.
Age Appropriateness: It is appropriate for use with adults over 45 years.

Summary: The CDR is a very widely used instrument and has been extensively cited. It is readily available and free to use, with the authors permission. It is administered by a clinician (either physician or other health professional). Training is required. Psychometric properties are very good and the instrument is available in numerous languages. The CDR has become one of the main global ratings of dementia used in studies investigating this disease. It is used in both research and clinical settings.

Reporter: Madeleine King and Siggi Zapart

Date of report: 17/1/2007

References


Appendix 5.3: Dementia Severity Rating Scale

AHOC INSTRUMENT REVIEW SHEET

Title: Dementia Severity Rating Scale.

Abbreviations: DSRS.

Author(s) Name: Dr Christopher M. Clark and Douglas C. Ewbank.

Author(s) Address: Dr Christopher M. Clark
Alzheimer’s Disease Core Center, and
Department of Neurology
University of Pennsylvania
Philadelphia, Pennsylvania, USA

Douglas C. Ewbank
Alzheimer’s Disease Core Center, and
Population Studies Center
University of Pennsylvania
Philadelphia, Pennsylvania, USA

Supplied by: Alzheimer’s Disease & Associated Disorders Association, USA.

Cost: Nil.

Training requirements: No training needed to be used by the caregiver, but staff may take several minutes to instruct caregiver.

Purpose: To provide a brief multiple-choice questionnaire for caregivers to assess the mildest to the most severe stages of Alzheimer Disease and other dementias in the major functional and cognitive domains. Its purpose is to identify how well the person functions in the home environment.

Administration time: 4-5 minutes.

Instrument Type: Short, multiple-choice rating scale completed by the caregiver.

Structure: The caregiver rates the subject in 11 categories. The first six categories address memory, orientation, judgment, social interaction, home activities and personal care. These mirror the items in the Clinical Dementia Rating (CDR) scale. The other five items address language, recognition, eating, incontinence and mobility.

Scoring: The total score is obtained from the summed score across all items, offering a quantitative rather than qualitative measure of dementia severity (Kane & Kane, 2000). The maximum score possible for each item ranges from 3 to 7; the minimum score for each item is 0. The minimum total score is 0 (no impairment) while the maximum total score is 51 (maximally impaired in every category). Item 11 (mobility/walking) is sometimes skipped by informants because a pre-existing mobility impairment in the subject can make it difficult to estimate how cognitive decline affects his or her mobility in the community. Users are advised to examine completion of individual items before using the composite scores.

Developed for: The DSRS was designed to be used by caregivers to provide an overall assessment of the cognitive and functional status of the person with dementia. It is suitable for multi-site collaborative studies and can be implemented with minimal staff time and training.

Normative Data: Clark & Ewbank (1996) obtained normative data from 24 control subjects.
Clinical Data: Clark & Ewbank (1996) recorded results from 165 patients with probable or definite Alzheimer's Disease.

Applications: The DSRS is used to obtain information from the caregiver about the subject's ability to function in their home environment. It is suitable for use as a first level screening measure by caregivers and clinicians, and has also been used as a baseline and outcome measure of progression of disease in clinical trials of 6-12 months duration. Tschanz, et al. (2000) used the DSRS to obtain information on clinical, medical and family history as well as ratings of impairment in functioning in a study examining the use of a neuropsychological algorithm in the classification of dementia.

Carer and/or Patient Use of Instrument: Caregiver and care staff use.

Psychometric Criteria

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal consistency</strong></td>
<td>Clark &amp; Ewbank (1996)</td>
<td>X Alpha &gt;0.70</td>
<td>Cronbach's alpha was 0.92, revealing a high degree of internal consistency.</td>
</tr>
<tr>
<td>The extent to which items in a (sub) scale are inter-correlated; a measure of the homogeneity of a (sub) scale.</td>
<td>□ Marginal or inadequate internal consistency (&lt;0.70)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cronbach's alpha should be between 0.70 and 0.90 for every dimension / sub-scale.</td>
<td>□ No information found on internal consistency</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Test – retest</strong></td>
<td>Clark &amp; Ewbank (1996)</td>
<td>X ICC &gt;.70 Time intervals and confidence intervals reported</td>
<td>High test-retest reliability coefficient of 0.90 on individual items when mean time interval between testing was 12.4 days. Kappa values of the individual items ranged from 0.46 to 0.79 and all were significantly different from 0 at the 5% level. Variance of difference between scores was 3.44 (95% CI of 1.40-5.19). With regard to its high test-retest reliability (0.90) - it is as high or higher than that of each individual test included in the Consortium to Establish a Registry for Alzheimer's disease (CERAD) battery and the MMSE.</td>
</tr>
<tr>
<td>The extent to which the same results are obtained on repeated administrations of the same questionnaire when no change in physical functioning has occurred</td>
<td>□ Marginal or inadequate internal consistency ICC&lt;.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calculation of an intraclass correlation coefficient (ICC); and an ICC &gt; 0.70 is desired</td>
<td>□ No information found on test-retest reliability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preferred if time interval and confidence intervals were presented</td>
<td><strong>Inter – rater</strong></td>
<td>Clark &amp; Ewbank (1996)</td>
<td>The DSRS has good reliability as measured by comparing caregiver responses to information collected by a doctor or an experienced</td>
</tr>
<tr>
<td>Limits of agreement, Kappa, or standard error of measurement (SEM) were</td>
<td>X Agreement reported and adequate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ Inadequate inter-rater agreement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VALIDITY</td>
<td>Studies Reported &amp; References</td>
<td>Adequacy Checks</td>
<td>Comment</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------------------</td>
<td>----------------</td>
<td>---------</td>
</tr>
<tr>
<td><strong>Content</strong></td>
<td>The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire</td>
<td>Clark &amp; Ewbank (1996)</td>
<td>□ No information provided</td>
</tr>
<tr>
<td></td>
<td>Harvey, et al. (2005)</td>
<td></td>
<td>Content of each item was evaluated highly by clinicians and caregivers when tested against the Clinical Dementia Rating Scale (DSRS score of 22 equivalent to 1.5 on CDR). In a cross sectional study with 183 caregiver/patient dyads from 12 clinical sites (in which caregivers administered 6 subscales including Instrumental ADL, Communication, Agitation, Memory and Disorganised thinking), patients and caregivers rated the usefulness of the DSRS. They found the items to be less useful in describing cognitive impairment and related function compared with the Dementia Severity Scale (DSS).</td>
</tr>
<tr>
<td><strong>Construct</strong></td>
<td>The extent to which scores on the questionnaire relate to other measures in a manner that is consistent with theoretically derived hypothesis concerning the domains that are measured.</td>
<td>Clark &amp; Ewbank (1996)</td>
<td>□ No evidence provided/failed a test of dimensionality X Some evidence provided to support internal structure □ Substantial evidence provided to support internal structure</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>The DSRS has adequate construct validity. The first 5 items of the DSRS (memory, orientation, judgement, social interaction and home activities) cover the same dimensions as the core of the CDR. Items on judgement, social interaction and home activities were most useful at the early stages of dementia. Feeding and incontinence items were most useful at later stages of dementia. Because the various items are useful at different stages, the total DSRS score is useful for people with a wide range of severity.</td>
</tr>
<tr>
<td><strong>Construct: Internal Structure</strong></td>
<td>Information provided on factor structure</td>
<td>Clark &amp; Ewbank (1996)</td>
<td>□ No evidence provided/failed a test of dimensionality X Some evidence provided to support internal structure □ Substantial evidence provided to support internal structure</td>
</tr>
<tr>
<td>Construct: Correlation with other measures</td>
<td>Karlawish, et al. (2003)</td>
<td>Correlations with other measures are reported</td>
<td>The MMSE and DSRS were strongly correlated in assessing the cognitive status and severity of dementia in community dwelling persons with dementia. The DSRS was highly correlated with the MMSE score (r = -0.77), and the combined CERAD score (r = -0.73) (p &lt; .001). As CDR ratings increased (indicating greater severity), so did DSRS scores, with the mean DSRS score for each CDR rating significantly higher than the mean DSRS score for the CDR rating below. DSRS was compared against items from the CERAD including verbal fluency, naming, word list, construction praxis, word delayed recall, word recognition and clock drawing test for patients with a MMSE of &gt;4 to reduce the floor effect on these scales. The correlation with the combined CERAD scores was -0.73.</td>
</tr>
<tr>
<td>□ Correlations not reported</td>
<td>Clark &amp; Ewbank (1996)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Construct: Discriminant Validity | Karlawish, et al. (2003) | Scale differentiates between relevant categories of respondents | Discriminates sensitively for varying degrees of dementia severity. Univariate analysis for caregivers who had a patient with more severe dementia: MMSE CI=1.07-4.90; DSRS CI = 1.39-4.64. Because the DSRS provides a quantitative rather than a qualitative assessment, the DSRS does not provide categories for differing severity of dementia, however, DSRS scores of less than or equal to 21 roughly correspond to a CDR indicating mild dementia, DSRS scores of 22-39 correspond to a CDR indicating moderate dementia, and scores of 40 or higher correspond to a CDR score indicating severe dementia. The items concerning judgement, social interaction and home activities were most useful when assessing persons in the early stages of Alzheimer’s Disease, while the |
| □ No information on discriminant validity | Clark & Ewbank (1996) |
| | Kane & Kane (2000) |
feeding and incontinence items were more useful in the later stages. This means that the DSRS is useful across a wide range of severity.

DSRS has all of the characteristics needed for a measure used in clinical trials of 6-12 months duration to score the progression of disease, by discriminating on all items relating to the absence and the severity of dementia.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Information on the relationship of scores to gold standard measures or clinical diagnosis is provided</th>
<th>Newberg, et al. (2003)</th>
<th>X Comparison made to criterion measures □ No comparison with criterion measures provided</th>
<th>DSRS and the MMSE scores correlated significantly with the Metabolic Imaging Severity Rating Scale (MISRS) which was developed for use with patients undergoing clinical assessment of cognitive impairment ($P &lt; 0.0001$). Qualitative aspects: Correlation between MISRS and MMSE ($r=0.47$) and DSRS ($r=-0.27$). Quantitative aspects: MISRS significantly correlated with MMSE ($r=0.52$) and DSRS ($r=-0.55$). When scores are compared with standard dementia measures such as the Mini-Mental State Examination (MMSE), the Consortium to Establish a Registry for Alzheimer’s disease (CERAD), and other short dementia measures, such as the Clock Drawing Test and the Word Recognition Test, there is a correlation of -0.73.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interpretability</td>
<td>The degree to which one can assign qualitative meaning to quantitative scores</td>
<td>Clark &amp; Ewbank (1996)</td>
<td>□ Authors provide 2 or more types of information on interpretability X Authors provide limited information to assist with interpretability □ No information provided</td>
<td>Caregivers find it relatively easy to rate level of severity by circling descriptors of subject’s perceived and the actual current abilities and behaviour, such as memory, mood and daily living activity. It is scored similarly by carers, doctors and nurses in a number of clinical studies and has high inter-rater reliability reported for every item. The DSRS is correlated with measures that assess severity of disease and impairment of cognitive function.</td>
</tr>
</tbody>
</table>
Information on the association between changes in scores and patients' global ratings of the magnitude of change they have experienced

<table>
<thead>
<tr>
<th>RESPONSIVENESS</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Floor and ceiling effects</strong></td>
<td>Clark &amp; Ewbank (1996)</td>
<td>X Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
<td>Except for incontinence scores, the DSRS discriminates on scores for the full range of cognitive impairment in dementia subjects (total range 2-24 for all 11 domains, average range is 0.68-0.83 for possible range of 0-5). No floor or ceiling effects noted.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information provided on floor and ceiling effects</td>
<td></td>
</tr>
<tr>
<td><strong>Sensitivity to change</strong></td>
<td>Clark &amp; Ewbank (1996)</td>
<td>□ Hypotheses were formulated and results were in agreement</td>
<td>The 11 items of the scale measure various functions that decline at different stages as the disease process, however feeding and incontinence items are more useful at later stages of illness.</td>
</tr>
<tr>
<td></td>
<td>Newberg, et al. (2003)</td>
<td>□ An adequate metric was used (ES, SRM, comparison with external standard)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information on sensitivity to change was provided</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ MCID - Information was provided about the magnitude of score differences which would be clinically meaningful</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ MCID – No information was provided.</td>
<td></td>
</tr>
</tbody>
</table>

**Cultural Applicability and Cultural Adaptations:** Mostly studies from the USA have been identified, but its simplicity may make it suitable for translation.

**Gender Appropriateness:** Appropriate for use with both genders.

**Age Appropriateness:** Appropriate for adults with mild to severe dementia.
Summary: The DSRS is a caregiver rating scale which has good psychometric properties and is also used by clinicians and researchers internationally. Its brevity and user-friendly approach make it one of the most suitable informant questionnaires to detect dementia and measure severity particularly in community settings. However, a comment made by Harvey, et al. (2005) that the “DSRS uses an inconsistent format and language, thus may be too complex for the average reader….by comparison, DSS is easily completed by informant”.

Reporter: Lynn Chenoweth and Emily Sansoni

Date of report: 18/01/07

References


Appendix 5.4: Blessed Dementia Scale

AHOC INSTRUMENT REVIEW SHEET

Title: Blessed Dementia Scale (incorporating the Blessed Information-Memory-Concentration Test and the Dementia Scale).

Abbreviations: BDS.

Author(s) Name: G Blessed, BE Tomlinson and Martin Roth.

Author(s) Address: G Blessed, BE Tomlinson and Martin Roth (1968): Medical Research Council Group on the Relation between Functional and Organic Psychiatric Illness. 11 Framlington Place, Newcastle upon Tyne 2 and The Department of Psychological Medicine, University of Newcastle upon Tyne.


Cost: Nil.

Training requirements: No specific training required, although familiarity with the assessment of function is desirable.

Purpose: Originally designed to assess the association between mental decline and pathological changes in the brain, the BDS aims to provide a quantitative assessment of the signs and severity of dementia (Lezak, et al. 2004).

Administration time: 20-30 minutes.

Instrument Type: Clinical rating scale based on a structured cognitive interview of the patient supported by an informant report for the activity and behaviour items.

Structure: The Blessed Dementia Scale (BDS) incorporates the Blessed Information-Memory-Concentration Test (BIMC) and the Dementia Scale (DS). It was designed to quantitatively assess the signs of dementia to enable comparisons to be made with pathological changes in the brain of the person with dementia. The domains assessed include orientation, long-term memory, recall, concentration and performance, as well as cognition, personality, apathy and basic self-care.

The Dementia Scale component assesses performance of everyday activities and habits over the past six months and the information is usually provided by caregivers (Lezak, et al. 2004). Four independent factors-cognition, (F1), personality (F2), apathy (F3) and basic self-care (F4) - describe constructs that may be more relevant to the clinician, patient, family caregivers, and care staff, since they describe specific behaviours/features of dementia, rather than just representing a general level of dysfunction.

The BIMC component aims to relate clinical functions of dementia to neuropathological change. This section is answered, if possible, by the patient but personal memory information must be obtained from a collateral source, such as a caregiver. It rates orientation, long-term memory, recall, concentration and performance. The six-item scale called the Blessed Orientation-Memory-Concentration Test (BOMCT) used to assess mental status is derived from the BIMC.
Scoring: The Dementia Scale assesses how well the patient copes with everyday tasks and activities (e.g. household tasks) as reported by a close relative or friend. Complete incompetence in a given activity results in a score of 1 for that item, while partial incapacity is given a score of ½ and 0 indicates full capacity. The scores for the 22 items are summed and the total can range between 0 (fully preserved capacity) and +28 (extreme incapacity). Generally, scores can be interpreted as follows: <4 = unimpaired, 4-9 = mild impairment, 10+ = moderate to severe impairment (Lezak, et al. 2004).

The BIMC contains 30 items which assess orientation, long-term memory, recent memory and concentration. A positive score is given for each item scored correctly and total scores can range between 0 (complete cognitive failure) and +37 (full cognitive capacity) (Blessed, et al. 1968).

Developed for: The BDS is used to assess both the cognitive status and the functional behaviour of dementia patients. It is useful for assessing those who can not usually be assessed with other neuropsychological measures (e.g. those who score very low, or are unable to be accurately assessed with more detailed and complex dementia measures). It is also used to provide a measure of competence in performance of daily living activities extending over relatively long periods of time. The BDS is reported to be suitable for use with patients/clients by nursing staff caring for persons with dementia in nursing homes, community and other health settings, and who have no neuropsychological training.

Normative Data: Heun, et al. (1998) obtained normative data from 287 subjects from a general population sample aged from 60 to 100 years.

Clinical Data: The studies cited below have all employed the BDS alone or in combination with the BIMC to score the cognitive function and severity of persons with suspected or actual dementia. Several other studies were located that used the BDS as one of the measures employed to test the reliability and validity, as well as the usefulness of recently developed measures of cognition, function and behaviour in persons with dementia and associated conditions, such as depression.

Villardita and Lomeo (1992) administered the BDS twice at four-week intervals to 41 patients with a diagnosis of probable AD.

Applications: The BDS differentiates between groups of demented and non-demented older persons and levels of dementia severity. Clearly identifies functional changes in performance of everyday activities, changes in habits, changes in personality, interests and drive, as well as cognitive changes occurring in processing information, total concentration, recent memory and orientation.

By tracking the behavioural changes that accompany cognitive deterioration, the BDS can measure dementia progression with repeated administrations over time (Lezak, et al. 2004). Landes, et al. (2005) found that incidence and severity of dementia would be more likely to be identified and scored using the BDS and other global measures, than measures specific to apathy, dysphoria and depression, and that the BDS would also identify these behaviours.

Carer and/or Patient Use of Instrument: Can be used as an interview with caregivers employing a structured interview method.
## Psychometric Criteria

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
</table>
| **Internal consistency** | Stern (1994) | □ Alpha >0.70  
X Marginal or inadequate internal consistency (<0.70)  
□ No information found on internal consistency | Assessing predictors of loss of social functioning and the course of AD in 233 patients in the early stages of the disease, the BDS shows marginal intra-class correlations with a Cronbach’s Alpha of 0.66. |
| | | | |
| **Test – retest** | Villardita & Lomeo (1992)  
Zilmer, et al. (1990)  
McDowell (2006) | X ICC >.70  
Time intervals and confidence intervals reported  
□ Marginal or inadequate test-retest reliability ICC<.70  
□ No information found on test-retest reliability | The test-retest reliability in a sample of 41 patients with probable AD was .89 for the BIMC and .88 for the DS at a four week interval, both significant at p<.001.  
The MMSE, BDS and BOMCT demonstrated excellent test-retest reliability (0.89 to 0.99) when used to identify cognitive disabilities and for describing changes in mental status over time in persons with dementia.  
For the DS, test-retest reliability over 4 weeks in 68 non-demented patients was .79.  
For the BIMC, test-retest reliability over 2 to 4 weeks in nursing home patients was found to be .88 and alpha was .93. |
| | | | |
| **Inter – rater** | Landes, et al. (2005)  
Stern (1994)  
McDowell (2006) | X Agreement reported and adequate  
□ Inadequate inter-rater agreement  
□ No information provided | 20 informants were independently interviewed by 2 raters. There was 100% agreement between 2 raters for dependence level.  
Comparisons of total scores across clinician raters (doctors and nurses) produced ICC of 0.90.  
For the DS only: inter-rater reliability was .59, with an intraclass correlation of .30. |
<table>
<thead>
<tr>
<th>VALIDITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Content</strong></td>
<td>Landes, et al. (2005)</td>
<td>X Patients and experts were involved during item selection and/or item reduction □ No patient involvement □ No information found on content validity</td>
<td>Interviews with caregivers, medical and nursing staff determined that the content of each item for the four factors of the DS (cognition (F1), personality (F2), apathy (F3) and self-care (F4)) compared favourably with the scores on scales measuring apathy (DAIR), behavioural symptoms (BRSD), and depression &amp; adaptive living skills (BDRS).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ There is an adequate coverage of relevant domains □ There is limited coverage of relevant domains</td>
<td></td>
</tr>
<tr>
<td><strong>Construct</strong></td>
<td>Landes, et al. (2005)</td>
<td>X Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used □ Limited/inadequate construct validity reported □ No information provided</td>
<td>BDS compares favourably for 4 factors/domains (Cognitive, Personality, Apathy/Withdrawal, Basic Self-Care) compared with MMMS and MMSE for 187 patients (aged 45-84 yrs) with dementia/probable AD.</td>
</tr>
<tr>
<td></td>
<td>Villardita &amp; Lomeo (1992)</td>
<td>□ Limited/inadequate construct validity reported □ No information provided</td>
<td>A close relationship was identified between AD patients’ scores for the BDS, Blessed IMC and the Mini Mental State Examination (MMSE), Cognitive Capacity Screening Examination (CCSE), Gottfries-Brane-Steen scale and Sandoz Clinical Assessment - Geriatric (SCAG).</td>
</tr>
<tr>
<td><strong>Construct: Internal Structure</strong></td>
<td>Lezak, et al. (2004)</td>
<td>□ No evidence provided/failed a test of dimensionality □ Some evidence provided to support internal structure □ Substantial evidence provided to support internal structure</td>
<td>A principle components analysis uncovered 4 factors in the DS: Cognitive, Personality Change, Apathy/Withdrawal, Basic Self-Care. When determining changes in cognitive function for patients with extra-pyramidal signs, a correlation of 0.5 was found between the modified MMSE (MMMS) and the DS for the 4 factors: cognition (F1), personality (F2), apathy (F3) and self-care (F4).</td>
</tr>
<tr>
<td></td>
<td>Clark, et al. (1997)</td>
<td>□ Some evidence provided to support internal structure □ Substantial evidence provided to support internal structure</td>
<td>Correlations between MMMS for all factors: Factor I r =- 0.44 (high) Factor 2 r =- 0.24 Factor 3 r =- 0.16 (weak)</td>
</tr>
</tbody>
</table>
Factor 4 $r = -0.51$ (high) (Note: Factor 3 is weak psychometrically).

Also, the order of decline over time in patients with dementia varied for the four factor scores in a sensible way, with deterioration in cognitive performance being observed much earlier than deterioration in self-care.

<table>
<thead>
<tr>
<th>Construct: Correlation with other measures</th>
<th>Davis, et al. (1990)</th>
<th>X Correlations with other measures are reported</th>
<th>Among well-characterised samples the BDS correlates well with the Clinical Dementia Rating scale (CDR) and the Short Portable Mental Status Questionnaire (SPMSQ) for all levels of dementia.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparisons made to other measures</td>
<td>Landes, et al. (2005)</td>
<td>□ Correlations not reported</td>
<td>BDS scores were compared with MMSE, CDR and medical and nursing staff’s clinical assessments to rate the cognitive &amp; functional status of 131 patients with probable/possible AD, as well as dysphoria (employing Dementia Apathy Interview Rating - DAIR) and apathy (employing Behaviour Rating Scale for Dementia - BRSD).</td>
</tr>
<tr>
<td></td>
<td>McDowell (2006)</td>
<td></td>
<td>BDS scores correlated moderately well with the MMSE and CDR (0.56) and were modestly associated with the BRSD and DAIR (0.21).</td>
</tr>
<tr>
<td></td>
<td>Villardita and Lomeo (1992)</td>
<td></td>
<td>The BIMCs correlation with: the Clinical Dementia Rating was .81, the Mental Status Questionnaire was .94, the MMSE was -.73 to -.88, and the Dementia Rating Scale was .82.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Correlation co-efficients were calculated between the BIMC and: the MMSE ($r = -.80$ to -.88), the CCSE ($r = -.73$ to -.81), the DS ($r = .58$ to .70), the Gottfries-Brane-Steen Scale (GBS) ($r = .53$ to .56) and the SCAG ($r = .47$ to .66).</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Correlation co-efficients were calculated between the DS and: the GBS ($r = .42$ to .48), the SCAG ($r = .56$ to .44), the MMSE ($r = -.67$ to -.73), the CCSE ($r = -.66$ to -.72) and the BIMC ($r = .58$ to .70).</td>
</tr>
<tr>
<td>Construct: Discriminant Validity</td>
<td>Landes, et al. (2005)</td>
<td>X Scale differentiates between relevant categories of respondents</td>
<td>The BDS discriminated well between function in Alzheimer’s disease and function in major depression, apathy and dysphoria in a series of studies with 131 participants. This was assessed by correlating scores of dementia severity (Cognitive &amp; Functional status) with the MMMS, Dementia Apathy Interview (DAIR), CERAD, Behaviour Rating Scale for Dementia (BRSD) and DSM-4. Cross sectional relationships between dependence level and severity indices comparing scores on MMMS ($r=0.27$), CDR ($r=0.34$), BDRS ($r=0.38$) and basic self-care ($r=0.26$), revealed good discrimination for these factors. Using Receiver Operating Characteristics (ROC) analysis a threshold score of ≤1 was found to be the best discriminator between dementia patients and non-demented controls, with 96% sensitivity and 82% specificity – considered to be adequate validity.</td>
</tr>
<tr>
<td>Heun, et al. (1998)</td>
<td>□ No information on discriminant validity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Criterion</td>
<td>Blessed, et al. (1968)</td>
<td>X Comparison made to criterion measures</td>
<td>Persons with dementia had statistically significant poorer mean scores on the BDS and in mean plaque counts post-mortem than patients with depression and functional psychosis. The correlation between plaque count and scores on the BDS is high ($r = .77$ for DS; $r = -.591$ for BIMC), thereby supporting the hypothesis that the BDS is a sensitive measure of dementia.</td>
</tr>
<tr>
<td>Information on the relationship of scores to gold standard measures or clinical diagnosis is provided</td>
<td>□ No comparison with criterion measures provided</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interpretability</td>
<td>Blessed, et al. (1968)</td>
<td>□ Authors provide 2 or more types of information on interpretability</td>
<td>Scores are quantified for all descriptive items on the BDS, including how well persons with dementia perform in activities of daily functioning in personal, domestic and social activities, and for performance in a number of simple psychological tests of orientation, remote memory, recent memory, and concentration.</td>
</tr>
<tr>
<td>The degree to which one can assign qualitative meaning to quantitative scores</td>
<td>X Authors provide limited information to assist with interpretability</td>
<td>□ No information provided</td>
<td></td>
</tr>
<tr>
<td>Do authors provide the following:</td>
<td>Landes, et al. (2005)</td>
<td></td>
<td>When assessing apathy and dysphoric symptoms, the mean score on the BDS was 4.92, compared with the MMSE mean score of 18.5.</td>
</tr>
<tr>
<td>Presentation of means and SD of scores before and after treatment</td>
<td>Information on the distribution of scores in relevant subgroups</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
relationship of scores to well-known functional measures or clinical diagnosis

| Information on the association between changes in scores and patients’ global ratings of the magnitude of change they have experienced |
| Zillmer, et al. (1990) |
| The standard deviation of the BDS was 2.59 compared with the MMSE standard deviation of 6.88. The six item BOMCT section of the BDS and the MMSE provide the same information about cognitive status for memory, orientation, attention and verbal comprehension, although both have limited incremental validity as an assessment of broad cognitive function. A BOMC mean score of 20.4 and SD of 8.4 is comparable with a MMSE mean score of 14.4 and SD of 10.8. |

### RESPONSIVENESS

<table>
<thead>
<tr>
<th>Floor and ceiling effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>The questionnaire fails to demonstrate a worse score in patients who clinically deteriorated and an improved score in patients who clinically improved</td>
</tr>
<tr>
<td>Authors should provide descriptive statistics of the distribution of scores</td>
</tr>
<tr>
<td>Davis, et al. (1990)</td>
</tr>
<tr>
<td>Blessed, et al. (1968)</td>
</tr>
<tr>
<td>□ Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
</tr>
<tr>
<td>X Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score</td>
</tr>
<tr>
<td>□ No information provided on floor and ceiling effects</td>
</tr>
<tr>
<td>Scores of persons with dementia assessed with the BDS and BDSC tend to cluster just below the top score for severity, and also tend to cluster at lower scores that are supposed to indicate very low or no cognitive impairment (18% of participants). Evaluating BDS scores against mean dementia scores in diagnostic groups, and mean plaque counts post-mortem, identified floor and ceiling effects.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sensitivity to change</th>
</tr>
</thead>
<tbody>
<tr>
<td>The ability to detect important change over time in the concept being measured</td>
</tr>
<tr>
<td>Zillmer, et al. (1990)</td>
</tr>
<tr>
<td>□ Hypotheses were formulated and results were in agreement</td>
</tr>
<tr>
<td>X An adequate metric was used (ES, SRM, comparison with external standard)</td>
</tr>
<tr>
<td>□ No information on sensitivity to change was provided</td>
</tr>
<tr>
<td>X MCID – Information was provided about the magnitude of score differences which would be clinically meaningful</td>
</tr>
<tr>
<td>□ MCID – No information was provided.</td>
</tr>
<tr>
<td>The BDS and the BOMCT component scored together and separately, scored strongly when detecting cognitive disabilities and changes over time, and in describing these changes, when compared with the MMSE. By tracking the behavioural changes that accompany cognitive deterioration, the BDS can measure the progression of dementia with repeated administrations over time.</td>
</tr>
</tbody>
</table>
Cultural Applicability and Cultural Adaptations: Intercultural application has been reported by a limited number of authors. Use of an appropriately trained translator is required to ensure the questions are asked as intended.

Gender Appropriateness: Appropriate for use with both genders.

Age Appropriateness: The BDS has been used with persons aged 55 and over and is appropriate for use with the elderly.

Summary: While the studies cited in this review provide convincing evidence for the utility of the BDS and the sub-component BIMC in accurately assessing the incidence and severity of dementia in a range of community and health care settings, an article by Holmes and Lovestone (2003) found in their study of 374 Alzheimer’s disease patients that the BDS and the MMSE have little value in detecting the rate of cognitive change when evaluating individual treatment responses. The BDS and its sub-components are widely used as a comparative measure when testing outcome measures. Stern (1990) also cautions against relying on the BDS to detect functional change in persons with dementia since disparate functional domains are assessed. Stern argues for the use of a multifactorial approach to the assessment of functional capacity for this reason. Davis et al (1990) found the six-item Short Blessed Test / BOMCT a better screening test than the BDS because of its brevity, ease of administration to the patient, inclusion of a learning task, reliability and neuropathologic validity. This finding was not supported in other studies identified.

Reporter: Lynn Chenoweth and Emily Sansoni

Date of report: 18/01/07

References


### Appendix 5.5: Sandoz Clinical Assessment – Geriatric

**AHOC INSTRUMENT REVIEW SHEET**

**Title:** Sandoz Clinical Assessment – Geriatric.

**Abbreviations:** SCAG.

**Author(s) Name:** Richard I. Shader, Jerold S. Harmatz and Carl Salzman.

**Author(s) Address:**
Massachusetts Health Center
74 Fernwood Road, Boston
Massachusetts 12115.


**Cost:** Nil.

**Training requirements:** The SCAG can be implemented by anyone familiar with the patient, such as clinical and care staff and/or the caregiver. Some training in its use is needed, although the descriptors defining each item are clear to the reader.

**Purpose:** The SCAG was developed to ensure the diagnostic differentiation between early dementia and depressive disorders in the older population by assessing early cognitive and related deterioration in the older person’s ability to engage in daily life. It is also used to assess changes in these areas following treatment.

**Administration time:** 15-20 mins.

**Instrument Type:** Clinical rating scale for the severity of cognitive and functional decline in elderly people.

**Structure:** The SCAG assesses 18 cardinal signs/symptoms of dementia, which cover agitation, cognitive dysfunction, depressed mood and withdrawal (McDowell, 2006), or, according to Venn (1983) the four main areas of impairment assessed are cognitive function, mood and behaviour, activities of daily living and somatic symptoms. However, factor analysis revealed five main factors rather than four (Venn, 1983). Item 19, which is completed after addressing the previous 18 items, represents a global severity rating used in many dementia clinical trials.

**Scoring:** The inventory of 18 target symptoms (items) of dementia is scored by severity. Each item is rated by a 7-point scale ranging from 1 (not present) to 7 (severe).

**Developed for:** The assessment of levels of cognitive and functional decline and the severity of dementia in older people.

**Normative Data:** Shader et al. (1974) provided normative data from 20 healthy volunteers, as well as data from 5 participants with mild cognitive deterioration.

**Clinical Data:** Data from 11 participants with primary affective disorders and 15 with senile dementia is provided in Shader, et al. (1974). Villardita and Lomeo (1992) administered the SCAG to 41 patients with a diagnosis of probable AD.

**Applications:** Suitable for use by clinicians and caregivers with those who have cognitive impairment of dementia, manifested by changes in mood, behaviour, self-
The SCAG has also experienced extensive use in drug research as an outcome measure (McDowell, 2006). Accordingly, the SCAG has been recommended as one of three tests that should be used in clinical trials of anti-dementia drugs, along with the Global Deterioration Scale and the Mini-Mental State Examination (Curran & Wattis, 1997).

Carer and/or Patient Use of Instrument: The SCAG can be used by caregivers and clinicians.

### Psychometric Criteria

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal consistency</strong></td>
<td>Grasel (2002)</td>
<td>X Alpha &gt;0.70</td>
<td>A parallel form of the SCAG was developed to document the symptoms of dementia in a sample of 720 primary caregivers for the non-professional health staff. The parallel form had a Cronbach’s alpha of 0.75 and a split-half reliability coefficient of 0.72.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Marginal or inadequate internal consistency (&lt;0.70)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information found on internal consistency</td>
<td></td>
</tr>
<tr>
<td><strong>Test – retest</strong></td>
<td>Venn (1983)</td>
<td>□ ICC &gt;.70</td>
<td>The Sandoz Manual was developed to ensure consistent &amp; accurate measurement and test-retest scores. If these guidelines are followed test-retest reliability will be &gt; .70.</td>
</tr>
<tr>
<td></td>
<td>Grasel (2002)</td>
<td>Time intervals and confidence intervals reported</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Villardita and Lomeo (1992)</td>
<td>X Marginal or inadequate test-retest reliability ICC&lt;.70</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information found on test-retest reliability</td>
<td></td>
</tr>
<tr>
<td><strong>Inter – rater</strong></td>
<td>Venn (1983)</td>
<td>X Agreement reported and adequate</td>
<td>Average value 0.78 for different groups of patients with different raters.</td>
</tr>
<tr>
<td></td>
<td>Shader, et al. (1974)</td>
<td>□ Inadequate inter-rater agreement</td>
<td>Inter-rater reliability resulted in an average intra-class correlation coefficient of 0.75 when tested with 51 older volunteers in four different groups (healthy, minimal dementia, depression and severe dementia).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information provided</td>
<td></td>
</tr>
<tr>
<td>VALIDITY</td>
<td>Studies Reported &amp; References</td>
<td>Adequacy Checks</td>
<td>Comment</td>
</tr>
<tr>
<td>----------</td>
<td>------------------------------</td>
<td>----------------</td>
<td>---------</td>
</tr>
<tr>
<td>Content</td>
<td>Grasel (2002)</td>
<td>X Patients and experts were involved during item selection and/or item reduction</td>
<td>Content of each item is distinguished clearly for different groups of patients when reviewed and employed by a range of clinicians with different levels of expertise and community dwelling caregivers.</td>
</tr>
<tr>
<td></td>
<td>Shader, et al. (1974)</td>
<td>□ Patients were consulted for reading and comprehension</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No patient involvement</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information found on content validity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ There is an adequate coverage of relevant domains</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ There is limited coverage of relevant domains</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Patients were consulted for reading and comprehension</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No patient involvement</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information found on content validity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ There is an adequate coverage of relevant domains</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ There is limited coverage of relevant domains</td>
<td></td>
</tr>
<tr>
<td>Construct</td>
<td>Grasel (2002)</td>
<td>□ Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used</td>
<td>The SCAG factor ‘mental-amnestic disturbance’ showed a significant correlation with the information and vocabulary subtest of the Wechsler Adult Intelligence Scale.</td>
</tr>
<tr>
<td></td>
<td>Shader, et al. (1974)</td>
<td>X Limited /inadequate construct validity reported</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information provided</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>X Limited /inadequate construct validity reported</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information provided</td>
<td></td>
</tr>
<tr>
<td>Construct: Internal Structure</td>
<td>Venn (1983)</td>
<td>□ No evidence provided/failed a test of dimensionality</td>
<td>Factor analyses have revealed 5 major factors: cognitive dysfunction, interpersonal relationships, affective disorders, apathy and somatic functioning. These factors accounted for 65% of the variance.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X Some evidence provided to support internal structure</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Substantial evidence provided to support internal structure</td>
<td></td>
</tr>
<tr>
<td>Construct: Correlation with other measures</td>
<td>Shader, et al. (1974)</td>
<td>X Correlations with other measures are reported</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Correlations not reported</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Correlations not reported</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Correlations not reported</td>
<td>The SCAG correlates highly with items on the Mental Status Examination Record (MSER) for depression variables $r = .84$, confusion $r = .80$, recent memory $r = .88$ and disorientation $r = .88$. It was identified as a valid and reliable instrument for assessing psychopathology in dementia.</td>
</tr>
<tr>
<td>Construct: Discriminant Validity</td>
<td>Grasel (2002)</td>
<td>X Scale differentiates between relevant categories of respondents □ No information on discriminant validity</td>
<td></td>
</tr>
<tr>
<td>---------------------------------</td>
<td>---------------</td>
<td>-------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>The scale differentiates between relevant categories of respondent e.g. sick vs. well, varying degrees of severity</td>
<td>Tedesco, et al. (1999)</td>
<td>The SCAG clearly distinguishes between healthy and depressed subjects, and subjects with mild and severe dementia. A score of less than 24 on the Mini Mental State Examination (MMSE) and more than 40 on the SCAG is predictive of cognitive impairment ($p &lt; .001$).</td>
<td></td>
</tr>
<tr>
<td>Criterion</td>
<td>Grasel (2002)</td>
<td>X Comparison made to criterion measures □ No comparison with criterion measures provided</td>
<td></td>
</tr>
<tr>
<td>Information on the relationship of scores to gold standard measures or clinical diagnosis is provided</td>
<td>Herrmann, et al. (1997)</td>
<td>When tested against the MMSE, Wechsler, Blessed and MSER on a number of key cognitive and functional domains, there were average coefficients of 0.75 for most domains. In a study to test the efficacy and safety of Nicergoline in 252 patients, the SCAG correlated highly with the Mini Mental State Examination (MMSE) for scores at baseline and at end of treatment with respect to baseline. SCAG scores at all time points were favourably compared against the scores on the Wechsler Adult Intelligence Scale and the Blessed Dementia Scale, the clinician’s assessment of daily living function, Montgomery Asberg Depression Rating Scale (MADRS), and Clinical Global Impression (CGI).</td>
<td></td>
</tr>
<tr>
<td>Interpretability</td>
<td>Studies Reported &amp; References</td>
<td>Adequacy Checks</td>
<td>Comment</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------------------------</td>
<td>----------------</td>
<td>---------</td>
</tr>
<tr>
<td>The degree to which one can assign qualitative meaning to quantitative scores</td>
<td>Shader, et al. (1974) Venn (1983)</td>
<td>□ Authors provide 2 or more types of information on interpretability</td>
<td>X Qualitative descriptors are used for all items and can be scored quantitatively using the named ratings in a range of 1-7.</td>
</tr>
<tr>
<td>Do authors provide the following:</td>
<td></td>
<td>□ Authors provide limited information to assist with interpretability</td>
<td></td>
</tr>
<tr>
<td>Presentation of means and SD of scores before and after treatment</td>
<td></td>
<td>□ No information provided</td>
<td></td>
</tr>
<tr>
<td>Comparative data on the distribution of scores in relevant subgroups</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information on the relationship of scores to well-known functional measures or clinical diagnosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information on the association between changes in scores and patients' global ratings of the magnitude of change they have experienced</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RESPONSIVENESS</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floor and ceiling effects</td>
<td></td>
<td>□ Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
<td>No information found.</td>
</tr>
<tr>
<td>The questionnaire fails to demonstrate a worse score in patients who clinically deteriorated and an improved score in patients who clinically improved</td>
<td></td>
<td>□ Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score</td>
<td></td>
</tr>
<tr>
<td>Authors should provide descriptive statistics of the distribution of scores</td>
<td></td>
<td>X No information provided on floor and ceiling effects</td>
<td></td>
</tr>
<tr>
<td>Sensitivity to change</td>
<td>Herrmann, et al. (1997)</td>
<td>X Hypotheses were formulated and results were in agreement</td>
<td>This study supported the hypothesis that the SCAG and other standard dementia measures/rating scales such as the MMSE are the preferred psychological methods for assessing cognitive and functional changes in dementia, as it has good sensitivity in detecting changes over time.</td>
</tr>
<tr>
<td>The ability to detect important change over time in the concept being measured</td>
<td></td>
<td>□ An adequate metric was used (ES, SRM, comparison with external standard)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information on sensitivity to change was provided</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ MCID - Information</td>
<td></td>
</tr>
</tbody>
</table>
Like the Mental Status Examination Record (MSER), the SCAG detects subtle changes in cognitive status and function over time, especially for depression, confusion, recent memory and orientation/disorientation.

The SCAG reveals subtle changes occurring in persons with dementia in comparison with clinically assessed symptoms of dementia.

However, Lezak, et al. 2004 and McDowell, 2006 provide evidence that questions this interpretation.

Cultural Applicability and Cultural Adaptations:
The SCAG has been employed in the USA, France and Germany. The French and German versions are used often in Europe and appear to have the same factor structure, providing further support for the construct validity and cross-cultural reliability of the SCAG (Venn, 1983).

Gender Appropriateness:
Appropriate for use with both genders.

Age Appropriateness:
The SCAG has been used with people 60 years and over and is appropriate for use with the elderly.

Summary:
The SCAG is a useful tool for caregivers and clinicians to detect changes in cognition and functioning in the older person by targeting 18 common symptoms associated with dementia, and it compares well with other measures such as the MMSE. It has also been widely used in clinical and psychopharmacological research. The major benefit of use in clinical practice lies in its utility to family caregivers and health staff with limited expertise in dementia assessment. However, the SCAG has also been referred to by Lezak, et al. (2004) as psychometrically deficient compared to the Alzheimer Disease Assessment Scale; and McDowell (2006) also noted that the SCAG has its critics (e.g. Salzman, 1983). Furthermore, Warburton and Rusted (1989 as cited in Curran & Wattis, 1997) felt that the SCAG should not be used alone in clinical trials as it is not precise or objective enough to pick up subtle changes in participants.

Reporter:
Lynn Chenoweth and Emily Sansoni

Date of report:
18/01/07

References


Appendix 6: Reviews of Dementia Specific Health Related Quality of Life Instruments
Appendix 6.1: Quality of Life in Alzheimer’s Disease

AHOC INSTRUMENT REVIEW SHEET

Title: Quality of Life in Alzheimer’s Disease.

Abbreviations: QoL-AD.

Author(s) Name: Rebecca G Logsdon, Laura E Gibbons, Susan M McCurry, Linda Teri.

Author(s) Address: Rebecca G Logsdon, PhD
Research Associate Professor
Psychosocial & Community Health
University of Washington
9709 3rd NE, Suite 597
Seattle, WA 98115-7263
USA.

Supplied by: The author. Written permission is required.

Cost: Free with author’s permission.

Training requirements: No formal training is required for interviewers. A detailed script for administration of the measure is available.

Purpose: To assess the quality of life of persons with dementia/Alzheimer’s disease.

Administration time: 10-15 minutes for patients, 5 minutes for caregivers.

Instrument Type: Self–report Questionnaire. The instrument has two versions: patient and proxy versions. Both are self administered and self-rated questionnaires. An interviewer can be used to oversee the administration and provide assistance and clarification as needed. In this case the questionnaire is interviewer administered, patient or carer rated. Patient and caregiver reports can also be combined.

Structure: The instrument contains 13 items, one for each dimension: physical health, energy, mood, living situation, memory, family, marriage, friends, self as a whole, ability to do chores around the house/room, ability to do things for fun, money, and life as a whole. Responses are rated on a 4 point ordered category scale, with options being “poor” (1), “fair” (2), “good” (3), and “excellent” (4).

Scoring: The item scores are summed to provide a global score with a minimum score of 13 and a maximum score of 52, with higher scores reflecting better QOL. The patient and caregiver reports can be combined into a single composite QOL score. In this case a weighted composite QoL-AD score is calculated by multiplying the patient score by 2, adding the caregiver score, and dividing the sum by 3.

Developed for: Patients with mild to moderate dementia (Alzheimer’s disease).

Normative Data: Normative data is not available.

Clinical Data: The instrument has been used in numerous studies investigating quality of life instruments, differences in patient and carer perspectives on quality of life, change in quality of life over time, and the effects of interventions including cognitive stimulation therapy, and drug treatment (see references below).

Carer and/or Patient Use of Instrument: Patient rated and or carer rated.

Psychometric Criteria

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td>Logsdon, et al. (1999) Logsdon, Gibbons, et al. (2002) Thorgrimsen, Selwood, et al. (2003)</td>
<td>X Alpha &gt;0.70 □ Marginal or inadequate internal consistency (&lt;0.70) □ No information found on internal consistency</td>
<td>Instrument shows good to excellent internal reliability with Cronbach’s alpha ranging from 0.78 to 0.94 for the patient version and 0.79 to 0.88 for the proxy version.</td>
</tr>
<tr>
<td>Test – retest</td>
<td>Logsdon, et al. (1999) Thorgrimsen, Selwood, et al. (2003)</td>
<td>X ICC &gt;.70 □ Marginal or inadequate test-retest reliability ICC&lt;.70 □ No information found on test-retest reliability</td>
<td>Good test-retest validity. Authors indicated ICC of 0.76 for patient version and 0.92 for proxy version at one week retest. Thorgrimsen, et al. indicated ICC of &lt; 0.60 for the patient version; the carer version was not evaluated.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sloane, Zimmerman, et al. (2005)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VALIDITY</td>
<td>Studies Reported &amp; References</td>
<td>Adequacy Checks</td>
<td>Comment</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------------------</td>
<td>----------------</td>
<td>---------</td>
</tr>
<tr>
<td>Content</td>
<td>Logsdon, Gibbons, et al. (2002)</td>
<td>X Patients were involved during item selection and/or item reduction</td>
<td>Patients, caregivers and experts were involved in item selection and item reduction to ensure an adequate coverage of the relevant domains.</td>
</tr>
<tr>
<td></td>
<td>Thorgrimsen, Selwood, et al. (2003)</td>
<td>□ No patient involvement</td>
<td></td>
</tr>
<tr>
<td></td>
<td>□ No information found on content validity</td>
<td>□ There is an adequate coverage of relevant domains</td>
<td></td>
</tr>
<tr>
<td></td>
<td>□ There is limited coverage of relevant domains</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Construct</td>
<td>Logsdon, Gibbons, et al. (2002)</td>
<td>X Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used</td>
<td>The patient version shows expected correlations with: Physical &amp; Instrumental Self-maintenance Scale (PIS-ADL), Alzheimer’s Disease Co-Operative Study (ACDS)-ADL, Revised Memory &amp; Behaviour Checklist (RMBPC)-depression, Cornell Scale for Depression in Dementia (CSDD), Geriatric Depression Scale (GDS), Rating Anxiety in Dementia (RAID), Mental Outcomes Study (MOS), Pleasant Events Schedule (PES-AD), Neuropsychiatric Inventory (NPI). Correlations with the MMSE were mixed.</td>
</tr>
<tr>
<td></td>
<td>Hoe, Katona, et al. (2005)</td>
<td>□ No information provided</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Selwood, Thorgrimsen, et al. (2005)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Winzelberg, Williams, et al. (2005)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sloane, Zimmerman, et al. (2005)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Smith, Lamping, et al. (2005)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hoe, Hancock, et al. (2006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fuh and Wang (2006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Construct: Internal Structure</td>
<td>Thorgrimsen, Selwood, et al. (2003) Edelman, Fulton, et al. (2005)</td>
<td>□ No evidence provided/failed a test of dimensionality X Some evidence provided to support internal structure □ Substantial evidence provided to support internal structure</td>
<td>Factor analysis supported the dimensions proposed by the instrument developers.</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------------------------------------------------</td>
<td>-----------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
Information on the relationship of scores to well-known functional measures or clinical diagnosis

Information on the association between changes in scores and patients' global ratings of the magnitude of change they have experienced

<table>
<thead>
<tr>
<th>RESPONSIVENESS</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floor and ceiling effects</td>
<td>Aisen, Schafer, et al. (2003)</td>
<td>X Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
<td>Authors have generally provided descriptive statistics of score distribution. Most studies have provided at least mean, SD and range or confidence intervals. No major floor or ceiling effects were detected in these studies.</td>
</tr>
<tr>
<td></td>
<td>Spector, et al. (2003)</td>
<td>□ No information provided on floor and ceiling effects</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chapman, Weiner, et al. (2004)</td>
<td>□ No information provided on floor and ceiling effects</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Orrell, Spector, et al. (2005)</td>
<td>□ No information provided on floor and ceiling effects</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Selwood, Thorgrimsen, et al. (2005)</td>
<td>□ No information provided on floor and ceiling effects</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sloane, Zimmerman, et al. (2005)</td>
<td>□ No information provided on floor and ceiling effects</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Smith, Lamping, et al. (2005)</td>
<td>□ No information provided on floor and ceiling effects</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teri, McCurry, et al. (2005)</td>
<td>□ No information provided on floor and ceiling effects</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Woods, Thorgrimsen, et al. (2006)</td>
<td>□ No information provided on floor and ceiling effects</td>
<td></td>
</tr>
<tr>
<td>Sensitivity to change</td>
<td>Woods, Thorgrimsen, et al. (2006)</td>
<td>X Hypotheses were formulated and results were in agreement</td>
<td>Change in self-reported QOL-AD showed small but significant correlations with change in cognition (MMSE, ADAS-Cog), depression (CSDD) and communication (HCS) (Woods, 2006, using data from Spector, 2003). Mixed results from RCTs of cognitive stimulation therapy (Spector paper supports responsiveness of QOL-AD,</td>
</tr>
<tr>
<td></td>
<td>Spector, Thorgrimsen, et al. (2003)</td>
<td>□ No information on sensitivity to change was provided</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Orrell, Spector, et al. (2005)</td>
<td>□ No information on sensitivity to change was provided</td>
<td></td>
</tr>
<tr>
<td></td>
<td>□ No information on sensitivity to change was provided</td>
<td>□ MCID - Information was provided about</td>
<td></td>
</tr>
</tbody>
</table>
the magnitude of score differences which would be clinically meaningful X MCID – No information was provided.

Cultural Applicability and Cultural Adaptations: Validated translations are available in French, Japanese, Mandarin, Portuguese, Danish, German, Italian, Spanish, Swedish and Greek languages.

Gender Appropriateness: Appropriate for use with both genders.

Age Appropriateness: It has been used with people 60 years and over and is appropriate for use with the elderly.

Summary: The instrument has both patient and carer versions which can be used either separately or together. It has been extensively cited and has very good psychometric properties. Access and administration are easy. The instrument has also been used as an outcome measure in numerous intervention studies.

Date of report: 17/1/2007

References


Appendix 6.2: DEMQOL

AHOC INSTRUMENT REVIEW SHEET

Title: DEMQOL (there is no extended title).

Abbreviations: DEMQOL.

Author(s) Name: SC Smith, DL Lamping, S Banerjee, R Harwood, B Foley, P Smith, JC Cook, J Murray, M Prince, E Levin, A Mann, and M Knapp.

Author(s) Address: Professor S Banerjee, PO Box 26 Section of Mental health and Ageing, Health Services Research Department Institute of Psychiatry, King’s College London London SE5 8AF UK.

Supplied by: Available on the website of the Institute of Psychiatry.

Cost: Free for academic use. Costs for commercial and pharmaceutical use are determined following discussions between developers and potential user.

Training requirements: Users' manual is available on the website.

Purpose: To assess quality of life in persons with mild to moderate dementia. DEMQOL Proxy can also be used for severe dementia.

Administration time: 10 – 20 minutes.

Instrument Type: DEMQOL is patient rated health related quality of life instrument, DEMQOL Proxy is carer rated; both are administered by an interviewer.

Structure: DEMQOL contains 28 items, covering 4 dimensions: Daily Activities, Memory, Negative Emotion, Positive Emotion. It also includes an additional item to assessing the patients feeling about their overall quality of life. Items are rated on a 4 point ordered category scale. Response options are “a lot”, “quite a bit”, “a little” “not at all”, except for the global question, which has the options “very good”, “good”, “fair”, “poor”.

DEMQOL Proxy contains 31 items covering 2 domains: Functioning and Emotion. It also includes an additional item to assess the patients feeling about their overall quality of life (as perceived by the carer) plus a global QOL item. Response options are the same as for DEMQOL.

Scoring: DEMQOL: Item scores are summed to provide a global score (minimum 28, maximum 112).

DEMQOL proxy: Item scores are summed to provide a global score (minimum 31, maximum 124).

For both versions, a higher score indicates better quality of life.

Developed for: Persons with mild to moderate dementia (DEMQOL); all stages (DEMQOL Proxy).

Normative Data: Normative data is not available.

Clinical Data: This is a very new instrument and has not been used in any interventions that have been published to date (see references below).
Applications: Assessment of quality of life in persons with dementia. DEMQOL is to be used for mild to moderate dementia; DEMQOL Proxy can be used for all stages.

Carer and/or Patient Use of Instrument: DEMQOL – Patient rated, DEMQOL Proxy – Carer rated. Use of both instruments together is recommended as the two systems are complementary.

Psychometric Criteria

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td>Smith, Lamping, et al. (2005)</td>
<td>X Alpha &gt;0.70 □ Marginal or inadequate internal consistency (&lt;0.70) □ No information found on internal consistency</td>
<td>Excellent internal consistency with Cronbach’s alpha of 0.87 for both patient and proxy versions.</td>
</tr>
<tr>
<td>Test – retest</td>
<td>Smith, Lamping, et al. (2005)</td>
<td>X ICC &gt;0.70 Time intervals and confidence intervals reported □ Marginal or inadequate test-retest reliability ICC&lt;.70 □ No information found on test-retest reliability</td>
<td>Good test-re-test reliability for both versions. Patient version: ICC’s were 0.84 when for whole sample was considered and 0.76 when only mild to moderate considered. Proxy version: ICC’s were 0.75, 0.67 and 0.84 for whole, mild to moderate and severe samples respectively.</td>
</tr>
<tr>
<td>Inter – rater</td>
<td>Smith, Lamping, et al. (2005)</td>
<td>X Agreement reported and adequate □ Inadequate inter-rater agreement □ No information provided</td>
<td>Correlations between DEMQOL and DEMQOL Proxy - overall scores moderate for mild/moderate dementia (r =0.36); low for severe dementia (r=-0.15).</td>
</tr>
<tr>
<td>VALIDITY</td>
<td>Studies Reported &amp; References</td>
<td>Adequacy Checks</td>
<td>Comment</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>------------------------------</td>
<td>----------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Content</strong></td>
<td>Smith, Murray, et al. (2005b)</td>
<td>X Patients and experts were involved during item selection and/or item reduction □ Patients were consulted for reading and comprehension □ No patient involvement □ No information found on content validity X There is an adequate coverage of relevant domains □ There is limited coverage of relevant domains</td>
<td>People with dementia, their carers and experts in the field were involved in item selection and item reduction to ensure an adequate coverage of the relevant domains.</td>
</tr>
<tr>
<td><strong>Construct</strong></td>
<td>Smith, Lamping, et al. (2005)</td>
<td>X Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used X Limited/inadequate construct validity reported □ No information provided</td>
<td>Findings show moderate construct validity in relation to hypothesised domains. Patient version: Low to moderate correlations with the Geriatric Depression Scale (GDS) and Barthel’s ADL. Correlations with GDS were expected to be higher. Proxy version: For people with mild to moderate dementia: moderate to high correlation with GDS, only low with Barthel’s ADL. For persons with severe dementia: high correlation with GDS. Univariate and multivariate analysis showed DEMQOL was significantly associated with the Neuropsychiatric Inventory (NPI) total score as well as the agitation, depression, anxiety, disinhibition and irritability subscales.</td>
</tr>
<tr>
<td><strong>Construct: Internal Structure</strong></td>
<td>Smith, Lamping, et al. (2005)</td>
<td>□ No evidence provided/failed a test of dimensionality X Some evidence provided to support internal structure □ Substantial evidence provided to support internal</td>
<td>Factor solution accounts for 33.3% of variance but authors say results should be considered exploratory due to the small sample size.</td>
</tr>
<tr>
<td>Construct: Correlation with other measures</td>
<td>Smith, Lamping, et al. (2005)</td>
<td>X Correlations with other measures are reported □ Correlations not reported</td>
<td>The patient and proxy versions show mild to moderate correlations with other QOL instruments as detailed below. Patient version: Moderate correlation with QOLAD and DQOL (both dementia specific instruments) and moderate to high correlation with SF-12 (generic health status instrument). Proxy version: For people with mild to moderate dementia: moderate correlation with QOLAD and low to moderate correlation with SF-12. For people with severe dementia: there were too few cases to evaluate correlations.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Construct: Discriminant Validity</td>
<td>Smith, Lamping, et al. (2005)</td>
<td>X Scale differentiates between relevant categories of respondents □ No information on discriminant validity</td>
<td>The patient version was able to discriminate by age indicating some support for discriminant validity.</td>
</tr>
<tr>
<td>Criterion</td>
<td>Information on the relationship of scores to gold standard measures or clinical diagnosis is provided</td>
<td>X Comparison made to criterion measures □ No comparison with criterion measures provided</td>
<td>There is no gold standard for QOL, so studies have evaluated the instrument against other QOL instruments and/or instruments measuring aspects of the dimensions covered (see above).</td>
</tr>
<tr>
<td>Interpretability</td>
<td>Smith, Lamping, et al. (2005)</td>
<td>X Authors provide 2 or more types of information on interpretability □ Authors provide limited information to assist with interpretability □ No information provided</td>
<td>The authors provide considerable information about interpretability with more than 2 or more types of information.</td>
</tr>
</tbody>
</table>
association between changes in scores and patients’ global ratings of the magnitude of change they have experienced

<table>
<thead>
<tr>
<th>RESPONSIVENESS</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floor and ceiling effects</td>
<td>Smith, Lamping, et al. (2005)</td>
<td>X Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
<td>Authors provide considerable information to indicate there are no major floor or ceiling effects.</td>
</tr>
<tr>
<td>Sensitivity to change</td>
<td></td>
<td>□ Hypotheses were formulated and results were in agreement □ An adequate metric was used (ES, SRM, comparison with external standard) X No information on sensitivity to change was provided □ MCID - Information was provided about the magnitude of score differences which would be clinically meaningful X MCID – No information was provided.</td>
<td>No responsiveness evaluations were conducted.</td>
</tr>
</tbody>
</table>

Cultural Applicability and Cultural Adaptations: This is a new instrument, so as yet there are no other translations.

Gender Appropriateness: Appropriate for use with both genders.

Age Appropriateness: Appropriate for use with elderly persons.
Summary: This is a very new instrument that has been developed by a team of world renowned experts in this field. It has both patient and carer versions that can be used either separately or together. Access and administration is easy. Due to its newness, the instrument has not as yet been widely cited, however the psychometric properties have been found to be good to very good. The authors acknowledge that more work needs to be done on validation, but the instrument is very promising.

Reporter: Madeleine King and Siggi Zapart

Date of report: 17/01/2007

References


Appendix 6.3: Quality of Life in Late-Stage Dementia

AHOC INSTRUMENT REVIEW SHEET

Title: Quality of life in Late-Stage dementia.

Abbreviations: QUALID.

Author(s) Name: Myron F Weiner, Kristin Martin-Cook, Doris A Svetlik, Kathleen Saine, Barbara Foster, Catherine S Fontaine.

Author(s) Address: Myron F Weiner, MD
Department of Psychiatry
University of Texas
Southwestern Medical centre
5323 Harry Hines Blvd
Dallas TX 75235-9070
USA

Supplied by: The author. Written permission is required.

Cost: Free for academics and non-profit research. Fees for commercial use are considered and charged on an individual basis.

Training requirements: No training is required. However, it is recommended the instrument be administered by at least a bachelor's level technician.

Purpose: To rate quality of life in persons with late stage Alzheimer's disease and other dementing illnesses.

Administration time: 5 minutes.

Instrument Type: Rating Scale - interviewer administered to an informant (family member or professional caregiver) based on observations made during 30 or more hours over the previous week.

Structure: The instrument contains 11 items describing observable behaviours and includes observations of affective state, behavioural signs of comfort, and engagement in activities and interactions with others. Items are rated on a 5 point ordered category scale. Window of observation is one week.

Scoring: The items scores are summed to provide a global score with a minimum score of 11 and a maximum score of 55. Items are not weighted. A lower score indicates better quality of life.

Developed for: Persons with late stage dementia residing in long-term care facilities.

Normative Data: Normative data is not available.

Clinical Data: The instrument has not been used in any interventions that have been published to date. It has however been adopted by the Care Keys Project and is part of the toolkit for the quality management of services for elderly people in Finland, Sweden, Germany, Estonia and England. No evaluations of this project have as yet been conducted.

Applications: Assessment of quality of life for patients with late stage dementia, and evaluation of change in quality of life due to therapeutic interventions.

Carer and/or Patient Use of Instrument: Administered by an interviewer to a caregiver (most likely a professional such as nurse, but could be a family member or other caregiver) who has sufficient familiarity with the patient (operationalised as 30 or more hours of
exposure to the patient in the past week).

Psychometric Criteria

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
</table>
| Internal consistency  
The extent to which items in a (sub) scale are inter-correlated; a measure of the homogeneity of a (sub)scale.  
Cronbach's alpha should be between 0.70 and 0.90 for every dimension / sub-scale.  
Weiner, Martin-Cook, et al. (2000)  
X Alpha >0.70 □ Marginal or inadequate internal consistency (<0.70) □ No information found on internal consistency | The instrument shows good internal consistency (0.77) and excellent internal consistency between items. |
| Test – retest  
The extent to which the same results are obtained on repeated administrations of the same questionnaire when no change in physical functioning has occurred  
Calculation of an intraclass correlation coefficient (ICC); and an ICC > 0.70 is desired  
Weiner, Martin-Cook, et al. (2000)  
X ICC >.70 Time intervals and confidence intervals reported □ Marginal or inadequate test-retest reliability ICC<.70 □ No information found on test-retest reliability | The instrument was administered twice over a 2 – 3 day period with an excellent ICC of 0.81, SEM = 0.08. |
| Inter – rater  
Limits of agreement, Kappa, or standard error of measurement (SEM) were presented  
Weiner, Martin-Cook, et al. (2000)  
X Agreement reported and adequate □ Inadequate inter-rater agreement □ No information provided | Excellent inter-rater reliability, ICC = 0.83, SEM = 0.07. |

<table>
<thead>
<tr>
<th>VALIDITY</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
</table>
| Content  
The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire  
Weiner, Martin-Cook, et al. (2000)  
□ Patients were involved during item selection and/or item reduction □ Patients were consulted for reading and comprehension  
X No patient involvement □ No information found on content validity  
X There is an | Developed in a series of consensus meetings of staff with extensive experience with late stage dementia.  
Patients are at a late stage and therefore are not able to be involved. |
<table>
<thead>
<tr>
<th>Construct</th>
<th>Weiner, Martin-Cook, et al. (2000) Luoma, Vaarama, et al. (2005)</th>
<th>X Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used □ Limited/inadequate construct validity reported □ No information provided</th>
<th>Shows expected correlations with geriatric Depression Scale (GDS), Neuropsychiatric Inventory (NPI), Cornell Scale for depression in Dementia (CSDD), and Philadelphia Geriatric Morale Scale (PGMS). Correlations with MMSE and The Physical Self Maintenance Scale (PSMS) were not significant as expected.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Construct</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The extent to which scores on the questionnaire relate to other measures in a manner that is consistent with theoretically derived hypothesis concerning the domains that are measured.</td>
<td>Weiner, Martin-Cook, et al. (2000) Luoma, Vaarama, et al. (2005)</td>
<td>X Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used □ Limited/inadequate construct validity reported □ No information provided</td>
<td>Shows expected correlations with geriatric Depression Scale (GDS), Neuropsychiatric Inventory (NPI), Cornell Scale for depression in Dementia (CSDD), and Philadelphia Geriatric Morale Scale (PGMS). Correlations with MMSE and The Physical Self Maintenance Scale (PSMS) were not significant as expected.</td>
</tr>
<tr>
<td><strong>Construct: Internal Structure</strong></td>
<td>Weiner, Martin-Cook, et al. (2000)</td>
<td>□ No evidence provided/failed a test of dimensionality X Some evidence provided to support internal structure □ Substantial evidence provided to support internal structure</td>
<td>Principal components analysis yielded a one factor solution.</td>
</tr>
<tr>
<td>Information provided on factor structure</td>
<td>Weiner, Martin-Cook, et al. (2000)</td>
<td>□ No evidence provided/failed a test of dimensionality X Some evidence provided to support internal structure □ Substantial evidence provided to support internal structure</td>
<td>Principal components analysis yielded a one factor solution.</td>
</tr>
<tr>
<td><strong>Construct: Correlation with other measures</strong></td>
<td></td>
<td>□ Correlations with other measures are reported X Correlations not reported</td>
<td>No other instrument assessing quality of life in late stage dementia is available.</td>
</tr>
<tr>
<td>Comparisons made to other measures</td>
<td>Valvanne, Luoma, et al. (2005)</td>
<td>X Scale differentiates between relevant categories of respondents □ No information on discriminant validity</td>
<td>Differentiates between patients with mild and moderate depression.</td>
</tr>
<tr>
<td><strong>Construct: Discriminant Validity</strong></td>
<td>Valvanne, Luoma, et al. (2005)</td>
<td>X Scale differentiates between relevant categories of respondents □ No information on discriminant validity</td>
<td>Differentiates between patients with mild and moderate depression.</td>
</tr>
<tr>
<td>The scale differentiates between relevant categories of respondent e.g. sick vs. well, varying degrees of severity</td>
<td>Valvanne, Luoma, et al. (2005)</td>
<td>X Scale differentiates between relevant categories of respondents □ No information on discriminant validity</td>
<td>Differentiates between patients with mild and moderate depression.</td>
</tr>
<tr>
<td><strong>Criterion</strong></td>
<td>Weiner, Martin-Cook, et al. (2000) Luoma, Vaarama, et al. (2005) Valvanne, Luoma, et al. (2005)</td>
<td>X Comparison made to criterion measures □ No comparison with criterion measures provided</td>
<td>There is no gold standard to measure against, so studies have evaluated the instrument against instruments measuring aspects of the domains covered.</td>
</tr>
<tr>
<td>Information on the relationship of scores to gold standard measures or clinical diagnosis is provided</td>
<td>Weiner, Martin-Cook, et al. (2000) Luoma, Vaarama, et al. (2005) Valvanne, Luoma, et al. (2005)</td>
<td>X Comparison made to criterion measures □ No comparison with criterion measures provided</td>
<td>There is no gold standard to measure against, so studies have evaluated the instrument against instruments measuring aspects of the domains covered.</td>
</tr>
<tr>
<td><strong>Interpretability</strong></td>
<td>Weiner, Martin-Cook, et al. (2000) Luoma, Vaarama, et al. (2005) Valvanne, Luoma, et al. (2005)</td>
<td>X Authors provide 2 or more types of information on interpretability □ Authors provide limited information to assist with interpretability □ No information provided</td>
<td>Authors provide information about distribution of scores as well as information about relationships to well known measures (MMSE PSMS, GDS, NPI).</td>
</tr>
<tr>
<td>The degree to which one can assign qualitative meaning to quantitative scores</td>
<td>Weiner, Martin-Cook, et al. (2000) Luoma, Vaarama, et al. (2005) Valvanne, Luoma, et al. (2005)</td>
<td>X Authors provide 2 or more types of information on interpretability □ Authors provide limited information to assist with interpretability □ No information provided</td>
<td>Authors provide information about distribution of scores as well as information about relationships to well known measures (MMSE PSMS, GDS, NPI).</td>
</tr>
</tbody>
</table>
**RESPONSIVENESS**

<table>
<thead>
<tr>
<th>Floor and ceiling effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>The questionnaire fails to demonstrate a worse score in patients who clinically deteriorated and an improved score in patients who clinically improved</td>
</tr>
<tr>
<td>Authors should provide descriptive statistics of the distribution of scores</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sensitivity to change</th>
</tr>
</thead>
<tbody>
<tr>
<td>The ability to detect important change over time in the concept being measured</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Martin-Cook, Hynan, et al. (2005)</td>
<td>□ Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected □ Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score X No information provided on floor and ceiling effects</td>
<td>No information about floor or ceiling effects was detected.</td>
</tr>
<tr>
<td>□ Hypotheses were formulated and results were in agreement □ An adequate metric was used (ES, SRM, comparison with external standard) □ No information on sensitivity to change was provided □ MCID - Information was provided about the magnitude of score differences which would be clinically meaningful X MCID – No information was provided.</td>
<td>Detected significant improvement in neuropsychiatric symptoms (agitated and or psychotic symptoms).</td>
<td></td>
</tr>
</tbody>
</table>
Cultural Applicability and Cultural Adaptations: The instrument is available in Swedish, Finnish, German, and Lithuanian. Details can be obtained from the author.

Gender Appropriateness: Appropriate for use with both genders.

Age Appropriateness: Appropriate for people with late stage dementia at any age.

Summary: This is the only instrument that has been developed specifically for use with patients who have late stage dementia. It has very good psychometric properties and is brief and easy to administer. The Care Keys Project, a program to improve management of quality of life in elderly person in Europe, has now adopted this instrument for use in its studies.

Reporter: Madeleine King and Siggi Zapart

Date of report: 13/11/06

References


Appendix 7: Reviews of Cognitive Assessment Instruments
Appendix 7.1: Modified Mini-Mental State Exam

AHOC INSTRUMENT REVIEW SHEET

Title: Modified Mini-Mental State Exam.

Abbreviations: 3MS.

Author(s) Name: Teng, EL & Chui, HC.

Author(s) Address: Professor Evelyn Teng
Department of Neurology, University of Southern California
Keck School of Medicine, Los Angeles, CA, USA
Phone: (626) 796-6057
E-mail: eteng@usc.edu

Professor Helena C. Chui
Chair, Department of Neurology
Raymond and Betty McCarron Chair in Neurology
University of Southern California
Keck School of Medicine
410 Keith Mayer Building, 1975 Zonal Avenue
Los Angeles, California 90033-4606
Phone: (323) 442-7686
E-mail: chui@usc.edu

Supplied by: Instrument and training aids are available from Dr E. Teng and H. Chui.
The 3 MS can also be reproduced from the original paper, with the authors’ permission, which explains administration and scoring methods for both the MMSE and the 3MS scores.

Cost: Free with authors’ permission (all qualified health-care professionals can use the 3MS free of charge in their research and clinical practice. Training aids include a manual, a record form, two forms of quizzes for qualifying users on the correct administering and scoring of the 3MS, and scoring keys on the quizzes. These are in WORD files. For parties interested in obtaining these materials, a modest contribution to the authors’ research fund is requested in order to help defray the development and handling costs).

Training requirements: No formal training is needed, however it is recommended that the interviewer gain mastery over the administration and scoring of the instrument based on the original paper (Teng and Chui, 1987) and Teng’s unpublished training aids.

Purpose: To assess a global cognitive function in adults including orientation, registration, recall, simple language, and construction. It was developed to address shortcomings of the Mini-Mental State Exam (MMSE) (Folstein, Folstein, et al. 1975)- to improve reliability and validity of the scores, minimise the floor/ceiling effect, and to enhance discrimination of various levels of cognitive abilities among people with cognitive impairment and dementia.

---

1 There are other versions of the MMSE also known as the Modified Mini-Mental State Exam. They are not necessarily the same as the 3MS.

2 This needs to be confirmed because the 3MS is derived from the MMSE Folstein MF, Folstein SE et al. (1975). "Mini-Mental State: A practical method for grading the cognitive state of patients for clinicians." Journal of Psychiatric Research 12: 189-198. which is copyrighted by Psychological Assessment Resources, Inc. and costs about US$1, including examination forms, guides and software, per test. It is yet to be confirmed how this impacts on the use of and the cost for the 3MS, which is a modified version of the MMSE.
Administration time: 10 minutes.

Instrument Type: A brief quantitative assessment of cognitive function assessed by patient responses to questions and answers rated by a skilled interviewer.

Structure: The 3MS consists of 27 items/questions1 (an extra 8 items have been added to the 19 items of the MMSE) under 15 domains, including date and place of birth (5 points), registration (3 points), mental reversal (7 points), first recall (9 points), temporal orientation (15 points), spatial orientation (5 points), naming (5 points), four-legged animals (10 points), similarities (6 points), repetition (3 points), read and obey “close your eyes” (5 points), three-stage command (3 points), writing (5 points), copying two pentagons (10 points), and second recall (9 points). The domains are designed to assess the individual's cognitive capacity in terms of orientation to time and place, attention, concentration, long- and short-term memory, language ability, constructional praxis, and abstract thinking.

Scoring: Each correct answer to the item yields a score (see above), and the item scores are summed to provide a global score ranging from 0-100 (compared to the MMSE ranging from 0-30). Higher scores indicate better cognitive performance, and cutting points range between 76 and 80. A single administration of the 3MS, with the addition of a few extra questions, can produce the scores for both the MMSE and the 3MS. The 3MS is a more finely graded scoring system than the MMSE scoring system, which allowed dichotomously scored responses only. This means there is room for attaining more marks for nearly accurate answers when using the 3MS.

Developed for: The original MMSE was developed to assist cognitive assessment of older patients in a clinical setting. The 3MS was developed to improve validity and reliability of the MMSE by adding items and extending the scoring precision to screen for both dementia and cognitive impairment. The 3MS test has been used extensively in both community and institutional settings.

Normative Data: Normative data, based on age (older populations) and education, have been reported in general population-based studies (Tombaugh, McDowell, et al. 1996, Jones, Schinka, et al. 2002) and, in particular, population focused studies such as for an elderly African American population (Brown, Schinka, et al. 2003) and for a non-demented elderly population (Bravo and Hebert, 1997; Tschanz, Welsh-Bohmer et al. 2002). Jones, et al.'s study (2002) also offered adjustments for age and education, which aimed to improve sensitivity and specificity in detecting dementia. Whilst adjustments for age, education and sensory impairment resulted in improved sensitivity and specificity to screen for dementia (Khachaturian, Gallo, et al. 2000; Hayden, Khachaturian, et al. 2003), findings from a large population-based study showed the use of age and education adjusted normative data resulted in reduced validity of the instrument as well as reducing sensitivity to dementia (O'Connell, Tuokko, et al. 2004).

Clinical Data: The 3MS has been used in numerous clinical studies in the following six categories2:

1) cognitive status/change in general populations or populations with physical or mental illness, without dementia: the primitive reflexes by electrophysiological assessments and their correlation with the cognitive and physical functioning of stroke patients (Chang, 2001); the association between stroke and cognitive function/incident cognitive decline (Suhr and Grace, 1999; Elkins, O'Meara, et al. 2004); left carotid artery disease and

---

1 The total number of the items/questions may be higher when some items are counted in a detailed manner; for example, serial abstracts and spelling “world” backward are counted as ten items, rather than two.
2 Some may overlap with other categories. Some of the studies cited used the 3MS as a baseline measure.

2) the effects of drugs, both prescribed and supplementary, on cognitive function: Impact of antidiabetic medications on physical and cognitive functioning (Wu, Haan, et al. 2003); the association of statin drug use on cognitive change (Bernick, Katz, et al. 2005); the association of antihypertensive agents with MRI white matter findings and with the 3MS in older adults (Heckbert, Longstreth, et al. 1997); the association between reported alcohol intake and cognition (Espeland, Gu, et al. 2005); the effect of hormone/hormone replacement therapy on cognition (Shumaker, Reboussin, et al. 1998; Steffens, Norton, et al. 1999; Yaffe, Haan, et al. 2000; Carlson, Zandi, et al. 2001; Rapp, Espeland, et al. 2003; Shumaker, Legault, et al. 2003; Whitmer, Haan, et al. 2003; Espeland, Rapp, et al. 2004; Shumaker, Legault, et al. 2004); the effect of Rivastigmine on Dementia with Lewy bodies (DLB) (Maclean, Collins, et al. 2001); the association between supplemental use of antioxidant vitamins and risk of significant cognitive decline (Maxwell, Hicks, et al. 2005); the effect of calcium-channel blockers and cognitive function (Maxwell, Hogan, et al. 1999); the effect of cognitive enhancement drug on cognition (Tariska and Paksy, 2000); use of herbal medicine and other dietary supplements (Nahin, Fitzpatrick, et al. 2006).

3) the effects of non-pharmacological interventions on cognitive status: the association between physical activity and cognitive function; the significance of music in the lives of senior individuals (Cohen, Bailey, et al. 2002); different types of CPR and cognitive outcome (Stiehl, Hebert, et al. 1996); the adverse cognitive effects of electroconvulsive therapy (Sobin, Sackeim, et al. 1995; Sackeim, Luber, et al. 2000); the effectiveness of cognitive nursing interventions (Abraham and Reel, 1992).

4) risk factors for dementia/cognitive impairment: the role of APOE genotype in modulating effects of other risk factors for cognitive decline (Haan, Shemanski, et al. 1999); the association between low folate status and impaired cognitive function dementia (Ramos, Allen, et al. 2005); the relation between total plasma homocysteine concentration and cognitive function (Miller, Green, et al. 2003); the predictive utility of olfactory identification deficits in patients with mild cognitive impairment for follow-up diagnosis of probable Alzheimer's disease (AD) (Devanand, Michaels-Marston, et al. 2000); the relationship between pantomime recognition and production in patients with AD (Dumon and Ska, 2000); the association between arm length and height and cognitive/functional abilities (Jeong, Kim, et al. 2005); socioeconomic differences in cognitive decline and the...
role of biomedical factors (Koster, Penninx, et al. 2005); the determinants of dementia (Kuller, Shemanski, et al. 1998; Kuller, Lopez, et al. 2003; Kuller, Lopez, et al. 2005); incidence, manifestations, and predictors of worsening white matter grade on serial imaging (Longstreth, Dulberg, et al. 2002; Longstreth, Arnold, et al. 2005); risk factors for mild cognitive impairment (Lopez, Jagust, et al. 2003); glucose tolerance and both AD and vascular dementia (Curb, Rodriguez, et al. 1999); comparison of dementia risks factors in terms of education and cognitive capacity between black and white populations (Shadlen, Siscovick, et al. 2006).


Applications:
It is used for the evaluation of cognitive function in both primary care/community dwelling and institutional care settings to detect change of cognitive status and cognitive impairment, and monitor response to treatment. People with various diagnostic criteria (e.g., dementia, AD, LBD, non-dementia/cognitively impaired, schizophrenia, depression, and cardiovascular disease) have been assessed using the 3MS. The 3MS has been used in both clinical and epidemiological studies. A couple of studies have been identified for a telephone adaptation of the 3MS (Norton, Tschanz, et al. 1999; Alexopoulos, Perneczyk, et al. 2006).

Carer and/or Patient Use of Instrument: Cognitive rating scale based on performance of set tasks. The 3MS is interviewer rated.
### Psychometric Criteria

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td>McDowell, Kristjansson, et al. (1997)</td>
<td>X Alpha &gt;0.70 □ Marginal or inadequate internal consistency (&lt;0.70) □ No information found on internal consistency</td>
<td>Excellent internal consistency with Cronbach’s alpha of 0.87 (compared to 0.78 for the MMSE). Split half reliability was 0.82 (0.76 for the MMSE). In a French version of the 3MS, Cronbach’s alpha of 0.80 was reported. Alpha was 0.90 for the 3MS (0.84 for the MMSE). Cronbach’s alpha for the Korean version of 3MS (K-mMMSE)² was 0.91, compared to 0.84 for the Korean version of MMSE (K-MMSE).</td>
</tr>
<tr>
<td>Test – retest</td>
<td>Teng and Chui (1987)</td>
<td>X ICC &gt;.70 Time intervals and confidence intervals reported □ Marginal or inadequate test-retest reliability ICC&lt;0.70 □ No information found on test-retest reliability</td>
<td>Excellent test-re-test reliability over delays between 52 and 98 days, with cutting point of 79/80, ranging from 0.91 to 0.93. (compared to 0.79 to 0.89 for the MMSE). One month stability coefficients were 0.8. (0.71 for the MMSE). Retest reliability was 0.92 (0.85 for the MMSE). The K-mMMSE also demonstrated excellent test-retest reliability (0.89) over mean interval delays of 26 days (range 19-32 days). In a French version of the 3MS a 14-day delay of the test-retest reliability coefficient was 0.87 A Canadian study of community-dwelling older persons with a diagnosis of dementia indicated a high intraclass correlation coefficient (ICC=0.85). In a two-phase community study.</td>
</tr>
<tr>
<td></td>
<td>Cappeliez, Quintal, et al. (1996)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Grace, Nadler, et al. (1995)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Correa, Perrault, et al. (2001)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

² K-mMMSE is NOT a Korean version of mMMSE. It is a slightly modified version of the 3MS designed to make the 3MS more suitable to Korean culture and language.
prevalence study the 3MS and the MMSE were implemented by a lay interviewer at first and by a nurse in the second phase (after a median delay of 49 days) an ICC for the 3MS was 0.87 while the MMSE had an ICC of 0.78.

A population-based longitudinal study of older people showed an intraclass correlation coefficient of 0.98, as was internal consistency (coefficient alpha=0.91). Test-retest reliability over 3 years was 0.78.

The intraclass correlation coefficient was 0.85 (95% confidence interval, CI: 0.81-0.88). Language (either tested in English or in French) did not make much difference as they both showed an ICC of 0.85.

<table>
<thead>
<tr>
<th>Inter – rater</th>
<th>Nadler, Relkin, et al. (1995)</th>
<th>X Agreement reported and adequate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bassuk and Murphy (2003)</td>
<td>□ Inadequate inter-rater agreement</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information provided</td>
</tr>
</tbody>
</table>

In a study conducted in a long-term care settings both the 3MS and the MMSE showed excellent inter-rater reliability (r=0.99).

Excellent inter-rater reliability, “free of rater bias” and an intraclass correlation coefficient=0.98).

<table>
<thead>
<tr>
<th>VALIDITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>McDowell (2006)</td>
<td>□ Patients were involved during item selection and/or item reduction</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Patients were consulted for reading and comprehension</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No patient involvement</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>X No information found on content validity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>X There is an adequate coverage of relevant domains</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ There is limited coverage of relevant domains</td>
<td></td>
</tr>
</tbody>
</table>

Given the MMSE was derived from existing instruments; it is safe to assume that the most domains of the 3MS originated from the existing theoretical premises.

Demonstrated in factor analysis described below, the 3MS appears to measure relevant domains of cognitive function.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Njegovans, Hing et al. (2001)</th>
<th>X Results were acceptable in accordance with the hypotheses and an</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Studies reported moderate to high construct validity in relation to hypothesised domains (functional capability)</td>
</tr>
</tbody>
</table>
other measures in a manner that is consistent with theoretically derived hypothesis concerning the domains that are measured.

<table>
<thead>
<tr>
<th>Construct: Internal Structure</th>
<th>Abraham, Manning, et al. (1993)</th>
<th>No evidence provided/failed a test of dimensionality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cappeliez, Quintal, et al. (1996)</td>
<td>Some evidence provided to support internal structure</td>
</tr>
<tr>
<td></td>
<td>Rapp, Espeland, et al. (2003)</td>
<td>Substantial evidence provided to support internal structure</td>
</tr>
</tbody>
</table>

A factor analytic study of the 3MS yielded five domains of psychomotor skills, memory, identification and association, orientation, and concentration and calculation. The solution explained 58.9% of the variance.

In the French version of the 3MS test, four factors accounting for 63.2% of the variance were reported.

In a clinical trial of hormone therapy for women, four factors (verbal memory with the heaviest loading, language and execution, orientation and language praxis) accounting for 37% of the total variance was reported.

<table>
<thead>
<tr>
<th>Construct: Correlation with other measures</th>
<th>Cappeliez, Quintal, et al. (1996)</th>
<th>Correlations with other measures are reported</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bassuk and Murphy (2003)</td>
<td>Correlations not reported</td>
</tr>
</tbody>
</table>

X Correlations with other measures are reported

Moderate to high correlations with other instruments testing cognition.

The 3MS was reported to be correlated with: the MMSE (0.90), the Blessed Dementia Scale (-0.80), the Cambridge Mental Disorders of the Elderly Examination (CAMDEX) Cognitive scale (CAMCOG) (0.85).

The correlation between the 3MS and the MMSE scores was 0.95.

Progressive cognitive decline is associated with a specific pattern of loss of functional tasks using instrumental activities of daily living (ADLs) and 14 Older American Resources and Services (OARS) items.

The K-mmMMSE showed significant correlations ($P < 0.001$ by Pearson’s correlation analyses) with Clinical Dementia Rating (CDR), Sum of Boxes of CDR (CDRSB), and Korean Instrumental Activities of Daily Living (K-IADL).

The correlation coefficient between K-mmMMSE and K-MMSE scores was 0.94. According to the CDR scores, the median values of the K-mmMMSE and K-MMSE changed significantly.


Adequate comparison measure was used

X Limited/inadequate construct validity reported

☐ No information provided

and activities of daily living).
The clinical utility study of the 3MS in the stroke population, in comparison with the MMSE, indicates that the 3MS yields consistently higher coefficients than the MMSE: Correlations with Boston Naming Test for language (0.61 for the 3MS and 0.55 for the MMSE); with Controlled Word Association for verbal fluency (0.81 for the 3MS and 0.59 for the MMSE); with the Logical Memory test (0.62 for the 3MS and 0.55 for the MMSE); and with the Functional Independence Measure (0.44 for the 3MS and 0.36 for the MMSE).

In a health, aging and body composition study, physical function measures (gait speed, chair stands, standing balance) were associated with both the 3MS and digit symbol substitution test (DSST) (p < 0.001).

### Construct: Discriminant Validity

The scale differentiates between relevant categories of respondent e.g. sick vs. well, varying degrees of severity

- Rosano, Simonsick, et al. (2005)
- McDowell (2006)
- Teng, Chui, et al. (1990)
- Rockwood, Tripp, et al. (1994)
- X Scale differentiates between relevant categories of respondents

No information on discriminant validity

Studies have shown moderate to high sensitivity and specificity of the 3MS in detecting dementia and severity of cognitive impairment in both community dwellings and long-term institutional settings. This indicates improved construct validity when compared with the MMSE.

At a specificity of 0.95, for people with 7 to 12 years of education the 3MS yielded sensitivity of 0.94 compared to sensitivity of 0.88 for the MMSE; for people with 13 or more years of education sensitivity was 0.91 for the 3MS and 0.86 for the MMSE.

Whilst the 3MS differentiated people with dementia from people without, it showed less competence for recognising people without cognitive impairment from those with Cognitively Impaired but No Dementia (CIND). This was improved, albeit not significantly, when Physical Function Measures (PFMs) were introduced, along with the 3MS.

The areas under the Receiver
operating characteristic (ROC) curves’ in identifying all levels of CIND or dementia were 0.91 for the K-mMMSE and 0.89 for the K-MMSE (P < 0.05). At the optimal cut-off score of 69/70 for a diagnosis of CIND using the K-mMMSE, a sensitivity of 0.86 (95% CI, 0.78–0.92) and a specificity of 0.79 (95%CI, 0.71–0.86) were reported, while, for a diagnosis of dementia, at the optimal cut-off score of 59/60, a sensitivity of 0.91 (CI, 0.79–0.98) and a specificity of 0.78 (95%CI, 0.72–0.84) were reported.

| Criterion | Information on the relationship of scores to gold standard measures or clinical diagnosis is provided | McDowell (2006) | X Comparison made to criterion measures □ No comparison with criterion measures provided | Various studies that examined the relationship of 3MS scores to clinical diagnosis of dementia showed high sensitivity, however this largely depends on cut-off points, use of normative data based on age, gender, education and ethnicity. Studies suggest the 3MS is a reasonable tool to screen for dementia.
Using standard cut-offs for impairment, the 3MS, MMSE, and Dementia Rating Scale (DRS) achieved high sensitivity (82% to 100%) but low specificity (33% to 52%) in the detection of dementia among nursing home residents (diagnosis was made based on DSM-III-R criteria by physicians specializing in geriatric medicine).
When the 3MS and the MMSE results were compared to a clinical diagnosis of dementia, sensitivity was 0.87 and specificity was 0.89. The area under the ROC curve was 0.94 for the 3MS compared to 0.89 for the MMSE. Analysis of a subset of the same study participants yielded slightly higher sensitivity (0.88) and specificity (0.90).

| Interpretability | The degree to which one can assign qualitative meaning to quantitative scores | McDowell (2006) | X Authors provide 2 or more types of information on interpretability □ Authors provide limited information to assist with interpretability □ No information | Studies that examined the 3MS have provided various cut-off points to screen for dementia and CIND as well as normative data for various age, gender and some ethnic groups, and for education levels. See construct and criterion validity.

1 used to determine the validity of the two screening tests graphically and statistically
<table>
<thead>
<tr>
<th>RESPONSIVENESS</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floor and ceiling effects</td>
<td>Teng and Chui (1987) McDowell (2006)</td>
<td>□ Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
<td>Large scaled studies have been conducted to obtain normative data for age, education and ethnic specific groups. Various studies demonstrated moderate to high sensitivity of the 3MS in detecting dementia and cognitive impairment. Psychometric properties, distribution and demographic correlates were developed for older people drawn from the Stirling County Study, which indicated the 3MS may be less prone to ceiling effects. Both the 3MS and the MMSE showed strongly skewed distributions, however, only 2.6% of the respondents scored perfectly on the 3MS compared to 22% on the MMSE.</td>
</tr>
<tr>
<td>□ No information provided on floor and ceiling effects</td>
<td>Bassuk and Murphy (2003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity to change</td>
<td>Maxwell, Hicks, et al. (2005)</td>
<td>X Hypotheses were formulated and results were in agreement □ An adequate metric was used (ES, SRM, comparison with external standard) □ No information on sensitivity to change</td>
<td>A longitudinal study that examined the association between supplemental use of antioxidant vitamins and risk of significant cognitive decline showed a possible protective effect for antioxidant vitamins in relation to cognitive decline (decrease in 3MS score of 10 points or more).</td>
</tr>
</tbody>
</table>

In a Canadian study of older community dwellers with dementia, individual score differences between a clinic assessment and a home assessment for the 3MS showed a normal distribution (mean of differences 0.2; SD 8.0; 95% CI: -16 to 16) which indicates the range of variability in a timeframe consistent with no change in cognition. The discrepancy between repeat 3MS scores can be as large as +/- 16.

The Women's Health Initiative Memory Study has provided descriptive statistics of the distribution of the 3MS baseline scores, and the associations of demographic information (i.e., age, education level and ethnicity).
The Modified Mini-Mental State Exam (3MS) is a highly recommended instrument in assessing a global cognitive status in older people applicable in both community and the institutional settings. It has superior psychometric properties than the MMSE and is extensively used in large scaled epidemiological studies internationally (mostly North American studies). An increasing number of studies use a translated version of the 3MS to achieve cultural appropriateness.

Reported: Yun-Hee Jeon

Date of report: 15/01/07

References


Teng EL, Chui HC and Gong A (1990) Comparisons between Mini-Mental State Examination (MMSE) and its modified version - the 3MS test In: Psychogeriatrics: Biomedical and Social Advances. Excerpta Medica, Amsterdam, pp.189-192.


Appendix 7.2: Alzheimer’s Disease Assessment Scale – Cognition

AHOC INSTRUMENT REVIEW SHEET

Title: Alzheimer’s Disease Assessment Scale – Cognition.

Abbreviations: ADAS-COG.

Author(s) Name: WG Rosen, RC Mohs, and KL Davis.

Author(s) Address: Dr KL Davis, M.D.
Dean, Mount Sinai School of Medicine and
Professor, Psychiatry and Professor Pharmacology and Biological
Chemistry
Mt Sinai School of Medicine
One Gustave L. Levy Place,
Box 1230,
New York, NY 10029
USA.

Supplied by: The authors, or from the Alzheimer’s Disease Cooperative Study (ADCS).
Ph. (858) 6225879, Fax. (858) 4520573.

Cost: The ADAS provides a packet that includes the instrument and instructions, free of charge to academic, non-profit, commercial, and pharmaceutical research groups. The ADAS test kits, which include the objects (e.g., flower, bed, and whistle) and word recognition and recall index cards with rings and stands, can be purchased if required.

Training requirements: The instrument is usually administered by a neuropsychologist or psychologist and training is required.

Purpose: To evaluate cognitive impairment in the assessment of Alzheimer’s disease.

Administration time: 30 – 45 minutes

Instrument Type: Interviewer administered cognitive rating scale.

Structure: The instrument consists of 11 items as follows:
- Memory (Orientation to time place and person, Word recall, Word Recognition, and Recall of test instructions on word recognition);
- Language (Naming objects and Fingers, Spoken language, Language);
- Comprehension (Word finding difficulty, and Following commands); and
- Praxis (Ideational praxis and Constructional praxis)

Scoring: Continuous rating scale with a total maximum score of 70
- Memory: Orientation (8 points), Word recall (10 points), Word recognition (12 points) and Recall of test instructions on word recognition (5 points);
- Language: Naming objects and Fingers, Spoken language, Language (5 points each)
- Comprehension: Word finding difficulty, and Commands (5 points each);
- Praxis: Ideational praxis and Constructional praxis (5 points each)

The ADAS-Cog total, which indicates level of impairment, is the sum of the scores on each of the items. High score indicates greater impairment. Item scores and domain scores are also used in research studies.
Developed for: Use across a range of settings to determine the level of cognitive impairment in Alzheimer's disease.

Normative Data: Limited normative data is available. Normative data for older adults is available (Graham, Cully, Snow, Massman, et al. 2004), and mean ADAS-Cog scores for age, educational level and Global Deterioration Scale stage have been reported (Doraiswamy, Bieber, Kaiser, Krishnan, et al. 1997). The Normacodem project has also adapted and normalised the instrument for use in Spanish speaking countries (Pena-Casanova, Aguilar, Santacruz, Bertran-Serra, et al. 1997; Pena-Casanova, Meza, Bertran-Serra, Manero, et al. 1997).


Applications: Evaluation of cognitive impairment in the assessment of Alzheimer's disease and other dementias, evaluation of change in cognitive impairment over time, outcome measure in intervention studies including clinical trials and other research studies evaluating associations between cognitive impairment and other variables.

Carer and/or Patient Use of Instrument: Interviewer administered; patient response.
## Psychometric Criteria

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test – retest</strong></td>
<td>Rosen, Mohs, et al. (1984) Kim, Nibbelink, et al. (1994) Pena-Casanova, Meza, et al. (1997) Weyer, Erzigkeit, et al. (1997) Chu, Chiu, et al. (2000) Liu, Teng, et al. (2002)</td>
<td>X ICC &gt;.70</td>
<td>Excellent test-re-test reliability with ICC’s of 0.86 - 0.96 (total score); 0.33 - 0.89 (individual items, 0.33 was for praxis in Liu, et al. 2002). Test re-test was also conducted on the factor structure (Kim, et al. 1994) and ICCs were 0.78 - 0.87.</td>
</tr>
<tr>
<td><strong>Inter – rater</strong></td>
<td>Mohs and Cohen (1988) Rosen, Mohs, et al. (1984) Chu, Chiu, et al. (2000) Liu, Teng, et al. (2002)</td>
<td>X Agreement reported and adequate □ Inadequate inter-rater agreement □ No information provided</td>
<td>Excellent inter-rater reliability with total ICCs ranging from 0.91 to 0.99 and ICCs for individual items ranging from 0.76 to 1.00.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VALIDITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Content</strong></td>
<td>Rosen, Mohs, et al. (1984)</td>
<td>□ Patients and experts were involved during item selection and/or item reduction X Patients were consulted for reading and comprehension □ No patient involvement</td>
<td>The instrument has been developed by a team of experts in the field. Patients were involved as participants in validation studies. Relevant domains are adequately covered.</td>
</tr>
</tbody>
</table>
### Construct
The extent to which scores on the questionnaire relate to other measures in a manner that is consistent with theoretically derived hypothesis concerning the domains that are measured.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Doraiswamy, Kaiser, et al. (2001)</th>
<th>X Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Suh, Ju, et al. (2004)</td>
<td>□ Limited / inadequate construct validity reported □ No information provided</td>
</tr>
<tr>
<td></td>
<td>Feldman, Van Baelen, et al. (2005)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lam, Lui, et al. (2005)</td>
<td></td>
</tr>
</tbody>
</table>

Expected correlations were found between ADAS-Cog the number of memory complaints, (Memscore) and with the DAD (Disability Assessment in Dementia).

### Construct: Internal Structure
Information provided on factor structure

<table>
<thead>
<tr>
<th>Construct: Internal Structure</th>
<th>Kim, Nibbelink, et al. (1994)</th>
<th>□ No evidence provided/failed a test of dimensionality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Talwalker, Overall, et al. (1996)</td>
<td>X Some evidence provided to support internal structure</td>
</tr>
<tr>
<td></td>
<td>Doraiswamy, Bieber, et al. (1997)</td>
<td>□ Substantial evidence provided to support internal structure</td>
</tr>
<tr>
<td></td>
<td>Doraiswamy, Bieber, et al. (1997)</td>
<td>□ No evidence provided/failed a test of dimensionality</td>
</tr>
<tr>
<td></td>
<td>Doraiswamy, Kaiser, et al. (2001)</td>
<td>□ No evidence provided/failed a test of dimensionality</td>
</tr>
<tr>
<td></td>
<td>Liu, Teng, et al. (2002)</td>
<td>□ No evidence provided/failed a test of dimensionality</td>
</tr>
</tbody>
</table>

Internal structure has been confirmed with studies citing significant item to item correlations for at least 50% of the items (0.17 to 0.55) and item total correlations of 0.52 to 0.90.

Significant correlations between domain scores (0.47 to 0.52) have also been reported, and the factor structure confirmed has been confirmed through factor analysis.

### Construct: Correlation with other measures
Comparisons made to other measures

<table>
<thead>
<tr>
<th>Construct: Correlation with other measures</th>
<th>Blessed, Tomlinson, et al. (1968)</th>
<th>X Correlations with other measures are reported</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cutler, Shrotriya, et al. (1993)</td>
<td>□ Correlations not reported</td>
</tr>
<tr>
<td></td>
<td>Burch and Andrews (1987)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ihl, Frolich, et al. (1992)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zec, Landreth, et al. (1992)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Doraiswamy, Bieber, et al. (1997)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Doraiswamy, Bieber, et al. (1997)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pena-Casanova, Meza, et al. (1997)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weyer, Erzigkeit, et al. (1997)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ihl, Grass-Kapanke, et al. (1999)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chu, Chiu, et al. (2000)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Doraiswamy, Kaiser, et al. (2001)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hannesdottir and Snaedal (2002)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Liu, Teng, et al.</td>
<td></td>
</tr>
</tbody>
</table>

The instrument has been found to correlate with the following instruments measuring cognitive impairment:

- Mini Mental State Examination (MMSE);
- Brief Cognitive Rating Scale (BCRF);
- Memory and Information Test (MIT);
- Cognitive Abilities Scoring Instrument (CASI);
- Nurses Observation Scale for Geriatric impairment (NOSGER);
- Syndrom-Kurz-Test (SKT);
- Geriatric Evaluation by Relatives Rating Scale (GERI).

It also correlates with The Clinical Dementia Rating Scale.
| Criterion | □ Comparison made to criterion measures X No comparison with criterion measures provided | No studies found. |
known functional measures or clinical diagnosis

Information on the association between changes in scores and patients’ global ratings of the magnitude of change they have experienced

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Silvestrini, Pasqualetti, et al. (2006)</td>
<td></td>
</tr>
</tbody>
</table>
### RESPONSIVENESS

<table>
<thead>
<tr>
<th>Floor and ceiling effects</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Sensitivity to change</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>The ability to detect important change over time in the concept being measured</td>
<td>Rosen, Mohs, et al. (1984) Farlow, Gracon, et al. (1992) Holford and Peace (1992) Weiner, Vobach, et al. (1993) Zemlan (1996) Rogers, Farlow, et al. (1998) Schmeidler, Mohs, et al. (1998) Burns, Rosso, et al. (1999) Imbimbo, Troetel, et al. (2000) Doraiswamy, Kaiser, et al. (2001) Tariot, Cummings, et al. (2001)</td>
<td>Hypotheses were formulated and results were in agreement □ An adequate metric was used (ES, SRM, comparison with external standard) □ No information on sensitivity to change was provided □ MCID - Information was provided about the magnitude of score differences which would be clinically meaningful □ MCID – No information was provided.</td>
<td>The instrument is sensitive to change over time. Studies show a significant change in scores over time (as expected) as the disease progresses. Sensitivity to treatment effects is also evident in clinical trials investigating the effectiveness of drug treatment in persons with Alzheimer’s disease and also with vascular dementia. ADAS-Cog scores improve significantly as do scores on other measures in the same studies, namely Clinical Dementia Rating (CDR), Mini Mental State Exam (MMSE) and Clinician’s Interview-Based Impression of Change (CIBIC).</td>
</tr>
</tbody>
</table>
Cultural Applicability and Cultural Adaptations: The ADAS-Cog has been translated into most European languages as well as Chinese, Turkish, Indian and Brazilian. Particular attention has been given to Spanish versions (see references below).

Gender Appropriateness: Appropriate for use with both genders.

Age Appropriateness: Appropriate for use with adults.

Summary: The ADAS-Cog is extensively used and cited. It is administered by a neuropsychologist or psychometrician and training is required. Administration time is very long (30 to 45 minutes), depending on level of impairment. Its' psychometric properties are excellent and the instrument is readily available in many languages. ADAS-Cog is widely used as an outcome measure in clinical trials as well as other research studies. However, the ADAS-Cog should be compared with newer instruments (using other memory recall and recognition items) which also provide a detailed assessment of cognitive function.

Reporter: Siggi Zapart

Date of report: 18/01/2007

References


Appendix 7.3: The General Practitioner Assessment of Cognition

AHOC INSTRUMENT REVIEW SHEET

Title: The General Practitioner Assessment of Cognition.

Abbreviations: GPCOG.

Author(s) Name: Henry Brodaty, Dimity Pond, Nicola M. Kemp, Georgina Luscombe, Louise Harding, Karen Berman, and Felicia A. Huppert.

Author(s) Address: Professor H. Brodaty
Academic Department for Old Age Psychiatry
Prince of Wales Hospital
Randwick NSW 2031
Australia.

Supplied by: The instrument can be readily obtained. It is attached to the original paper outlining its development.

Cost: Free.

Training requirements: None.

Purpose: To assist General Practitioners in detecting dementia.

Administration time: 4 to 5 minutes.

Instrument Type: Cognitive rating scale based on interviewer administered questionnaire (patient and informant reported).

Structure: The instrument has 9 items covered in two sections: Cognitive testing (patient examination) of four items (Word recall, Time orientation, Clock drawing, Reporting a recent event); and historical (informant interview) of 6 items (Patient's memory of recent conversations, Misplacing objects, Word finding difficulties, Ability to manage money, Ability to manage medication, and need for travel assistance).

Scoring: The patient section score is the total number of correct responses (maximum score of 9). A score of 9 indicates no cognitive impairment, a score of 4 or lower suggests cognitive impairment. For the patient section if scores are in the range of 5 – 8 then cognitive impairment is regarded as being doubtful / uncertain. Then the informant section should be completed to obtain more information.

The informant section is the total number of "no" responses (maximum score of 6). A score of 3 or lower suggests cognitive impairment.

Developed for: General Practitioners - to assist them in detecting cognitive impairment/dementia in the primary care setting.

Normative Data: This is a new instrument and normative data is not yet available.

Clinical Data: This is a new instrument and clinical data is not yet available.

Applications: Detection of cognitive impairment/dementia in the primary care setting.

Carer and/or Patient Use of Instrument: Interviewer administered, patient and informant responses.
## Psychometric Criteria

### RELIABILITY

<table>
<thead>
<tr>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td>Only one study but very good internal consistency with $\alpha = 0.84$ for both the patient and informant sections.</td>
</tr>
<tr>
<td>Test – retest</td>
<td>Test-retest reliability very good with ICC of 0.87 for both patient and informant sections.</td>
</tr>
<tr>
<td>Inter – rater</td>
<td>Inter-rater reliability good for patient section with ICC of 0.75, and satisfactory for informant section (ICC of 0.56).</td>
</tr>
</tbody>
</table>

#### Studies Reported & References

- Brodaty, Pond, et al. (2002)

#### Adequacy Checks

- X Alpha >0.70
- X ICC >.70
- X Agreement reported and adequate

#### Comment

- □ Marginal or inadequate internal consistency (<0.70)
- □ Marginal or inadequate test-retest reliability ICC<.70
- □ No information found on internal consistency
- □ No information found on test-retest reliability
- □ Inadequate inter-rater agreement
- □ No information provided

### VALIDITY

<table>
<thead>
<tr>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>Developed by a team of experts. Items were derived from other instruments measuring cognitive and physical functioning abilities and other psychogeriatric assessments.</td>
</tr>
</tbody>
</table>

#### Studies Reported & References

- Brodaty, Pond, et al. (2002)

#### Adequacy Checks

- □ Patients and experts were involved during item selection and/or item reduction
- □ Patients were consulted for reading and comprehension
- □ No patient involvement
- □ No information found on content validity
- □ There is an adequate coverage of

#### Comment

- □ No information on content validity
- □ No information on patient involvement

---

Centre for Health Service Development

Final Report: Dementia Outcomes Measurement Suite Project
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Construct: Internal Structure</strong></td>
<td>Information provided on factor structure</td>
<td>Correlations with other measures are reported</td>
<td>Scale significantly correlated with other instruments measuring cognitive functioning: Mini-Mental State Exam (MMSE) and Global Deterioration Scale (GDS).</td>
</tr>
<tr>
<td><strong>Construct: Correlation with other measures</strong></td>
<td>Comparisons made to other measures</td>
<td>Correlations not reported</td>
<td>The diagnostic accuracy with DSM-IV criteria of the instruments attest to it’s discriminant validity (Brodaty, et al. 2002).</td>
</tr>
<tr>
<td><strong>Construct: Discriminant Validity</strong></td>
<td>The scale differentiates between relevant categories of respondent e.g. sick vs. well, varying degrees of severity</td>
<td>Scale differentiates between relevant categories of respondents</td>
<td>Patient section: Area under ROC was 0.86. Sensitivity and specificity were 82% and 70% respectively. Positive and negative predictive value were 0.53 and 0.90. Informant section: Area under ROC was 0.84. Sensitivity and specificity were 89% and 66% respectively. Positive and negative predictive values were 0.52 and 0.94 (Thomas, et al. 2006). These results were similar to those for MMSE. Scores also differentiated between patients with and without dementia (Brodaty, et al. 2004).</td>
</tr>
<tr>
<td><strong>Criterion</strong></td>
<td><strong>Comparison made to criterion measures</strong></td>
<td><strong>See the studies described above in the Discriminant Validity section.</strong></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>------------------------------------------</td>
<td>-------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Information on the relationship of scores to gold standard measures or clinical diagnosis is provided</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Interpretability</strong></th>
<th><strong>Interpretability</strong></th>
<th><strong>Interpretability</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>The degree to which one can assign qualitative meaning to quantitative scores</td>
<td>Brodaty, Pond, et al. (2002) Brodaty, Kemp, et al. (2004) Thomas, Lalloue, et al. (2006)</td>
<td>X Authors provide 2 or more types of information on interpretability Authors provide limited information to assist with interpretability No information provided</td>
</tr>
<tr>
<td>Do authors provide the following:</td>
<td>Studies provide means, standard deviations and confidence intervals. They also provide information on the relationship of scores to other measures (including the GDS, SF-12, MMSE and AMT).</td>
<td></td>
</tr>
<tr>
<td>Presentation of means and SD of scores before and after treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparative data on the distribution of scores in relevant subgroups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information on the relationship of scores to well-known functional measures or clinical diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information on the association between changes in scores and patients' global ratings of the magnitude of change they have experienced</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>RESPONSIVENESS</strong></th>
<th><strong>Studies Reported &amp; References</strong></th>
<th><strong>Adequacy Checks</strong></th>
<th><strong>Comment</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Floor and ceiling effects</td>
<td></td>
<td>X Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected X No information provided on floor and ceiling effects</td>
<td>No studies found.</td>
</tr>
<tr>
<td>Sensitivity to change</td>
<td>□ Hypotheses were formulated and results were in agreement</td>
<td>□ An adequate metric was used (ES, SRM, comparison with external standard)</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>---------------------------------------------------------</td>
<td>-----------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>The ability to detect important change over time in the concept being measured</td>
<td>□ No information on sensitivity to change was provided</td>
<td>X No information on sensitivity to change was provided</td>
<td></td>
</tr>
<tr>
<td></td>
<td>□ MCID - Information was provided</td>
<td>X MCID – No information was provided.</td>
<td></td>
</tr>
</tbody>
</table>

| No studies found. |

**Cultural Applicability and Cultural Adaptations:** No information published regarding culture or language bias. The instrument has also not been translated into other languages as yet.

**Gender Appropriateness:** Appropriate for use with both genders.

**Age Appropriateness:** Appropriate for use with adults.

**Summary:** This is a new instrument designed to be used in the primary care setting. The instrument combines cognitive and informant data. It is quick (four to five minutes) and easy to administer and its’ psychometric properties are very good. Due to its newness, it has not yet been widely used but the evidence to date suggests that is very suitable for use by general practitioners for the detection of dementia in this setting.

**Reporter:** Siggi Zapart

**Date of report:** 19/01/07

**References**


Appendix 7.4: Rowland Universal Dementia Assessment Scale

AHOC INSTRUMENT REVIEW SHEET

Title: Rowland Universal Dementia Assessment Scale.

Abbreviations: RUDAS.

Author(s) Name: Joella E Storey, Jeffrey T.J. Rowland, David A Conforti and Hugh G Dickson.

Author(s) Address: Joella E Storey,
Aged Care Research
Liverpool Hospital
Locked Bag 7103
Liverpool BC NSW 1871
Australia

Supplied by: The instrument is available from the authors, with permission.

Cost: $15.00 (includes training video, books and scoring sheets).

Training requirements: Approximately 40 minutes of training (using videotape).

Purpose: Short cognitive screening tool, for the assessment of dementia.

Administration time: 10 minutes.

Instrument Type: Interviewer administered, patient response questionnaire.

Structure: Six item questionnaire covering the following cognitive domains:
- Memory (memorise and delayed recall of 4 shopping items);
- Visuo-spatial orientation (naming part of the body);
- Praxis (hand fist exercise);
- Visuo-constructional drawing (cube drawing);
- Judgement (person describes what they would do if they need to cross a busy street with no crossing or traffic lights);
- Language (number of animals named in 1 minute).

Scoring: The instrument is scored out of 30 with scores below 23 suggesting dementia. Item scores are summed to give a total score. Individual items scores are as follows:
- Memory: 2 points for each item recalled. Total possible score = 8.
- Visuo-spatial orientation: 1 point for each body part correctly identified, once 5 correct parts are identified, this section is discontinued. Total possible score = 5.
- Praxis: 3 point scale – 0 = failed, 1 = partially/adequate, 2 = normal. Total possible score = 2.
- Visuo-constructional drawing: 1 point for each of base drawn, all internal lines appear, and all external lines appear. Total possible score = 3.
- Judgement Items: 2 points for each for: look for traffic, additional safety proposal. Total possible score = 4.
- Language: 1 point for each animal named. This section is discontinued after 8 animals have been named. Total possible score = 8.

**Developed for:**
Assessment of cognitive impairment/dementia in culturally diverse populations.

**Normative Data:**
This is a relatively new instrument and normative data is not available at this stage.

**Clinical Data:**
Storey, et al. (2004) provide clinical data on 166 geriatric medicine outpatients when developing the scale.

**Applications:**
Assessment of cognitive status, at diagnosis stage, over time, and as an outcome measure.

**Carer and/or Patient Use of Instrument:**
Interviewer administered; patient response questionnaire.

### Psychometric Criteria

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td>The extent to which items in a (sub) scale are inter-correlated; a measure of the homogeneity of a (sub)scale. Cronbach’s alpha should be between 0.70 and 0.90 for every dimension / sub-scale.</td>
<td>□ Alpha &gt;0.70 □ Marginal or inadequate internal consistency (&lt;0.70) X No information found on internal consistency</td>
<td></td>
</tr>
<tr>
<td>Test – retest</td>
<td>The extent to which the same results are obtained on repeated administrations of the same questionnaire when no change in physical functioning has occurred. Calculation of an intraclass correlation coefficient (ICC); and an ICC &gt; 0.70 is desired. Preferred if time interval and confidence intervals were presented</td>
<td>Storey, Rowland, et al. (2004) X ICC &gt;.70 Time intervals and confidence intervals reported □ Marginal or inadequate test-retest reliability ICC&lt;.70 □ No information found on test-retest reliability</td>
<td>Excellent test-retest reliability with ICC of 0.98.</td>
</tr>
<tr>
<td>Inter – rater</td>
<td>Limits of agreement, Kappa, or standard error of measurement (SEM) were presented</td>
<td>Storey, Rowland, et al. (2004) X Agreement reported and adequate □ Inadequate inter-rater agreement □ No information provided</td>
<td>Excellent inter-rater reliability with ICC of 0.99.</td>
</tr>
<tr>
<td>VALIDITY</td>
<td>Studies Reported &amp; References</td>
<td>Adequacy Checks</td>
<td>Comment</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Content</td>
<td>Storey, Rowland, et al. (2004)</td>
<td>□ Patients and experts were involved during item selection and/or item reduction</td>
<td>Developed by a team of experts the field of dementia care, in consultation with representatives from 22 cultural and linguistic groups.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>X Patients were consulted for reading and comprehension</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No patient involvement</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information found on content validity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>X There is an adequate coverage of relevant domains</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ There is limited coverage of relevant domains</td>
<td></td>
</tr>
<tr>
<td>Construct</td>
<td>Storey, Rowland, et al. (2004)</td>
<td>□ Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Limited /inadequate construct validity reported</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>X No information provided</td>
<td></td>
</tr>
<tr>
<td>Construct: Internal</td>
<td></td>
<td>□ No evidence provided/failed a test of dimensionality</td>
<td>Item – total correlations ranged from 0.35 to 0.50.</td>
</tr>
<tr>
<td>Structure</td>
<td></td>
<td>X Some evidence provided to support internal structure</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Substantial evidence provided to support internal structure</td>
<td></td>
</tr>
<tr>
<td>Construct: Correlation</td>
<td>Rowland, Basic, et al. (2006b)</td>
<td>X Correlations with other measures are reported</td>
<td>Scores significantly correlated with Mini-Mental State Exam (MMSE).</td>
</tr>
<tr>
<td>with other measures</td>
<td></td>
<td>□ Correlations not reported</td>
<td></td>
</tr>
<tr>
<td>Construct: Discriminant</td>
<td>Storey, Rowland, et al. (2004)</td>
<td>X Scale differentiates between relevant categories of respondents</td>
<td>Instrument has good diagnostic accuracy. Studies show Area under the receiver operated curves (ROC) figures ranging from 0.86 to 0.94, sensitivity and specificity ranging from 72 to 89% and 76 to 100%. These are better than for MMSE and GPCOG.</td>
</tr>
<tr>
<td>Validity</td>
<td>Rowland, Basic, et al. (2006a)</td>
<td>□ No information on discriminant validity</td>
<td></td>
</tr>
<tr>
<td></td>
<td>lype, Ajitha, et al. (2006)</td>
<td>X Scale differentiates between relevant categories of respondents</td>
<td></td>
</tr>
<tr>
<td>Criterion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------------------------------------</td>
<td>----------------------</td>
<td>----------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Information on the relationship of scores to gold standard measures or clinical diagnosis is provided</td>
<td>Comparison made to criterion measures</td>
<td>X No comparison with criterion measures provided</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interpretability</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>The degree to which one can assign qualitative meaning to quantitative scores</td>
<td>Storey, Rowland, et al. (2004)</td>
<td>Rowland, Basic, et al. (2006a)</td>
<td>Rowland, Basic, et al. (2006b)</td>
</tr>
<tr>
<td>Do authors provide the following:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presentation of means and SD of scores before and after treatment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparative data on the distribution of scores in relevant subgroups</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information on the relationship of scores to well-known functional measures or clinical diagnosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information on the association between changes in scores and patients’ global ratings of the magnitude of change they have experienced</td>
<td>X Authors provide 2 or more types of information on interpretability</td>
<td>□ Authors provide limited information to assist with interpretability</td>
<td>□ No information provided</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RESPONSIVENESS</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Floor and ceiling effects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The questionnaire fails to demonstrate a worse score in patients who clinically deteriorated and an improved score in patients who clinically improved</td>
<td>□ Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
<td>□ Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score</td>
<td>X No information provided on floor and ceiling effects</td>
</tr>
<tr>
<td>Authors should provide descriptive statistics of the distribution of scores</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Studies provide means, standard deviations and confidence intervals. They also provide information on relationship of scores to other measures.
Sensitivity to change
The ability to detect important change over time in the concept being measured

- □ Hypotheses were formulated and results were in agreement
- □ An adequate metric was used (ES, SRM, comparison with external standard)
- X No information on sensitivity to change was provided
- □ MCID - Information was provided about the magnitude of score differences which would be clinically meaningful
- X MCID – No information was provided.

No studies available.

Cultural Applicability and Cultural Adaptations:
The RUDAS items can be directly translated and are relevant to most cultures. It can be easily used with persons from Non English Speaking Backgrounds with the help of an interpreter. It can also be readily translated into other languages without the need to change the structure or the format of any item. One item may not be appropriate for use with Indigenous people in remote locations and its applicability for use with Indigenous peoples needs to be assessed.

Gender Appropriateness: Appropriate for use with both genders.
Age Appropriateness: Appropriate for use with adults.

Summary:
The RUDAS is a short multicultural cognitive screening tool for the assessment of dementia. It was developed and validated in an area where 40% of the population are born in non-English speaking countries and more than 80 languages are spoken. Developers included experts in the field of dementia care and representatives from 22 cultural and linguistic groups. The items are culturally fair and easily translated. The instrument is interviewer administered and takes about 10 minutes to complete. Training is required but is readily available at a low cost of $15.00. Evidence relating to psychometric properties is limited as the instrument is new, but existing data is promising with results indicating the instrument is valid and reliable.

Reporter: Madeleine King and Siggi Zapart
Date of report: 19/1/07

References


Appendix 7.5: **Minimum Data Set Cognition Scale**

**AHOC INSTRUMENT REVIEW SHEET**

**Title:** Minimum Data Set Cognition Scale.  
InterRAI – Long Term Care Facility (LTCF) – Cognitive section.

**Abbreviations:** MDS-COG.

**Author(s) Name:** The authors of the original MDS-Cog are Susan L Hartmeier, Philip D Sloane, Harry A Guess and Gary G Koch. This version has been incorporated into the INTER-RAI - Nursing Home Edition developed by InterRAI Australia.

**Author(s) Address:** Philip D Sloane  
Department of Family Medicine,  
School of Family Medicine  
C.B.7595, University of North Carolina  
Chapel Hill NC 27599.

InterRAI equivalent:  
c/- The University of Queensland  
School of Medicine  
Academic Unit in Geriatric Medicine  
Princess Alexandra Hospital  
Building 1, Level 4  
Ipswich Rd  
Woolloongabba  
Queensland 4102.

**Supplied by:** InterRAI protocols are copyright and permission to use these tools must be obtained from interRAI. For research and development purposes, approval may be obtained from the interRAI Australia Coordinating Centre. For ongoing use, a formal agreement must be made with interRAI central office. This can be facilitated through the Australian Centre.

Email: interrai@soms.uq.edu.au

**Cost:** Use of the MDS-Cog/interRai LTCF, in the nursing home setting, is computerised. The fee is therefore tied up with the licence for the computer software and the cost is negotiated with the software provider. Use of the paper version by researchers, for a limited period, can be negotiated with interRAI and is free.

**Training requirements:** Training that incorporates the use of the instrument and the software applications is recommended.

**Purpose:** Cognitive screening tool to assess level of cognitive impairment and or dementia status.

**Administration time:** 10 to 20 minutes.

**Instrument Type:** Interviewer administered – data routinely collected by staff on entry into care facilities.

**Structure:** The MDS-COG combines 8 items from the Minimum Data Set (MDS), in use at all care facilities, into a simple 10 point additive scale. Items cover the following domains:

- Cognitive patterns – 6 items: short term memory, long term memory, location of own room, knows he/she is in a nursing home, no orientation items recalled and decision making.
- Communication patterns – 1 item: making self understood.
- Physical Functioning – 1 item: dressing self performance.

**Scoring:**
The instrument is scored as an additive scale that ranges from 0 = no cognitive impairment to 10 = very severe cognitive impairment. Individual items are scored as follows:
- Items 1 to 5 are scored either 0 or 1.
- Items 6 and 7 are on a 4 point scale of 0, 1, 2, or 3.
- Item 8 is on a 5 point scale of 0, 1, 2, 3 or 4.

**Developed for:**
Use in long term care facilities.

**Normative Data:**
There is no normative data available.

**Clinical Data:**
There is no clinical data for the MDS-COG itself. However there is data available for the MDS-CPS. The instrument has been used in studies evaluating the effect of long term analgesics (Won, Lapane, Vallow, Schein, et al. 2004), and estrogen use (Fernandez & Lapane, 2000) in nursing home residents, investigating clinical correlates of residents with stroke (Quilliam & Lapane, 2001), and comparing nursing home residents with multiple sclerosis and dementia to those with multiple sclerosis only (Buchanan, Martin, Moore, Wang, et al. 2005).

The MDS-CPS has also been used in studies assessing psychological or behavioural problems. One study examined the relationship between anhedonic and dysphoric symptoms with diagnosed depression and anti-depressant treatment in institutionalised adults (Stones, Clyburn, Gibson & Woodbury, 2006). Another investigated resident characteristics associated with physical aggression and verbal abuse (Leonard, Tinetti, Allore & Drickamer, 2006). The validity of using the MDS-CPS for assessing the cognitive status of persons with schizophrenia in nursing homes was has evaluated (Bowie, Fallon & Harvey, 2006).

**Applications:**
Cognitive screening for dementia in long term care facilities.

**Carer and/or Patient Use of Instrument:**
Interviewer administered.

**Psychometric Criteria**

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td>Gruber-Baldini, Zimmerman, et al. (2000)</td>
<td>X Alpha &gt;0.70</td>
<td>□ Marginal or inadequate internal consistency (&lt;0.70)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information found on internal consistency</td>
<td>This study cites an excellent internal consistency of α = 0.85.</td>
</tr>
<tr>
<td>Test – retest</td>
<td></td>
<td>□ ICC &gt;.70</td>
<td>Time intervals and confidence intervals reported</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Marginal or inadequate test-retest reliability ICC&lt;.70</td>
<td></td>
</tr>
<tr>
<td>Functioning has occurred</td>
<td>X No information found on test-retest reliability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calculation of an intraclass correlation coefficient (ICC); and an ICC &gt; 0.70 is desired</td>
<td>Preferred if time interval and confidence intervals were presented</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inter- rater</td>
<td>□ Agreement reported and adequate □ Inadequate inter-rater agreement X No information provided</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limits of agreement, Kappa, or standard error of measurement (SEM) were presented</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## VALIDITY

<table>
<thead>
<tr>
<th>Content</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire</td>
<td>□ Patients and experts were involved during item selection and/or item reduction □ Patients were consulted for reading and comprehension □ No patient involvement</td>
<td>X No information found on content validity X There is an adequate coverage of relevant domains □ There is limited coverage of relevant domains</td>
<td></td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Construct</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
</table>
| The extent to which scores on the questionnaire relate to other measures in a manner that is consistent with theoretically derived hypothesis concerning the domains that are measured. | Gruber-Baldini, Zimmerman, et al. (2000) | □ Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used □ Limited/inadequate construct validity reported □ No information provided | The instrument showed expected correlations with the following instruments:  
Moderate correlation with staff rating on the Psychogeriatric Dependency Rating Scale (PGDRS) - orientation and behaviour scales ($r = 0.66$ and 0.31).  
Katz Activities of Daily Living (ADL) ($r = -0.50$). |

<table>
<thead>
<tr>
<th>Construct: Internal Structure</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information provided on factor structure</td>
<td>Gruber-Baldini, Zimmerman, et al. (2000)</td>
<td>□ No evidence provided/failed a test of dimensionality X Some evidence provided to support internal structure □ Substantial evidence provided to support internal</td>
<td>Findings from the study provide evidence to support internal structure with Item-total correlations ranging from 0.32 to 0.81</td>
</tr>
<tr>
<td>Construct: Correlation with other measures</td>
<td>Hartmaier, Sloane, et al. (1994) Cohen-Mansfield, Taylor, et al. (1999) Gruber-Baldini, Zimmerman, et al. (2000)</td>
<td>X Correlations with other measures are reported □ Correlations not reported</td>
<td>Construct validity is supported by correlations with the other measures of cognitive ability: Correlations with the Global Deterioration Scale (GDS) ranged from kappa = 0.82 - 0.88; r = 0.92 - 0.77. Percentage agreement with GDS stages 5 and 7 was 60 and 80%. Correlations with MMSE were also strong ranging from r = -0.68 to -0.88. Correlations with the minimum Data Set Cognitive Performance Scale (MDS-CPS) were also high ranging from r = 0.92 - 0.93.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Construct: Discriminant Validity</td>
<td>Hartmaier, Sloane, et al. (1994) Gruber-Baldini, Zimmerman, et al. (2000)</td>
<td>X Scale differentiates between relevant categories of respondents □ No information on discriminant validity</td>
<td>Hartmeier, et al. showed the instrument has very good diagnostic accuracy. The total area under the Receiver Operated Curve (ROC) was 0.94 and sensitivity and specificity were 89 and 98%. Results for ROC, sensitivity and specificity for the two sub samples investigated (133 and 67 participants) were 0.96, 95, 98%; and 0.97, 94%, 100%. Positive and negative predictive values for the two samples were 0.96, 0.85; and 1.00. 0.82.</td>
</tr>
<tr>
<td>Criterion</td>
<td>Information on the relationship of scores to gold standard measures or clinical diagnosis is provided</td>
<td>□ Comparison made to criterion measures X No comparison with criterion measures provided</td>
<td></td>
</tr>
</tbody>
</table>
Information on the relationship of scores to well-known functional measures or clinical diagnosis

Information on the association between changes in scores and patients’ global ratings of the magnitude of change they have experienced

<table>
<thead>
<tr>
<th>RESPONSIVENESS</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floor and ceiling effects</td>
<td>The questionnaire fails to demonstrate a worse score in patients who clinically deteriorated and an improved score in patients who clinically improved</td>
<td>□ Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected  □ Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score  X No information provided on floor and ceiling effects</td>
<td>No studies found.</td>
</tr>
<tr>
<td>Sensitivity to change</td>
<td>The ability to detect important change over time in the concept being measured</td>
<td>□ Hypotheses were formulated and results were in agreement  □ An adequate metric was used (ES, SRM, comparison with external standard)  X No information on sensitivity to change was provided  □ MCID - Information was provided about the magnitude of score differences which would be clinically meaningful  X MCID – No information was provided.</td>
<td>No studies found.</td>
</tr>
</tbody>
</table>

Cultural Applicability and Cultural Adaptations: There is no information available about cultural applicability. Translations are not available as yet.

Gender Appropriateness: Appropriate for use with both genders.
Age Appropriateness: Not age specific but is used to assess persons of senior age residing in nursing home care facilities.

Summary: The MDS-COG is a scale that forms part of the MDS data set, and now, in Australia part of the interRAI-Long Term Care Facility: Comprehensive Assessment and Care Planning for Long Term Care Facilities. It is administered as part of the data routinely collected by staff on entry into care facilities and takes about 10 to 20 minutes to complete. Psychometric evidence is limited but available data does support the scales reliability and validity.

Reporter: Siggi Zapart

Date of report: 19/1/07

References


Appendix 8: Reviews of Multi-attribute Utility Measures
Appendix 8.1: EQ-5D Index

AHOC INSTRUMENT REVIEW SHEET

Title: EQ-5D index (formerly the EuroQol).

Abbreviations: EQ-5D.

There is some confusion in the literature about the name of the EQ-5D index. There are three types of measurement, all of which are commonly referred to as the EQ-5D:

1. The EQ-5D health status measure. This describes where the 5 items of the EQ-5D are completed and presented as a health profile measure.
2. The EQ-5D VAS which is where the respondent rates their health state on a visual analog scale, where the endpoints are 0.00 (worst imaginable health) and 1.00 (best imaginable health).
3. The EQ-5D index, which is where the responses to the EQ-5D are weighted and combined to form a utility measure for use in economic evaluation.

The EQ-5D index is reviewed here. This review sheet does not review the EQ-5D VAS. It only reviews the EQ-5D health status measure where this is directly related to the EQ-5D index in dementia or Alzheimer’s disease. For a discussion of the differences between the EQ-5D measurement types, the reader is referred to Krabbe & Weijnen (2003).

Author(s) Name: The original EuroQol was developed by the EuroQol Group, although much of the work can be attributed to Professor Alan Williams, Centre for Health Economics, University of York. Membership of the EuroQol Group can be found in The EuroQol Group (1990).

Author(s) Address: Enquiries should be directed to the EQ-5D website (http://www.euroqol.org/).
Dr. Frank de Charro
EuroQol Business Manager
PO BOX 4443
3006 AK Rotterdam
The Netherlands
Tel: +31 10 408 1545
Fax: +31 10 452 5303
E-mail: fdecharro@compuserve.com

Supplied by: A copy of the EQ-5D can be found in Rabin & De Charro (2001).

Cost: The EQ-5D is available to public health researchers free of charge and is in the public domain. Where use is commercial, costs apply. Enquiries should be directed to the EQ-5D website (http://www.euroqol.org/) where additional assistance can be found.

Training requirements: Nil.

Purpose: The EQ-5D was designed to provide a standardized, generic health-related quality of life measure for both describing and valuing health status. It was designed to be used alongside other measures rather than replace them. It was designed to be used in cross-cultural postal surveys.

Administration time: 2-5 minutes.

Instrument Type: Multi-attribute utility instrument.
The EQ-5D comprises 5 dimensions, each with 1 item. The dimensions are: mobility, self-care, usual activities, pain/discomfort and anxiety/depression.

Each item has 3 responses levels, from a normal health state to a very poor health state. It should be noted that the different response levels are simply different levels of health; they do not possess interval or cardinal properties.

The EQ-5D index is computed using an econometric regression model; essentially it is an additive model.

The English-language utility weights are from a British population random sample \((n = 3395\) respondents, response rate 56\%) based on the TTO for 42 marker health states using a 10 year timeframe (Dolan, 1997). Other utility values were regression modelled (MVH Group, 1995; Dolan, 1997; Dolan, et al. 1996). The upper boundary is 1.00, and the lower boundary is −0.59: it permits health state values worse than death. English-language utility weights for the EQ-5D are also available for the US. (Fryback, 2005; Shaw, et al. 2005; Johnson, et al. 1998).

Regarding the generalisability of the UK utility weights, a Swedish study showed similar values when compared with both British and Dutch values using a visual analogy scale (VAS) to rank selected health states (Brooks, et al. 1991; Bjork and Norinder, 1999). A recent Swedish study showed differential weights by age group (Burstrom, et al. 2006). A Spanish study (Badia, et al. 2001b) using time trade-off, however, showed there were significant differences between Spanish/British values on 35% of the EQ-5D health states valued. Similar rankings of values were found for Zimbabweans when compared with the British weights, but there were absolute differences on 32/38 health states valued and it was recommended that Zimbabwean weights should used in Zimbabwean studies (Jelsma, et al. 2003). A study aimed at producing New Zealand weights, based on a population random sample and using a VAS reported that 79% of respondents provided inconsistent rankings of health states (Devlin, et al. 2003).

Polsky, et al. (2001) compared the values of the general public with those of patients, reporting \(R^2 = 0.57\). It was also reported in a Spanish study that there were significant differences in values between those who were ill and the healthy, based on rankings of health states (Badia, et al. 2001b).

The EQ-5D was developed to enable cross-cultural comparisons of health state valuations through providing a standardized, generic health-related quality of life measure for both describing and valuing health status. It was designed to be used alongside other measures (including postal surveys) rather than replace them.

British population norms were published by Kind, et al. (1998), broken down by gender and age group. The overall population norm was 0.86 \((n = 3392; SD = 0.23)\). For males it was 0.86 \((n = 1467; SD = 0.24)\) and for females it was 0.85 \((n = 1925; SD = 0.22)\). The range, depending on population age group, was between 0.73 - 0.86.

Preliminary Australian population values have been published \((n = 343)\), giving a range of between 0.79–0.92 depending upon age group (Hawthorne and Richardson, 2001).

The studies included here are representative of public health studies using the EQ-5D. Because of the number of published studies, no claim is made for comprehensiveness.

Clinical Data:

The studies included here are representative of clinical studies using the EQ-5D. Due to the number of published studies, no claim is made for comprehensiveness.

Applications:

There are three applications for the EQ-5D. As a generic instrument, it is applicable to all public health and clinical conditions and interventions.

---

1. It can be used as a simple HRQoL measure, providing profile scores on all five items. Used like this, the EQ-5D can be used in program evaluation to assess the benefits associated with health interventions.

2. When utilities are computed, the EQ-5D can be used as a HRQoL global index outcome measure in program evaluation to assess the benefits associated with health interventions.

3. When utilities are computed, the EQ-5D can be used in economic evaluation, specifically in cost-utility analysis requiring the computation of quality-adjusted life years (QALYs). These evaluations may directly compare the cost-per-QALY gained between different health interventions for different health conditions, where the cost-utility analysis is intended to provide information assisting with decisions regarding resource allocation.

Carer and/or Patient Use of Instrument: Self-administered. It can also be administered by interviewer and over the telephone. The EQ-5D was designed for self-completion.

Proxy completion is also common, although the agreement between patient and proxy scores on items is moderate (kappa range = 0.18-0.73 (Pickard, et al. 2004) and others have reported poor agreement (Coucill, et al. 2001; Ankri, et al. 2003).

In general, patients obtain higher utility scores than do proxies, although less reliable scores (Dorman, et al. 1998). Coucill, et al. (2001) reported that the mean scores obtained by dementia case and caregiver were 0.80 and 0.58 respectively. Overall there was no more than ‘fair’ agreement between cases and proxies. This result is similar to that of Naglie et al (2006) who reported that among Alzheimer’s disease patients the mean EQ-5D utility was 0.86 compared with 0.62 for proxies – a difference of 0.24 utilities.

In a study of Alzheimer’s disease Jonsson, et al. (2006) reported on the proportions of cases and proxies obtaining each of the three levels for each EQ-5D item. There were substantial disagreements on all items (e.g. for usual activities 76% of cases reported no impairment compared with just 23% of proxies). For all items, proxies’ endorsements were lower than cases.

Bryan, et al. (2005) reported the correlation between caregivers and clinicians as proxies, noting the correlations were generally fair or poor.

Psychometric Criteria

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td>Shaheen and Lindholm (2006)</td>
<td>□ Alpha &gt;0.70</td>
<td>Not reported in dementia studies.</td>
</tr>
<tr>
<td></td>
<td>Riazi, et al. (2006)</td>
<td>X Marginal or inadequate internal consistency (&lt;0.70)</td>
<td>In other studies Cronbach’s α = 0.58 - 0.69.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information found on internal consistency</td>
<td></td>
</tr>
<tr>
<td>Test – retest</td>
<td>Ankri, et al. (2003)</td>
<td>X ICC &gt;.70</td>
<td>Test-retest at 3-days among patients suffering dementia was reported within the range kappa = 0.34 – 0.59; based on the overall (unweighted) score</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Time intervals and confidence intervals reported</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>X Marginal or</td>
<td></td>
</tr>
</tbody>
</table>

Final Report: Dementia Outcomes Measurement Suite Project
the same questionnaire when no change in physical functioning has occurred

Calculation of an intraclass correlation coefficient (ICC); and an ICC > 0.70 is desired

Preferred if time interval and confidence intervals were presented

<table>
<thead>
<tr>
<th>Study Reference</th>
<th>Test-Retest Reliability</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naglie, et al. (2006)</td>
<td>inadequate test-retest reliability ICC&lt;.70</td>
<td>No information found on test-retest reliability</td>
</tr>
<tr>
<td>van Agt, et al. (1994)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marra, et al. (2005a)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brazier, et al. (1999)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stavem (1999)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stavem, et al. (2005)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fransen and Edmonds (1999)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dorman, et al. (1998)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schweikert, et al. (2006)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

the ICC was 0.74. The mean difference in scores between test-retest was 0.11 utilities. In a study of those with Alzheimer’s disease, at 2-week test-retest, the reliability coefficient for the EQ-5D was 0.79, although this varied by MMSE level. For those with MMSE 19-26 the ICC was 0.70 compared with 0.83 for those with MMSE 10-18.

Reliability of the EQ-5D items at 10-month test-retest was assessed using G-study, which showed that 82% of the variance was explained by respondents’ health state; when expressed as a test-retest reliability correlation coefficient this was reported as 0.36 at the group level and 0.90 at the individual level.

Marra, et al. reported test-retest at 5-weeks for those with rheumatoid arthritis where the ICC = 0.45, which was deemed unacceptable.

Brazier cites two earlier studies of the EQ-5D involving older women and those with chronic obstructive pulmonary disease. Test-retest at 6-months among those who stated their health had not changed was 0.67 and 0.83.

Stavem (1999) reported Spearman correlations at 2-week test-retest of $r_s = 0.73$ for those with chronic obstructive pulmonary disease. In a study of patients with HIV/AIDS Stavem et al (2005) reported the test-retest to be ICC = 0.78.

The ICC was reported to be 0.70 for those with osteoarthritis.

Using test-retest among stroke patients, Dorman et al reported Kappa = 0.63 to 0.80 for EQ-5D items.

Schweikert, et al. reported a test-retest Kappa of 0.54.

**Inter – rater**

Limits of agreement, Kappa, or standard error of measurement (SEM) were presented

- □ Agreement reported and adequate
- □ Inadequate inter-rater agreement
- X No information provided
<table>
<thead>
<tr>
<th>VALIDITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td></td>
<td>□ Patients and experts were involved during item selection and/or item reduction</td>
<td>Patients or community members were not involved in the construction of the EQ-5D descriptive system or items. This construction work was carried out by the EuroQol Group of 22 experts.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Patients were consulted for reading and comprehension X No patient involvement</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information found on content validity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ There is an adequate coverage of relevant domains X There is limited coverage of relevant domains</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EuroQol Group (1990)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Brooks (1996)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Buckingham (1995)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Krabbe, et al. (1999)</td>
<td></td>
<td>Buckingham reported that it only covers 39% of the concepts regarded by the public as salient to health.</td>
</tr>
<tr>
<td></td>
<td>Hawthorne and Richardson (2001)</td>
<td></td>
<td>Krabbe, et al. argued that the absence of a cognitive function attribute was a major limitation of the EQ-5D.</td>
</tr>
<tr>
<td></td>
<td>Hawthorne, et al. (2000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burstrom, et al. (2006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Essink-Bot, et al. (1997)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vitale, et al. (2001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Construct</td>
<td></td>
<td>□ Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used X No information provided</td>
<td>Not reported in the dementia literature.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Limited/inadequate construct validity reported</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No evidence provided/failed a test of dimensionality</td>
<td>Not reported in the dementia literature.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Some evidence provided to support internal structure</td>
<td>Essink-Bot, et al. used exploratory factor analysis (EFA) with varimax rotation in a Dutch sample comprising</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Substantial</td>
<td></td>
</tr>
<tr>
<td>Construct: Internal Structure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information provided on factor structure</td>
<td>Essink-Bot, et al. (1997)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Construct: Correlation with other measures

Comparisons made to other measures

<table>
<thead>
<tr>
<th>Study</th>
<th>Correlations with other measures are reported</th>
<th>Correlations not reported</th>
<th>Not reported in the dementia literature.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vitale, et al. (2001)</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hawthorne, et al. (2001a)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vitale, et al. (2001b)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holland, et al. (2001b)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bosch and Hunink (2000)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lubetkin and Gold (2003)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luo, et al. (2003a)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luo, et al. (2005)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hawthorne, et al. based on a stratified sample of community, outpatient and inpatient participants, reported that the EQ-5D correlated with the AQoL ($r = 0.73$), HUI3 ($0.64$), the 15D ($0.76$) and the SF6D ($0.75$). It should be noted, however, that correlations between measures are not good tests of agreement on actual scores as shown in the Hawthorne, et al. (2004) and Holland, et al. (2001b) studies.

In a study of treatment for claudication, it was reported that at baseline the intraclass reliability coefficient between the EQ-5D and HUI3 was 0.49, rising to 0.68 at 12-month follow-up.

These coefficients were within the range reported for the EQ-5D and HUI3 for patients at health centres ($r = 0.69$) and for those with rheumatic disease ($r = 0.45$) and in a
<table>
<thead>
<tr>
<th>Study</th>
<th>Correlation Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schulz, et al. (2002)</td>
<td>However, a study of prostate cancer reported a correlation of $r = 0.33$. Overall, it would seem that the correlation with other quality of life measures may be a function of disease condition and severity.</td>
</tr>
<tr>
<td>Brazier, et al. (2004)</td>
<td>Regarding correlation with the SF6D, Brazier, et al. reported that on the instrument similar dimensions the correlations ranged from 0.45 to 0.60. The correlations between the two utility scores were between 0.28 and 0.55 for different patient groups.</td>
</tr>
<tr>
<td>van Stel and Buskens (2006)</td>
<td>Elsewhere the correlation between the EQ-5D and SF6D was ICC = 0.45.</td>
</tr>
<tr>
<td>Franks, et al. (2006a)</td>
<td>In another study of health risk factors in a US population sample it was reported that the correlation between the EQ-5D and SF6D was 0.93. As noted by Bryan and Longworth these discrepancies are a function of different valuations for similar health states.</td>
</tr>
<tr>
<td>Riazi, et al. (2006)</td>
<td>Correlations with SF-36 scales have been reported in the range of 0.13 to 0.70. In general, higher correlations have been reported with the SF36 physical scales than the mental health scales.</td>
</tr>
<tr>
<td>Vitale, et al. (2001)</td>
<td></td>
</tr>
</tbody>
</table>

### Construct: Discriminant Validity

The scale differentiates between relevant categories of respondent e.g. sick vs. well, varying degrees of severity

<table>
<thead>
<tr>
<th>Study</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ankri, et al. (2003)</td>
<td>X Scale differentiates between relevant categories of respondents</td>
</tr>
<tr>
<td></td>
<td>□ No information on discriminant validity</td>
</tr>
<tr>
<td>Ankri, et al. (2003)</td>
<td>In a study of dementia where the EQ-5D was administered by interviewer, the EQ-5D items of mobility, self-care and usual activities systematically varied by activities of daily living, whereas there was no association with the anxiety item.</td>
</tr>
<tr>
<td>Andersen, et al. (2004)</td>
<td>When examined by MMSE score, the only EQ-5D item that systematically varied was the anxiety/depression item.</td>
</tr>
</tbody>
</table>

More encouraging results were reported by Andersen, et al. in a study of dementia. Based on MMSE classifications into mild (MMSE >20), moderate (10-19) and severe (<10) dementia they reported that EQ-5D scores systematically varied, being 0.64, 0.60 and 0.49, respectively although these were not statistically
Serrano-Aguilar, et al. (2006) compared the EQ-5D item scores from those with Alzheimer’s disease with those from the general community, reporting greater problems on all items.

In the recent literature the EQ-5D has been shown to discriminate between known groups. This has been reported in (representative studies only): arthritis, cancer survival, cardiac rehabilitation, caregiving, claudication, migraine, neuralgia, pain, sleep disturbance, social class, and stroke.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Interpretability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information on the relationship of scores to gold standard measures or clinical diagnosis is provided</td>
<td>The proportion of missing data in dementia studies is about 20% of cases.</td>
</tr>
</tbody>
</table>

There are no gold standard measures for utility instruments.

- □ Comparison made to criterion measures
- X No comparison with criterion measures provided

- □ Authors provide 2 or more types of information on interpretability
- Authors provide limited information to assist with interpretability
- □ No information provided

The proportion of missing data in dementia studies is about 20% of cases.

In another study of Alzheimer’s disease-sufferers, the missing data rates, as defined by missing 2+ items from the EQ-5D were 0.5% for those with mild dementia and 10% for those with moderate dementia. The proportions varied by MMSE level, with EQ-5D scores available for 84% of participants with MMSE 19-26, and 77% for those with MMSE 10-18.

In other health areas the proportion of cases with missing data has varied from 1% to 20%.

Regarding interpretability assessed against an external criterion, correlations of 0.28 with an ergometer have been reported.
The correlation with the Barthel Index has been reported at $r = 0.70$. 

<table>
<thead>
<tr>
<th>RESPONSEIVENESS</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Floor and ceiling effects</strong></td>
<td>The questionnaire fails to demonstrate a worse score in patients who clinically deteriorated and an improved score in patients who clinically improved</td>
<td>Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
<td>Ceiling effects in quality of life measures are normal where the samples are drawn from the population (this is because in a population it is expected that most people are well, health and enjoy a good quality of life). Vitale, et al. argued that under these conditions any multi-dimensional instrument with ≥50% of respondents obtaining a ceiling score would be invalid.</td>
</tr>
<tr>
<td></td>
<td>Authors should provide descriptive statistics of the distribution of scores</td>
<td>No information provided on floor and ceiling effects</td>
<td>In clinical samples, however, ceiling effects should not occur unless it can be shown that treatment has returned patients to a 'normal' quality of life.</td>
</tr>
<tr>
<td></td>
<td>Vitale, et al. (2001)</td>
<td>□ Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
<td>Ceiling effects were not reported in Ankri, et al.’s dementia study, where 15% of respondents obtained perfect scores.</td>
</tr>
<tr>
<td></td>
<td>Ankri, et al. (2003)</td>
<td>X Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score</td>
<td>In contrast, 31/64 dementia cases (23 with mild dementia and 7 with moderate dementia) obtained ceiling scores, yet only 2 of these cases were classified by their clinicians as having no problems on all of the EQ-5D items.</td>
</tr>
<tr>
<td></td>
<td>Coucill, et al. (2001)</td>
<td>Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
<td>Similarly, Naglie, et al. in their study of Alzheimer’s patients reported that 43% of respondents at the first interview and 57% at a second interview obtained ceiling scores on the EQ-5D, yet only for proxy ratings these data were 7% and 2%, respectively.</td>
</tr>
<tr>
<td></td>
<td>Naglie, et al. (2006)</td>
<td>No information provided on floor and ceiling effects</td>
<td>Elsewhere ceiling effects have been widely reported, and have ranged from 14% to 50%. One of the key reasons for this situation is that very few respondents will endorse the lowest level of the EQ-5D response scales (typically in the range 1-4%). Although these scales have just 3-levels, the lowest levels report</td>
</tr>
<tr>
<td></td>
<td>Brazilian, et al. (2004)</td>
<td>□ Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
<td>Ceiling effects were not reported in Ankri, et al.’s dementia study, where 15% of respondents obtained perfect scores.</td>
</tr>
<tr>
<td></td>
<td>Brazilian, et al. (1993)</td>
<td>X Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score</td>
<td>In contrast, 31/64 dementia cases (23 with mild dementia and 7 with moderate dementia) obtained ceiling scores, yet only 2 of these cases were classified by their clinicians as having no problems on all of the EQ-5D items.</td>
</tr>
<tr>
<td></td>
<td>Hawthorne, et al. (2001b)</td>
<td>Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
<td>Similarly, Naglie, et al. in their study of Alzheimer’s patients reported that 43% of respondents at the first interview and 57% at a second interview obtained ceiling scores on the EQ-5D, yet only for proxy ratings these data were 7% and 2%, respectively.</td>
</tr>
<tr>
<td></td>
<td>Bharmaal and Thomas (2006)</td>
<td>No information provided on floor and ceiling effects</td>
<td>Elsewhere ceiling effects have been widely reported, and have ranged from 14% to 50%. One of the key reasons for this situation is that very few respondents will endorse the lowest level of the EQ-5D response scales (typically in the range 1-4%). Although these scales have just 3-levels, the lowest levels report</td>
</tr>
<tr>
<td></td>
<td>Macran, et al. (2003)</td>
<td>□ Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
<td>Ceiling effects were not reported in Ankri, et al.’s dementia study, where 15% of respondents obtained perfect scores.</td>
</tr>
<tr>
<td></td>
<td>Rutter et al. (2006)</td>
<td>X Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score</td>
<td>In contrast, 31/64 dementia cases (23 with mild dementia and 7 with moderate dementia) obtained ceiling scores, yet only 2 of these cases were classified by their clinicians as having no problems on all of the EQ-5D items.</td>
</tr>
</tbody>
</table>
van Stel and Buskens (2006)  
Schweikert, et al. (2006)  
Vitale, et al. (2001)  
Scalone, et al. (2006)  
Luo, et al. (2005)  
van der Zanden, et al. (2006)

extremely severe health states (e.g. confined to bed). For example, van der Zanden et al dichotomized the response categories because of this sparse data distribution. The implication is that the EQ-5D will be insensitive to health states close to full health.

### Sensitivity to change

The ability to detect important change over time in the concept being measured

<table>
<thead>
<tr>
<th>Sensitivity to change</th>
<th>Hypotheses were formulated and results were in agreement</th>
<th>No information on sensitivity to change was provided</th>
<th>Information on sensitivity to change was provided about the magnitude of score differences which would be clinically meaningful</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lammers, et al. (2006)</td>
<td>An adequate metric was used (ES, SRM, comparison with external standard)</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Li, et al. (2006)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Marra, et al. (2005a)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Hayhurst, et al. (2006)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Mehta, et al. (2006)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Bosch and Hunink (2000)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>van Stel and Buskens (2006)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Lenzen, et al. (2006)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Langfitt, et al. (2006)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>van der Roer, et al. (2006)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>McDermott, et al. (2006)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ryan, et al. (2006)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Michaels, et al. (2006)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Langfitt, et al. (2006)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Marra, et al. (2005b)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Keating, et al. (2006)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Holst, et al. (2006)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

Not reported in the dementia literature.

Elsewhere the recent literature has shown the EQ-5D to be sensitive to change over time in areas such as (representative studies only): anxiety disorders, arthritis, bipolar disorder, Claudication, coronary heart disease, epilepsy, pain, stroke, and varicose vein surgery.

It has been estimated that the minimum important difference detectable by the EQ-5D is 0.05 utilities.

In other studies it has not been responsive over time.

### Cultural Applicability and Cultural Adaptations:

The EQ-5D is available in the following languages: Afrikaans, Armenian, Basque, Bulgarian, Catalan, Chinese, Croatian, Czech, Danish, Dutch, English, Estonian, Filipino, Finnish, French, German, Greek, Hebrew, Hungarian, Icelandic, Indonesian, Italian, Japanese, Korean, Latvian, Lithuanian, Malay, Norwegian, Polish, Portuguese, Romanian, Russian, Sesotho, Shona, Slovenian, Spanish, Swedish, Thai, Tongan, Turkish, Xhosa, Zulu.

The EQ-5D was developed cross-culturally across the UK, the Netherlands, Norway, Finland and Sweden (Anderson, et al. 1993). It was developed for use in a battery of other measures and for use in postal surveys hence its shortness (Brazier, et al. 1993).

### Gender Appropriateness:

It appears to be appropriate for use with both genders. No information has been published suggesting a gender bias.

### Age Appropriateness:

It appears to be appropriate for use with adults including elderly persons. Ankri, et al. (2003) reported that among those with dementia, except for pain, all items on the EQ-5D deteriorated with age.
Summary: The literature on the measurement properties of the EQ-5D in dementia is extremely sparse and far from reassuring. Generally, it suggests that self-completion is preferred and that there are very large differences between case and proxy reports. There are issues around ceiling effects and some indication of missing data. More generally, although the EQ-5D has been shown to be sensitive in many conditions and is prized for its brevity, there are issues around the distribution of scores, the weighting procedure and the insensitivity of it to good health states.

Reporter
Graeme Hawthorne

Date of Report
January 2007

References


Bergstrom KG, Arambula K and Kimball AB (2003) Medication formulation affects quality of life: a randomized single-blind study of clobetasol propionate foam 0.05% compared with a combined program of clobetasol cream 0.05% and solution 0.05% for the treatment of psoriasis. *Cutis*. Vol. 72, No.5, pp.407-11.


hypopituitary patients on maintenance growth hormone replacement. J Clin Endocrinol Metab. Vol. 91, pp. 3773-3779


Hart HE, Redekop WK, Berg M, Bilo HJ and Meyboom-de Jong B (2005a) Factors that predicted change in health-related quality of life were identified in a cohort of diabetes mellitus type 1 patients. Journal of Clinical Epidemiology. Vol. 58, No.11, pp.1158-64.


Hawthorne G, Richardson J and Day N (2000b) A comparison of the Assessment of Quality of Life (AQoL) with four other generic utility instruments. XII Medical Symposium "Quality of Life Measurement in Clinical Studies", Helsinki, Finland, pp.358-3760.


Appendix 8.2: Assessment of Quality of Life

AHOC INSTRUMENT REVIEW SHEET

Title: Assessment of Quality of Life.

Abbreviations: AQoL.

Author(s) Name: Hawthorne G & Richardson J.

Author(s) Address: Hawthorne G, Department of Psychiatry, The University of Melbourne, Level 1 North, Royal Melbourne Hospital, Grattan Street, Parkville, Victoria, Australia, 3050. Telephone: + 61 3 8344 5467 Email: graemeeh@unimelb.edu.au

Richardson J, Centre for Health Economics, Building 75, Monash University, Clayton, Victoria, Australia, 3800 Telephone: +61 3 9905 0733 Email: jeff.richardson@buseco.monash.edu.au


Cost: Free. Registration is required. To register contact A/Professor Graeme Hawthorne (graemeeh@unimelb.edu.au).

The AQoL is copyright and may not be used or reproduced without permission. Permission to use the AQoL can be obtained from A/Professor Graeme Hawthorne.

Website: http://www.acpmh.unimelb.edu.au/whogol_agol.html

Training requirements: Nil.

Purpose: The AQoL is a generic health-related quality of life instrument which provides a profile of four dimensions of life, or it can be used as a weighted multi-attribute instrument in economic evaluation, specifically cost-utility studies where the calculation of quality adjusted life years (QALYs) is needed.

Administration time: 5-10 minutes.

Instrument Type: Multi-attribute utility instrument.

Structure: The AQoL comprises 5 dimensions, each with 3 items. The dimensions and items are: Illness (medicine use, medical aids, and medical treatment), Independent Living (self-care, daily activities, mobility), Social Relationships (intimacy, friends, family role), Physical Senses (vision, hearing, communication) and Psychological Wellbeing (sleep, anxiety, pain). Each item has 4 responses levels, from a normal health state to a very poor health state. It should be noted that the different response levels are simply different levels of health; they do not possess interval or cardinal properties.

Scoring: There are two ways the AQoL can be scored.

A. It can be scored by simply adding up the responses of respondents. This summative model of scoring will provide profile scores for each dimension,
and an overall AQoL score. There are no published papers where this scoring system has been reported.

B. It can be scored as a utility instrument providing an index that represents a person’s quality of life. When scored like this, only the last four dimensions of the AQoL are used in obtaining the utility scores. Endorsed responses are substituted with values obtained from a stratified sample of the general Australian community using time trade-off (TTO; n = 225). Within each dimension, the three items are combined using a multiplicative model. Dimension scores are then combined into a single index, again using a multiplicative model, and the resulting score is then transformed onto a life-death scale. The upper boundary, 1.00, represents the best HRQoL state, 0.00 represents HRQoL states that are death equivalent, and the lower boundary, -0.04, represents HRQoL states worse than death. HRQoL states worse than death are necessary to explain people who have either committed suicide or euthanasia.

Developed for:
The AQoL was designed for the economic evaluation of health interventions, specifically cost-utility analysis. The descriptive system was developed by the researchers based on the World Health Organization’s classification of disease, impairment, disability and handicap. A review of HRQoL instruments published since 1970 suggested the dimensions, and actual items were written in focus groups with clinicians. The pilot involved inpatients and a random sample of the general community. The utility weights used in scoring the AQoL were obtained from a representative sample of the Australian population, using the time trade-off (Hawthorne, et al. 1999, Richardson and Hawthorne, 1998).

Normative Data:
Normative data are reported in Hawthorne and Osborne (2005). The general population norm for the AQoL utility, based on the 1998 South Australian Health Omnibus Survey (n = 3000 adults) is 0.83 (SD = 0.20). A smaller stratified sample of the Victorian population (n = 334) provided estimates from 0.71–0.84, depending upon age group (Hawthorne, et al. 2001).

Public Health Data:

Clinical Data:

Applications:
There are three applications for the AQoL. As a generic instrument, it is applicable to all public health and clinical conditions and interventions.
1. It can be used as a simple additive HRQoL measure, providing profile scores on independent living, social relationships, physical senses, and psychological wellbeing. Used like this, the AQoL can be used in program evaluation to assess the benefits associated with health interventions.
2. When utilities are computed, the AQoL can used as a HRQoL global index outcome measure in program evaluation to assess the benefits associated with health interventions.
3. When utilities are computed, the AQoL can used in economic evaluation,
specifically in cost-utility analysis requiring the computation of quality-adjusted life years (QALYs). These evaluations may directly compare the cost-per-QALY gained between different health interventions for different health conditions, where the cost-utility analysis is intended to provide information assisting with decisions regarding resource allocation.

**Carer and/or Patient Use of Instrument:** Self-administered. The AQOL can also be administered by interviewer and over the telephone. Hawthorne showed that there was no significant effect by administration mode (mail versus telephone) (Hawthorne, 2003). Both the Wlodarczyk et al and Sturm et al studies reported on self-completion, although Sturm et al did report that a proportion of cases were assisted (Wlodarczyk, et al. 2004, Sturm, et al. 2004).

Wlodarczyk et al (2004) reported that the correlation between case and proxy AQoL utility scores was 0.37, although proxy scores were significantly lower than case scores; for the individual dimensions the correlations were 0.17 for physical senses, 0.20 for psychological wellbeing, 0.26 for social relationships and 0.29 for independent living.

Herrman et al reported a Pearson $r = 0.55$ of proxy (case manager) and self-completion by those suffering long–term psychosis; and suggested that self-completion was to be preferred (Herrman, et al. 2002).

**Psychometric Criteria**

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td>Richardson and Hawthorne (1998) Osborne, et al. (2003) Osborne, et al. (2000) Hawthorne, et al. (2004) Hogan, et al. (2001) Manser, et al. (2006)</td>
<td>X Alpha &gt;0.70 □ Marginal or inadequate internal consistency (&lt;0.70) □ No information found on internal consistency</td>
<td>Based on community and hospital samples, the range of internal consistency estimates for the AQoL descriptive system is Cronbach $\alpha = 0.73 – 0.84$.</td>
</tr>
<tr>
<td>Test – retest</td>
<td>Hawthorne (2003)</td>
<td>X ICC &gt;.70 Time intervals and confidence intervals reported □ Marginal or inadequate test-retest reliability ICC &lt;.70 □ No information found on test-retest reliability</td>
<td>Based on a random community sample, two-week test-retest reliability was ICC = 0.83.</td>
</tr>
<tr>
<td>Inter – rater</td>
<td>Studies Reported &amp; References</td>
<td>Adequacy Checks</td>
<td>Comment</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------------------------</td>
<td>----------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Limits of agreement, Kappa, or standard error of measurement (SEM) were presented | □ Agreement reported and adequate  
□ Inadequate inter-rater agreement  
X No information provided | No studies found. | |

<table>
<thead>
<tr>
<th>VALIDITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
</table>
| Content             | Hawthorne and Richardson (2001)  
Hawthorne, et al. (2001b)  
Hawthorne, et al. (1999) | □ Patients and experts were involved during item selection and/or item reduction  
□ Patients were consulted for reading and comprehension  
□ No patient involvement  
X No information found on content validity  
X There is an adequate coverage of relevant domains  
□ There is limited coverage of relevant domains | Coverage of the HRQoL construct was assessed as being adequate by comparison with the content of other HRQoL measures. |
| Construct           | Hawthorne, et al. (2001b) | X Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used  
□ Limited/Inadequate construct validity reported  
□ No information provided | Spearman correlations with other multi-attribute utility instruments have been reported. For the EQ-5D $r = 0.73$; the HUI3 $r = 0.74$; the 15D $r = 0.80$; for the SF6D $r = 0.74$. |
Osborne, et al. (2003)  
Hawthorne, et al. (2001a) | □ No evidence provided/failed a test of dimensionality  
X Some evidence provided to support internal structure  
□ Substantial evidence provided to support internal structure | The underlying model of the AQoL and its dimensions was developed through exploratory factor analysis, and subsequently confirmed in a community sample of older adults with chronic health conditions.  
The internal structure was subsequently confirmed using structural equation modelling (CFI $= 0.91$). |
| Construct: Correlation with other measures | Sturm, et al. (2002)  
Osborne, et al. (2003)  
Watson, et al. (2005) | X Correlations with other measures are reported  
□ Correlations not reported | Spearman correlations with health status in stroke patients have been reported. For the SF36 Physical Component Score $r = 0.37 - 0.81$ and $r = 0.36 - 0.43$ for the mental health. |
| Construct: Discriminant Validity | Ackerman, et al. (2006) | Spearman correlations with the WHOQOL-BREF domains has been reported at \( r = 0.67 \) (Physical), \( r = 0.71 \) (Psychological), \( r = 0.33 \) (Social) and \( r = 0.44 \) (Environment). |
| X Scale differentiates between relevant categories of respondents | Sturm, et al. (2004). | In a study of post-stroke at 2-year follow-up, mean AQoL scores for those with dementia and non-dementia were 0.12 and 0.50, respectively. |
| Wlodarczyk, et al. (2004) | Hawthorne and Richardson (2001) | Wlodarczyk et al reported that the AQoL systematically varied by MMSE status; for those with severe Alzheimer’s disease (MMSE = 10) the mean utility was 0.52, and 0.71 for those with mild Alzheimer’s (MMSE 200-24). There was a monotonic relationship for intermediate states. |
| Goldney and Bain (2006) | Whitfield, et al. (2006) | Hawthorne & Richardson examined five MAU-instruments, and based on pooled data by utility quartile reported that the AQoL was the most sensitive of the five instruments. |
| Chua, et al. (2006) | Herrman, et al. (2002) | Sensitivity to group status has been reported for age groups, community/outpatient/inpatient status, depression and double-depression; New York Heart Association classification; osteoarthritis; psychosis; 6-minute walk test, social isolation. |

### Interpretability

The degree to which one can assign qualitative meaning to quantitative scores.

- Do authors provide the following:
  - Presentation of means and SD of scores before and after treatment.
  - Comparative data on the distribution of scores in relevant subgroups.
  - Information on the relationship of scores to well-known functional measures or clinical diagnosis.
  - Information on the association between changes in scores and patients' global ratings of the magnitude of change they have experienced.

<table>
<thead>
<tr>
<th>Authors provide 2 or more types of information on interpretability</th>
<th>Authors provide limited information to assist with interpretability</th>
<th>No information provided</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hawthorne, et al. (2006)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Although no studies reporting interpretation of the AQoL have been published, unpublished re-analysis of data from Hawthorne et al's WHOQOL-Bref norms paper, based on self-rated quality of life of random population samples (n = 849) suggest the following modelled interpretations may be considered:

- 0.90 – 1.00 = Excellent
- 0.77 – 0.89 = Very good
- 0.64 – 0.76 = Good
- 0.45 – 0.63 = Fair
- 0.21 – 0.44 = Poor
- -0.04 – 0.20 = Very poor

No cutpoints against clinical criteria have been published.

### RESPONSIVENESS

<table>
<thead>
<tr>
<th>Floor and ceiling effects</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>The questionnaire fails to demonstrate a worse score in patients who clinically deteriorated and an improved score in patients who clinically improved.</td>
<td>Hawthorne and Osborne (2005)</td>
<td>X Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected.</td>
<td>Population samples, where people are generally healthy and enjoy a good quality of life show that ~45% of respondents will obtain scores in the top 10% of the utility scale. This, it could be argued, is to be expected as the ceiling represents good to excellent HRQoL.</td>
</tr>
<tr>
<td>Authors should provide descriptive statistics of the distribution of scores.</td>
<td>Ackerman, et al. (2006)</td>
<td>□ Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score. □ No information provided on floor and ceiling effects</td>
<td>In a study of hip and knee replacement, 2% of participants obtained floor scores on the AQoL.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sensitivity to change</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>The ability to detect important change over time in the concept being measured</td>
<td>Sturm, et al. (2004)</td>
<td>□ Hypotheses were formulated and results were in agreement. X An adequate metric was used (ES, SRM, comparison with external)</td>
<td>Baseline dementia state predicted AQoL score at 2-year follow-up among stroke victims; for those with baseline dementia the AQoL utility score was 0.02 compared to non-dementia cases of 0.27.</td>
</tr>
<tr>
<td>Reference</td>
<td>Sensitivity to change over time</td>
<td>MCID - Information was provided</td>
<td>MCID - No information was provided</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------------------</td>
<td>---------------------------------</td>
<td>-----------------------------------</td>
</tr>
</tbody>
</table>

**Cultural Applicability and Cultural Adaptations:** The AQoL has been translated into Canadian French and Danish. No particular difficulties were reported. However, there have been no tests of this in the reported literature.

**Gender Appropriateness:** The AQoL is appropriate for use with both genders. No reports have been published suggesting that there are significant gender issues.

**Age Appropriateness:** The AQoL has been used in studies involving older adults, and no particular difficulties have been reported (Harris, et al. 2001; Lim, et al. 2003; Osborne, et al. 2003; Holland, et al. 2004; Hawthorne, 2006; Wlodarczyk, et al. 2004).

**Summary:** Although the AQoL has been reported in a small literature (the number of published reports was 39 at the time of writing), only two papers were identified reporting its use among those with dementia or Alzheimer’s disease. Neither paper explicitly examined its measurement properties.

In general, the available evidence suggests that the AQoL is a stable measure that has a nomological net of evidence supporting its reliability and validity. It is the only multi-attribute instrument weighted with Australian values.

**Reporter** A/Prof Graeme Hawthorne

**Date of Report** 31/01/2007

**References**


Appendix 9: Reviews of Social Isolation Measures
Appendix 9.1: Loneliness Scale

AHOC INSTRUMENT REVIEW SHEET

Title: Loneliness Scale (usually referred to as the De Jong Gierveld Loneliness Scale to distinguish it from other scales with the title ‘Loneliness Scale’).

Abbreviations: N/A.

Author(s) Name: J. De Jong Gierveld and F. Kamphuis.

Author(s) Address: Professor Dr Jenny De Jong Gierveld Netherlands Institute for Advanced Study in the Humanities and Social Sciences & the Faculty of Social Sciences, Department of Sociology The Free University De Boelelaan 1081, NL-1081 HV Amsterdam, The Netherlands Email: gierveld@nidi.nl


Cost: No copyright restrictions were identified in the published literature or on the internet. Requests for permission to use should be addressed to directly to Professor Jenny De Jong Gierveld and/or Applied Psychological Measurement.

No fees for use of the scale were identified in the published literature or on the internet.

Training requirements: None. The manual for the full 11-item version can be downloaded from: http://fome.fsw.vu.nl/tg.van.tilburg/manual_loneliness_scale_1999.html


Purpose: The De Jong Gierveld Loneliness Scale was developed in response to the need for a short, valid measure of loneliness.

Administration time: 5-10 minutes.

Instrument Type: Self-administered psychometric instrument. Can also be administered by interviewer and over the telephone. The manual suggests that there are differences by administration mode.

Structure: There are two sub-scales measuring emotional loneliness (all negative items; 6 items) and social loneliness (all positive items, 5 items). The response categories are: yes/ yes/ more or less/ no/ no!.

For the 6 item version, there are 3 items on each of the sub-scales.
Scoring: Scoring the De Jong Gierveld Loneliness Scale is recommended through reversing positive items, dichotomizing the item responses (yes!/ yes/ more or less/ no/ no!) using the category ‘no’ on the ground that the interrogation point (the point of indifference; ‘more or less’) was not neutral – essentially it was considered a positive response. Thus yes!, yes, more or less/ no, no!, and then summing the resultant values for each item.

Where the scale is telephone administered or used with older adults the recommended response scale is yes/ more or less/ no, with dichotomization giving yes, more or less/ no.

Developed for: The De Jong Gierveld Loneliness Scale was developed in response to the need for a short, valid measure of loneliness.

Normative Data: No normative data have been reported. Limited data are reported in the user manual for those aged over 54 years (De Jong Gierveld and Tilburg, 1999).

Clinical Data: N/A. However, the De Jong Gierveld Loneliness Scale has been used in studies of employment (De Jong Gierveld and Van Tilburg, 1987), the living arrangements of older adults (De Jong Gierveld and Tilburg, 2006; De Jong Gierveld and van Tilburg 1999), relationships (Dykstra and De Jong Gierveld, 2004), aging (Dykstra, et al. 2005; van Tilburg, et al. 2004), and the Dutch general population (De Jong Gierveld and Tilburg, 2006).

Applications: The De Jong Gierveld Loneliness Scale is a psychometric measure designed to assess loneliness.

Carer and/or Patient Use of Instrument: The De Jong Gierveld Loneliness Scale was designed for self-completion. No information was available on proxy completion.

### Psychometric Criteria

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td>De Jong Gierveld and Tilburg (1999) Dykstra, et al. (2005) Moorer and Suurmeijer, (1993) van Baarsen, et al. (1999)</td>
<td>X Alpha &gt;0.70 □ Marginal or inadequate internal consistency (&lt;0.70) □ No information found on internal consistency</td>
<td>Across different samples the reliability (whether Cronbach’s α or Mokken’s ρ) falls within the range 0.71 to 0.90.</td>
</tr>
<tr>
<td>Test – retest</td>
<td>□ ICC &gt;.70 Time intervals and confidence intervals reported □ Marginal or inadequate test-retest reliability ICC&lt;.70 X No information found on test-retest reliability</td>
<td>N/A.</td>
<td></td>
</tr>
</tbody>
</table>
correlation coefficient (ICC); and an ICC > 0.70 is desired
Preferred if time interval and confidence intervals were presented

<table>
<thead>
<tr>
<th>Inter – rater</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limits of agreement, Kappa, or standard error of measurement (SEM) were presented</td>
</tr>
<tr>
<td>□ Agreement reported and adequate</td>
</tr>
<tr>
<td>□ Inadequate inter-rater agreement</td>
</tr>
<tr>
<td>X No information provided</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VALIDITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Studies Reported &amp; References</td>
</tr>
<tr>
<td>Adequacy Checks</td>
</tr>
<tr>
<td>Comment</td>
</tr>
<tr>
<td>Content</td>
</tr>
<tr>
<td>The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire</td>
</tr>
<tr>
<td>De Jong Gierveld and Kamphuis (1985)</td>
</tr>
<tr>
<td>□ Patients and experts were involved during item selection and/or item reduction</td>
</tr>
<tr>
<td>□ Patients were consulted for reading and comprehension</td>
</tr>
<tr>
<td>X No patient involvement</td>
</tr>
<tr>
<td>□ No information found on content validity</td>
</tr>
<tr>
<td>X There is an adequate coverage of relevant domains</td>
</tr>
<tr>
<td>□ There is limited coverage of relevant domains</td>
</tr>
<tr>
<td>De Jong Gierveld and Tilburg (2006)</td>
</tr>
<tr>
<td>The content validity of the De Jong Gierveld Loneliness Scale appears to be excellent. There is evidence that the items were based on a sound theoretical model of social isolation, based on the experiences of the isolated. Importantly, the construction and validation samples were population-based samples from the Dutch community, stratified by loneliness level. This suggests that the content of the scale is probably reflective of the concerns of the lonely and socially isolated. That the items were empirically drawn from a larger pool of items using modern test theory (Rasch modelling) along with logical criteria to ensure fidelity and coverage of the theoretical model is strong evidence of content validity. The evidence for the 6 item version is probably just as strong since the construction sample was based on a stratified sample of older adults from three regions in the Netherlands (n = 3,987, response rate 62%). Care was taken during construction to maintain fidelity to the structure of the original scale through selection of items that met both logical and psychometric criteria.</td>
</tr>
<tr>
<td>Construct</td>
</tr>
<tr>
<td>The extent to which scores on the questionnaire relate to</td>
</tr>
<tr>
<td>□ Results were acceptable in accordance with the hypotheses and an</td>
</tr>
</tbody>
</table>
other measures in a manner that is consistent with theoretically derived hypothesis concerning the domains that are measured | adequate comparison measure was used □ Limited / inadequate construct validity reported X No information provided

<table>
<thead>
<tr>
<th>Construct: Internal Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information provided on factor structure</td>
</tr>
<tr>
<td>De Jong Gierveld and Kamphuis (1985)</td>
</tr>
<tr>
<td>De Jong Gierveld and Tilburg (1999)</td>
</tr>
<tr>
<td>De Jong Gierveld and Tilburg (2006)</td>
</tr>
<tr>
<td>De Jong Gierveld and van Tilburg (1999)</td>
</tr>
<tr>
<td>Dykstra and De Jong Gierveld (2004)</td>
</tr>
<tr>
<td>van Baarsen, et al. (2001)</td>
</tr>
<tr>
<td>□ No evidence provided/failed a test of dimensionality</td>
</tr>
<tr>
<td>X Some evidence provided to support internal structure</td>
</tr>
<tr>
<td>□ Substantial evidence provided to support internal structure</td>
</tr>
</tbody>
</table>

The underlying model of the Loneliness Scale postulated unidimensionality; yet at the same time it was reported that under factor analysis the items loaded on positive and negative factors. This finding was explained by De Jong Gierveld and Kamphuis as a methodological artefact ‘response set’ problem such that the obtained Rasch model fit – which assumes unidimensionality – could not be claimed as proof of unidimensionality for the scale; i.e. the homogeneity of the scale was not very strong.

The evidence for this was that the Loevinger H was in the range 0.30 to 0.50 for different samples.

More recently it has been reported that there are two sub-scales measuring Emotional (negative items) and Social (positive items) loneliness.

Other researchers have also reported a 2-dimensional structure.

<table>
<thead>
<tr>
<th>Construct: Correlation with other measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparisons made to other measures</td>
</tr>
<tr>
<td>□ Correlations with other measures are reported</td>
</tr>
<tr>
<td>X Correlations not reported</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Construct: Discriminant Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>The scale differentiates between relevant categories of respondent e.g. sick vs. well, varying degrees of severity</td>
</tr>
<tr>
<td>Dykstra, et al. (2005)</td>
</tr>
<tr>
<td>Dykstra and De Jong Gierveld (2004)</td>
</tr>
<tr>
<td>De Jong Gierveld and Tilburg (1999)</td>
</tr>
<tr>
<td>van Baarsen, et al. (1999)</td>
</tr>
<tr>
<td>Sadler, et al. (2006)</td>
</tr>
<tr>
<td>Stevens and Westerhof (2006)</td>
</tr>
<tr>
<td>van Baarsen, et al. (2001)</td>
</tr>
<tr>
<td>X Scale differentiates between relevant categories of respondents</td>
</tr>
<tr>
<td>□ No information on discriminant validity</td>
</tr>
</tbody>
</table>

Regarding responsiveness, the Loneliness Scale was sensitive to differences among older adults’ increasing age, depression, gender, household composition, physical condition and number of chronic illnesses, relationship status, self esteem or life satisfaction, self reported loneliness, social anxiety, social network size and social participation/support. National differences have also been reported among older adults between Canada, Italy and the Netherlands.
| Criterion | Information on the relationship of scores to gold standard measures or clinical diagnosis is provided | De Jong Gierveld and Tilburg 1999 | □ Comparison made to criterion measures  
X No comparison with criterion measures provided | There are no ‘gold’ standard for loneliness or social isolation instruments. Criterion validation evidence for the De Jong Gierveld Loneliness Scale has been published in Dutch, but this evidence was not accessible to the reviewer. In the user manual this was reported as being an area where there is insufficient data due to the absence of research. |

| Interpretability | The degree to which one can assign qualitative meaning to quantitative scores | Authors provide 2 or more types of information on interpretability  
X Authors provide limited information to assist with interpretability  
X No information provided | No studies reporting interpretation of the De Jong Gierveld Loneliness Scale have been published in English. |

<table>
<thead>
<tr>
<th>RESPONSIVENESS</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
</table>
| Floor and ceiling effects | The questionnaire fails to demonstrate a worse score in patients who clinically deteriorated and an improved score in patients who clinically improved | □ Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected  
□ Descriptive statistics of the distribution of scores were presented and more than 15% of | No evidence of floor or ceiling effects appears to have been published. |
<table>
<thead>
<tr>
<th>Sensitivity to change</th>
<th>van Baarsen, et al. (1999)</th>
<th>□ Hypotheses were formulated and results were in agreement</th>
<th>□ MCID – No information was provided.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Kremers, et al. (2006)</td>
<td>X An adequate metric was used (ES, SRM, comparison with external standard)</td>
<td>X MCID - Information was provided about the magnitude of score differences which would be clinically meaningful</td>
</tr>
<tr>
<td></td>
<td>Martina and Stevens (2006)</td>
<td>□ No information on sensitivity to change was provided</td>
<td>□</td>
</tr>
</tbody>
</table>

**Responsiveness over time among older adults who between administrations lost their partner by death was reported by van Baarson, et al. Kremers et al reported that scores over time significantly changed for both the treatment and control cohorts in self-management for older women, and in friendship enrichment there was a significant decline in scores.**

**Cultural Applicability and Cultural Adaptations:** The De Jong Gierveld Loneliness Scale is available in English, Dutch, French, Italian.

**Gender Appropriateness:** The instruments is appropriate for use with both genders. No reports have been published suggesting that there are significant gender issues.

**Age Appropriateness:** No reports have been published suggesting that there are significant age issues. It has been used with older adults.

**Summary:** The De Jong Gierveld Loneliness Scale is an excellent scale which has been carefully conceived over a very substantial period of time, developed in population samples (including older adults), and there is there is a very substantial body of evidence supporting its reliability and validity. It will require a linguistic validation study before it is used with Australian samples.

**Reporter:** A/Professor Graeme Hawthorne

**Date of report:** July, 2007

**References**


Appendix 10: Reviews of Associated Symptom Measures
Appendix 10.1: Global Assessments of Behavioural and Psychological Symptoms of Dementia
Appendix 10.1.1: Neuropsychiatric Inventory

AHOC INSTRUMENT REVIEW SHEET

Title: Neuropsychiatric Inventory.

Abbreviations: NPI.

Author(s) Name: Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J.

Author(s) Address: Neurological Unit, Psychiatry Service (116F), West Los Angeles, VAMC, 11301, Wilshire Boulevard, Los Angeles, CA 90073, USA.


Supplied by: Neurological Research Centre, Department of Neurology and Psychiatry and Behavioural Sciences, UCLA School of Medicine, Los Angeles, USA.

Cost: Nil.

Training requirements: A training pack is available to assist staff recording caregivers’ responses, however, specific training is not required. The questions can be answered by anyone familiar with the person with dementia, such as a regular caregiver, who speaks the same language as the interviewer.

Purpose: To assess psychopathology in the person with dementia and to help distinguish between different causes of dementia. It includes symptoms known to be rare in Alzheimer’s disease, but are characteristic of fronto-temporal dementias. It also assesses the level of caregiver distress engendered by each of the neuropsychiatric disorders.

Administration time: 10-20 minutes.

Instrument Type: Clinical rating scale.

Structure: The Neuropsychiatric Inventory contains 12 items. These comprise 10 sub-sections examining behavioural areas and 2 types of neuro-vegetative change, each with 5-8 items. The 12 items address: delusions; hallucinations; agitation; depression; anxiety; euphoria; apathy; disinhibition; irritability; aberrant behaviours; night-time behaviours; and appetite and eating disorders. A screening question is asked first for each item, followed by sub-questions if the response in the screening question suggests the presence of abnormalities in the neuropsychiatric domain. After answering the sub-questions, the informal caregiver rates the frequency and severity of each abnormality. The Neuropsychiatric Inventory with Caregiver Distress Scale has an additional question on each domain specifically addressing the level of distress caused to carers by each specific symptom.

The Neuropsychiatric Inventory-Nursing Home version is a modified version of the original instrument and designed to measure psychiatric symptoms in patients with dementia. Modified wording on each NPI question enables care staff to act as the informant, rather than obtaining the information from the informal carer of a community dwelling person with cognitive impairment.
The NPI-Questionnaire, a shorter version of the NPI, is useful for the purpose of surveying the surface of neuropsychiatric symptoms in dementia, and is considered suitable for use with caregivers and care staff, as well as with General Medical Practitioners.

**Scoring:**

Both frequency and severity of behavioural symptoms are scored with possible scores ranging from 0 to 144. For each domain there are four scores: frequency, severity, total (frequency x severity), and caregiver distress.

Frequency is rated as 1 (occasionally-less than once per week) to 4 (very frequently-daily or essentially continuously present).

Severity is rated as: 1 (mild-produces little distress in the person) to 3 (severe-very disturbing to the person and difficult to redirect).

Distress for the caregiver is scored as: 0 (no distress) to 5 (very severe or extreme) on the NPI Caregiver Distress Scale. The same scoring occurs for the Nursing Home version.

**Developed for:**

The NPI is a clinical rating scale used to assess the behavioural symptoms associated with dementia. It is employed in clinical settings such as the hospital, and is also routinely used in the community setting and in research, including drug and treatment trials.

**Normative Data:**

The NPI was administered to 40 persons without dementia. It was found that the symptom groupings on the NPI are specific to dementia, and do not discriminate among elderly individuals spread across a range of possible scores. Only a few of the NPI categories show a minimal response in non-demented controls, indicating that the older person without cognitive impairment/dementia has hardly any of the symptoms identified by the NPI (Cummings, JL., et al. 1994: Mega, et al. 1996). NPI items are influenced very little by the normal ageing process, a finding confirmed by family carers of persons with dementia (Mega, Cummings, Fiorello, et al. 1996). The NPI was also administered to a Japanese population sample of 1162 people (Ikeda, et al. 2004).

**Clinical Data:**

Levy, et al. (1996) have used the NPI to measure behavioural distinctions associated with frontotemporal dementias and Alzheimer's disease, Craig, et al. (1996) has used the NPI to measure behavioural correlates of cerebral blood flow in Alzheimer's disease, Litvan, et al. (1996) and Mega, et al. (1996) have used the NPI to track the neurological disease process, White, et al. (2004) used the NPI-NH to identify the link between behavioural disturbance in dementia and body mass index, and Kaufer, et al. (1996; Kaufer, et al. 1998) used the NPI to evaluate the efficacy of pharmacological interventions. The NPI, therefore, is effective at measuring change in regards to drug and other treatments/therapies.

**Applications:**

NPI, NPI-Q and the Caregiver Distress Scale are routinely employed in clinical practice by clinicians to gain reports from caregivers of their experiences of the frequency and severity of behavioural disturbances over the previous four weeks, as well as the distress each of these behaviours has caused the caregiver.

The NPI-Questionnaire, a shorter version of the NPI, is useful for surveying the surface of neuropsychiatric symptoms in dementia, especially for caregivers, but is not suitable for use in medication trials because of its brevity and continued reliance on caregiver report.

**Carer and/or Patient Use of Instrument:**

The NPI is intended for use by clinicians and other health professionals, with reference to the regular caregiver’s experiences of neuropsychiatric symptoms in the person with dementia. Regular caregivers are chosen to convey information on the frequency and severity of symptoms on the NPI because the person with dementia may not exhibit these behavioural
abnormalities during the course of a clinical visit, or assessment period. Thus, changes would be underestimated if the ratings were based on the clinician’s observations. It is not usually used by family caregivers themselves, as they are the source of the information required, however, this is a possible application when caregivers need to report these experiences to clinicians via telephone, or other media. It is not suitable for use by the person with dementia as they are unlikely to be able to recall or describe their own symptoms and the level of severity of these. The NPI-Caregiver Distress Scale can be used by the caregiver to score the impact of the neuropsychiatric symptoms on themselves. The shorter version, the NPI-Short Questionnaire is suitable for use by informal caregivers. The NPI-Nursing Home version is employed by regular care staff to rate their own experiences of neuropsychiatric symptoms of dementia to assist with care planning and treatment.

Psychometric Criteria

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td>Cummings, Mega, Gray, et al. (1994)</td>
<td>X Alpha &gt;0.70</td>
<td>A high level of internal consistency (Cronbach’s alpha = 0.88) was found between the items/sub-scales.</td>
</tr>
<tr>
<td></td>
<td>Cummings (1997)</td>
<td>□ Marginal or inadequate internal consistency (&lt;0.70)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Politis, Mayer, Passa, Maillis (2004)</td>
<td>□ No information found on internal consistency</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lange, Hopp, Kang (2004)</td>
<td></td>
<td>An expert panel participated in a Delphi method to rate the scale items. Internal consistency was established (α = 0.75-0.89) for each sub-scale of the NPI.</td>
</tr>
<tr>
<td></td>
<td>Bidzan &amp; Bidzan (2005)</td>
<td></td>
<td>Cronbach’s alpha for the H-NPI (Hellenic version) total score was 0.76 and varied from 0.69 to 0.76 for individual domains. This is indicative of a high degree of reliability. Cronbach’s alpha for the total score on the English version was 0.88, with a range of 0.87 to 0.88 on the sub-domain scores.</td>
</tr>
<tr>
<td></td>
<td>Cummings (1997)</td>
<td>X ICC &gt;.70</td>
<td>The internal consistency reliability of the NPI-NH in nursing home patients with dementia, estimated using Cronbach’s alpha, was 0.67.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Time intervals and confidence intervals reported</td>
<td>Internal consistency of the Polish translated version of the NPI-NH was good. Cronbach’s alpha was 0.85 for both the frequency and severity of symptoms.</td>
</tr>
</tbody>
</table>

Test – retest

<table>
<thead>
<tr>
<th>Tests Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cummings (1997)</td>
<td>X ICC &gt;.70</td>
<td>Test-retest reliability was established by conducting a second round of NPI interviews within three weeks of the first interviews. One half</td>
</tr>
<tr>
<td>Study</td>
<td>Test-retest reliability</td>
<td>Notes</td>
</tr>
<tr>
<td>-------</td>
<td>------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Cummings, et. al. (1994)</td>
<td>Marginal or inadequate internal consistency ICC&lt;.70</td>
<td>Calculation of an intraclass correlation coefficient (ICC); and an ICC &gt; 0.70 is desired. Recommended if time interval and confidence intervals were presented. Test-retest scores of all items were significantly correlated, with overall correlations of 0.79 for frequency ( (p=0.0001) ), and 0.86 for severity ( (p=0.0001) ). The reliability of telephone interviews did not differ from face-to-face interviews.</td>
</tr>
<tr>
<td>Choi, Na, Kwon, et al. (2000)</td>
<td></td>
<td>NPI testing was repeated after 32 days with a sub-group (29) of caregivers of 49 controls and 92 patients with Alzheimer's Disease (43), vascular dementia (32), frontotemporal lobar degeneration (11) and other causes (6). Prevalence was highly correlated with the first interview scores for all 12 items.</td>
</tr>
<tr>
<td>Iverson, et al. (2002)</td>
<td></td>
<td>72 hour test-retest reliability coefficients ranged from ( r = 0.55-0.88 ) for each of the individual symptoms in a geriatric neuropsychiatric sample using the NPI-NH. The test-retest reliability of the total score was ( r = 0.76 ). However, with repeated testing at different time points these scores either declined or improved to become more like the average group score.</td>
</tr>
<tr>
<td>Bidzan &amp; Bidzan (2005)</td>
<td></td>
<td>Test-retest reliability was 0.75 for the total score of the Polish translation of the NPI-NH and ranged from 0.40 to 0.85 for each symptom assessed.</td>
</tr>
<tr>
<td>Boada, et al. (2002)</td>
<td></td>
<td>Test-retest reliability of the Spanish translation of the NPI-Q (brief questionnaire form) compared with the original NPI-Q, using Pearson's correlation index was ( r = 0.89 ) for the total symptom scale and ( r = 0.90 ) for the Caregiver Distress Scale.</td>
</tr>
</tbody>
</table>

**Inter-rater**

<table>
<thead>
<tr>
<th>Study</th>
<th>Agreement</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cummings (1997)</td>
<td>Agreement reported and adequate</td>
<td>Inter-rater reliability was determined by two raters blind to the rating chosen by the other for 54 patients with dementia, with a mean MMSE score of 17.4. Inter-rater reliability varied from 0.96</td>
</tr>
</tbody>
</table>

Marginal or inadequate internal consistency ICC<.70 □ No information found on test-retest reliability □ Preferred if time interval and confidence intervals were presented
<table>
<thead>
<tr>
<th>Validity</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>Cummings (1997)</td>
<td>X Patients were involved during item selection and/or item reduction □ Patients were consulted for reading and comprehension □ No patient involvement □ No information found on content validity X There is an adequate coverage of relevant domains There is limited coverage of relevant domains</td>
<td>There is no &quot;gold standard&quot; instrument, with which NPI can be compared, for apathy, irritability, dis-inhibition and euphoria, so content validity was established with a Delphi panel of 10 highly regarded international experts in geriatric psychiatry, behavioural neurology and neuropsychiatry. Each panel member rated the screening and the sub-questions between 1 (well assessed) and 4 (poorly assessed) with criteria for scoring. In developing the NPI, 40 spouses of persons diagnosed with dementia (mean MMSE score of 28.4) were interviewed to identify behaviours of their relative that were different to usual behaviours observed in older persons without dementia. The NPI items were influenced very little by the normal ageing process, and elevated scores on the NPI were found to be evidence of psychopathology for 88% of 50 persons with Alzheimer's Disease, as</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cummings, et al. (1994)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wood, Cummings, Hsu (2000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bidzan &amp; Bidzan. (2005)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Inter-rater reliability established as over 0.90 with 45 separate patient assessments. Correlations were moderate but significant for all domains of the NPI-NH except anxiety and appetite disturbance when comparing Licensed Vocational Nurses ratings with researchers’ observations. Certified Nurses’ Aids ratings correlated moderately well, especially for residents with high levels of neuropsychiatric disturbance. Findings support the use of the patient’s primary nurse to rate the NPI-NH in formal care settings. Inter-rater reliabilities were 0.86 for the total score and ranged from 0.71 to 1.00 for each of the symptoms reported for the Polish translation of the NPI-NH.
<table>
<thead>
<tr>
<th>Construct</th>
<th>Cummings, et al. (1994)</th>
<th>X Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>□ Limited /inadequate construct validity reported</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information provided</td>
</tr>
</tbody>
</table>

| Devoid of a "gold standard" measure of key behavioural symptoms of dementia, the NPI subscales were compared with the BEHAVE-AD and Hamilton Depression Rating Scale (HAM-D). Highly significant correlations were found between the subscales of all three instruments. All but one correlation reached 0.001 significance level. |
| The Hellenic version of the NPI sub-scale scores correlate highly with the Brief Psychiatric Rating Scale (BPRS), and caregiver distress compares favourably with the Caregiver Emotional Distress Scale (EDS) for Greeks diagnosed with dementia. |
| The convergent and |

---

Ikeda, Fukuhara, Shigendobu, et al. (2004) observed and reported by caregivers and clinicians. Two domains not covered in similar instruments of psychopathology in dementia were added as result: night time behaviours and eating behaviours.

In a population sample of 1,162 Japanese people assessed with the NPI, the most common symptoms were apathy/indifference (56.7%), agitation/ aggression (35%), aberrant motor behaviour (31.7%), and irritability (31.7%). The NPI findings compared favourably with clinician assessment using health and social history, physical and neurological examination, and use of the Clinical Dementia Rating Scale (CDR).

The Hellenic version of -NPI findings compared favourably with: social worker assessment of behaviours of concern that evoked embarrassment and/or fear in caregivers via telephone interview, expert clinician health and social history, MMSE test scores, neurological examination, rating of behavioural disturbance with the Brief Psychiatric Rating Scale (BPRS), and of caregiver distress with the Caregiver Emotional Distress Scale (EDS).
discriminant validity of the NPI-NH was within an acceptable range when correlated with other behavioural measures of agitation, mania, depression, and various psychiatric symptoms of dementia.

Correlations across all five factors identified in the NPI, compared highly with different factors in validated measures, including: the agitation factor in the Cohen-Mansfield Agitation Inventory; sleep/aberrant motor activity factor, aggression and compliance factor and the psychosis factor in the Geri-SNAP; elevated behaviour factor in the Mania Rating Scale; mood in the Cornell Depression Scale and the Brief Psychiatric Rating Scale.

Convergent validity between NPI-Q and NPI using Pearson correlation index was $r = 0.99$ and $r = 0.92$ for the Caregiver Distress Scale.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>□ No evidence provided/failed a test of dimensionality</td>
<td>X Some evidence provided to support internal structure</td>
<td>□ Substantial evidence provided to support internal structure</td>
</tr>
</tbody>
</table>

Comparison of NPI-NH study results with other NPI study factor solutions reveals a co-concurrence of some behavioural symptoms across studies: agitation, aggression, irritability and lability consistently form one factor; and depression, dysphoria and anxiety consistently form another factor.

The 12 NPI items load into five main factors which are well defined, with factor loadings ranging from moderate to high.

In the NPI, a five factor solution (agitation, mood, psychosis, sleep/motor activity and elevated behaviour) accounted for 63% of the variance.

Developing and testing the NPI with 162 Italians with Alzheimer’s Disease caregiver reports revealed a clustering of all but two behavioural symptoms into three factors: mood, psychosis and frontal syndromes. Apathy and aberrant motor behaviour had high loadings on all three factors, indicating their poor specificity for one single dimension. The “mood” factor
had high loadings of anxiety and depression subscales, the “psychosis” factor had high loadings of agitation, hallucinations, delusions, and irritability, while the “frontal behaviour” factor had high loadings of disinhibition and euphoria.

| Construct: Correlation with other measures | Ikeda (2004) I | X Correlations with other measures are reported | The NPI sub-scale and total scores correlate highly with the Clinical Dementia Rating Scale (CDR) and the Mini Mental State Exam (MMSE) in community-dwelling Japanese diagnosed with dementia. All correlations between the Occupational Disruptiveness scale (OD), a proxy version of the NPI- Caregiver Distress Scale, and the domain scores on the NPI were significant in a Rivastigmine drug trial in nursing home residents with dementia. |
| Comparisons made to other measures | Haloum, et al. (2005) | □ Correlations not reported | |
| Construct: Discriminant Validity | Cummings, et al. (1994) Mega, Cummings, Fiorello, et al. (1996) | X Scale differentiates between relevant categories of respondents | Forty spouses of non-demented aged control subjects with MMSE scores of 25 or above were interviewed as well as caregivers of 50 patients with dementia with a mean MMSE score of 16.2. A few of the NPI categories showed minimal responses for non-demented controls, thus, older people without cognitive impairment/dementia have few symptoms identified by the NPI. Conversely, elevation of NPI scores was present in the patients with dementia, indicating the presence of psychopathology. Age and psychopathology were not related, however, apathy (72%), agitation (60%), anxiety (48%), irritability (42%), dysphoria and aberrant motor behaviour (38% each), hallucinations (10%) and euphoria (8%) were present in dementia and increased with dementia severity. Also, the screening questions of the NPI have been shown to have a false-negative rate of less than 5% (i.e. failure to detect the psychopathology that is revealed if all of the sub-questions are asked). Persons with frontotemporal dementias exhibited significantly more apathy, disinhibition, euphoria, and |
| | Levy, Miller, Cummings, et al. (1996) | □ No information on discriminant validity | |
Choi, et al. (2000)  
The Korean version of the NPI correlates positively with dementia levels and severity as assessed by the Korean MMSE. NPI scores for depression, anxiety, irritability, night time behaviour, and eating changes were identified at zero or very low rates in controls and significantly less than in persons with dementia.

Politis (2004)  
The Hellenic version of the NPI was able to distinguish patients referred for different reasons, such as apathy, delusions, aggression, irritability and aberrant motor behaviour. Those with more severe dementia exhibited slightly more severe neuropsychiatric symptoms as identified by the H-NPI.

However, it has also been found that there is no difference in NPI scores for those diagnosed with Alzheimer’s Disease and Vascular Dementia.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Information on the relationship of scores to gold standard measures or clinical diagnosis is provided</th>
<th>Cummings, et al. (1994)</th>
<th>X Comparison made to criterion measures  □ No comparison with criterion measures provided</th>
<th>No “Gold standard” measure exists for behavioural symptoms of dementia, however the NPI has tested favourably against the BEHAVE-AD and the Hamilton Depression Rating Scale (HAM-D). Regional cerebral blood flow to the brain, as measured by single photon emission computed tomography (SPECT) reveal that changes in pre-frontal and anterior temporal perfusion are most highly correlated with NPI apathy scores. NPI scores compared favourably with diagnosis confirmed by computed tomography, blood tests including serum vitamin B-12 and thyroid function tests and by reference to DSM-III-R criteria.</th>
</tr>
</thead>
</table>

<p>| Interpretability | The degree to which one can assign qualitative meaning to quantitative scores | Cummings (1997) | X Authors provide 2 or more types of information on interpretability  □ Authors provide limited information to |  |</p>
<table>
<thead>
<tr>
<th>Do authors provide the following:</th>
<th>assist with interpretability</th>
<th>deviations and a range of scores are reported in most studies, as well as an association between a change in sub-scores and the total score or global ratings on comparable measures.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presentation of means and SD of scores before and after treatment</td>
<td>□ No information provided</td>
<td>When 40 controls were tested, all NPI sub-scores were zero from a possible score of 12 (frequency x severity), except for depression (mean 0.25), disinhibition (mean 0.13) and irritability (mean 0.05). By comparison most of the 50 subjects with Alzheimer’s Disease had significantly elevated scores for all NPI sub-scales.</td>
</tr>
<tr>
<td>Comparative data on the distribution of scores in relevant subgroups</td>
<td></td>
<td>NPI provides valuable insight into the magnitude and characteristics of patient’s responses in clinical drug trials of Tacrine (cholinesterase inhibitor).</td>
</tr>
<tr>
<td>Information on the relationship of scores to well-known functional measures or clinical diagnosis</td>
<td></td>
<td>The NPI sub-question scores correlate highly with expert clinician assessment, including reference to DSM-III-R criteria, detailed history, physical and neurological examination, use of the Clinical Dementia Rating Scale (CDR), and dementia diagnosis by computed tomography, blood tests including serum vitamin B-12 and thyroid function tests.</td>
</tr>
<tr>
<td>Information on the association between changes in scores and patients’ global ratings of the magnitude of change they have experienced</td>
<td></td>
<td>A clinically meaningful change score on the NPI has been identified as plus or minus nine points, or a 50% positive or negative movement from the average baseline NPI scores of study participants.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Estimates of change in NPI-NH scores were calculated by considering the degree of measurement error on each sub-scale on two occasions of testing. This identified that a rather large change of total scores is necessary to conclude that the person with dementia has reliably improved or deteriorated over time. Using an 80% CI, the person’s total score would have to improve or decline by 22 or more points to exceed the level of change that could be attributed to measurement error.</td>
</tr>
</tbody>
</table>

- Mega, Cummings, Fiorello, et al. (1996)
- Kaufer, et al. (1998)
- Kaufer, et al. (1998)
- Iverson, et al. (2002)
## RESPONSIVENESS

<table>
<thead>
<tr>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Floor and ceiling effects</strong></td>
<td>Mega, Cummings, Fiorello, et al. (1996)</td>
<td>X Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected. □ Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score. □ No information provided on floor and ceiling effects.</td>
</tr>
<tr>
<td><strong>Sensitivity to change</strong></td>
<td>Mega, Cummings, Fiorello, et al. (1996)</td>
<td>□ Hypotheses were formulated and results were in agreement. X An adequate metric was used (ES, SRM, comparison with external standard). □ No information on sensitivity to change was provided. □ MCID - Information was provided about the magnitude of score differences which would be clinically meaningful. X MCID – No information was provided.</td>
</tr>
<tr>
<td>Frisoni, et al. (1999)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaufer, et al. (1998)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cummings, et al. (1994)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
and NPI scores in dose related improvements in behaviour. With the NPI-NH relatively large change scores (22 or more points) are necessary to conclude that the person with dementia has improved or deteriorated over time (statistical significance).

Cultural Applicability and Cultural Adaptations: NPI, NPI-NH, NPI-Q and the Caregiver Distress Scale are employed in the USA, Canada, UK and other English speaking countries, and have been translated successfully for use in many other countries. Cross-cultural reliability and validity have been established in translations of the English versions. These include: Greek (Politis, et al. 2004); Italian (Frisoni, 1999); Japanese (Hirono, et al. 1998; Ikeda 2004); Korean (Choi 2000); Mexican (Dias, et al. 1996a); Polish (Bidzan & Bidzan 2005); Spanish (Boada, et Al. 1992); and Dutch (Kat & De Jongle 2002).

Gender Appropriateness: Appropriate for use with both genders.

Age Appropriateness: The NPI is generally used with people over 60 but no age limitation has been specified by the authors.

Summary: The Neuropsychiatric Inventory (NPI) and the Neuropsychiatric Inventory-Nursing Home (NPI-NH) are comprehensive assessment tools for a range of behavioural dysfunctions, including sleep patterns, euphoric states, agitation, aggression and apathy. Behaviours are identified and rated according to frequency, severity and the level of distress these cause for both formal and informal caregivers. The scale is in questionnaire form and is widely used in all health care settings and can be used by the caregiver when reporting on behaviours over time. While it has 12 items to answer, because of an initial screening item for each domain, if any of the behaviours are not present, then the interviewer can fast track to the next domain. A major benefit of the Caregiver Distress Scale component of the NPI and the NPI-NH is the caregiver’s opportunity to report the effect that the dysfunctional behaviours have on them. In addition to documenting neuropsychiatric symptoms as part of a care plan, the original NPI has been used as an outcome measure in several research studies. The NPI-Questionnaire, a shorter version of the NPI, is useful for the purpose of surveying the surface of neuropsychiatric symptoms in dementia, but is not suitable for use in medication trials because of its brevity and continued reliance on caregiver report.

Levy, et al. (1996) used the NPI to measure behavioural distinctions associated with frontotemporal dementias and Alzheimer’s disease, Craig, et al. (1996) has used the NPI to measure behaviour correlates of cerebral blood flow in Alzheimer’s disease, Litvan, et al. (1996) and Mega, et al. (1996) have used the NPI to track the neurological disease process, and Kaufer, et al. (1996; 1998) have used the NPI to evaluate the efficacy of pharmacological interventions. The NPI, therefore, is effective at measuring change in relationship to treatment.

The NPI, the NPI-NH and the Caregiver Distress Scale have high test-retest reliability and established inter-rater reliability in several cultures. The short version, NPI-Q has also been validated and is considered useful in General Medical practice, for use by qualified nurses in formal care settings and for use by caregivers in the community setting (Boada, et al. 2002). The benefits of using the NPI and its different versions internationally is that reliability and validity have been established in translated versions, such as
Greek (Politis, et al. 2004), Spanish (Boada, et al. 2002), Italian (Frisoni, et al. 1999), Japanese (Ikeda 2004), Korean (Choi 2000), Polish (Bidzen & Bidzan 2005), Mexican (Dias, et al. 1996) and Dutch (Kat & de Jongle 2002). The NPI-NH (nursing home) version is widely used to document neuropsychiatric symptoms for care planning and to assess longitudinal changes in behavioural and psychological symptoms in aged care residents. The NPI is also used to evaluate behavioural distinctions associated with fronto-temporal and Alzheimer’s dementia, and assess the effects of various pharmacological treatments in dementia populations. The threshold of >4 score on the NPI is often relied on in clinical trials involving persons with neuropsychiatric symptoms of dementia (Lyketsos, 2007)

Nevertheless, Frisoni’s, et al. (1999) study with 162 Italians cautions that the reliance on the NPI for measuring apathy and aberrant motor behaviour may not identify them as separate factors in dementia, since they loaded onto more than one factor when factorial analysis was conducted on the NPI. It was hypothesised that these particular symptoms may represent separate dimensions of psychopathology in Alzheimer’s disease that, being assessed by only one sub-scale each, failed to be isolated as separate factors. These authors also caution that some of the behavioural symptoms might be secondary to others, such as aggression which can result from hallucinations and delusions. Alternatively these symptoms can arise from a common pathological and neuro-chemical substrate, which might give rise to agitation, hallucinations and delusions. Consequently, Frisoni, et al. (1999) recommend the identification of a few behavioural syndromes (such as mood, psychosis and frontotemporal syndromes) in place of the larger number of different symptoms identified with the NPI, as this is more likely to lead to greater ease in identifying the underlying aetiologies of dementia.

Iverson, et al. (2002) also felt that a change of scores of less than 22 points on the NPI should be interpreted with caution, as this difference may not reflect a meaningful change between two NPI scores, citing the possibility that a halo effect may occur when scoring a person who has previously had a positive or negative effect on the caregiver. Iverson, et al. (2002) also warned that while statistically meaningful scores increase the likelihood that a change in symptom scores will be due to factors other than measurement error, it does not guarantee clinically meaningful change, e.g. a statistically significant drop in aggressive behaviour does not mean that the lower level of aggression is tolerated by care staff/carers. Conversely, real change can occur even if the difference score does not exceed measurement error that can occur in the clinical setting.

When using the NPI-NH with a heterogeneous population, Lange, et al. (2004) recommend scoring and interpreting the individual items or factors (agitation, mood, psychosis, sleep/motor activity, and elevated behaviour), as opposed to total scores.

Lyketsos (2007) agrees with Rosenberg, et al. (2005) in the suggestion that NPI ratings are most vulnerable to the effect of caregiver variables, due to its reliance on caregiver assessment without reference to the clinical judgement of experienced clinicians and the patient. A revised NPI is recommended which includes these additional inputs to assessment. As well, expanding the NPI could address not only measurement of the frequency and severity of neuropsychiatric symptoms in dementia, it could also include more detail of specific symptoms. These may include a broader spectrum of items in the existing domains; two sets of NPI items, with a limited set across all domains to be used as a broad spectrum tool and an expanded set in each domain to be used as a narrow spectrum tool; retention of the current NPI scoring system; and two administration methods, the current caregiver method and a clinician-driven method which employs clinical judgement based on staff and caregiver input. Lyketsos (2007) also recommends that all raters be systematically trained to use the NPI in a consistent manner.
Despite these suggestions, the NPI meets most of the key ingredients for a robust and accessible behavioural assessment instrument, including:
comprehensive and detailed description of the behaviours being measured;
a weighting of severity of these behaviours; frequency, duration and effect on caregivers; objective and uniform measurement; and practical for use in the clinical area and in the community setting; and translated versions are reliable and valid in different cultures.

**Reporter:** Lynn Chenoweth

**Date of report:** 8/05/07

**References**


Appendix 10.1.2: Behavioural Pathology in Alzheimer's Disease Rating Scale

AHOC INSTRUMENT REVIEW SHEET

Title: Behavioural Pathology in Alzheimer's Disease Rating Scale.

Abbreviations: BEHAVE-AD.

Author(s) Name: Barry Reisberg.

Author(s) Address: Aging and Dementia Research Centre New York University School of Medicine 500 First Avenue, New York, NY10016


A version of the updated scale with a frequency component is available in the appendix of the journal by Monteiro and colleagues (2001) *Addition of a frequency-weighted score to the Behavioral Pathology in Alzheimer's Disease Rating Scale: the BEHAVE-AD-FW: methodology and reliability.* *European Psychiatry: the Journal of the Association of European Psychiatrists.* Vol. 16 Suppl 1, pp. 5s-24s.

Cost: Free.

Training requirements: None, no manual available. Note that the psychiatric language in the scale makes it unsuitable for completion by persons without some clinical training.

Purpose: To measure behavioural and psychological symptoms of dementia in persons with Alzheimer's disease.

Administration time: 20 minutes (estimate from Burns, et al. 2003), although it may be faster if used in self-completion form by a suitable informant.

Instrument Type: Clinical rating scale based on the interview of the informant (though it has been used in self-complete form by residential care nurses).

Structure: There are 25 items grouped into 7 major categories: paranoid and delusional ideation, hallucinations, activity disturbance, aggressiveness, diurnal rhythm disturbances, affective disturbances and anxieties and phobias.

At the end of the scale there is a 4-point global assessment of the overall magnitude of the behavioural symptoms in terms of disturbance to the caregiver and/or dangerousness to the patient.

The new version (Monteiro, Boksay, et al. 2001) includes a frequency score for each item.

Scoring: Each symptom is rated on a 4-point scale of severity over the previous 2 weeks:

0 = not present.
1 = present.
2 = present, generally with an emotional component.
3 = present, generally with an emotional and physical component.
**Frequency** is score over the previous 2 weeks:
  - If not present, frequency is automatically scored 0.
  - 1 = once.
  - 2 = every several days.
  - 3 = daily.
  - 4 = more than once daily.

Note that item number 19 ‘day/night disturbance’ is assessed using a three rating point frequency score.

Symptom category severity scores can be obtained, and all items except the global rating are summed to produce the BEHAVE-AD total score. In the revised BEHAVE-AD-FW, a frequency-weighted severity score for each item is calculated by multiplying the severity score for each item by its frequency score. Symptom category and total scores are then summed.

When using the scale in a residential care sample (Brodaty, Ames et al. 2003) did not include item 2 ‘one’s house is not one’s home’ and item 4 ‘delusion of abandonment’ as affirmation of these items was often true.

**Developed for:** Persons with Alzheimer’s disease.

**Normative Data:** None.

**Clinical Data:**

The scale has been used as the outcome measure in trials of different models of nursing home care (Brodaty, Draper, et al. 2003), **Risperidone** (De Deyn, Rabheru, et al. 1999; Chan, Lam, et al. 2001; Brodaty, Ames et al. 2003; Brodaty, Ames, et al. 2005), **Rivastigmine** (Burns, Spiegel et al. 2004), **Clozapine** (Chacko, Hurley, et al. 1995), **Haloperidol** (De Deyn, Rabheru et al. 1999; Chan, Lam, et al. 2001), and **Donepezil** (Cummings, McRae, et al. 2006).

The scale has been validated for administration by telephone (Monteiro, Boksay, et al. 1998).

**Applications:** Measurement of BPSD including change as a result of interventions.

**Carer and/or Patient Use of Instrument:** May be completed by professional care staff.

**Psychometric Criteria**

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| The extent to which items in a (sub)scale are intercorrelated; a measure of the homogeneity of a (sub)scale. | □ Alpha >0.70
□ Marginal or inadequate internal consistency (<0.70) | X No information found on internal consistency |
### Test – retest

The extent to which the same results are obtained on repeated administrations of the same questionnaire when no change in physical functioning has occurred.

- Calculation of an intraclass correlation coefficient (ICC); and an ICC > 0.70 is desired.
- Preferred if time interval and confidence intervals were presented.

<table>
<thead>
<tr>
<th>Study</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sclan, Saillon, et al. (1996)</td>
<td>ICC &gt; .70</td>
<td>Time intervals and confidence intervals reported.</td>
</tr>
</tbody>
</table>
| Monteiro, Boksay, et al. (2001) | Marginal or inadequate internal consistency ICC < .70 | No information found on test-rest reliability.

### Inter–rater

Limits of agreement, Kappa, or standard error of measurement (SEM) were presented.

- For the BEHAVE-AD original version: N = 18, ICC > .70 for all symptom categories except anxieties and phobias where ICC = 0.65.
- For the BEHAVE-AD-FW: N = 28, ICC > .75 for frequency-weighted severity scores for all symptom categories except diurnal rhythm disturbances where ICC = 0.69.

### VALIDITY

**Content**

The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire.

- Patients and experts were involved during item selection and/or item reduction.
- Patients were consulted for reading and comprehension.
- No patient involvement.
- No information found on content validity.
- There is an adequate coverage of relevant domains.
- There is limited coverage of relevant domains.

- Items were derived from a retrospective chart review to identify the specific nature of behavioural disturbances in Alzheimer's disease. The authors sought to develop an assessment tool that does not include symptoms that primarily result from cognitive or functional impairments.
- It should be noted that the BEHAVE-AD was developed to measure BPSD in AD. It has been suggested that the instrument is suitable without revision for measurement of BPSD in VaD.
- The instrument has also been used in other types of dementia including Lewy Body dementia and fronto-temporal dementia. However, the instrument does not contain items measuring some behavioural changes common in FTD such as...
<table>
<thead>
<tr>
<th>Construct</th>
<th>Reisberg, Auer, et al. (1996)</th>
<th>X Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used □ Limited/ inadequate construct validity reported □ No information provided</th>
<th>All BEHAVE-AD symptoms peak in occurrence and magnitude at stages before the final stage of Alzheimer’s disease as defined using the GDS. Pharmacological and psychological intervention produce changes in BEHAVE-AD symptom categories independent of cognitive scores.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construct: Internal Structure</td>
<td>Schreinzer, Ballaban, et al. (2005) Harwood, Ownby, et al. (1998)</td>
<td>□ No evidence provided/failed a test of dimensionality X Some evidence provided to support internal structure □ Substantial evidence provided to support internal structure</td>
<td>One study found 3 factors – agitation, affectivity and day/night disturbances. Another found 5 factors – agitation/anxiety, psychosis, aggression, depression and activity disturbance.</td>
</tr>
<tr>
<td>Construct: Correlation with other measures</td>
<td>Auer, Monteiro, et al. (1996) Finkel, Lyons, et al. (1992) Finkel, Lyons, et al. (1993) Cummings, McRae, et al. (2006)</td>
<td>X Correlations with other measures are reported □ Correlations not reported</td>
<td>ICC of 0.54 between the BEHAVE-AD and an observer rated version of the scale, the E-BEHAVE-AD. Correlates with the CMAI in nursing home residents. Correlates with the Brief Agitation Rating Scale (BARS) in nursing home residents. Excellent concurrent validity with the NPI.</td>
</tr>
<tr>
<td>Construct: Discriminant Validity</td>
<td></td>
<td>□ Scale differentiates between relevant categories of respondents X No information on discriminant validity</td>
<td>N/A (no formal definition of different levels of BPSD severity).</td>
</tr>
<tr>
<td>Criterion</td>
<td></td>
<td>X Comparison made to criterion measures □ No comparison with criterion measures provided</td>
<td>No gold standard measure available. This was the first scale developed to measure only behavioural disturbance in AD, many other scales for behavioural disturbance have been compared with the BEHAVE-AD.</td>
</tr>
<tr>
<td>Interpretability</td>
<td>Sclan, Saillon, et al. (1996)</td>
<td>X Authors provide 2 or more types of</td>
<td>Studies provide information on the percentage of</td>
</tr>
</tbody>
</table>
The degree to which one can assign qualitative meaning to quantitative scores
Do authors provide the following:
- Presentation of means and SD of scores before and after treatment
- Comparative data on the distribution of scores in relevant subgroups
- Information on the relationship of scores to well-known functional measures or clinical diagnosis
- Information on the association between changes in scores and patients' global ratings of the magnitude of change they have experienced

<table>
<thead>
<tr>
<th>Authors</th>
<th>Information on interpretability</th>
<th>Persons who have symptoms within each cluster and the mean and standard deviation of total scores, across different levels of dementia severity and MMSE levels.</th>
</tr>
</thead>
</table>

**RESPONSIVENESS**

<table>
<thead>
<tr>
<th>Floor and ceiling effects</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>The questionnaire fails to demonstrate a worse score in patients clinically deteriorated and an improved score in patients who clinically improved</td>
<td>Sclan, Saillon, et al. (1996) Monteiro, Boksay, et al. (2001)</td>
<td>X Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected □ Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score □ No information provided on floor and ceiling effects</td>
<td>The possible range of scores on the total BEHAVE-AD is 0-45 and for the BEHAVE-AD-FW is 0-297. Mean total BEHAVE-AD score in persons with AD of different levels of severity ranged from 2.64 ± 3.26 to 11.17 ± 8.38. Mean total BEHAVE-AD-FW in persons with AD of different levels of severity ranged from 9.4 ± 11.9 (range 0-30) to 37.3 ± 12.0 (range 20-46). Therefore ceiling effects are unlikely. Floor effects may occur, though it is possible that a score of 0 may represent the true absence of any BPSD, rather than a floor effect.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sensitivity to change</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>The ability to detect important change over time in the concept being measured</td>
<td>De Deyn, Rabheru, et al. (1999) Katz, Jeste, et al. (1999) Brodaty, Ames, et al. (2005)</td>
<td>X Hypotheses were formulated and results were in agreement □ An adequate metric was used (ES, SRM, comparison with external standard) □ No information on sensitivity to change was provided</td>
<td>Different criteria have been applied for clinically significant change. Katz defined a clinical response as ≥ 50% reduction in BEHAVE-AD total score. De Deyn et al. defined clinical response as ≥ 30% reduction in BEHAVE-AD total score.</td>
</tr>
</tbody>
</table>
### Cultural Applicability and Cultural Adaptations:

The BEHAVE-AD has been translated into French (see Sclan, 1996), Swedish (Midlov, Bondesson et al. 2002), German (Auer, Hampel et al. 2000), Dutch (Engelborghs, Maertens et al. 2005), Spanish (Boada, Tarraga et al. 2006), Chinese (Chan, Lam et al. 2001), and Korean (Suh, Son et al. 2004).

### Gender Appropriateness:

Appropriate for use with both genders.

### Age Appropriateness:

Generally used with elderly persons 65+ but no age limitation is specified by the authors.

### Summary:

This was one of the first instruments developed for measuring BPSD. It has been used extensively in clinical and residential care settings and translated into many languages. It has also been successfully used in clinical trials to measure change.

### Reporter:

Lee-Fay Low

### Date of report:

April 2007

## References


Appendix 10.1.3:  Dementia Behaviour Disturbance Scale

AHOC INSTRUMENT REVIEW SHEET

Title: Dementia Behaviour Disturbance Scale.

Abbreviations: DBD scale (DBDS).

Author(s) Name: Baumgarten, M., Becker, R. & Gauthier, S.

Author(s) Address: Associate Professor Mona Baumgarten, Ph.D.
University of Maryland at Baltimore
Department of Epidemiology & Preventive Medicine
Division of Gerontology Howard Hall, room 212
660 W. Redwood St.
Baltimore, MD 21201
Phone: (410)706-1531
Fax: (410)706-4433
Email: mbaumgar@epi.umaryland.edu
Web address: http://medschool.umaryland.edu/epidemiology/baumgart.asp

Supplied by: Can be reproduced from Baumgarten, M., Becker, R., & Gauthier, S.

Cost: Free, no permission is required so long as the source is acknowledged.

Training requirements: No formal training required. No manual available.

Purpose: Assessment of behavioural disturbance in persons with dementia.

Administration time: 15 minutes (may be shorter if self-completed).

Instrument Type: Clinical rating scale/checklist based on interview with the caregiver. The checklist may also be self completed by the caregiver.

Structure: The frequencies, during the past week, of 28 items behavioural items are rated:
0 = never
1 = rarely
2 = sometimes
3 = often
4 = all the time

Scoring: All item scores are summed to produce a total score.

Developed for: Persons with dementia.

Normative Data: The DBDS has been used in the caregiver substudy population based Canadian Study of Health and Ageing. However average values for this population could not be located in the published literature.

Clinical Data: The scale has been used in outpatient settings (Ott, Tate, et al. 1996; Coen, Swanwick, et al. 1997), residential care facilities (Kurita, Katayama, et al. 1997; Draper and Turner, 2003; Neville and Byrne, 2007) and in the community (Neville and Byrne, 2007).

Applications: Refer to studies cited in the tables below.

Carer and/or Patient Use of Instrument: Interview with caregiver/self report by caregiver.
### Psychometric Criteria

#### RELIABILITY

<table>
<thead>
<tr>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>X Alpha &gt;0.70</td>
<td>Coefficient of internal consistency &gt;0.80.</td>
</tr>
</tbody>
</table>

**Internal consistency**
The extent to which items in a (sub)scale are intercorrelated; a measure of the homogeneity of a (sub)scale.

- Cronbach's alpha should be between 0.70 and 0.90 for every dimension / sub-scale.
- Baumgarten, Becker, et al. (1990)
- X Alpha >0.70
- □ Marginal or inadequate internal consistency (<0.70)
- □ No information found on internal consistency

**Test – retest**
The extent to which the same results are obtained on repeated administrations of the same questionnaire when no change in physical functioning has occurred.

- Calculation of an intraclass correlation coefficient (ICC); and an ICC > 0.70 is desired
- Preferred if time interval and confidence intervals were presented
- Baumgarten, Becker, et al. (1990)
- Neville and Byrne (2002)
- X ICC >.70
- Time intervals and confidence intervals reported
- □ Marginal or inadequate internal consistency ICC<.70
- □ No information found on test-rest reliability

**Inter– rater**
Limits of agreement, Kappa, or standard error of measurement (SEM) were presented

- Neville and Byrne (2002)
- X Agreement reported and adequate
- □ Inadequate inter-rater agreement
- □ No information provided

#### VALIDITY

**Content**
The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire

- The DBD items were chosen from a literature review supplemented by a listing of common behavioural disturbances in the practices of two of the authors.
- □ Patients and experts were involved during item selection and/or item reduction
- □ Patients were consulted for reading and comprehension
- □ No patient involvement
- □ No information found on content validity
- □ There is an adequate coverage of
<table>
<thead>
<tr>
<th>Construct</th>
<th>Relevant Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>The extent to which scores on the questionnaire relate to other measures in a manner that is consistent with theoretically derived hypothesis concerning the domains that are measured.</td>
<td>□ There is limited coverage of relevant domains</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Construct: Internal Structure</th>
<th>Relevant Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information provided on factor structure</td>
<td>X Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used</td>
</tr>
<tr>
<td></td>
<td>□ Limited/ inadequate construct validity reported</td>
</tr>
<tr>
<td></td>
<td>□ No information provided</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Construct: Correlation with other measures</th>
<th>Relevant Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparisons made to other measures</td>
<td>X Correlations with other measures are reported</td>
</tr>
<tr>
<td></td>
<td>□ Correlations not reported</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Construct: Discriminant Validity</th>
<th>Relevant Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>The scale differentiates between relevant categories of respondent e.g. sick vs. well, varying degrees of severity</td>
<td>X Scale differentiates between relevant categories of respondents</td>
</tr>
<tr>
<td></td>
<td>□ No information on discriminant validity</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Relevant Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information on the relationship of scores to gold standard measures or clinical diagnosis is provided</td>
<td>□ Comparison made to criterion measures</td>
</tr>
<tr>
<td></td>
<td>X No comparison with criterion measures provided</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interpretability</th>
<th>Relevant Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>The degree to which one can assign qualitative meaning to quantitative scores</td>
<td>□ Authors provide 2 or more types of information on interpretability</td>
</tr>
<tr>
<td>Do authors provide the following:</td>
<td>□ Authors provide limited information to assist with interpretability</td>
</tr>
<tr>
<td>Presentation of means and SD of scores before and after treatment</td>
<td>X No information provided</td>
</tr>
</tbody>
</table>

| | | Correlations were in the hypothesized direction with the MMSE, Older Americans Research and Service ADL scale and the Rapid Disability Rating Scale. |
| | | Persons with dementia score higher on the scale than persons without dementia. |
Comparative data on the distribution of scores in relevant subgroups

Information on the relationship of scores to well-known functional measures or clinical diagnosis

Information on the association between changes in scores and patients' global ratings of the magnitude of change they have experienced

<table>
<thead>
<tr>
<th>RESPONSIVENESS</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floor and ceiling effects</td>
<td>Coen, Swanwick, et al. (1997)</td>
<td>X Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
<td>Coen, Swanwick, et al. (1997)</td>
</tr>
<tr>
<td>Sensitivity to change</td>
<td></td>
<td>□ Hypotheses were formulated and results were in agreement</td>
<td></td>
</tr>
</tbody>
</table>

In a clinical sample of persons with Alzheimer's disease mean DBD scores were 20.92 ± 11.43 (range 4-50).

Gender Appropriateness: Appropriate for use with both genders.

Age Appropriateness: Generally used with elderly persons 65 years and over but no age limitation is specified by the authors.

Summary: The Dementia Behavior Rating Scale is a valid and reliable scale for measuring behavioural disturbance in persons with dementia in community, clinical and residential care settings. It was originally designed for completion by clinician following caregiver interview, but has also been administered as a self-completion instrument by carers and nursing home staff. There have been relatively few research studies using this instrument so data on sensitivity to change and interpretation of scores is not available.

Reporter: Lee-Fay Low

Date of report: May 2007

References


Appendix 10.1.4: The Neurobehavioural Rating Scale

AHOC INSTRUMENT REVIEW SHEET

Title: The Neurobehavioural Rating Scale.

Abbreviations: NRS, NBRS, NRS-R (revised).

Author(s) Name: Harvey S Levin.

Author(s) Address: Departments of Physical Medicine and Rehabilitation, Psychiatry and Behavioural Sciences and Neurosurgery Baylor College of Medicine Houston, Texas, USA.


Cost: Free.

Training requirements: No training requirements are specified and no manual is available. The scale was originally completed after structured interview and clinical observation which suggests that the scale should administered by persons with suitable clinical experience.

Purpose: Assessment of behavioural disturbance in patients with head injury.

Administration time: 15-20 minutes (McCauley, Levin et al. 2001).

Instrument Type: Clinician rating scale based on a structured interview and clinical observation.

Structure: 27 items measuring cognitive deficits, psychiatric symptoms and behavioural disturbances.

Note that an additional item ‘fluent aphasia’ was added when the reliability and validity of this scale was tested in persons with dementia (Sultzer, Levin et al. 1992).

There is also a revised version, the NRS-R (McCauley, Levin et al. 2001) which comprises the following changes to the original 27 item version:

- The addition of 2 items: “difficulties in mental flexibility” and “irritability”
- the combination of two previous items “tension” and “anxiety” into “anxiety”
- the separation of “inattention” into “reduced alertness” and “attention”

Scoring: Each item is rated on a 7 point scale:

0 = not present
1 = very mild
2 = mild
3 = moderate
4 = moderately severe
5 = severe
6 = extremely severe

Items are summed to produce the total score. For the NRS-R version scoring points were changed:

0 = absent
1 = mild  
2 = moderate  
3 = severe  

**Developed for:**  
Patients with head injury.

**Normative Data:**  
None.

**Clinical Data:**  

The NRS has also been used for persons with Parkinson’s disease (Mathias, 2003) and stroke (Frank, Schlote, et al. 2006).

In patients with dementia the NRS has been used in clinical settings (Sultzer, Levin, et al. 1992; Sultzer, Levin, et al. 1993; Sultzer, Berisford, et al. 1995; Sultzer, Mahler, et al. 1995; Kastango, Kim, et al. 2002), the community (Kelly, Todd, et al. 2006) and in residential care (Rosen, Bobys, et al. 1999). In persons with dementia NRS has been used to demonstrate change on behavioural disturbance after Citalopram treatment, (Foglia, Pollock, et al. 1997) and Citalopram and Perpethazine treatment (Pollock, Mulsant, et al. 2002). The insight item has also been used to measure impaired insight in persons with dementia (Harwood, Sultzer, et al. 2000; Harwood, Sultzer, et al. 2005).

**Applications:**  
Assessment of neuropsychiatric symptoms in traumatic brain injury and dementia.

**Carer and/or Patient Use of Instrument:**  
Clinical rating of the patient.

**Psychometric Criteria**

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal consistency</strong></td>
<td>Corrigan, Dickerson, et al.</td>
<td>( \times ) Alpha &gt; 0.70</td>
<td>Cronbach’s alpha ranged from 0.97 to 0.89 in patients with head injury for NRS.</td>
</tr>
<tr>
<td></td>
<td>(1990) McCauley, Levin, et al. (2001)</td>
<td>□ Marginal or inadequate internal consistency (&lt; 0.70)</td>
<td>Cronbach’s alpha ranged from 0.62 to 0.88 for the NRS-R factors in patients with head injury.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information found on internal consistency</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cronbach’s alpha should be between 0.70 and 0.90 for every dimension / sub-scale.</td>
<td></td>
</tr>
</tbody>
</table>

| Test – retest                | van Baalen, Odden, et al.     | \( \times \) ICC > 0.70 | CC of 0.87 with brain injury patients. |
|------------------------------| (2006)                        | Time intervals and confidence intervals |                           |
|                              |                               | Cronbach's alpha ranged from 0.75 to 0.87 in patients with head injury for NRS. |                                       |
results are obtained on repeated administrations of the same questionnaire when no change in physical functioning has occurred.

Calculation of an intraclass correlation coefficient (ICC); and an ICC > 0.70 is desired

Preferred if time interval and confidence intervals were presented

Pollock, Mulsant, et al. (2002) reported
- Marginal or inadequate internal consistency ICC < 0.70
- No information found on test-retest reliability

Pollock, Mulsant, et al. (2002) reported

- Inadequate inter-rater agreement
- No information provided


<table>
<thead>
<tr>
<th>VALIDITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
</table>
| Content | Levin, High, et al. (1987) | □ Patients and experts were involved during item selection and/or item reduction
- Patients were consulted for reading and comprehension
- No patient involvement
- No information found on content validity
X There is an adequate coverage of relevant domains
- There is limited coverage of relevant domains | Many of the NRS items were obtained from the Brief Psychiatry Rating Scale and additional items measuring cognitive function were included. Therefore it should be noted that the NRS-R does not provide a pure measure of behavioural disturbance as there are 8 cognitive items such as “memory difficulties”, “difficulty in planning”, “difficulty in mental flexibility” and “disorientation”.

| Construct | Sultzer, Levin, et al. (1992) | Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used
- Limited/ inadequate construct validity reported
- No information provided | For most items, scores were greater in persons with more severe dementia.

X Some evidence provided to support internal structure | In persons with dementia, principle components analysis revealed 6 factors: cognition/insight, agitation/disinhibition, behavioural retardation,
<table>
<thead>
<tr>
<th>Construct: Correlation with other measures</th>
<th>Sultzer, Levin, et al. (1992)</th>
<th>X Correlations with other measures are reported</th>
<th>The anxiety/depression factory correlated with the Hamilton depression scale ( r = 0.54 ) as did the psychosis factor ( r = 0.40 ) and the agitation/disinhibition factor ( r = 0.35 ) with brain injury patients. NRS-R correlates with the Overt Behavioural Scale ( r = 0.42 ) in brain injury patients.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparisons made to other measures</td>
<td>Kelly, Todd, et al. (2006)</td>
<td>☐ Correlations not reported</td>
<td></td>
</tr>
<tr>
<td>Construct: Discriminant Validity</td>
<td>□ Scale differentiates between relevant categories of respondents</td>
<td>X No information on discriminant validity</td>
<td>N/A (no formal definition of different levels of BPSD severity).</td>
</tr>
<tr>
<td>The scale differentiates between relevant categories of respondent e.g. sick vs. well, varying degrees of severity</td>
<td>Rosen, Bobys, et al. (1999)</td>
<td>☐ Comparison made to criterion measures No comparison with criterion measures provided</td>
<td>For a consensus diagnosis of disruptive behaviours in nursing homes, that meet the threshold for neuroleptic use, the optimal cut–off score was 60 (using ROC analysis).</td>
</tr>
<tr>
<td>Criterion</td>
<td>Pollock, Mulsant, et al. (2002)</td>
<td>X Authors provide 2 or more types of information on interpretability Authors provide limited information to assist with interpretability No information provided</td>
<td>A cut-off score of 60 is recommended for a diagnosis of disruptive behaviours for patients in nursing homes. This information could be used to support clinical decision making. In hospitalised patients with dementia, average NRS total scores changed from 53.5 ±10.2 to 43.6 ±12.1 after Citalopram treatment for up to 17 days, from 57.1 ±14.0 to 49.9 ±14.2 after Perphenazine treatment and 58.3 ±11.9 to 56.0 ±15.2 after placebo treatment.</td>
</tr>
<tr>
<td>Interpretability</td>
<td>Rosen, Bobys, et al. (1999)</td>
<td>☐ Do authors provide the following: Presentation of means and SD of scores before and after treatment Comparative data on the distribution of scores in relevant subgroups Information on the relationship of scores to well-known functional measures or clinical diagnosis Information on the association between changes in scores and patients’ global ratings of the magnitude of change they have experienced</td>
<td></td>
</tr>
</tbody>
</table>
### RESPONSIVENESS

<table>
<thead>
<tr>
<th>Floor and ceiling effects</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>The questionnaire fails to demonstrate a worse score in patients clinically deteriorated and an improved score in patients who clinically improved</td>
<td>Sultzer, Levin, et al. (1993)</td>
<td>□ Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score</td>
<td>The total possible score range of the 28-item NRS used in persons with dementia is 0 to 168. In patients with Vascular dementia, mean NRS total score was 37.4 ± 10.9, and in patients with Alzheimer’s disease, mean NRS total score was 31.7 ± 10.8.</td>
</tr>
<tr>
<td>Authors should provide descriptive statistics of the distribution of scores</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sensitivity to change</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>The ability to detect important change over time in the concept being measured</td>
<td>McCauley, Levin, et al. (2001) Pollock, Mulsant, et al. (2002)</td>
<td>□ Hypotheses were formulated and results were in agreement</td>
<td>Total and factor scores of the NRS-R are sensitive to recovery 3-6 months after head injury.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ An adequate metric was used (ES, SRM, comparison with external standard)</td>
<td>Total and factor scores of the NRS-R are sensitive to change in behavioural disturbance in hospitalised persons with dementia after treatment with Citalopram and Perphehazine.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information on sensitivity to change was provided</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ MCID – Information was provided about the magnitude of score differences which would be clinically meaningful</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>X MCID – No information was provided.</td>
<td></td>
</tr>
</tbody>
</table>

### Cultural Applicability and Cultural Adaptations:

The NRS has been translated into French (Soury, Mazaux, et al. 2005) and Mandarin (Chiu, Lin, et al. 1993).

### Gender Appropriateness:

Appropriate for use with both genders.

### Age Appropriateness:

Generally used with elderly persons 65 years and over but no age limitation is specified by the authors.

### Summary:

The NRS was developed for use in patients with head/brain injury, the majority of studies validating and utilising this scale have been in patients with head injury. Nonetheless a few studies have rated the NRS in persons with dementia with good reliability and validity and change in behavioural disturbance was successfully measured using this scale.

### Reporter:

Lee-Fay Low
References


Appendix 10.1.5: Consortium to Establish a Registry for Alzheimer’s Disease – Behavioral Rating Scale for Dementia

AHOC INSTRUMENT REVIEW SHEET

Title: Consortium to Establish a Registry for Alzheimer’s Disease – Behavioral Rating Scale for Dementia.

Abbreviations: CERAD–BRSD, sometimes referred to as C-BRSD.

Author(s) Name: Tariot PN, Mack JI, Patterson MB, Edland SD, Weiner MF, Fillenbaum G, Blazina L, Teri L, Rubin E, Mortimer JA, and Stern Y and the Behavioral Pathology Committee of the Consortium to Establish a Registry for Alzheimer’s Disease.

Author(s) Address: Gerda Fillenbaum, PhD Dept of Psychiatry and Behavioural Sciences Box 3003 DUMC Durham, NC 27710 USA.

Supplied by: With author’s written permission.

Cost: The instrument and instruction manual can be purchased for a cost of $US85.00. A training video is also available at a cost of $US75.00.

Training requirements: Training is required and a training video is available.

Purpose: To evaluate the extent and severity of behavioural pathology in demented or cognitively impaired persons.

Administration time: 20 to 30 minutes.

Instrument Type: Clinical rating scale based on semi-structured interview administered to an informant familiar with the person to be rated.

Structure: The instrument has 46 questions, most rated on a severity scale but some with a yes/no response. Ratings are based on the frequency with which the behaviour has occurred during the month prior to the interview. The original instrument had eight domains: depressive features, psychotic features, defective self-regulation, irritability/agitation, vegetative features, apathy, aggression, and affective lability.

The current version has six domains:
- depressive features,
- inertia,
- psychotic features,
- vegetative features,
- irritability/aggression,
- behavioural dysregulation.

A 17 item abbreviated version is also available.

Scoring: The eight yes/no items are rated 1 and 0. The 38 frequency items are rated as follows:
0 = not occurred since illness began
1 = 1 - 2 days in the past month
2 = 3 - 8 days in the past month (up to twice a week)
3 = 9 - 15 days in the past month
4 = 16 days or more in the past month
For both the yes/no and frequency items there is also the option of rating either 8 (occurred since illness began but not in past month) or 9 (unable to rate). Total BRSD scores can range from 0 to 164 with higher scores representing greater behavioural disturbance.

**Developed for:**

The CERAD-BRSD is used in both clinical and research settings to describe the level of behavioural disturbance in persons with dementia or with cognitive impairment.

**Normative Data:**

Normative data for the instrument is available in the CERAD-BRSD manual. Additionally the CERAD database is available to researchers in the form of a CD-ROM. It contains data for 1094 patients with a clinical diagnosis of Alzheimer’s disease and 463 control subjects evaluated annually between 1987 and 1996. The data includes clinical findings and neuropsychological test scores, behavioural manifestations of dementia, time to death or admission to a nursing home and neuropathological findings. The CD–ROM costs $US600.00 to purchase.

**Clinical Data:**

The CERAD-BRSD has been used in numerous clinical and intervention studies. It has been used as an outcome measure in studies evaluating the effectiveness of drug treatment (Martinon-Torres, Fioravanti & Grimley, 2004; Teri, Logsdon, Peskind, Raskind, et al. 2000; Weiner, Martin-Cook, Foster, Saine, et al. 2000) and in a study assessing an activities based adult dementia care program (Higgins, Koch, Hynan, Carr, et al. 2005). The instrument has been used in studies investigating comorbidity in community dwelling persons with Alzheimer’s disease (AD) (Tractenberg, Weiner, Patterson, Teri, et al. 2003), predicting psychosis onset (Wilkosz, Miyahara, Lopez, Dekosky, et al. 2006) and investigating subtypes of psychosis (Perez-Madrinan, Cook, Saxton, Miyahara, et al. 2004), and in a longitudinal study examining the effects over time of depressive symptoms in persons with AD on depression in their family caregivers (Neundorfer, McIlandon, Smyth, Stuckey, et al. 2001). It has also been used in studies investigating the association between white matter changes and neuropsychiatric symptoms (Lee, Choo, Kim, Jhoo, et al. 2006), and the relationship between nursing home placement and measures of change (Knopman, Berg, Thomas, Grundman, et al. 1999).

The CERAD-BRSD has also been used in a pilot study of a potential new outcome, expected emergence (Tractenberg, Gamst, Thomas, Patterson, et al. 2002).

**Applications:**

Common applications include research studies evaluating behavioural disturbance in persons with dementia or cognitive impairment, studies evaluating the effectiveness of drug treatments or other interventions, and studies investigating psychopathology.

**Carer and/or Patient Use of Instrument:**

Carers provide information to a trained interviewer.
### Psychometric Criteria

#### RELIABILITY

<table>
<thead>
<tr>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha &gt; 0.7</td>
<td>Mixed findings for internal consistency. It is very good for the total scale with α of 0.87, and for the Depressive Symptoms, Irritability/Aggression and Psychotic Symptoms subscales with α’s of 0.77, 0.75 and 0.80. Internal consistency is inadequate for the Inertia, Vegetative Symptoms and Behavioural Dysregulation subscales with α’s of 0.48, 0.56 and 0.51.</td>
</tr>
</tbody>
</table>

#### Internal consistency

The extent to which items in a (sub) scale are inter-correlated; a measure of the homogeneity of a (sub)scale. Cronbach’s alpha should be between 0.70 and 0.90 for every dimension / sub-scale.

- Mack, Patterson & Tariot (1999)

#### Test – retest

The extent to which the same results are obtained on repeated administrations of the same questionnaire when no change in physical functioning has occurred. Calculation of an intraclass correlation coefficient (ICC); and an ICC > 0.70 is desired. Preferred if time interval and confidence intervals were presented.

- Patterson, Mack, Mackell, Thomas, et al. (1997)

#### Inter – rater

Limits of agreement, Kappa, or standard error of measurement (SEM) were presented.

- Tariot, Mack, Patterson, Edland, et al. (1995)

#### VALIDITY

<table>
<thead>
<tr>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agreement reported and adequate</td>
<td>Excellent inter-rater reliability with kappas ranging from 0.77 to 1.00.</td>
</tr>
</tbody>
</table>

#### Content

The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire.

- Tariot, Mack, Patterson, Edland, et al. (1995)
- Patterson, Mack, Mackell, Thomas, et al. (1997)
- Tariot, Mack, Patterson & Tariot, (1999)
<p>| Construct | Jacobs, Strauss, Patterson &amp; Mack (1998) Weiner, Koss, Patterson, Jin, et al (1998) Weiner, Tractenberg, Teri, Logsdon, et al (2000) Tractenberg, Weiner, Patterson, Gamst, et al (2002) | X Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used. □ Limited /inadequate construct validity reported □ No information provided | Showed expected correlations with the following measures of physical and cognitive functioning: Functional Assessment Staging (FAST); Activities of Daily Living (ADL-Functional status, Galago et al 1997); Revised Memory and Behavior Problems Checklist (RMBPC). Scores have also been shown to be associated with indicators of depression. |
| Construct: Internal Structure | Tariot, Mack, Patterson, Edland, et al. (1995) Mack, Patterson &amp; Tariot (1999) | □ No evidence provided/failed a test of dimensionality □ Substantial evidence provided to support internal structure | Internal structure has been confirmed with one study citing inter-item consistency ranging from 0.49 to 0.80. Factor analysis for the original 51 item scale suggests an eight factor solution. For the current 46 item version eight factors were again identified with 6 factors common to mildly and moderately demented persons. |
| Construct: Discriminant Validity | Tariot, Mack, Patterson, Edland, et al. (1995) Whitehouse, Patterson, Strauss, Geldmacher, et al. (1996) Patterson, Mack, Mackell, Thomas, et al. (1997) Weiner, Koss, Patterson, Jin, et al. (1998) | X Scale differentiates between relevant categories of respondents □ No information on discriminant validity | The instrument has been shown to discriminate between different levels of dementia severity and between demented and non-demented persons. CERAD-BRSD scores have also been shown to be associated with white matter changes in the brain (Dong Yung-Lee, et al. 2006). |</p>
<table>
<thead>
<tr>
<th>Criterion</th>
<th>Interpretability</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ Comparison made to criterion measures X No comparison with criterion measures provided</td>
<td>Authors provide 2 or more types of information on interpretability □ Authors provide limited information to assist with interpretability □ No information provided</td>
<td>Most studies report at least means and standard deviations. Most also report either comparative data in relevant subgroups, information on the relationship of scores to well-known functional measures or clinical diagnosis and or relationship of scores to well-known measures. Some also include information about the association between changes in scores and global ratings.</td>
</tr>
</tbody>
</table>

**Criterion**

Information on the relationship of scores to gold standard measures or clinical diagnosis is provided

**Interpretability**

The degree to which one can assign qualitative meaning to quantitative scores

Do authors provide the following:

- Presentation of means and SD of scores before and after treatment
- Comparative data on the distribution of scores in relevant subgroups
- Information on the relationship of scores to well-known functional measures or clinical diagnosis
- Information on the association between changes in scores and patients’ global ratings of the magnitude of change they have experienced

*Authors provide 2 or more types of information on interpretability □ Authors provide limited information to assist with interpretability □ No information provided*

Most studies report at least means and standard deviations. Most also report either comparative data in relevant subgroups, information on the relationship of scores to well-known functional measures or clinical diagnosis and or relationship of scores to well-known measures. Some also include information about the association between changes in scores and global ratings.
## RESPONSIVENESS

<table>
<thead>
<tr>
<th>Floor and ceiling effects</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
</table>

### Sensitivity to change

The ability to detect important change over time in the concept being measured.

<table>
<thead>
<tr>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patterson, Mack, Mackell, Thomas, et al. (1997) Teri, Logsdon, Peskind, Raskind, et al. (2000) Weiner, Martin-Cook, Foster, Saine, et al. (2000) Higgins, Koch, Hynan, Carr, et al. (2005)</td>
<td>□ Hypotheses were formulated and results were in agreement</td>
<td>Sensitivity is mixed. The instrument has been shown to be sensitive to the effects of drug treatment. There is also moderate evidence for sensitivity over time with Patterson et al (1997) reporting significant change but only for persons with mild to moderate dementia.</td>
</tr>
</tbody>
</table>

Score on the CERAD-BRSD did change significantly as a result of a one year weekly activity program but not in the expected direction.

---

**Cultural Applicability and Cultural Adaptations:** The 48 item version of the instrument has been translated into French and Spanish.

**Gender Appropriateness:** Appropriate for use with both genders.

**Age Appropriateness:** Generally used with elderly persons 65 years and over but no age limitation is specified by the authors.

**Summary:**

The CERAD-BRSD is one of the assessment instruments that make up the CERAD battery. It is quite long and takes a trained interviewer about 20 to 30 minutes to complete. The instrument and an instruction manual are available from the authors at a cost of $85.00. A training video is also available. Psychometric properties are very good. The CERAD-BRSD has been used in both clinical and research settings and extensive normative data is available.
References


Appendix 10.2: Measures to Assess Delirium
Appendix 10.2.1: Confusion Assessment Method

AHOC INSTRUMENT REVIEW SHEET

Title: Confusion Assessment Method.

Abbreviations: CAM.

Author(s) Name: Inouye, SK, van Dyck, CH, Alessi, CA, Balkin, S, Siegal, AP, & Horwitz, RI.

Author(s) Address: Sharon K. Inouye, M.D., MPH
Professor of Medicine, Beth Israel Deaconess Medical Center
Harvard Medical School
Milton and Shirley F. Levy Family Chair
Director, Aging Brain Center
Institute for Aging Research
Hebrew Senior Life
1200 Centre Street
Boston, MA 02131
Telephone (617) 363-8020
Fax (617) 363-8901
Email: agingbraincenter@hrca.harvard.edu


Cost: Free for researchers and clinicians, however, when the CAM is used commercially fees will be charged.

Training requirements: Some training is required for optimal use of the CAM. The CAM manual is available from the website (Inouye, 2003)\(^1\). Given the variance of the sensitivity depending on the training of the rater/interviewer, it is recommended more extensive training could reduce the problem. It is also recommended the CAM is used as a screening tool, which means when the patient shows positive CAM results a proper diagnostic investigation of delirium may be necessary.

Purpose: To detect delirium for use in older people who are at high risk for the development of delirium (e.g. older medical and surgical in-patients).

Administration time: 5 minutes (10-15 minutes when combined with the MMSE and the Digit Span).

Instrument Type: Clinical rating scale based on a brief structured questionnaire completed by a clinician and based on observations during a brief, formal cognitive testing\(^2\) (e.g. MMSE) and information provided by a carer.

Structure: The CAM is based on the operational application of DSM-III-R\(^3\) and expert opinion, and consists of nine features of delirium (acute onset and

---

2 It is strongly recommended the CAM be scored based on observation during a formal cognitive testing as its sensitivity is compromised significantly when informal clinical observation is the only source of information (refer to Inouye al. 2001).
3 CAM was developed before DSM-IV was produced, however according to the manual the CAM criteria agree more closely with the DSM-IV criteria than they did with the previous DSM-III-R criteria (Inouye, 2003).
fluctuating course, inattention, disorganised thinking, altered level of consciousness, disorientation, memory impairment, perceptual disturbances, psychomotor agitation, psychomotor retardation and altered sleep-wake cycle).

**Scoring:**

The CAM provides a diagnostic algorithm for delirium based on its four cardinal features (referred to as ‘Short CAM’). The presence of two cardinal features (acute onset and fluctuating course, and inattention), and at least one of the two secondary features (disorganised thinking and altered level of consciousness) indicate delirium. The remaining five features have been shown as non significant to the diagnosis of delirium, however some still use the entire nine questions (referred as ‘Long CAM’) to fulfil the DSM definition of delirium. An additive score of the four cardinal features, ranging between 0-7, is used to measure the severity of delirium (the higher, the more severe) (refer to Inouye, et al. 1999). This method of rating the severity of delirium is yet to be validated.

**Developed for:**

Non psychiatrically trained clinicians to identify delirium quickly and accurately in both clinical and research settings.

**Normative Data:**

Normative data is not available.

**Clinical Data:**

Studies use the CAM as a screening, diagnostic and/or an outcome measure in a variety of clinical settings including both acute and post acute, nursing home/long-term care facilities, albeit mostly for acute inpatients pre-/post-operatively. Clinical studies examine occurrence of delirium, clinical features and other predictive factors of delirium (e.g., depression, activities of daily living, dementia, cognitive impairment, co-morbidity, disability, educational level, apolipoprotein E genotypes, melatonin levels, drug metabolism, anaemia/metabolic disorders, changes in plasma large neutral amino acid, concentrations, nutritional status, fluid and electrolyte imbalance, nitrous oxide, serum anticholinergic activity, postoperative pain, care environment), and relationships between delirium and various outcomes (e.g., mortality, functional outcomes, developing complications, institutionalisation, length of hospital stay; use of restraints, costs). They are summarised in the following three categories:

**Study settings:**


---

1 Some may overlap with other categories. Some of the studies cited used the CAM as a baseline measure.
Delirium in pre/postoperative care setting:


**Pharmacological impact** (Agostini, Leo-Summers, & Inouye, 2001; Kalisvaart, et al. 2005; Kudoh, Takase, Takahira, & Takazawa, 2004; Lipitzin, Laki, Garb, Fingeroth, & Krushell, 2005);

**Depressive symptoms** (Leung, Sands, Mullen, Wang, & Vaurio, 2005);

**Melatonin levels** (Shigeta, et al. 2001);

**Nitrous oxide level** (Leung, Sands, Vaurio, & Wang, 2006).

Psychometric properties of the CAM and its comparisons with other delirium measures:


Applications:

Assessment of delirium in older patients, with or without dementia. The CAM is used either alone or in conjunction with other delirium rating scales such as: the Delirium Symptom Interview (DSI), the Delirium Rating Scale (DRS) and the Memorial Delirium Assessment Scale (MDAS); cognitive test such as the Mini-Mental State Exam (MMSE) and Digit Span; and pathophysiologic examinations. The CAM has been used in both clinical and epidemiological studies conducted in various care settings, acute, sub-acute, community and long-term care settings. Some studies reported results based on a telephone interview CAM (Marcantonio, Michaels, et al. 1998; Nelson, et al. 2006). To detect delirium for patients in ICU who are non-verbal and mechanically ventilated, the CAM-ICU version has been developed and validated. The CAM-ICU showed a high sensitivity and specificity, and excellent inter-rater reliability when used by trained physicians and nurses (Ely, Inouye, et al. 2001; Ely, Margolin, et al. 2001). The CAM has also been adopted to suit emergency care settings (Lewis, et al. 1995) and measure the severity of delirium (Jones, et al. 2006; McCusker, et al. 1998), however, neither of those two adoptions appears to have been widely utilised or sufficiently validated.
Carer and/or Patient Use of Instrument: The CAM is interviewer administered.

**Psychometric Criteria**

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal consistency</strong></td>
<td></td>
<td>□ Alpha &gt;0.70 □ Marginal or inadequate internal consistency (&lt;.0.70) X No information found on internal consistency</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ ICC &gt;.70 □ Time intervals and confidence intervals reported □ Marginal or inadequate internal consistency ICC&lt;.70 X No information found on test-retest reliability</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Agreement reported and adequate □ Inadequate inter-rater agreement □ No information provided</td>
<td>Mixed results have been reported. When the CAM is used by sufficiently trained clinicians/researchers inter-rater reliability is high. For assessing the presence or absence of delirium agreement was 100% (kappa = 1.0); for rating all nine items agreement was 88% (kappa = 0.67); for assessing the 4 cardinal features of the CAM agreement was 93% (kappa = 0.81). The strength of agreement for the four individual CAM features was substantial, ranging from 84% (kappa = 0.56) to 100% (kappa = 1.0). The clinical diagnosis by a psychiatrist had an agreement kappa coefficient of 0.58 (95% CI: 0.41-0.74). The nurse clinician diagnosis using the CAM had an agreement kappa coefficient of 0.86 (95%CI: 0.75-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inouye, et al. (1990)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Zou, et al. (1998)</td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>Sensitivity</td>
<td>Specificity</td>
<td>Note</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>Rolfson, et al. (1999)</td>
<td>0.97)</td>
<td></td>
<td>When the comparison was made between the raters (physician vs. nurse) the nurse rated sensitivity was significantly lower (0.13) than that of the physician’s (1.0), while there was no difference in the specificity (1.0 for both groups). It is noteworthy that the same physician who had made the initial diagnosis of delirium using DSM-III-R conducted the CAM, naturally resulting in complete agreement, and the potential for measurement bias.</td>
</tr>
<tr>
<td>Inouye, et al. (2001)</td>
<td></td>
<td></td>
<td>Sensitivities of untrained nurses’ ratings for delirium (using the CAM based on their observations during routine clinical care without formal cognitive testing) were low (0.15-0.32) but specificities were high (0.91-0.99) when compared with trained researchers (diagnosis of delirium using the CAM in conjunction with the MMSE). Four risk factors for under-recognition of delirium by nurses were hypoactive delirium (adjusted odds ratio [OR], 7.4; 95%CI, 4.2-12.9), age 80 years and older (OR, 2.8; 95%CI, 1.7-4.7), vision impairment (OR, 2.2; 95%CI, 1.2-4.0), and dementia (OR, 2.1; 95%CI, 1.2-3.7). When all four of these risk factors were present, the possibility of under-recognition increased more than 20-fold.</td>
</tr>
<tr>
<td>Monette, et al. (2001)</td>
<td></td>
<td></td>
<td>Extensive training appears to be one of the key factors in improving reliability of the CAM. When compared the CAM results between a geriatrician and a trained lay interviewer (in the emergency care setting) the CAM demonstrated a moderate sensitivity (0.86), high specificity (1.0), and kappa (0.91; 95%CI, 0.81-1.01).</td>
</tr>
<tr>
<td>Gaudreau, et al. (2005)</td>
<td></td>
<td></td>
<td>The agreement between psychiatrists and research nurses who were properly trained for the CAM was high (kappa = 0.89).</td>
</tr>
<tr>
<td>Lemiengre, et al. (2006)</td>
<td></td>
<td></td>
<td>Inter-rater reliability test between two research nurses, trained to conduct the CAM and the 12 item MMSE, showed complete agreement (kappa = 1.0; p &lt; 0.001).</td>
</tr>
<tr>
<td>Simon, et al. (2006)</td>
<td></td>
<td></td>
<td>In a reliability testing study of main delirium assessment tools</td>
</tr>
</tbody>
</table>

When a Spanish version of the CAM showed high inter-rater reliability (kappa = 0.89) and moderate to high concordance for each item (kappa ranging between 0.56-0.95).

---

### VALIDITY

<table>
<thead>
<tr>
<th>VALIDITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
</table>
| Content  | Inouye, et al. (1990)        | □ Patients and experts were involved during item selection and/or item reduction  
□ Patients were consulted for reading and comprehension  
X No patient involvement  
□ No information found on content validity  
X There is an adequate coverage of relevant domains  
□ There is limited coverage of relevant domains |
|          |                              | High face and content validity reported. The CAM is based on the operational application of DSM-III-R\(^1\) and expert opinion. |

\(^1\) CAM was developed before DSM-IV was produced, however according to the manual the CAM criteria agree more closely with the DSM-IV criteria than they did with the previous DSM-III-R criteria (Inouye, 2003).

| Construct | Jones, et al. (2006)       | X Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used  
□ Limited/inadequate construct validity reported |
|-----------|---------------------------|---------------------------------------------------------------|
|           |                           | Various studies reported below demonstrate the CAM has moderate to high construct validity, e.g., high agreement with gold standard diagnostic criteria such as the DSM-III-R and the DSM-IV and cognitive tests such as the MMSE.  
Based on a secondary analysis of two hospital-based studies, the findings demonstrate the correlations between the level of |

---
<table>
<thead>
<tr>
<th>Construct: Internal Structure</th>
<th>Inouye, et al. (1990)</th>
<th>□ No evidence provided/failed a test of dimensionality X Some evidence provided to support internal structure □ Substantial evidence provided to support internal structure</th>
<th>Internal structure was tested against psychiatric diagnostic criteria (refer above) to determine the sensitivity, specificity and likelihood ratio of each clinical feature. Four cardinal features used in the diagnostic algorithm showed the highest likelihood ratios.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construct: Correlation with other measures</td>
<td>Inouye, et al. (1990)</td>
<td>X Correlations with other measures are reported □ Correlations not reported</td>
<td>Moderate to high correlations with other instruments testing cognition and confusional status. The CAM ratings were found to agree moderately or substantially (p &lt; 0.001) with other mental status indexes examining cognition and attention: For MMSE, kappa = 0.64; for story recall, kappa = 0.59; for the visual analogue Scale for Confusion, kappa = 0.82; and for the digit span test, kappa = 0.66). The CAM showed only nominal correlation with a change in the MMSE score (0.25, p = 0.03) and a change in the Clock score (0.35, p &lt; 0.01) using Spearman’s rank correlation coefficient. The overall correlations between the Delirium Observation Screening (DOS) scale and the CAM for the group of hip fracture patients were moderate to good (0.63; p = 0.001), in which the highest correlations were found on items regarding memory and orientation. The CAM had a strong negative correlation coefficient (r = -0.73) with the NEECHAM Confusion Scale. The Pearson’s correlation between the CAM(^1) and the MMSE (Spanish versions) was high (r = -0.84; p &lt; 0.01). The overall correlations between the Delirium Observation Screening (DOS) scale and the CAM for the group of hip fracture patients were moderate to good (0.63; p = 0.001), in which the highest correlations were found on items regarding memory and orientation. The CAM had a strong negative correlation coefficient (r = -0.73) with the NEECHAM Confusion Scale. The Pearson’s correlation between the CAM(^1) and the MMSE (Spanish versions) was high (r = -0.84; p &lt; 0.01). The overall correlations between the Delirium Observation Screening (DOS) scale and the CAM for the group of hip fracture patients were moderate to good (0.63; p = 0.001), in which the highest correlations were found on items regarding memory and orientation. The CAM had a strong negative correlation coefficient (r = -0.73) with the NEECHAM Confusion Scale. The Pearson’s correlation between the CAM(^1) and the MMSE (Spanish versions) was high (r = -0.84; p &lt; 0.01).</td>
</tr>
<tr>
<td></td>
<td>Rolfson, et al. (1999)</td>
<td>□ Correlations not reported</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Schuurmans, Shortridge-Baggett, et al. (2003)</td>
<td>□ Correlations not reported</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Milisen, et al. (2005)</td>
<td>□ Correlations not reported</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gonzalez, et al. (2004)</td>
<td>□ No information provided</td>
<td>education of elderly hospitalised patients and the occurrence of delirium, rated by the CAM. In both studies, risk of delirium was higher among persons with fewer years of education.</td>
</tr>
</tbody>
</table>

\(^1\) In this study, the CAM adaptation included two additional strategies: two interview questions with carer and four interview questions with patients (discourse, orientation in time, orientation in space and focusing attention). To test convergent validity of the CAM against the MMSE, these four questions were used (Gonzalez, et al. 2004).
| Construct: Discriminant Validity | Inouye, et al. (1990) | X Scale differentiates between relevant categories of respondents | High sensitivity and specificity of the CAM in detecting delirium among patients with varying degrees of severity were reported. Sensitivity ranges from 94% to 100% and specificity from 90% to 95%. The Positive Predictive Value (PPV) ranges from 91% to 94% while the Negative Predictive Value (NPV) ranges from 90% to 100%.

When four screening instruments were assessed for the test characteristics in detecting delirium, against the diagnosis of delirium made by clinicians based on DSM-III-R, the CAM and the CAC (Clinical Assessment of Confusion) performed better than the performance based tools (Digit Span Test and Vigilance "A" Test). The sensitivity and the specificity of the CAM were 0.46 and 0.92 respectively.

When the areas under the receiver operating characteristic (ROC) curves (AUC) were compared between the CAM, MMSE, Clock Tests, the AUC for the CAM (0.81, 95%CI = 0.69-0.94) was found to be significantly greater than the AUC for the MMSE (0.64, 95%CI = 0.50-0.78, p = 0.28) and Clock Test (0.67, 95%CI = 0.53-0.81, p = 0.46).

In the study of ICU patients who are verbal, unrestrained and unintubated, the overall agreement between the CAM and the CAM-ICU was 82% (kappa = 64; 95%CI, 0.32-0.94).

A Spanish version of the CAM demonstrated high sensitivity (90%) and specificity (100%), with the PPV of 1.0 and the NPV of 0.97. About 50% of the total sample had dementia (N=62) in the study. When the CAM scores of the dementia group were considered against those of the non-dementia group (N=61), the dementia group showed slightly lower sensitivity (87%) than the non-dementia group (sensitivity = 93%). |
<p>| Rolfson, et al. (1999) |
| McNicoll, et al. (2005) |</p>
<table>
<thead>
<tr>
<th>Criterion</th>
<th>X Comparison made to criterion measures</th>
<th>Sensitivity rates may vary depending on the rater: nurse rated tests resulted in poor performance, sensitivity ranging between 0.13 and 0.89; research assistant rated test resulted in sensitivity of 0.46; physician rated resulted in sensitivity of 0.81 and 1.0.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information on the relationship of scores to gold standard measures or clinical diagnosis is provided</td>
<td>Studies have shown varying degrees (low to high) of sensitivity and specificity in detecting delirium using the CAM, when measured against clinical measures. Training status of the interviewer/observer implementing the CAM appears to have influenced heavily on the sensitivity and the specificity.</td>
<td>Compared to the consensus diagnosis, the clinical diagnosis by a psychiatrist had a sensitivity of 0.73 (95% confidence interval [CI]: 0.61–0.85), a specificity of 0.93 (95% CI: 0.79–1.0). The nurse clinician diagnosis had a sensitivity of 0.89 (95% CI: 0.81-0.97), a specificity of 1.00.</td>
</tr>
<tr>
<td>Schuurmans, Deschamps, et al. (2003)</td>
<td>Sensitivity rates may vary depending on the rater: nurse rated tests resulted in poor performance, sensitivity ranging between 0.13 and 0.89; research assistant rated test resulted in sensitivity of 0.46; physician rated resulted in sensitivity of 0.81 and 1.0.</td>
<td>When compared with the DSM-III-R based diagnosis of delirium, the sensitivity of the CAM was 0.70 and the specificity was 1.00 when rated by both physician and nurse. Similar results were found in a study conducted in general hospital where the nursing CAM had a sensitivity of 0.68 and a specificity of 0.97.</td>
</tr>
<tr>
<td>Zou, et al. (1998)</td>
<td>Studies have shown varying degrees (low to high) of sensitivity and specificity in detecting delirium using the CAM, when measured against clinical measures. Training status of the interviewer/observer implementing the CAM appears to have influenced heavily on the sensitivity and the specificity.</td>
<td>Based on the geriatrician's assessment, 19% of the study population (older people presented in the emergency care setting) met the CAM criteria for delirium, while 20% and 24% of the population met the DSM-IV and DSM-III-R criteria for delirium respectively, and 21% were detected for delirium based on clinical impression. Kappa statistics indicate: CAM vs. DSM-IV (0.97; 95%CI: 0.78–1.16); CAM vs. DSM-III-R (0.86; 95%CI: 0.43–1.12); and CAM vs. clinical impression (0.94; 95%CI: 0.76–1.13).</td>
</tr>
</tbody>
</table>
| Rolfson, et al. (1999) | In a cross-sectional study with blinded assessments, conducted in acute geriatric hospitals (43% of the study | }
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>The degree to which one can assign qualitative meaning to quantitative scores</td>
<td>Authors provide 2 or more types of information on interpretability</td>
<td>The CAM algorithm (four cardinal criteria) is a binary instrument, hence there is no studies presenting means and SD of scores. However, when the severity of delirium is assessed using an additive score (0-7, higher the more severe) of the four cardinal CAM features, it is possible to present means and SD. For example, in two studies of elderly hospitalised patients with medical conditions and/or dementia, the relationship between the education level and the severity of delirium was examined. No relationship between them was found in terms of mean [SD] severity of delirium for high vs. low education (in Study 1, 3.4 [1.3] vs. 3.5 [1.3], p = 0.70; in Study 2, 3.7 [1.0] vs. 4.2 [1.6], p=0.24). It is notable that the severity of delirium using the CAM has not well been validated.</td>
</tr>
<tr>
<td>Do authors provide the following:</td>
<td>Authors provide limited information to assist with interpretability</td>
<td></td>
</tr>
<tr>
<td>Presentation of means and SD of scores before and after treatment</td>
<td>No information provided</td>
<td></td>
</tr>
<tr>
<td>Comparative data on the distribution of scores in relevant subgroups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information on the relationship of scores to well-known functional measures or clinical diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information on the association between changes in scores and patients' global ratings of the magnitude of change they have experienced</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Population had dementia), the sensitivity rates of the CAM were moderate (0.81-0.86) and the specificity rates were low (0.63-0.84) when measured against all formal criteria of delirium (DSM-III, DSM-III-R, DSM-IV and ICD-10). The CAM had the highest agreement with DSM-IV (sensitivity of 0.81 and specificity of 0.84), and lowest agreement with ICD-10 (sensitivity of 0.88 and specificity of 0.63).
postoperative delirium (coefficient of determination=0.15). Age was an independent predictive factor for delirium.

There are two methods to interpret the scores of the CAM: the specific (SPEC) where delirium is detected when criteria 1 and 2, and either 3 or 4 exist; and the sensitive (SENS) the 1st criterion “acute onset and fluctuation” is modified to “acute onset or fluctuation”. When the CAM-SPEC and the CAM-SENS were used by bedside nurses based on their clinical observations only, against the CAM scores used by research nurses based on clinical observations and the 12 item MMSE (as the criterion standard), the sensitivity and the specificity of both SENS and SPEC were low (the sensitivity (23.8%) and the specificity (97.7%) of the SPEC; the sensitivity (66.7%) and the specificity (90.8%) of the SENS). For both methods NPV was high (94.8% for the SPEC and 97.5% for the SENS); and the PPV was moderate to low (41.7% for the SPEC and 33.7% for the SENS) (kappa = 0.33-0.51).

<table>
<thead>
<tr>
<th>RESPONSIVENESS</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floor and ceiling effects</td>
<td>The questionnaire fails to demonstrate a worse score in patients clinically deteriorated and an improved score in patients who clinically improved</td>
<td>□ Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Authors should provide descriptive statistics of the distribution of scores</td>
<td>□ Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score</td>
<td></td>
</tr>
<tr>
<td>Sensitivity to change</td>
<td>The ability to detect important change over time in the concept being</td>
<td>□ Hypotheses were formulated and results were in agreement</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NB: The CAM is a binary tool (scoring produces either ‘yes’ or ‘no’ indicating the presence of delirium) hence it is not applicable to test</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Cultural Applicability and Cultural Adaptations:


### Gender Appropriateness:

Appropriate for use with both genders.

### Age Appropriateness:


### Summary:

The Confusion Assessment Method (CAM) is the most widely utilised screening / diagnostic tool for delirium, in particular older people with or without dementia, in hospital care settings. It is simple, easy to master and implement, and easily accessible, with no cost involved when used non-commercially. It is increasingly well recognised amongst clinicians in Australia, as well as internationally. Its successful adaptation in various different languages and countries indicate some degree of cultural appropriateness. Whilst there exists some limitations in detecting delirium superimposed to dementia this review indicates the CAM is by far the most efficient way of screening delirium in both clinical and research contexts.
References


54, No.4, pp.685-689.


Appendix 10.2.2: Delirium Rating Scale-Revised-98

AHOC INSTRUMENT REVIEW SHEET

Title: Delirium Rating Scale-Revised-98.

Abbreviations: DRS-R-98.

Author(s) Name: Trzepacz, PT., Mittal, D, Torres, R, Kanary, K, Norton, J & Jimerson, N.

Author(s) Address: Dr Paula T. Trzepacz, MD
Lilly Corporate Center
Drop code 4133
Indianapolis
IN 46285, USA
Email: TRZEPACZ_PAULA_T@LILLY.COM

Supplied by: DRS-R-98 and DRS original versions are available in the following publications. Both are copyrighted by Paula T. Trzepacz. Permission required for reproduction through Dr Trzepacz.


Cost: Free for use by researchers working in a not-for-profit setting or for research funded by a public/not-for-profit funding source, but there is charge for use in, for example, pharmaceutical trials.

Training requirements: Training is required, in particular for those without psychiatric background, for optimal use of the DRS-R-98 / the DRS.

Purpose: To diagnose a broad range of delirium symptoms and measure the severity of those symptoms.

Administration time: Rated over a 24-hour period, actual time required to implement the tool is not specified.

Instrument Type: Clinical rating scale including a criterion-based symptom rating

Structure: The DRS-R-98 is a 16-item scale, consisting of the 13-item severity section (sleep-wake cycle disturbance, perceptual disturbances and hallucinations, delusions, lability of affect, language, thought process abnormalities, motor agitation, motor retardation, orientation, attention, short-term memory, long-term memory, and visuospatial ability) and the 3-item diagnostic section (temporal onset of symptoms, fluctuation of symptom severity and physical disorder). The DRS-R-98 also has a score sheet where additional specific characteristics observed as well as a summary of scores are noted. An earlier version, the DRS, is a 10-item rating scale. The items are temporal onset of symptoms, perceptual disturbances, hallucination type, delusions, psychomotor behavior, cognitive status during formal testing, physical disorder, sleep-wake cycle disturbance, lability of mood, and variability of symptoms.

Scoring: The DRS-R-98 assessment should be based on all available information from the patient interview, medical status examination, medical history and
tests, medical and nursing observations, family reports, etc. Each of the 13-item severity section is scored from 0 to a maximum of 3; and each of the 3-item diagnostic section is scored from 0 to a maximum either of 2 or 3. The total summed scores range between 0-46 points (includes the three diagnostic items) and a maximum severity score of 39 points, used separately for repeated measures. Higher scores indicate more severity in delirium. Whenever an item of the DRS-R-98 cannot be rated, often caused by a patient being uncooperative it can be noted and later scored midway, that is as 1.5. The DRS items also have specific descriptors that can be scored from 0 to a maximum either of 2, 3 or 4 points, depending on the item, with a maximum possible score of 32 points.

Developed for:
The DRS was developed to measure the severity of the symptoms of delirium. It is intended to be completed by a clinician with psychiatric training to more sensitively detect the range of psychiatric symptoms assessed by the scale. Revision of the DRS was intended to address some of the shortcomings of the original scale, for instance limited its usefulness in evaluating various aspects of cognitive function, measuring repeated ratings, assessing motor subtypes of delirium, and conducting broad phenomenological and longitudinal intervention research. With appropriate training to understand psychiatric phenomenology, other physicians, nurses and psychologists can also use the DRS. (For more information refer to Trzepacz (1999) and Trzepacz, et al. (2001).

Normative Data:
Not available.

Clinical Data:
The DRS-R-98 has been largely used in clinical trials to measure pharmacological effectiveness on severity of delirium. Given its short history the range of studies using the DRS-R-98 is small in its quantity. Earlier version, the DRS, has been widely used to assess patients with medical, surgical or psychiatric illness, with or without dementia, in a variety of clinical settings including geriatric psychiatric, general hospital (Trzepacz, 1999). Currently available reports containing the DRS-R-98 are categorized into the following three categories:


Applications:
The DRS-R-98 can be applied to both clinical practice and research on the phenomenology, pathophysiology and treatment of delirium among patients with or without dementia. It is recommended that the ICD or the DSM criteria be used along with the DRS-R-98 to increase sensitivity when measuring delirium superimposed to dementia. Assessment of stupor or coma is not part of the scale application. Given the substantial changes made to the revised scale, in some research settings the DRS-R-98 can be used along with the DRS, which appears to be more useful for patients coming out of stupor (Trzepacz, et al. 2001). It was originally suggested that the DRS be used in conjunction with a cognitive test such as the

---

1 Some may overlap with other categories. Some of the studies cited used the CAM as a baseline measure.
**Psychometric Criteria**

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
</table>
| **Internal consistency** | Trzepacz, et al. (2001)       | X Alpha >0.70  | High degrees of internal consistency.  
The Cronbach’s alpha coefficient in the delirium group was 0.90 for the DRS-R-98 total scale and 0.87 for the DRS-R-98 severity scale. When the effect of each item was deleted from the scale coefficients ranged from 0.88 to 0.90 for the total and 0.84 to 0.87 for the severity scale.  
The Spanish version of the DRS-R-98 had Cronbach’s alpha (r = 0.78).  
The Dutch version of the DRS-R-98 severity scale proved to have high internal consistency: Cronbach’s alpha coefficient was 0.94 (the range of the Cronbach’s alpha coefficient if one item was deleted: 0.94–0.95). |
|                        | Fonseca, et al. (2005)         | □ Marginal or inadequate internal consistency (<0.70) |
|                        | de Rooij, et al. (2006)        | □ No information found on internal consistency |
| **Test – retest**      |                              | □ ICC >.70  | Whilst the DRS-R-98 can be repeated with 24 hours; test-retest reliability is not available, which may be due to fluctuating nature of delirium. |
|                        |                              | Time intervals and confidence intervals reported |
|                        |                              | □ Marginal or inadequate internal consistency ICC<.70 |
|                        |                              | X No information found on test-rest reliability |
| **Inter– rater**       | Trzepacz, et al. (2001)       | X Agreement reported and adequate  | Excellent inter-rater reliability. Excellent ICC when administered by three trained psychiatrists. ICC for the DRS = 0.99; for the DRS-R-98 = 0.96; for the DRS-R-98 severity only = 0.99. When each combination of pairs of raters were compared for each rating scale the ICCs ranged from 0.98 to 0.99.  
The Spanish version of the DRS-R-98 showed high inter-rater reliability for the total scale and the severity scale = 0.96 each.  
The Dutch version of the DRS-R-98 |
<table>
<thead>
<tr>
<th>VALIDITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>de Jonghe, et al. (2007)</td>
<td>□ Patients and experts were involved during item selection and/or item reduction □ Patents were consulted for reading and comprehension □ No patient involvement X No information found on content validity □ There is an adequate coverage of relevant domains □ There is limited coverage of relevant domains</td>
<td>No information available, however it is assumed that the scale is based on the DSM-IV delirium criteria as it was developed to enable measuring of the severity of delirium that is not clearly described in the DSM-IV. A recent study showed fairly limited ability of the DRS-R-98 in measuring early symptoms of delirium. There is ‘no gold standard’ to detect early symptoms in the prodromal phase of delirium and further prospective study is needed to validate the DRS-R-98 severity scale.</td>
</tr>
<tr>
<td>Construct</td>
<td>Trzepacz, et al. (2001)</td>
<td>X Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used □ Limited/ inadequate construct validity reported □ No information provided</td>
<td>Adequate construct validity has been established when compared with measures mainly for cognitive function and other delirium tests. After treatment for delirium there were significant improvements in mean scores for CTD (Cognitive Test for Delirium), DRS, DRS-98 severity and CGI (Clinical Global Impression scale): the DRS-R-98 severity improved (i.e. declining scores) from a mean of 21.5 (SD 5.6) to 5.2 (SD 3.5) (t = 7.13, df = 5, p &lt; 0.001); the DRS from a mean of 18.3 (SD 3.9) to 3.5 (SD 2.1) (t = 10.6, df = 5, p &lt; 0.001); the CGI from ‘moderate/marked impairment’ to ‘much/very improved’ (t = 6.3, df = 5, p = 0.001).</td>
</tr>
<tr>
<td>Construct: Internal Structure</td>
<td>Meagher, et al. (2007) Grassi, et al. (2001) Trzepacz &amp; Dew, (1995)</td>
<td>□ No evidence provided/failed a test of dimensionality X Some evidence provided to support internal structure □ Substantial evidence provided to support internal structure</td>
<td>The relationships between the key domains of the DRS-R-98, such as cognitive and non-cognitive domains, and other criteria such as the DSM-IV and the CTD, have been reported. Factor structure of the DRS has been reported, for instance, Grass et al. (2001) suggest a three factor structure consisting of vigilance and attention disturbances, psychotic symptoms and time course and course of condition.</td>
</tr>
<tr>
<td>Construct: Correlation with other measures</td>
<td>Trzepacz, et al. (2001)</td>
<td>X Correlations with other measures are reported</td>
<td>Significant and strong correlations were reported within and between the DRS-R-98 and other relevant</td>
</tr>
</tbody>
</table>
### Comparisons made to other measures

<table>
<thead>
<tr>
<th>Study</th>
<th>Rating scales</th>
<th>Correlations not reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meagher, et al. (2007)</td>
<td>DRS-R-98 total and severity (r = 0.99, p &lt; 0.001); DRS with the DRS-R-98 total (r = 0.83, p &lt; 0.001) and the DRS-R-98 severity (r = 0.80, p &lt; 0.001); DRS-R-98 with the CTD (r = 0.01 for the total scale; r = 0.63, p = 0.001 for the severity scale); DRS-R-98 with the CGI (r = 0.62, p = 0.001 for the total scale; r = 0.61, p = 0.001 for the severity scale).</td>
<td>Correlations not reported</td>
</tr>
<tr>
<td>Fonseca, et al. (2005)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>de Jonghe, et al. (2005)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Domains of CTD items (attention, orientation, memory and comprehension) were assessed against relevant DRS-R-98 items to see the correlations, showing high correlations.

Significant correlations (P< 0.001) were found between the Spanish DRS-R-98 and other cognitive measures such as the MMSE (r = -0.67 for the severity score, r = -0.64 for the total score), the MEC (Mini Examen Cognoscitivo, the Spanish version of the MMSE) (r = -0.62 for the severity score; r = -0.59 for the total score) and the OS (Orientation Scale) (r = 0.73 for the severity score; r = 0.72 for the total score).

In a study of validating a new delirium severity measure, the Delirium-O-Meter (DOM), DOM total scores were highly related to the DRS-R-98.

### Construct: Discriminant Validity

The scale differentiates between relevant categories of respondent e.g. sick vs. well, varying degrees of severity

<table>
<thead>
<tr>
<th>Study</th>
<th>Scale differentiates between relevant categories of respondents</th>
<th>No information on discriminant validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trzepacz, et al. (2001)</td>
<td>X</td>
<td>Moderate to high sensitivity and specificity of the DRS-R-98 total and the DRS-R-98 severity scales. Using ROC analyses comparing the delirium group vs. all other diagnostic groups, sensitivity and specificity of the DRS-R-98 total scale were 92% and 95% respectively (cut-off score of 17.75); for the DRS-R-98 severity scale 92% and 93% (cut-off score of 15.25). When the delirium group was compared with the dementia group, using the best cut-off scores (17.75 for total and 15.25 for severity), 92% sensitivity for both total and severity scales, but higher specificity for the total scale (85%) than the severity scale (77%).</td>
</tr>
<tr>
<td>de Rooij, et al. (2006)</td>
<td></td>
<td>The Dutch version of the DRS-R-98 was able to differentiate patients with delirium from demented as well as from non-psychiatric patients. The MMSE score for the delirium group was significantly lower compared to the patients with dementia and the patients without any psychiatric illness, while the DRS-R-98 and the IQCODE (Dutch version of the Informant Questionnaire Cognitive Decline in the Elderly) were</td>
</tr>
</tbody>
</table>
significantly higher in delirious patients compared to comparison groups and the demented patients (p<0.05). The DRS-R-98 score did not differ among dementia or comparison groups.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Information on the relationship of scores to gold standard measures or clinical diagnosis is provided</th>
<th>Trzepacz, et al. (2001)</th>
<th>X</th>
<th>Comparison made to criterion measures □ No comparison with criterion measures provided</th>
<th>Patients with delirium were assessed before and after treatment when delirium was detected based on the DSM-IV delirium criteria, there were significant improvements in mean scores for CTD, DRS, DRS-98 severity and CGI after the treatment. In the most recent study of phenomenology of delirium in palliative care patients, 100 cases of DSM-IV based delirium were assessed using the CTD and the DSR-R-98 to test the relationship between cognitive and non-cognitive delirium symptoms.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meagher, et al. (2007)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interpretability</th>
<th>The degree to which one can assign qualitative meaning to quantitative scores</th>
<th>Trzepacz, et al. (2001)</th>
<th>X</th>
<th>Authors provide 2 or more types of information on interpretability □ Authors provide limited information to assist with interpretability □ No information provided</th>
<th>There is sufficient evidence demonstrating interpretability of the DRS-R-98 in numerous studies. Mean scores and standard deviations for each of the five diagnostic groups (delirium, dementia, schizophrenia, depression, and other) for DRS, DRS-R-98 total and severity, CTD (Cognitive Test for Delirium) and CGI (Clinical Global Impression scale) indicated a highly significant difference among the groups: e.g., the DRS-R-98 total scale (F = 47.9, df = 4.63, p &lt; 0.001); the severity scale (F = 35.0, df = 4.63, p &lt; 0.001). With pairwise comparisons: the mean DRS-R-98 total score was significantly higher in the delirium group (p &lt; 0.001) and distinguished dementia from schizophrenia and depressed group (p &lt; 0.05); the mean severity score was significantly higher in the delirium group (p &lt; 0.001) and distinguished dementia from both depression and ‘other’ groups (p &lt; 0.05). On Kruskal-Wallis comparisons, significant difference among groups was reported (p &lt; 0.001) for the DRS, DRS-R-98 total and severity scales. In a study of the Dutch version of the DRS-R-98 severity scale, within 48 hours after admission to the hospital, 54 patients with DSM-IV based delirium diagnosis consecutively were assessed for delirium subtypes (hypoactive delirium-'no hyperactive symptoms' and non hypoactive delirium-'4 hyperactive or mixed') using the DRS-R-98: Positive scores of DRS-R-98 item 4 (affect liability).</th>
</tr>
</thead>
</table>
and item 7 (motor agitation) predicted the presence of non-hypoactive delirium, with a specificity of 89% and a sensitivity of 57%.

Numerous studies have presented means and SD of scores and/or medians and ranges demonstrating significant improvements before and after pharmacological treatment.

<table>
<thead>
<tr>
<th>RESPONSIVENESS</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Floor and ceiling effects</strong></td>
<td></td>
<td>□ Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
<td>No information identified.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>X No information provided on floor and ceiling effects</td>
<td></td>
</tr>
<tr>
<td><strong>Sensitivity to change</strong></td>
<td>Pae, et al. (2004)</td>
<td>X Hypotheses were formulated and results were in agreement</td>
<td>The DRS-R-98 scores were significantly reduced after treatment from a mean of 21.8 (SD 3.2) to 9.3 (SD 3.8) (p &lt; 0.0001). A significant reduction in the CGI was also observed, from a mean of 4.9 (SD 0.8) to 2.1 (SD 1.1) (p &lt; 0.0001). Another study using the same tools to measure the effectiveness of treatment showed similar results, significant improvement of delirium using the DRS-R-98 or/and the CGI.</td>
</tr>
<tr>
<td></td>
<td>Kalisvaart, et al. (2005)</td>
<td>□ An adequate metric was used (ES, SRM, comparison with external standard)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Straker, et al. (2006)</td>
<td>□ No information on sensitivity to change was provided</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Takeuchi, et al. (2007)</td>
<td>□ MCID - Information was provided about the magnitude of score differences which would be clinically meaningful</td>
<td></td>
</tr>
<tr>
<td></td>
<td>de Jonghe, et al. (2005)</td>
<td>X MCID – No information was provided.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information on floor and ceiling effects provided</td>
<td></td>
</tr>
</tbody>
</table>
Cultural Applicability and Cultural Adaptations: The DRS-R-98 has been used successfully in various non-English background countries, including Japan (Takeuchi, et al. 2007); Korea (Pae, et al. 2004); Spain (Fonseca, et al. 2005); and the Netherlands (de Jonghe, et al. 2005; de Jonghe, et al. 2007; de Rooij, et al. 2006; Kalisvaart, et al. 2005). The Spanish (Fonseca, et al. 2005) and Dutch (de Rooij, et al. 2006) versions of the DRS-R-98 have been the subject of validation demonstrating various psychometric properties of the translated versions. The original DRS is available in 11 languages, French, Italian, Spanish, Dutch, Mandarin Chinese, Korean, Swedish, Japanese, German, and Indian-language translations, which have been successfully applied in a variety of ethnicities and countries.

Gender Appropriateness: Appropriate for use with both genders.

Age Appropriateness: Mostly used in the older population however the DRS-R-98 is not age specific (the DRS was used to evaluate delirium in the paediatric population (Turkel, Braslow, Tavare, & Trzepacz, 2003).

Summary: The DRS-R-98 is a reasonably new tool developed to enable repeated measures of severity of delirium as well as for diagnosing the syndrome, and it can be applied in both clinical and research settings. Whilst content validity of the DRS-R-98 severity scale has yet to be established, studies so far have demonstrated its high validity and reliability, including moderate sensitivity to dementia, and applicability in a variety of institutional settings among diverse groups of people with medical and/or psychiatric conditions. It is a revised version of the DRS, which has been used widely internationally in the last two decades with good reliability and validity demonstrated. Limitations, for both versions, are that they are time taxing as they are based on detailed clinical observations over a 24 hour period. They both require special training, especially for those without a psychiatric background, due to some of the psychiatric specific terminologies and descriptors used in the tool.

Reporter: Yun-Hee Jeon

Date of report: 15 May 2007

References


Appendix 10.3: Individual Symptom Measures for Associated Symptoms
Appendix 10.3.1: Rating Scale for Aggressive Behaviour in the Elderly

AHOC INSTRUMENT REVIEW SHEET

Title: Rating Scale for Aggressive Behaviour in the Elderly.

Abbreviations: RAGE.

Author(s) Name: Patel V and Hope RA.

Author(s) Address: Institute of Psychiatry, London. SES AZ, UK and Department of Psychiatry, Warneford Hospital, Oxford, England.


Cost: Nil.

Training requirements: Brief training and written instructions are recommended to assist staff in understanding how to record responses, and in particular, to remind staff of the need to assess aggressive behaviours objectively and not overlook “minor” behaviours that signify aggression. The questions can be answered by anyone who provides the day to day care for the person with dementia, such as informal and formal caregivers and nurses.

Purpose: To identify and measure aggressive behaviour within psychogeriatric populations.

Administration time: 5 minutes.

Instrument Type: Clinical rating scale: assessment of aggressive behaviour within psychogeriatric populations.

Structure: The scale consists of 21 items: 19 items assess specific kinds of aggressive behaviour, including: demanding/argumentative; shouting/screaming; swearing/abusive; disobeying ward rules; un-cooperativeness/resistant to help; bad mood/irritability; critical/derogatory; impatient; threatening harm; antisocial; pushing/shoving; throwing things; anger with self; attempt to kick; attempt to hit; attempt to bite; lashing out; injury to others. Item 20 assesses the consequences of staff response to the aggressive behaviour; and item 21 asks the rater to make an overall assessment of aggressive behaviour using a 4-point scale.

Scoring: The original scoring represents the rater’s perceptions/estimations over a three-day period. For 20 items, the incidences and consequences of aggression are rated as:

0 – not once in the past three days (never)
1 – at least once in the past three days (occasionally)
2 – at least once every day in the past three days (often)
3 – more than once every day in the past three days (frequently)

Each point is thus precisely defined and the points are in sequential progression.

Shah, et al. (1997, 1998) found RAGE can also be employed to score the incidence and severity of aggression over the preceding week with little change in scores compared to a three day observation period.

Developed for: The RAGE was initially developed so nurses could measure aggressive
behaviour in in-patient psychogeriatric populations. The developers also aimed to employ RAGE to enable research to be carried out on the effects of potential treatments of aggressive behaviour, and on the relationships between aggressive behaviours and other factors.

**Normative Data:**

There are no reports of use with persons without cognitive impairment or mental illness, however, data was obtained from a number of patients with different forms of cognitive impairment (schizophrenia and intellectual disability) over 7 and 14 days (Patel & Hope, 1992) and over 18 weeks (Shah, 1998a) for different forms and levels of aggression. This was also done for the Chinese version (C-RAGE) by Lam, Chui & Ng (1997) where C-RAGE scores were obtained from populations with different forms of cognitive impairment in different health contexts. In treatment studies data has been obtained from controls.

**Clinical Data:**

Clinical data was obtained from 13 patients with dementia displaying aggressive behaviours. The RAGE has also been administered to community dwelling clients with moderate to high levels of aggression associated with various forms of dementia (Shah, A., Evans, H., Parkash, N. 1998). Data was also obtained from dementia patients 6 hours, 7 days and 14 days after their first assessment (Patel & Hope, 1992).

**Applications:**

The RAGE is commonly used by family caregivers and/or health staff who frequently care for the person with dementia to identify frequencies, types and severity of aggressive behaviours, patterns in aggression occurring over the previous three days, the use of restraint to manage aggressive behaviour, and to rate overall assessment of the frequency of aggressive behaviours (Patel, et al.1995; Lam, et al. 1997; Shah, et al. 1998; 2000). The RAGE is also employed in research to identify baseline behaviours and also in drug and treatment/therapy trials (Shah, et al. 2000; Gormley, Lyons. & Howard, 2001).

**Carer and/or Patient Use of Instrument:**

The RAGE is intended for use by regular caregivers, including nurses, or by reference to the experiences of regular caregivers. Regular caregivers are chosen to convey information on frequency, types and severity of aggressive behaviours because the person with dementia may not exhibit aggression during a single assessment or over a shorter assessment period. Thus, aggressive episodes would be underestimated if the ratings were based on observations occurring per nurse shift, or on separate days (Patel & Hope, 1992a; Patel & Hope, 1992b). It is not suitable for use by the person with dementia as they are unlikely to acknowledge aggressive behaviours, and be able to recall and thus, estimate the frequency of aggressive symptoms over the past three days.

**Psychometric Criteria**

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td>Shah, Evans, Parkash (1998)</td>
<td>X Alpha &gt;0.70</td>
<td>Internal consistency as measured by Cronbach’s alpha for the RAGE was 0.89 when rated with 13 patients with dementia displaying aggressive behaviours. RAGE subscales compare highly with relevant subscales of the CMAI (rho = 0.73, p = 0.005) and the BARS (rho = 0.72, p = 0.006).</td>
</tr>
<tr>
<td>The extent to which items in a (sub) scale are inter-correlated; a measure of the homogeneity of a (sub)scale.</td>
<td></td>
<td>□ Marginal or inadequate internal consistency (&lt;0.70)</td>
<td></td>
</tr>
<tr>
<td>Cronbach’s alpha should be between 0.70 and 0.90 for every dimension / sub-scale.</td>
<td></td>
<td>□ No information found on internal consistency</td>
<td></td>
</tr>
<tr>
<td>Test – retest</td>
<td>Lam, Chui, Ng (1997)</td>
<td>Internal consistency of the Chinese translation of the RAGE (C-RAGE) is high with an alpha coefficient of 0.74. The split half reliability of the C-RAGE is 0.79.</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------------------------</td>
<td>----------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Test-retest reliability was established by conducting three rounds of RAGE assessments with 90 patients after three days of observation: 6 hours after their first assessment; 7 days after their initial assessment and then 14 days after the initial assessment. Test-retest scores of all items were over 0.75.</td>
<td>Patel &amp; Hope (1992a)</td>
<td>□ Marginal or inadequate internal consistency ICC&lt;.70 □ No information found on test-retest reliability</td>
<td></td>
</tr>
<tr>
<td>Patel &amp; Hope (1992b)</td>
<td>X ICC &gt;.70</td>
<td>□ Marginal or inadequate internal consistency ICC&lt;.70 □ No information found on test-retest reliability</td>
<td></td>
</tr>
<tr>
<td>Patel, Hope, Hall, Fairburn (1995)</td>
<td></td>
<td>□ Marginal or inadequate internal consistency ICC&lt;.70 □ No information found on test-retest reliability</td>
<td></td>
</tr>
<tr>
<td>Shah, Evans, Parkash (1998)</td>
<td></td>
<td>□ Marginal or inadequate internal consistency ICC&lt;.70 □ No information found on test-retest reliability</td>
<td></td>
</tr>
<tr>
<td>Shah (1999)</td>
<td></td>
<td>□ Marginal or inadequate internal consistency ICC&lt;.70 □ No information found on test-retest reliability</td>
<td></td>
</tr>
<tr>
<td>Lam, Chui, Ng (1997)</td>
<td></td>
<td>□ Marginal or inadequate internal consistency ICC&lt;.70 □ No information found on test-retest reliability</td>
<td></td>
</tr>
<tr>
<td>Test-retest reliability in a psychogeriatric ward was found to be at least 0.75 for most items, and 0.94 for the total score ($p&lt;0.00001$).</td>
<td></td>
<td>□ Marginal or inadequate internal consistency ICC&lt;.70 □ No information found on test-retest reliability</td>
<td></td>
</tr>
<tr>
<td>The total RAGE score and all its sub-item scores do not spontaneously decline with repeated testing at different points in time (3 days and 7 days).</td>
<td></td>
<td>□ Marginal or inadequate internal consistency ICC&lt;.70 □ No information found on test-retest reliability</td>
<td></td>
</tr>
<tr>
<td>Three week test-retest reliabilities of the Chinese translation of RAGE (C-RAGE) total score ($r = 0.83, p = 0.01$) and the global aggressiveness score ($r = 0.92, p = 0.001$) were high.</td>
<td></td>
<td>□ Marginal or inadequate internal consistency ICC&lt;.70 □ No information found on test-retest reliability</td>
<td></td>
</tr>
<tr>
<td>Inter – rater</td>
<td>Patel &amp; Hope (1992a)</td>
<td>□ Agreement reported and adequate □ Inadequate inter-rater agreement □ No information provided</td>
<td></td>
</tr>
<tr>
<td>Patel &amp; Hope (1992b)</td>
<td></td>
<td>□ Agreement reported and adequate □ Inadequate inter-rater agreement □ No information provided</td>
<td></td>
</tr>
<tr>
<td>Patel, Hope, Hall, Fairburn (1995)</td>
<td></td>
<td>□ Agreement reported and adequate □ Inadequate inter-rater agreement □ No information provided</td>
<td></td>
</tr>
<tr>
<td>Shah, Evans, Parkash (1998)</td>
<td></td>
<td>□ Agreement reported and adequate □ Inadequate inter-rater agreement □ No information provided</td>
<td></td>
</tr>
<tr>
<td>Shah (1999)</td>
<td></td>
<td>□ Agreement reported and adequate □ Inadequate inter-rater agreement □ No information provided</td>
<td></td>
</tr>
<tr>
<td>Lam, Chui, Ng (1997)</td>
<td></td>
<td>□ Agreement reported and adequate □ Inadequate inter-rater agreement □ No information provided</td>
<td></td>
</tr>
<tr>
<td>50 residents were rated by nurses with access to a checklist of usual aggressive behaviours occurring in the wards where they worked, and 40 residents were assessed by nurses who had no checklist. The correlation was 0.94 ($p&lt;0.001$) when using a checklist and 0.54 without a checklist. The values for the individual items of weighted Kappa for inter-rater reliability ranged from 0.61 and 0.92.</td>
<td></td>
<td>□ Agreement reported and adequate □ Inadequate inter-rater agreement □ No information provided</td>
<td></td>
</tr>
<tr>
<td>RAGE assessments were</td>
<td></td>
<td>□ Agreement reported and adequate □ Inadequate inter-rater agreement □ No information provided</td>
<td></td>
</tr>
<tr>
<td>VALIDITY</td>
<td>Studies Reported &amp; References</td>
<td>Adequacy Checks</td>
<td>Comment</td>
</tr>
<tr>
<td>----------------</td>
<td>---------------------------------</td>
<td>----------------</td>
<td>---------</td>
</tr>
<tr>
<td><strong>Content</strong></td>
<td>The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire</td>
<td>Patel &amp; Hope (1992a) Patel &amp; Hope (1992b) Patel, Hope, Hall, Fairburn (1995) Shah, Evans, Parkash (1998)</td>
<td>X Patients and experts were involved during item selection and/or item reduction □ Patients were consulted for reading and comprehension No patient involvement □ No information found on content validity □ There is an adequate coverage of relevant domains There is limited coverage of relevant domains</td>
</tr>
</tbody>
</table>
| Construct | Shah, Evans, Parkash (1998) | X Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used  
□ Limited /inadequate construct validity reported  
□ No information provided | When rated with 13 patients with dementia displaying aggressive behaviours, the RAGE subscales compare highly against relevant subscales of the CMAI (rho = +0.73, p =0.005) and the Brief Agitation Rating Scale BARS (rho = 0.72, p = 0.006). RAGE scores correlate closely with the aggressiveness subscale scores of the Behavioural Pathology in Alzheimer’s Disease Scale (BEHAVE-AD) in a predominantly community-based sample.  
There is a strong and significant correlation on the total and sub-scale scores of RAGE and Staff Observation Aggression Scale (SOAS) (r = +0.81, p = 0.00001). |
X Some evidence provided to support internal structure  
□ Substantial evidence provided to support internal structure | Comparisons of a number of study results group RAGE items into three factors: verbal aggression, physical aggression and anti-social behaviour, reflecting the most commonly occurring aggressive behaviours observed in persons with dementia by family caregivers, care staff and clinicians.  
Correlation coefficients were computed for all items with the total scores. All items showed correlations of 0.4 to 0.6, signifying the strength of how each item correlated with the construct of aggression. The intra-class correlation coefficient was 0.89, confirming content validity. |
### Construct: Correlation with other measures

**Comparisons made to other measures**

- Lam LCW, Chui HFK, Ng (1997)
- Oye, Loke, Chan, Kwok (2005)
- Bathareethan & Shah (2000)

**X Correlations with other measures are reported**

- Correlations not reported

When validating the Chinese version of the RAGE with the Chinese version of the MMSE, C-RAGE aggression scores were highly correlated with C-MMSE scores of 10-15.

Construct validity of the Chinese version of RAGE (C-RAGE) is established and the impact on caregiver items of C-RAGE correlate highly with relevant items on the Chinese version of the Cost of Care Index ($r[s] = 0.26, p <0.01$).

There is a high correlation coefficient between the Checklist of regularly occurring aggressive behaviours in persons with dementia constructed by dementia care staff, and the C-RAGE total score ($r = 0.84, p = 0.001$), confirming content validity of the scale.

There is a strong correlation between RAGE total score and the Modified Alienation Scale (MAS) total score ($r = +0.67, p = 0.00001$) and 4 of its sub-scales: feeling distant from the patient ($r = +0.44$); deliberate use of symptoms by the patient ($r = +0.36$); alienation of the patient ($r = +0.85$); and variable mood of the patient ($r = +0.77$). The RAGE total score also had moderate to high correlations with the MAS sub-items.

### Construct: Discriminant Validity

**The scale differentiates between relevant categories of respondent e.g. sick vs. well, varying degrees of severity**

- Patel & Hope (1992a)
- Patel & Hope (1992b)
- Shah & De (1998a)
- Shah, Evans, Parkash (1998)
- Lam, Chui, Ng (1997)

**X Scale differentiates between relevant categories of respondents**

- No information on discriminant validity

The RAGE differentiates between different categories of patients with impaired cognition (such as schizophrenia, intellectual disability and dementia) over 7 and 14 days (Patel & Hope 1992) and over 18 weeks (Shah 1998a) for different forms and levels of aggression.

The RAGE differentiates between acute in-patient populations, nursing home residents and community dwelling clients with moderate to high levels of aggression associated with various forms of dementia.

C-RAGE scores vary in populations with different forms of cognitive impairment in different health contexts. Staff rated patients of long-
stay wards with schizophrenia as more uncooperative to ward rules, and residents of nursing home with dementia rated as more irritable. Lam et al. suggested that Chinese care staff’s greater tolerance towards aggressiveness in persons with dementia, in comparison with aggression in patients with schizophrenia, may explain this finding.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Information on the relationship of scores to gold standard measures or clinical diagnosis is provided</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Shah, Evans, Parkash (1998) X Comparison made to criterion measures □ No comparison with criterion measures provided</td>
</tr>
<tr>
<td></td>
<td>Shah, et al. (2000)</td>
</tr>
<tr>
<td></td>
<td>Gormley, Rizwan, Lovestone (1998)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interpretability</th>
<th>The degree to which one can assign qualitative meaning to quantitative scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Do authors provide the following:</td>
</tr>
<tr>
<td></td>
<td>Presentation of means and SD of scores before and after treatment</td>
</tr>
<tr>
<td></td>
<td>Comparative data on the distribution of scores in relevant subgroups</td>
</tr>
<tr>
<td></td>
<td>Information on the relationship of scores to well-known functional measures or clinical diagnosis</td>
</tr>
<tr>
<td></td>
<td>Information on the association between changes in scores and patients’ global</td>
</tr>
</tbody>
</table>

While no gold standard measure exists for assessing aggressive behaviours in the population with dementia, the RAGE tests favourably against other validated and often-used measures of aggression in dementia, and with clinicians’ diagnosis of aggression.

RAGE subscales compare highly with relevant subscales of the Cohen Mansfield Agitation Inventory (CMAI) and the Brief Agitation Rating Scale (BARS).

There is a strong and significant correlation on the total and sub-scale scores of RAGE and Staff Observation Aggression Scale (SOAS).

RAGE subscales correlate closely with the aggressiveness subscale of the Behavioural Pathology in Alzheimer's Disease Scale (BEHAVE-AD).

In a number of studies, means, standard deviations and a range of RAGE scores are reported, and commonly in comparison to other validated measures of aggression in dementia. The aggression incidences recorded on the RAGE concur with checklists and clinical notes of aggressive behaviours in persons with dementia that staff routinely observe and report in psychogeriatric wards and nursing homes. The individual and total scores also concur with expert clinician assessment of aggression using case history, psychological examination, and clinical observation. Clinical trials of drug and other treatments/therapies reveal strong correlations between RAGE scores and treatment.
ratings of the magnitude of change they have experienced  

<table>
<thead>
<tr>
<th>RESPONSIVENESS</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floor and ceiling effects</td>
<td>Lam, Chui, Ng (1997)</td>
<td>X Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected □ Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score □ No information provided on floor and ceiling effects</td>
<td>Only one study identified the floor and ceiling effects of RAGE, and there were no floor or ceiling effects found in C-RAGE.</td>
</tr>
<tr>
<td>Sensitivity to change</td>
<td>Patel &amp; Hope (1992a) Patel &amp; Hope (1992b) Shah &amp; De (1998a) Shah, Evans, Parkash (1998) Patel &amp; Hope (1992a) Patel &amp; Hope (1992b) Gormley, Lyons, Howard (2001)</td>
<td>X Hypotheses were formulated and results were in agreement X An adequate metric was used (ES, SRM, comparison with external standard) □ No information on sensitivity to change was provided X MCID - Information was provided about the magnitude of score differences which would be clinically meaningful □ MCID – No information was provided.</td>
<td>The RAGE is sensitive to change over 7 and 14 days (Patel &amp; Hope 1992) and over 18 weeks (Shah 1998a) for different forms and levels of aggression in patients with different reasons for cognitive impairment, such as schizophrenia, intellectual disability and dementia. RAGE is sensitive to change in persons with dementia in acute in-patient populations, in nursing home residents and in community dwelling clients who exhibit moderate to high levels of aggression. For 14 subjects independently rated to have shown a decrease in aggressive behaviours, mean RAGE scores for all items fell from 17.8 to 6.5 (p &lt; 0.05). For 7 subjects independently rated to have shown an increase in aggressive behaviours, mean RAGE scores for all items rose (insignificantly) from 6.7 to 16 (p &gt;0.05). In a Randomised Controlled Trial of a behaviour management program for aggressive behaviours in 62 patients with dementia, a reduction in RAGE scores</td>
</tr>
</tbody>
</table>
corresponded with a reduction in aggressive behaviours in the treatment group.

Cultural Applicability and Cultural Adaptations: The RAGE is employed largely in English speaking countries, such as the USA, Canada, and UK. Cross-cultural reliability and validity has been established for the Chinese population (Lam, Chui & Ng, 1997) and in Scandinavia (Patel & Hope, 1992b). Translations of the English version of the RAGE are reliable and valid (Lam, Chiu & Ng, 1996; Patel & Hope, 1992b).

Gender Appropriateness: Appropriate for use with both genders.

Age Appropriateness: Generally RAGE is used with elderly persons 65 years and over but no age limitation is specified by the authors.

Summary: The Rating Scale for Aggressive Behaviour in the Elderly (RAGE) addresses the need to identify the presence and severity of aggression in dementia, given it is one of the seven most common behavioural symptoms occurring worldwide (Alzheimer’s Disease International, 2003). Aggression was defined by the instrument developers, Patel and Hope, as “an overt act involving the delivery of noxious stimuli to (but not necessarily aimed at) another organism, object or self which is clearly not accidental”.

RAGE was specifically designed to be completed by nurses and care staff to record aggression. However, the questions can be answered by anyone who provides the day to day care for the person with dementia, such as informal and formal caregivers and nurses, to assist in care planning and treatment regimens. The original RAGE measured aggressive behaviour for the preceding three days. However, this was adapted for use over the preceding week because the original reliability study conducted by Patel and Hope (1992) reported little change over 7 days and 14 days in early studies in nursing homes, and in acute and continuing care psycho-geriatric wards (Shah and De, 1998; Shah, et al. 1998). The RAGE has been extensively tested, is sensitive to change and is suitable as an outcome measure in treatment studies.

It is a reliable and valid measure of aggression not only in dementia, but also in mental illness and intellectual disability, in all different care contexts and in the community. Test-retest and inter-rater reliability studies show high agreement when RAGE and C-RAGE are used by clinicians, care staff and researchers. None of the studies reviewed identified any flaws in the psychometric properties of the instrument and it compares highly favourably with other widely-used and validated measures of aggression in dementia, such as the Cohen-Mansfield Agitation Inventory (CMAI), the Brief Agitation Rating Scale (BARS), and the Behavioural Pathology in Alzheimer’s Disease Scale (BEHAVE-AD).

The Chinese translated version (C-RAGE) has been validated in the Chinese population, although the scores for two items never reached the levels reported in Caucasian populations. When validating the C-RAGE with the Chinese version of the MMSE, Lam, Chiu and Ng (1997), found that the C-RAGE aggression scores were highly correlated with C-MMSE scores of 10-15. On the other hand, in an Australian study reported by Martin, McKenzie and Ames (1994) investigating disturbed behaviour in nursing home clients with dementia, no correlation was identified between the incidence and level of aggressive behaviour as identified on the RAGE and level of cognitive impairment or performance. Nor was there an association between RAGE scores of aggressive behaviour and environmental characteristics (Shah, Chiu and Ames, 2000).
The most salient features of RAGE as a measure of aggression in dementia are that it is valid and reliable in English and some other languages; it compares favourably with other validated measures of aggression; it focuses specifically on aggression incidence and severity; it is accessible to caregivers and care staff; it can be relied on in clinical intervention studies; and it captures information about use of restraint associated with aggression, which is important in care planning and care monitoring.

**References**


Appendix 10.3.2: Cohen Mansfield Agitation Inventory – Long Form

AHOC INSTRUMENT REVIEW SHEET

Title: Cohen Mansfield Agitation Inventory – Long Form.

Abbreviations: CMAI.

Author(s) Name: J Cohen-Mansfield, MS Marx and AS Rosenthal.

Author(s) Address: Jiska Cohen-Mansfield
Director of Research
Hebrew Home
Greater Washington
6121 Montrose Road
Rockville MD 20852
USA.

Supplied by: The authors.

Cost: Free with authors permission.

Training requirements: Training is recommended. An instruction manual and a training video are available from the authors. The manual is free.

Purpose: To assess the frequency of manifestations of agitated behaviours in elderly persons.

Administration time: 10 – 15 minutes.

Instrument Type: Clinical rating questionnaire completed by caregiver.

Structure: The instrument comprises 29 items describing agitated behaviours that can be summarised into three domains or factors: aggressive behaviour, physically nonaggressive behaviour, and verbally aggressive behaviour. Each item is rated on a 7 point scale based on the frequency with which the person has engaged in the behaviour in the previous two weeks. The CMAI is used in the nursing home population.

Several other versions of the instrument are also available. The 14 item (CMAI-Short), also used in the nursing home population, has the same domains as the CMAI and is rated on a 5 point scale. The 37 item CMAI-Community (CMAI-C) is available for use in the community and can be used by both professional and family caregivers. This instrument can be summarised into 4 domains or factors: physically nonaggressive behaviour, physically aggressive behaviour, verbally nonaggressive behaviour and verbally aggressive behaviour. It consists of 36 frequency items (rated on a 7 point scale) and one item to determine the time of day the behaviour occurred. In addition to these versions there is also the Long form with expanded definitions (provides additional examples of each behaviour) and the disruptiveness form in which the disruptiveness of the behaviour is rated along with the frequency. This is available in the Long (i.e. the normal 29 item version), Short and Community forms.

Scoring: Ratings for the CMAI and CMAI-C range from 1 (never) to 7 (several time an hour). Ratings for the CMAI-Short range from 1 (never) to 5 (a few times an hour or continuous for half an hour or more). Item scores can be summed to give a total score of 29-203 for the CMAI, 14-56 for the CMAI-Short, and 36-216 for the CMAI-
C. Alternatively, or in addition, items relating to specific behaviours of interest only, or items relating to each of the domains can be summed. Regardless of the scoring used, higher scores indicate greater agitation or behavioural disruption.

Developed for: The CMIAI is used in clinical and research settings to investigate agitated behaviour in persons with cognitive impairment, both in the nursing home setting, and in the community.

Normative Data: There is no normative data currently available for the CMIAI.


**Applications:**

Common applications include studies evaluating the effectiveness of drug treatments or other interventions, and research studies evaluating agitated behaviour in persons with dementia or cognitive impairment.

**Carer and/or Patient Use of Instrument:**

The instrument is administered by a carer, familiar with the patient, such as a nursing home staff member, family caregiver, social worker, or activity director of a day care centre. Ratings are based on observations of the patient during the previous two week period.

### Psychometric Criteria

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal consistency</strong></td>
<td>The extent to which items in a (sub) scale are inter-correlated; a measure of the homogeneity of a (sub)scale.</td>
<td></td>
<td>The CMAI has excellent internal consistency with α for the total scale ranging from 0.75 to 0.91 and for the subscales. Cronbach’s alpha for the sub scales has generally also been above 0.70 except for 2 studies that reported alphas ranging from 0.62 - 0.78 for Physically Non Aggressive Behaviour, and 0.59 - 0.78 for Verbally Agitated Behaviour. In summary ranges for alpha were: Aggressive Behaviour α = 0.81 – 0.82. Physically Nonaggressive Behaviour α = 0.67 to 0.78. Verbally Agitated Behaviour α = 0.59 to 0.78. Internal consistency for the CMAI-C and CMAI-Short has not been reported.</td>
</tr>
<tr>
<td><strong>Test – retest</strong></td>
<td>The extent to which the same results are obtained on repeated administrations of the same questionnaire when no change in physical functioning has occurred Calculation of an intraclass correlation coefficient (ICC); and an ICC &gt; 0.70 is desired</td>
<td>X ICC &gt; .70 Time intervals and confidence intervals reported □ Marginal or inadequate internal consistency ICC &lt; .70 □ No information found on test-retest reliability</td>
<td>Test-retest is excellent with correlations ranging from 0.79 to 0.97 for the CMAI, and 0.83 for the CMAI-C. Test-retest for the CMAI-Short has not been reported.</td>
</tr>
</tbody>
</table>
Prefered if time interval and confidence intervals were presented

<table>
<thead>
<tr>
<th>Inter – rater</th>
<th>Suh (2004)</th>
<th>X Agreement reported and adequate □ Inadequate inter-rater agreement □ No information provided</th>
<th>Excellent inter-rater reliability with correlations ranging from 0.76 to 0.96 for CMAI, 0.71 to 0.92 for CMAI-C and 0.82 to 0.92 for CMAI-Short.</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>VALIDITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>Cohen-Mansfield (1986) Cohen-Mansfield, Marx &amp; Rosenthal (1989) Werner, Cohen-Mansfield, Koroknay &amp; Braun (1994) Cohen-Mansfield, Werner, Watson &amp; Pasis (1995)</td>
<td>□ Patients and experts were involved during item selection and/or item reduction □ Patients were consulted for reading and comprehension X No patient involvement □ No information found on content validity X There is an adequate coverage of</td>
<td>The CMAI was developed by a team of experts in the field based on information obtained from interviews with nursing home staff members and a review of the literature. The CMAI-Short and CMAI-C were adapted from this. Patients were involved as participants in validation studies. Relevant domains are</td>
</tr>
</tbody>
</table>

The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire
<table>
<thead>
<tr>
<th>Construct: Internal Structure</th>
<th>Relevant Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Construct</strong></td>
<td></td>
</tr>
<tr>
<td>The extent to which scores on the questionnaire relate to other measures in a manner that is consistent with theoretically derived hypothesis concerning the domains that are measured.</td>
<td>□ There is limited coverage of relevant domains</td>
</tr>
<tr>
<td>Cohen-Mansfield (1986)</td>
<td>X Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used</td>
</tr>
<tr>
<td>Weiner, Tractenberg, Teri, Logsdon, et al. (2000)</td>
<td>□ No information provided</td>
</tr>
<tr>
<td>Villanueva, Smith, Erickson, Lee, et al. (2003)</td>
<td>The CMAI shows expected correlations with the Rapid Disability Rating Scale (RDRS) and the Pain Assessment for the Dementing Elderly (PADE).</td>
</tr>
<tr>
<td>O'Leary, Jyringi &amp; Sedler (2005)</td>
<td>It is also associated with the presence of psychotic symptoms such as delusions, and paranoia as determined by nurses' ratings.</td>
</tr>
<tr>
<td></td>
<td>The CMAI-C is associated with the presence of delusions and hallucinations.</td>
</tr>
<tr>
<td></td>
<td>Substantial evidence provided to support internal structure</td>
</tr>
<tr>
<td></td>
<td>In the main, studies have continually confirmed a three factor structure (Aggressive Behaviour, Physically Nonaggressive Behaviour, and Verbally Agitated Behaviour) for the CMAI and CMAI-Short.</td>
</tr>
<tr>
<td></td>
<td>Some studies, however, have reported a 4 factor solution that included Hiding and Hoarding (Schreiner, et al. 2000; Suh, et al. 2004 and Rabinowitz, et al. 2005). One Dutch study (Zuidema et al 2007) reports both a restricted 3 factor solution (as reported in other studies) and an unrestricted 6 factor solution (Aggressive Behaviour, Physically Nonaggressive Behaviour, Verbally Agitated Behaviour, Hiding and Hoarding, Vocal Agitation and a Miscellaneous Items factor (repetitious mannerisms, spitting).</td>
</tr>
<tr>
<td></td>
<td>Findings for CMAI-C are mixed. Cohen-Mansfield, et al. (1991) suggests a four factor solution, (Physically Nonaggressive Behaviour, Physically Aggressive Behaviour, Verbally Bonaggressive Behaviour, and Verbally Aggressive Behaviour). Cohen-Mansfield, et al. (1995) suggest both a three (as for the CMAI) and a four factor solution may be appropriate. Weiner, et al. (2002) stated that the 4 factor solution was not appropriate and that a total score was best suited to describe behaviour.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Construct: Internal Structure</th>
<th>Internal Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information provided on factor structure</td>
<td>□ No evidence provided/failed a test of dimensionality</td>
</tr>
<tr>
<td>Cohen-Mansfield (1986)</td>
<td>X Some evidence provided to support internal structure</td>
</tr>
<tr>
<td>Cohen-Mansfield, Marx &amp; Rosenthal (1989)</td>
<td>□ Substantial evidence provided to support internal structure</td>
</tr>
<tr>
<td>Cohen-Mansfield (1991)</td>
<td></td>
</tr>
<tr>
<td>Miller, Snowdon &amp; Vaughan (1995)</td>
<td></td>
</tr>
<tr>
<td>de Jonghe &amp; Kat (1996)</td>
<td></td>
</tr>
<tr>
<td>Schreiner, Yamamoto &amp; Shiotani (2000)</td>
<td></td>
</tr>
<tr>
<td>Choy, Lam, Chan, Li, et al. (2001)</td>
<td></td>
</tr>
<tr>
<td>Vespa, Gori, Bonaiuto, Cruciani, et al. (2002)</td>
<td></td>
</tr>
<tr>
<td>Suh (2004)</td>
<td></td>
</tr>
<tr>
<td>O'Leary, Jyringi &amp; Sedler (2005)</td>
<td></td>
</tr>
<tr>
<td>Rabinowitz, Davidson, De Deyn, Katz, et al. (2005)</td>
<td></td>
</tr>
<tr>
<td>Zuidema, de Jonghe, Verhey &amp; Koopmans (2007)</td>
<td></td>
</tr>
</tbody>
</table>
### Construct: Correlation with other measures

Comparisons made to other measures

<table>
<thead>
<tr>
<th>Study</th>
<th>Correlations with other measures are reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finkel, Lyons &amp; Anderson (1992)</td>
<td></td>
</tr>
<tr>
<td>Miller, Snowden &amp; Vaughan (1995)</td>
<td></td>
</tr>
<tr>
<td>de Jonghe &amp; Kat (1996)</td>
<td></td>
</tr>
<tr>
<td>Weiner, Williams &amp; Risser (1997)</td>
<td></td>
</tr>
<tr>
<td>Weiner, Koss, Patterson, Jin, et al. (1998)</td>
<td></td>
</tr>
<tr>
<td>Ramadan &amp; Naughton (1999)</td>
<td></td>
</tr>
<tr>
<td>Volicer, Camberg, Hurley, Ashley, et al. (1999)</td>
<td></td>
</tr>
<tr>
<td>Choy, Lam, Chan, Li, et al. (2001)</td>
<td></td>
</tr>
<tr>
<td>Suh, 2004</td>
<td></td>
</tr>
<tr>
<td>(Nagels, Engelborghs, Vloeberghs, Van Dam, et al. (2006))</td>
<td></td>
</tr>
</tbody>
</table>

X Correlations with other measures are reported

CMAI shows expected correlations with the following well known measures of agitation and aggression:
- Agitated Behaviors Mapping Instrument (ABMI);
- Brief Agitation Rating Scale (BARS);
- Rating Scale for Aggressive behaviour in the Elderly (RAGE).

It also correlated with other indicators of agitation such as: daily verbalisation scores (VS) and actigraphic recordings made using an octagonal motionlogger.

Correlations with the following global measures of behavioural disturbance are also reported:
- Behavioral and Emotional Activities Manifested in Dementia (BEAM-D);
- Behavioral Pathology in Alzheimer’s Disease Scale (BEHAVE-AD);
- Behavioral Syndrome Scale for Dementia (BSSD);
- Dutch Behaviour Rating Scale for Psychogeriatric Inpatients (GIP);
- Nursing Home Problem Behaviour Scale (NHPBS);
- Revised Memory and Behavior Problems Checklist (RMBPC).

The CMAI-C showed expected correlations with Agitated Behaviour in Dementia (ABID) and the global scale, Consortium to establish a Registry for Alzheimer’s Disease – Behavior Rating Scale dementia (CERAD-BRSD).

The CMAI-Short correlated with an agitation indicator: Visual analogue scale (VAS) – agitation.

### Construct: Discriminant Validity

The scale differentiates between relevant categories of respondent e.g. sick vs. well, varying degrees of severity

<table>
<thead>
<tr>
<th>Study</th>
<th>Scale differentiates between relevant categories of respondents</th>
<th>No information on discriminant validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ponce, Molinari, Kunik, Orengo, et al. (1998)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weiner, Koss, Patterson, Jin, et al. (1998)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O’Leary, Jyringi &amp; Sedler (2005)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schreiner, Ballaban,</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

X Scale differentiates between relevant categories of respondents

Both the CMAI and CMAI-C have been shown to discriminate between different levels of dementia severity.
<table>
<thead>
<tr>
<th><strong>Criterion</strong></th>
<th>Brannath, Lang, et al. (2005)</th>
<th>□ Comparison made to criterion measures</th>
<th>X No comparison with criterion measures provided</th>
<th>No studies found.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information on the relationship of scores to gold standard measures or clinical diagnosis is provided</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Interpretability</strong></th>
<th>Cohen-Mansfield (1986)</th>
<th>X Authors provide 2 or more types of information on interpretability</th>
<th>□ Authors provide limited information to assist with interpretability</th>
<th>□ No information provided</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>The degree to which one can assign qualitative meaning to quantitative scores</td>
<td>Cohen-Mansfield, Marx &amp; Rosenthal (1989)</td>
<td>Most studies include at least means, standard deviations and or range. Many provide comparative data in relevant subgroups, and or relationship to other measures. Some provide information about the association between the change in score and global ratings.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do authors provide the following:</td>
<td>Cohen-Mansfield, Werner, Watson &amp; Pasis (1995)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presentation of means and SD of scores before and after treatment</td>
<td>Miller, Snowdon &amp; Vaughan (1995)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparative data on the distribution of scores in relevant subgroups</td>
<td>Buettner, Lundegren, Lago, Farrell, et al. (1996)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information on the relationship of scores to well-known functional measures or clinical diagnosis</td>
<td>de Jonghe &amp; Kat (1996)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information on the association between changes in scores and patients’ global ratings of the magnitude of change they have experienced</td>
<td>Koss, Weiner, Ernesto, Cohen-Mansfield, et al. (1997)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>McGee, Orengo, Kunik, Molinari, et al. (1997)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ponce, Molinari, Kunik, Orengo, et al. (1998)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weiner, Koss, Patterson, Jin, et al. (1998)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kunik, Graham, Snow-Turek, Molinari, et al. (1998)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Weiner, Koss, Patterson, Jin, et al. (1998)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ramadan &amp; Naughton (1999)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Volcic, Camberg, Hurley, Ashley, et al. (1999)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Whall, Black, Yankou, Groh, et al. (1999)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RESPONSIVENESS</td>
<td>Studies Reported &amp; References</td>
<td>Adequacy Checks</td>
<td>Comment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Floor and ceiling effects</td>
<td>The questionnaire fails to demonstrate a worse score in patients clinically deteriorated and an improved score in patients who clinically improved. Authors should provide descriptive statistics of the distribution of scores.</td>
<td>□ Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected. □ Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score.</td>
<td>Studies did not provide information about floor or ceiling effects.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Sensitivity to change

The ability to detect important change over time in the concept being measured

<table>
<thead>
<tr>
<th>Study</th>
<th>Hypotheses were formulated and results were in agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ An adequate metric was used (ES, SRM, comparison with external standard)</td>
<td></td>
</tr>
<tr>
<td>□ No information on sensitivity to change was provided</td>
<td></td>
</tr>
<tr>
<td>□ MCID - Information was provided about the magnitude of score differences which would be clinically meaningful</td>
<td>X MCID – No information was provided.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>McCallister &amp; Reisberg (1987)</th>
<th>X The CMAI has been shown to be sensitive to drug treatment with scores improving significantly as a result of treatment.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goddaer &amp; Abraham (1994)</td>
<td>It is also sensitive to non pharmacological treatments such as music therapy, air mat therapy and a motor and occupational intervention that included music and other social and occupational activities. A couple of studies with very small sample sizes of 10 (Richeson, et al. 2003 and Skjerve, et al. 2002) reported sensitivity of the instrument to the effects of animal assisted therapy and bright light therapy.</td>
</tr>
<tr>
<td>Buettner, Lunegren, Lago, Farrell, et al. (1996)</td>
<td>The CMAI has been shown to be sensitive to drug treatment.</td>
</tr>
<tr>
<td>Calkin, Kunik, Oreno, Molinari, et al. (1997)</td>
<td>Findings regarding sensitivity to change over time are not strong. Koss, et al. (1997) reported a non-significant change in score in a sample of moderate to severely demented persons. Weiner, et al. (1998) found no significant difference over time in a sample of persons with mild dementia.</td>
</tr>
<tr>
<td>Koss, Weiner, Ernesto, Cohen-Mansfield, et al. (1997)</td>
<td>The CMAI-C has been shown to be sensitive to drug treatment.</td>
</tr>
<tr>
<td>Ramadan &amp; Naughton (1999)</td>
<td></td>
</tr>
<tr>
<td>De Deyn, Rabheru, Rasmussen, Bocksberger, et al. (1999)</td>
<td></td>
</tr>
<tr>
<td>Tariot, Schneider, Mintzer, Cutler, et al. (2001)</td>
<td></td>
</tr>
<tr>
<td>Vespa, Gori &amp; Spazzafumo (2002)</td>
<td></td>
</tr>
<tr>
<td>Richeson (2003)</td>
<td></td>
</tr>
<tr>
<td>Suh, Son, Ju, Jcho, et al. (2004)</td>
<td></td>
</tr>
<tr>
<td>Hicks-Moore (2005)</td>
<td></td>
</tr>
<tr>
<td>Suh, Greenspan &amp; Choi (2006)</td>
<td></td>
</tr>
<tr>
<td>van Diepen, Bailon, Redman, Rooke, et al. (2006)</td>
<td></td>
</tr>
</tbody>
</table>

### Cultural Applicability and Cultural Adaptations:

The CMAI has been translated into numerous languages. It is available in the following European languages: Dutch, Danish, French, German, Greek, Norwegian and two Spanish versions.
Asia and the Middle East it is available in Chinese, Korean and Japanese and Hebrew. Information about how to obtain these translations is provided in the instruction manual available from the authors.

**Gender Appropriateness:** Appropriate for use with both genders.

**Age Appropriateness:** Generally used with elderly persons 65 years and over but no age limitation is specified by the authors.

**Summary:** The CMAI is a widely used rating scale with versions suitable for use in the nursing home and community settings. The instrument is readily available and free to use with the authors' permission. It is administered by a caregiver and takes about 10-15 minutes to complete. Training is recommended. Psychometric properties are very good and the instrument is available in numerous languages.

**References**


### Pittsburgh Agitation Scale

**Title:** Pittsburgh Agitation Scale.

**Abbreviations:** PAS.

**Author(s) Name:** J Rosen, L Burgio, M Kollar, M Cain, M Allison, M Fogleman, M Michael and GS Zubenko.

**Author(s) Address:** Jules Rosen  
Western Psychiatric Institute and Clinic  
3811 O’Hara Street  
Pittsburgh PA  15213  
USA.

**Supplied by:** The instrument can be obtained by contacting the author. It is also available on the internet.

**Cost:** Free.

**Training requirements:** Minimal.

**Purpose:** To assess agitation in patients with dementia.

**Administration time:** Less than one minute.

**Instrument Type:** The PAS is a simple rating scale completed by staff members during the course of their direct observation and documentation.

**Structure:** The instrument measures the severity of agitation in four general categories: aberrant vocalisation, motor agitation, aggressiveness and resisting care. It comprises one item for each of these categories, with each item rated on a 4 point scale.

**Scoring:** Each item is rated from 0 (not present) to 4 (indicates the most disruptive or unsafe behaviour). Ratings are based on behaviours during a rating period which is typically 4 to 8 hour. Scores for ‘vocalisation’ and ‘motor agitation’ are determined by the intensity and disruptiveness within the environment, and the ease with which the behaviour can be redirected. Scores for the rating of ‘aggressiveness’ are based on a general description of the behaviour. Scores for ‘resisting care’ are based on the behaviour associated with specific identified activities such as washing, dressing etc. Scores of 3 or 4 reflect behaviours that are not responsive to redirection, distraction, or other behavioural interventions. Total scores range from 0 to 16, where a higher score indicates more agitation.

**Developed for:** The PAS was developed for use during routine nursing care situations.

**Normative Data:** None.

**Clinical Data:** The PAS has been used in a study investigating pain and agitation in long term care (Zieber, Hagen, Armstrong-Esther & Aho 2005). It has also been used as an outcome measure in intervention studies evaluating the effectiveness of aromatherapy (Holmes, Hopkins, Hensford, MacLaughlin, et al. 2002), age and stage appropriate activities programs (Mahoney, 2003), enhanced
dining programs (Perivolaris, LeClerc, Wilkinson & Buchanan, 2006) and the provision of individualised advice to patients on non-pharmacological strategies (Mador, Giles, Whitehead & Crotty, 2004). It has also been used in a study assessing the effectiveness of a training program for nurses (Wells, Dawson, Sidani, Craig, et al. 2000).

Applications:
The PAS has been used to assess the level of agitation in persons with dementia both in the nursing home and inpatient setting. It has also been used as an outcome measure in intervention studies and in studies.

Carer and/or Patient Use of Instrument:
Carers (nursing staff) assess the patient and rate behaviour based on direct observation over a one to eight hour period.

<table>
<thead>
<tr>
<th>Psychometric Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RELIABILITY</strong></td>
</tr>
<tr>
<td><strong>Studies Reported &amp; References</strong></td>
</tr>
<tr>
<td>Internal consistency</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Test – retest</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Inter – rater</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>VALIDITY</td>
</tr>
<tr>
<td>--------------------</td>
</tr>
</tbody>
</table>
| Content            | Rosen, Burgio, Kollar, Cain, et al. (1994) | □ Patients and experts were involved during item selection and/or item reduction  
□ Patients were consulted for reading and comprehension  
X No patient involvement  
□ No information found on content validity  
X There is an adequate coverage of relevant domains  
□ There is limited coverage of relevant domains | The PAS has been developed by experts in the area and field tested by researchers. It was tested with inpatients in the psychiatric setting and in the nursing home setting. Patients only involved in validation studies. Relevant domains are adequately covered. |
|                    | Zieber, Hagen, Armstrong-Esther & Aho (2005) | X Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used  
□ Limited/inadequate construct validity reported  
□ No information provided | Expected correlations are found with the following pain measures: Discomfort Scale for Dementia of the Alzheimer’s Type (DS-DAT); Pain ratings made by palliative care and facility nurses. |
|                    | Rosen, Bobys, Mazumdar, Mulsant, et al. (1999) | X No evidence provided/failed a test of dimensionality  
□ Some evidence provided to support internal structure  
□ Substantial evidence provided to support internal structure | No studies found. |
|                    | Rosen, Burgio, Kollar, Cain, et al. (1994)  
Rosen, Bobys, Mazumdar, Mulsant, et al. (1999) | X Correlations with other measures are reported  
□Correlations not reported | PAS correlated significantly with the following measures of agitation and aggression:  
Direct observation for vocalisation, motor agitation and aggressiveness (in acute psychiatric setting); Need for restraints (either chemical or physical) in nursing home setting.  
PAS did not show significant correlations with the Neurobehavioral Rating Scale (NBRS). |
### Discriminant Validity

The scale differentiates between relevant categories of respondent e.g. sick vs. well, varying degrees of severity.

<table>
<thead>
<tr>
<th>Rosen, Bobys, Mazumdar, Mulsant, et al. (1999)</th>
<th>X Scale differentiates between relevant categories of respondents</th>
<th>No information on discriminant validity</th>
</tr>
</thead>
</table>

The diagnostic accuracy of the PAS attests to its discriminant validity. Area under the curve (AUC) showed sensitivity and specificity of 78% and 95%.

### Criterion

Information on the relationship of scores to gold standard measures or clinical diagnosis is provided.

| Rosen, Burgio, Kollar, Cain, et al. (1994) | X Authors provide 2 or more types of information on interpretability | No comparison with criterion measures provided |
| Rosen, Bobys, Mazumdar, Mulsant, et al. (1999) | X Authors provide 2 or more types of information on interpretability | No studies found. |
| Wells, Dawson, Sidani, Craig, et al. (2000) | X Authors provide 2 or more types of information on interpretability | No studies found. |
| Zieber, Hagen, Armstrong-Esther & Aho (2005) | X Authors provide 2 or more types of information on interpretability | No studies found. |

### Interpretability

The degree to which one can assign qualitative meaning to quantitative scores.

| Rosen, Burgio, Kollar, Cain, et al. (1994) | X Authors provide 2 or more types of information on interpretability | No comparison with criterion measures provided |
| Rosen, Bobys, Mazumdar, Mulsant, et al. (1999) | X Authors provide 2 or more types of information on interpretability | No comparison with criterion measures provided |
| Wells, Dawson, Sidani, Craig, et al. (2000) | X Authors provide 2 or more types of information on interpretability | No comparison with criterion measures provided |
| Zieber, Hagen, Armstrong-Esther & Aho (2005) | X Authors provide 2 or more types of information on interpretability | No comparison with criterion measures provided |

Studies report at least 2 types of information that include combinations of means and standard deviations, comparative data in relevant subgroups and relationship of scores to well known measures.

### Floor and ceiling effects

The questionnaire fails to demonstrate a worse score in patients clinically deteriorated and an improved score in patients who clinically improved.

| Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected | X Authors should provide descriptive statistics of the distribution of scores |

Authors should provide descriptive statistics of the distribution of scores.

No information was found on this aspect of the scale.
<table>
<thead>
<tr>
<th>Sensitivity to change</th>
<th>Possible score</th>
<th>Sensitivity to change</th>
<th>The ability to detect important change over time in the concept being measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>X No information provided on floor and ceiling effects</td>
<td>X Hypotheses were formulated and results were in agreement</td>
<td>□ An adequate metric was used (ES, SRM, comparison with external standard)</td>
<td>□ No information on sensitivity to change was provided</td>
</tr>
<tr>
<td>Wells, Dawson, Sidani, Craig, et al. (2000) Holmes, Hopkins, Hensford, MacLaughlin, et al. (2002) Mahoney (2003) Perivolaris, LeClerc, Wilkinson &amp; Buchanan (2006)</td>
<td>□ No information on sensitivity to change was provided</td>
<td>□ MCID - Information was provided about the magnitude of score differences which would be clinically meaningful</td>
<td>X MCID – No information was provided.</td>
</tr>
<tr>
<td>The instrument was shown to be sensitive to the effects of the following interventions: stage and age based activities program; an abilities focussed program for nurses aimed at assisting them with the care of patients. Evidence of sensitivity to other behavioural interventions is not strong. A small study by Holmes, et al. (2002) reported a change in score as a result of aromatherapy treatment. Another small study by Perivolaris, et al. (2006) showed no significant effect resulting from an enhanced dining program.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Cultural Applicability and Cultural Adaptations:** Translations are not available as yet.

**Gender Appropriateness:** Appropriate for use with both genders.

**Age Appropriateness:** Generally used with elderly persons 65 years and over but no age limitation is specified by the authors.

**Summary:** The PAS is a user friendly very short rating scale completed by staff members during their direct observation and documentation. It takes less than one minute to complete, is readily available at no cost and training is minimal. Evidence relating to psychometric properties is not extensive but available data indicate they are good. No translations are available as yet but its simplicity makes it suitable for translation.

**Reporter:** Siggi Zapart

**Date of report:** 18/4/07

**References**


Appendix 10.3.4: Rating Anxiety in Dementia

AHOC INSTRUMENT REVIEW SHEET

Title: Rating Anxiety in Dementia.

Abbreviations: RAID.

Author(s) Name: Dr. K. K. Shankar, M. Walker, D. Frost and Dr. M. Orrell.

Author(s) Address: Dr Martin Orrell (Corresponding Author)
Department of Mental Health Sciences
University College London
Wolfson Building
48 Riding House Street
London W1W 7EY
United Kingdom

E-mail: m.orrell@ucl.ac.uk

Supplied by: A copy of the RAID is available in the original publication by Shankar, et al. (1999).

Cost: Need to seek permission to use from Dr. Orrell.

Training requirements: Little information is provided. Authors argue that “little additional training was needed to administer the scale” (page 41).

Users need to be familiar with the administration and scoring guidelines founding in the original publication by Shankar, et al. (1999).

Purpose: To measure anxiety in patients suffering dementia (Burns, et al. 2004).

Administration time: 20 minutes (10 minutes with staff member or carer + 10 minutes with patient) (Shankar, et al. 1999).

Instrument Type: Clinical Rating Scale – based on staff or carer report, patient interview and clinical notes (Gibbons, et al. 2006). The timeframe for the observations to be rated is two weeks prior to the assessment.

Structure: 18 item scale with four subscales: worry; apprehension and vigilance; motor tension; and autonomic hypersensitivity. Plus two items on phobias and panic attacks not included in the total score.

Scoring: Each item is scored according to the following response format: Absent, Mild or Intermittent, Moderate and Severe (0 – 3). Items 1 – 18 are added to give a total score.

A total score of 11 or more indicates significant anxiety symptoms (Shankar, et al. 1999).

RAID contains 5 of the 6 DSM-IV criteria for Generalised Anxiety Disorder. The “Difficulty concentrating or mind going blank” criteria was omitted (Gibbons, et al. 2006).

Developed for: To counter the paucity of research on anxiety in dementia. This is the first scale in this area. The corollary scale in depression, the Cornell Scale for Depression in Dementia was used as a model for the RAID (Burns, et al. 2004).

Normative Data: Not available.
Clinical Data: RAID is starting to be used as an outcome measure in clinical studies. Three are identified below:


Unmet needs in Nursing Home residents: Hancock, et al. (2006).

Applications: Shankar & Orrell (2000) provide an extremely useful review of the issues regarding the measurement of anxiety and depression in dementia.

Other important features of the new scale are noted in Shankar, et al. (1999):

- The autonomic hyperactivity subscale items tend to score less, while the restlessness item is the most frequently endorsed.

- Scores are not related to dementia severity level (CDR) or level of cognitive impairment (MMSE).

- People with one or more physical illness scored higher on the scale.

- People with insight into their condition also scored higher than those without insight.

By removing the seven common items between the CSDD and RAID a correlation of only 0.2 (Spearman) was found by Shankar et al 1999. But “Conceptually, these items could not be separated from the RAID scale as it would make RAID an incomplete anxiety scale. Also clinically, there was known to be a significant co-morbidity between anxiety and depression.” (page 46).

Carer and/or Patient Use of Instrument: The RAID is a Clinical Rating Scale – based on staff or carer report, patient interview and clinical notes (Gibbons, et al. 2006).

Psychometric Criteria

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal consistency</strong></td>
<td>Shankar, et al. (1999)</td>
<td>X Alpha &gt;0.70</td>
<td>High internal consistency reliability (Cronbach’s alpha) = 0.83. Though the four subscales were lower 0.51 - 0.74 (Shankar, et al. 1999).</td>
</tr>
<tr>
<td></td>
<td>Shankar &amp; Orrell (2000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gibbons, et al. (2006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Twelftree &amp; Qazi (2006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Marginal or inadequate internal consistency (&lt;0.70)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information found on internal consistency</td>
<td></td>
</tr>
<tr>
<td><strong>Test – retest</strong></td>
<td>Shankar, et al. 1999</td>
<td>X ICC &gt;.70</td>
<td>Moderate test-retest reliability, Kappa range of between 0.53 – 1.00 for individual items (repeat interviews 7 -10 days apart) (Shankar, et al. 1999).</td>
</tr>
<tr>
<td></td>
<td>Shankar &amp; Orrell (2000)</td>
<td>Time intervals and confidence intervals reported</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burns, et al. (2004)</td>
<td>□ Marginal or inadequate test-retest reliability ICC&lt;.70</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gibbons, et al. (2006)</td>
<td>□ No information found on test-retest</td>
<td></td>
</tr>
</tbody>
</table>
Calculation of an intraclass correlation coefficient (ICC); and an ICC > 0.70 is desired
Preferred if time interval and confidence intervals were presented

<table>
<thead>
<tr>
<th>Calculation of an intraclass correlation coefficient (ICC); and an ICC &gt; 0.70 is desired Preferred if time interval and confidence intervals were presented</th>
<th>reliability</th>
<th>Inter – rater</th>
<th>Shankar, et al. (1999) Shankar &amp; Orrell (2000) Burns, et al. (2004) Gibbons, et al. (2006)</th>
<th>X Agreement reported and adequate □ Inadequate inter-rater agreement □ No information provided</th>
<th>Moderate to high inter-rater reliability - overall agreement of between 82 - 100 % for individual items (two raters of 33 patients) (Shankar, et al. 1999).</th>
</tr>
</thead>
</table>

**VALIDITY**

<table>
<thead>
<tr>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
</table>

**Content**
The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire

Shankar, et al. (1999) X Patients and experts were involved during item selection and/or item reduction □ Patients were consulted for reading and comprehension □ No patient involvement □ No information found on content validity X There is an adequate coverage of relevant domains □ There is limited coverage of relevant domains

The RAID was developed from the literature and major diagnostic manuals and structured clinical interviews (e.g. DSM-IV, PSE, GMS). Then it was sent to experienced clinicians and carer groups for comment / input.

Phobia and panic attack items were improved.

Need for clearer guidelines was noted in the original paper (Shankar, et al. 1999).

**Construct**
The extent to which scores on the questionnaire relate to other measures in a manner that is consistent with theoretically derived hypothesis concerning the domains that are measured

Shankar, et al. (1999) Burns, et al. (2004) X Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used □ Limited / inadequate construct validity reported □ No information provided

83 inpatients and outpatients with a DSM-IV diagnosis of dementia. Scores were found not related to sex, age, cognitive impairment, type of dementia but associated with presence of physical illness, as well as the preservation of insight.

**Construct: Internal Structure**
Information provided on factor structure

Shankar, et al. (1999) □ No evidence provided/failed a test of dimensionality X Some evidence provided to support internal structure □ Substantial evidence provided to support internal structure

A 5 factor structure emerged accounting for 63.8% of variance. But 3 factors are defined by less than 3 items.

Selwood, et al. (2005) also found that the RAID correlates highly with QoL measures (QoL-AD, DQoL).

Correlation with anxiety measures (Rating scale):
NPI Anxiety subscale = 0.40 (Gibbons, et al. 2006).

Correlations with self-report anxiety questionnaires:
Clinical Anxiety Scale = 0.54 (Shankar, et al. 1999);
Anxiety Status Inventory = 0.62 (Shankar, et al. 1999);
HADS (anxiety) = 0.32 (Cheston, et al. 2003);
STAI-S = 0.60 (Twelftree & Qazi, 2006); Participant Anxiety Scale = 0.27 Gibbons, et al. 2006.

Correlations with depression measures (rating scales and self report):
CSDD = 0.66 (Gibbons, et al. 2006); 0.55 (Cheston, et al. 2003); NPI Depression subscale = 0.28 (Gibbons, et al. 2006); RMBPC Depression subscale (proxy rating) = 0.40 (Gibbons, et al. 2006); GDS (self-report) = 0.20 (Gibbons, et al. 2006).

Correlation with the Cohen-Mansfield Agitation Inventory = 0.41 (Spearman) (single overlapping item on restlessness removed)(Twelftree & Qazi, 2006).


Sensitivity (90%) and specificity (78.5%) using a cut score of 11 (see Shankar, et al. 1999).

The RAID and CSDD have a number of items in common reflecting symptom overlap / co-morbidity issues that require further refinement |
### Criterion

Information on the relationship of scores to gold standard measures or clinical diagnosis is provided

<table>
<thead>
<tr>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shankar, et al. (1999)</td>
<td>X</td>
<td>Comparison made to criterion measures</td>
</tr>
<tr>
<td>Shankar &amp; Orrell (2000)</td>
<td></td>
<td>□ No comparison with criterion measures provided</td>
</tr>
<tr>
<td>Burns, et al. (2004)</td>
<td></td>
<td>In the absence of a “gold standard” the RAID correlates highly with an independent consultant psychiatrist’s rating and the carer’s rating (visual analogue scales) (Shankar, et al. 1999).</td>
</tr>
</tbody>
</table>

### Interpretability

The degree to which one can assign qualitative meaning to quantitative scores

Do authors provide the following:

- Presentation of means and SD of scores before and after treatment
- Comparative data on the distribution of scores in relevant subgroups
- Information on the relationship of scores to well-known functional measures or clinical diagnosis
- Information on the association between changes in scores and patients’ global ratings of the magnitude of change they have experienced

<table>
<thead>
<tr>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shankar, et al. (1999)</td>
<td>X</td>
<td>Authors provide 2 or more types of information on interpretability</td>
</tr>
<tr>
<td>Gibbons, et al. (2006)</td>
<td></td>
<td>□ Authors provide limited information to assist with interpretability</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information provided</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A total score of 11 or over indicates significant anxiety symptoms (Shankar, et al. 1999).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RAID contains 5 of the 6 DSM-IV criteria for Generalised Anxiety Disorder (Gibbons, et al. 2006).</td>
</tr>
</tbody>
</table>

### RESPONSIVENESS

**Floor and ceiling effects**

The questionnaire fails to demonstrate a worse score in patients who clinically deteriorated and an improved score in patients who clinically improved

Authors should provide descriptive statistics of the distribution of scores

<table>
<thead>
<tr>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shankar, et al. (1999)</td>
<td>X</td>
<td>Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information provided on floor and ceiling effects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Histogram of the distribution of scores provided by Shankar, et al. (1999).</td>
</tr>
</tbody>
</table>
Sensitivity to change
The ability to detect important change over time in the concept being measured

<table>
<thead>
<tr>
<th>Qazi, et al. 2003</th>
<th>X Hypotheses were formulated and results were in agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>□ An adequate metric was used (ES, SRM, comparison with external standard)</td>
</tr>
<tr>
<td></td>
<td>□ No information on sensitivity to change was provided</td>
</tr>
<tr>
<td></td>
<td>□ MCID - Information was provided about the magnitude of score differences which would be clinically meaningful</td>
</tr>
<tr>
<td></td>
<td>X MCID – No information was provided.</td>
</tr>
</tbody>
</table>

Qazi, et al. (2003) using a series of single case studies looking at Anxiety management in dementia patients. Found decreases in RAID and CSDO scores. Suggests RAID would be useful in RCTs.

Cultural Applicability
and Cultural Adaptations:
Information concerning cultural adaptations was not found.

Gender Appropriateness:
Appropriate for use with both genders.

Age Appropriateness:
Appropriate for use with adults but it has mainly been used with elderly people.

Summary:
The RAID is a promising new rating scale for anxiety in people with dementia. However, it requires wider application especially in the areas of cultural and language adaptation, and sensitivity to change. Also the RAID requires better training resources.

Due to the co-morbidity of depression and anxiety, and symptom overlap issues (plus factor structure limitations), further work is required to refine and better understand individual items in the RAID. As Gibbons, et al. (2006) said, “Further refinement of the definition and measurement of anxiety in dementia will result in a better understanding of anxiety and its complex relationship with depression and other aspects of dementia” (page 207).

Reporter:
Nicholas Marosszeky

Date of report:
August 2007

References


Appendix 10.3.5: Apathy Evaluation Scale

AHOC INSTRUMENT REVIEW SHEET

Title: Apathy Evaluation Scale.

Abbreviations: AES.

Author(s) Name: Robert S. Marin, MD.

Author(s) Address: Robert S. Marin
Department of Psychiatry
University of Pittsburgh School of Medicine
Western Psychiatric Institute & Clinic
3811 O’Hara Street
Pittsburgh, PA 15213
USA.

Supplied by: A copy of the AES is available at the COMBI web-site:
http://www.tbims.org/combi/aes/index.html or in a recent review paper by

Cost: Need to seek permission for use from Dr. Marin.

Training requirements: Users need to be familiar with the administration and scoring guidelines. These are provided in the original paper by Marin, et al. (1991) (Also provided on the COMBI web-site).

Little information is available in the scientific literature on the training of users of the AES. Review papers by Malloy & Boyle (2005) and Clarke, et al. (2007) also point this out.

Purpose: To quantify and characterize apathy in adult patients.

Administration time: 20 minutes (reviewer’s estimate).

Instrument Type: The AES comes in three versions, crossing instrument category types - AES-S (Self report), AES-C (Clinician rated), AES-I (Informant / proxy rated). They each share the same items.

The AES-C is based on a semi-structured interview with the patient. As well as clinical ratings, some items for the AES-C are quantitative (e.g. number of friends) while others involve patient self evaluation during the interview.

Structure: 18 item scale with behavioural, cognitive and emotional domains.

Scoring: Items combined into a total score (score range 18 – 72), with higher scores indicating greater apathy.

Response categories: Not at all Characteristic, Slightly Characteristic, Somewhat Characteristic, Very Characteristic OR Not at all True, Slightly True, Somewhat True, Very True (1 – 4 scoring).

Developed for: To address the lack of appropriate assessment tools for apathy, a "complex neuro-behavioural syndrome" which is manifest across a range of conditions (e.g. dementia, TBI and HIV-AIDS; Clarke, et al. 2007).

Normative Data: Some normative data has been provided by Marin, et al. (1991) (n = 31). However, there have been no published studies examining the AES in a large sample of normal control subjects stratified for age or educational levels(Malloy & Grace, 2005) This kind of work is very necessary for self-report measures like the AES-S.
Clinical Data: Clinical studies have provided information on Alzheimer’s disease including Starkstein, et al. (1995), Starkstein, et al. (2001), Starkstein, et al. (2005a) and Clarke, et al. (2007).

Besides dementia, the AES is used in a number of neurological groups where apathy symptoms can be found:

- Huntington's disease: Chatterjee, et al. 2005
- Late-Life Depression: Lavretsky, et al. (1999); Lavretsky, et al. (2004); Lampe & Heeren (2004)
- Parkinson’s Disease: Van Horn, et al. (2001); Remy, et al. (2005); Drapier, et al. (2006); Kirsch-Darrow, et al. (2006)
- PTSD: Ramirez, et al. (2001)
- Schizophrenia: Kiang, et al. (2003); Van Horn, et al. (2001)
- Stroke: Brodaty, et al. (2005); Starkstein, et al. (1993)
- TBI: Andersson, et al. (1999); Andersson, et al. (2002); Glenn, et al. (2002); Newburn & Newburn (2005); Muller, et al. (2006); Rao, et al. (2007)

An abridged version of the AES was developed by Starkstein, et al. (1992) and used with Parkinson’s disease patients (But it was also noted by Starkstein, et al. (1992) that this abridged scale has been piloted with Alzheimer’s disease, Huntington’s disease and stroke). It is based on the original Marin (1990) paper.

The AES has had a significant influence of the development of other apathy measures in this new field. Instruments include the Lille Apathy Rating Scale (LARS) (Sockeel, et al. 2006); the Apathy Inventory (AI) (Robert, et al. 2001) (which also cites the NPI as a source); and the Structured Interview for Apathy (SIA) (Starkstein, et al. 2005b).

Applications: A Children’s Motivation Scale has also been derived from the AES (see Marin & Wilkosz (2005) for details)

Carer and/or Patient Use of Instrument: Clinician, informant/carer and patient self report forms are available.

Psychometric Criteria

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td>Marin, et al. (1991) Malloy &amp; Grace (2005) Clarke, et al. (2007)</td>
<td>X Alpha &gt;0.70</td>
<td>Marin et al (1991) reported Cronbach’s Alpha across the three versions: AES-C = 0.90; AES-I = 0.94; AES-S = 0.86.</td>
</tr>
</tbody>
</table>
### Test – retest
The extent to which the same results are obtained on repeated administrations of the same questionnaire when no change in physical functioning has occurred.

- Calculation of an intraclass correlation coefficient (ICC); and an ICC > 0.70 is desired.
- Preferred if time interval and confidence intervals were presented.

<table>
<thead>
<tr>
<th>Study</th>
<th>Adequacy Checks</th>
<th>Test-retest reliability: AES-C = 0.88; AES-I = 0.94; AES-S = 0.76 (Mean interval = 25 days).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marin, et al. (1991)</td>
<td>X ICC &gt;.70 Time intervals and confidence intervals reported</td>
<td></td>
</tr>
<tr>
<td>Malloy &amp; Grace (2005)</td>
<td>□ Marginal or inadequate test-retest reliability ICC&lt;.70</td>
<td></td>
</tr>
<tr>
<td></td>
<td>□ No information found on test-retest reliability</td>
<td></td>
</tr>
</tbody>
</table>

### Inter – rater
Limits of agreement, Kappa, or standard error of measurement (SEM) were presented.

<table>
<thead>
<tr>
<th>Study</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marin, et al. (1991)</td>
<td>X Agreement reported and adequate</td>
<td>Marin, et al. (1991) found inter-rater reliability of 0.94 for the AES-C (Two clinical raters).</td>
</tr>
<tr>
<td></td>
<td>□ Inadequate inter-rater agreement</td>
<td></td>
</tr>
<tr>
<td></td>
<td>□ No information provided</td>
<td></td>
</tr>
</tbody>
</table>

### Validity

<table>
<thead>
<tr>
<th>Content</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire</td>
<td>Marin, et al. (1991)</td>
<td>□ Patients and experts were involved during item selection and/or item reduction</td>
<td>Developed from a pool of items from the literature and the author's observations. A revised list of 56 items was further reduced to 18 using item total correlations and examining the items relationship to the Hamilton Rating Scale for Depression. (This was done to help develop a scale which has items that discriminate between apathy and depression).</td>
</tr>
</tbody>
</table>
|                               | Malloy & Grace (2005)         | □ Patients were consulted for reading and comprehension | Inter-correlations between the three versions range from 0.43 – 0.72.  
AES-C + AES-I = 0.62  
AES-C + AES-S = 0.72  
AES-I + AES-S = 0.43 |
|                               |                               | X No patient involvement |                                                                                                                                         |
|                               |                               | □ No information found on content validity |                                                                                                                                         |
|                               |                               | X There is an adequate coverage of relevant domains |                                                                                                                                         |
|                               |                               | □ There is limited coverage of relevant domains |                                                                                                                                         |

### Construct
The extent to which scores on the questionnaire relate to other measures in a manner that is consistent with theoretically derived hypothesis concerning the domains that are measured.

<table>
<thead>
<tr>
<th>Study</th>
<th>Adequacy Checks</th>
<th>Some highlights:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eling, et al. (2006)</td>
<td>X Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used</td>
<td>Relationship to the presentation of novel stimuli to apathetic dementia patients; patients were relatively irresponsible to the stimuli (though sensitive to the emotional content) (Eling, et al. 2006).</td>
</tr>
<tr>
<td></td>
<td>□ Limited /inadequate construct validity reported</td>
<td></td>
</tr>
<tr>
<td></td>
<td>□ No information provided</td>
<td></td>
</tr>
<tr>
<td>Construct: Internal Structure</td>
<td>Marin, et al. (1991)</td>
<td>Single factor of apathy accounting for 32 to 53% of the variance. Additional smaller factors about interest and insight were also found, together these three factors accounted for 50 to 65% of total variance, across versions. Clarke, et al. (2007) found a two factor structure for AES-C and AES-I. However, the amount of variance accounted for by the apathy factor differed markedly between them, while the AES-S produced only a single reliable factor of apathy. This finding suggests that further work is required to examine the AES’s factor structure across versions.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Information provided on factor structure</td>
<td>Marin, et al. (1991)</td>
<td>X Some evidence provided to support internal structure</td>
</tr>
<tr>
<td>Clarke, et al. (2007)</td>
<td>□ No evidence provided/failed a test of dimensionality</td>
<td></td>
</tr>
<tr>
<td>Construct: Correlation with other measures</td>
<td>Clarke, et al. (2007)</td>
<td>Significant correlations (for each version) with the NPI Apathy subscale was found by Clarke, et al. (2007). Using the Hamilton Rating Scale for Depression and Anxiety (clinical), as well as Zung Rating Scale for Depression and Anxiety (informant and self-report), Marin, et al. (1991) found using the multitrait-multimethod matrix criteria some supportive evidence for convergence and discriminate validity of all versions of the AES. Though there is less evidence for the AES-I Correlation with Hamilton Rating Scale becomes non-significant when apathy items from Hamilton</td>
</tr>
<tr>
<td>Comparisons made to other measures</td>
<td>Marin, et al. (1991)</td>
<td>□ Correlations not reported</td>
</tr>
<tr>
<td></td>
<td>Clarke, et al. (2007)</td>
<td>X Correlations with other measures are reported</td>
</tr>
<tr>
<td>Muller, et al. (2006)</td>
<td>Relationship to self-initiated locomotor activity (TBI sample, Muller, et al. (2006)). Clarke, et al. (2007) provides evidence regarding the correlation with relevant subscales of the NPI (apathy and depression but not others subscales e.g. delusions). Lampe &amp; Heeren (2004) found no difference in AES scores between late onset and early onset depression; and scores were related to the negative symptoms of depression, as measured by the MADRS.</td>
<td></td>
</tr>
</tbody>
</table>
Construct: Discriminant Validity

The scale differentiates between relevant categories of respondents e.g. sick vs. well, varying degrees of severity

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Reference</th>
<th>Information on the relationship of scores to gold standard measures or clinical diagnosis is provided</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eslinger, et al. (2005)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Discriminated between well elderly, major depression, AD and stroke (right hemisphere) (Marin, et al. 1991). This was confirmed by Marin, et al. (1994) finding that the relationship between apathy and depression differs across diagnostic groups (health elderly, left or right hemisphere stroke, probable AD and depression).

AES did not correlate with the Structured Clinical Interview for DSM-IV (SCID-Depression section) (Clarke, et al. 2007).

Need research to apply the AES to people with frontal systems damage (Malloy & Grace, 2005).

Other related studies:

- AES scores related to chronic vs. nonchronic / major late-life depression (Lavretsky, et al. 1999; Lavretsky, et al. 2004) (also associated with neuroimaging data).
- Using the AES and other instruments, Eslinger et al (2005) found differences in self and carer profiles between AD and frontal-temporal dementia (FTD). In AD, self and carer discrepancies (i.e. self-awareness)were more focused on apathy and memory, while in FTD discrepancies were more general.

Criterion

Information on the relationship of scores to gold standard measures or clinical diagnosis is provided

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Reference</th>
<th>Information on the relationship of scores to gold standard measures or clinical diagnosis is provided</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resnick, et al. (1998)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Resnick, et al. (1998) found that the AES (and a short 7 item version) as measures of motivation were significant predictors of discharge functioning after a geriatric rehabilitation program.

(Resnick (1998) conducted a similar study in nursing home residents and found that motivation was one of two factors that predicted function).
### Interpretability

The degree to which one can assign qualitative meaning to quantitative scores.

Do authors provide the following:

- Presentation of means and SD of scores before and after treatment
- Comparative data on the distribution of scores in relevant subgroups
- Information on the relationship of scores to well-known functional measures or clinical diagnosis
- Information on the association between changes in scores and patients' global ratings of the magnitude of change they have experienced

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Information Provided</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marin, et al. (1993)</td>
<td>X Authors provide 2 or more types of information on interpretability</td>
</tr>
<tr>
<td>Glenn, et al. (2002)</td>
<td>□ Authors provide limited information to assist with interpretability</td>
</tr>
<tr>
<td>Clarke, et al. (2007)</td>
<td>□ No information provided</td>
</tr>
</tbody>
</table>

Clarke, et al. (2007) reports that Marin, et al. (1993) found a cut-off score of 37.5 points for apathetic and non apathetic individuals (using a clinical diagnosis of apathy).

Glenn, et al. (2002) report that the AES has poor sensitivity and specificity in relation to a clinical diagnosis of apathy (in a TBI sample).

Clarke, et al. (2007) provided up to date sensitivity and specificity information (using a clinician diagnosis of apathy). They found a cut-off score of 40.5 points for the AES-C. They recommend using the AES-I in the absence of highly trained raters, as they found superior sensitivity and equivalent specificity for this informant version. However, this is undermined by the fact that the two versions have markedly different factor loadings for apathy.

### RESPONSEIVENESS

**Floor and ceiling effects**

The questionnaire fails to demonstrate a worse score in patients who clinically deteriorated and an improved score in patients who clinically improved.

Authors should provide descriptive statistics of the distribution of scores.

<table>
<thead>
<tr>
<th>Adequacy Checks</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
</tr>
<tr>
<td>□ Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score</td>
</tr>
<tr>
<td>X No information provided on floor and ceiling effects</td>
</tr>
</tbody>
</table>

No information found.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>The ability to detect important change over time in the concept being measured</td>
<td>□ Hypotheses were formulated and results were in agreement □ An adequate metric was used (ES, SRM, comparison with external standard) □ No information on sensitivity to change was provided □ MCID - Information was provided about the magnitude of score differences which would be clinically meaningful</td>
<td>X Hypotheses were formulated and results were in agreement</td>
<td>□ An adequate metric was used (ES, SRM, comparison with external standard) □ No information on sensitivity to change was provided □ MCID - Information was provided about the magnitude of score differences which would be clinically meaningful</td>
<td>□ An adequate metric was used (ES, SRM, comparison with external standard) □ No information on sensitivity to change was provided □ MCID - Information was provided about the magnitude of score differences which would be clinically meaningful</td>
<td></td>
</tr>
<tr>
<td>Gender Appropriateness:</td>
<td>The instrument is appropriate for use with both genders. Lavretsky, et al. (2004) has found a sex difference in patients with major late-life depression on the AES – where more severe scores were found for men than women (a non dementia sample).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age Appropriateness:</td>
<td>The AES is appropriate for adults over 55 years.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summary:</td>
<td>The AES is a highly influential instrument in the emerging area of apathy syndrome measurement. However, there is some concern noted in the literature about the lack of adequate training and guidance for the clinical and informant ratings (Malloy &amp; Boyle, 2005; Clarke, et al. 2007). Here the abridged version of the AES by Starkstein, et al. (1992) may be a useful alternative. Also, following the work of Clarke, et al. (2007), the informant</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

X Hypotheses were formulated and results were in agreement □ An adequate metric was used (ES, SRM, comparison with external standard) □ No information on sensitivity to change was provided □ MCID - Information was provided about the magnitude of score differences which would be clinically meaningful X MCID – No information was provided.
and self-report versions of the AES require further examination of their factor structure, as well as more normative data

**Reporter:** Nicholas Marosszeky

**Date of report:** August 2007

**References**


Appendix 10.3.6: Geriatric Depression Scale

AHOC INSTRUMENT REVIEW SHEET

Title: Geriatric Depression Scale.

Abbreviations: GDS.

Author(s) Name: Jerome A. Yesavage, M.D.
T.L. Brink, PhD.

Author(s) Address: Jerome A. Yesavage, M.D.
Director
Aging Clinical Research Center
Department of Psychiatry and Behavioral Sciences
Stanford University School of Medicine
Stanford CA, 94305 – 5548.

Supplied by: Visit the following web-site:
http://www.stanford.edu/~eyesavage/GDS.html

Cost: In the public domain (see above).

Training requirements: Minimal.

Purpose: To assess and screen depression in elderly people, using an instrument that was simple to administer and did not require interviewer training (Burns, et al. 2004).

Administration time: 5 – 10 minutes.

Instrument Type: Self-administered or interviewer administered questionnaire.

Structure: 30 questions with dichotomous (Yes / No) response items producing a total score.

The dichotomous (Yes / No) response format of the GDS is contentious, as it is different to other depression scales. This has generated a number of research papers which are summarised below.

Olin, et al. (1992) reports that GDS produces similar results to that of the BDI for older adults; while being simpler for them to complete as it is based on dichotomous responses not multiple responses. Dunn & Sacco (1989) also report a lower complication rate for the GDS than the Zung Self Rating Depression Scale, because of it dichotomous response format. Lyness, et al. (1997) also comments on the easier administration format.

However, Fischer, et al. (1996) in their content analysis found that older people found the Yes / No format restrictive and changed question descriptors. They also tried to fill in the context e.g. my mood is affected by my health, personal relationships, and talked about their personal style rather than depression. Fischer, et al. (1996) suggest that certain personality profiles and situations may influence reporting.

A paper by Dunn & Sacco (1988) randomly assigned a community sample of older people into four groups, changing the reference group instructions for the GDS (no instruction, age group peers, adults in general, themselves when younger) and found that their responses did not change greatly.

Cannon, et al. (2002) found significant test-retest correlations between the oral and written administration formats of the GDS for cognitively intact participants. The same could not be said for those who were cognitively...
impaired (testing was completed over one session).

**Scoring:**

There is 0 or 1 scoring of the Yes / No responses to produce a total score out of 30. Answers indicating depression are scored 1 and those not indicating depression are scored 0 (can be either yes or no responses).

Score in the following ranges suggest:

- 0 – 9 = Not depressed
- 10 – 19 = Mild depression
- 20 – 30 = Severe depression

(Source: PROQOLID website)

Scores can also be prorated if items are missing.

A 15 item version (GDS-15) was developed to reduce the chance of test fatigue in physically ill or demented patients (Shiekh & Yesavage, 1986). It takes 5 minutes to use and has a cut-score of 5 (see Web-site, PROQOLID and Bijl, et al. 2005). The GDS-15 has high correlations with (0.84 – 0.89), and similar properties to, the GDS-30 (Lesher & Beeryhill 1994, Wall, et al. 1999, Aikman & Oehlert, 2001). However, Bowling (2005) and McDowell (2006), cite a paper by Alden, et al. (1989) which found a low correlation of 0.66 between the two versions in a community sample. Ingram (1996) also found poor agreement between the GDS-30 and GDS-15 and lower test-retest reliability for the GDS-15 (r = 0.67) in the community sample (GDS-15 was extracted).

This work highlights an emerging issue with the GDS that choosing the site of administration is very important. Blank, et al. (2004) suggests that the GDS works best in residential care settings.

**Developed for:**

The GDS was developed from a pool of 100 items generated by clinicians and researchers (Bowling, 2005) which was reduced to 30 items on the grounds of high item-total correlations.

An advantage of the GDS is that it does not include somatic symptoms but focuses on the affective aspects of depression (PROQOLID; Bowling, 2005; McDowell, 2006). As McDowell (2006) explains: “Symptoms indicative of depression in young people (e.g. sleep disturbance, weight loss, pessimism about the future), may also occur in the elderly as normal effects of aging or as a result of a physical illness.” Using an instrument with somatic symptoms items may result in false positive cases (Bowling, 2005).

*It should be noted however that the use of somatic items is open to some debate in the literature. The view that by excluding somatic items that the GDS is a better measure of depression in the elderly is supported by Bolla-Wilson & Bleecker (1989) and Salamero & Marcos (1992). However, Norris & Woehr (1998) using the BDI, CES-D and GDS found that some somatic items (diminished energy, sleep disturbance and health worries) were consistent with depression.*

**Normative Data:**

Recent normative data from those aged 75 years and over in the United Kingdom is provided by Osborn, et. al (2002). McDowell reports a mean of 5.6 (SD= 4.4) for a group of healthy seniors (60 – 95 years of age).

**Clinical Data:**

The GDS has been used widely in many clinical studies, applicable to older adults. Below are some highlights:

- Abdominal surgery: Zalon 2004
Also there is some literature that clinicians have problems recognising depression in elderly patients and this has been shown in studies using the GDS as a measure of depression symptoms (Rapp, et al, 1988; Pond, et al, 1990; Jackson & Baldwin, 1993; Garrard, et al, 1998; Bagley, et al, 2000; Peach, et al, 2001 and Ruchinskas, 2002). This may be confounded by dementia severity / cognitive status (Snowdon & Lane, 1999; Ruchinskas, 2002). This co-morbid situation (i.e. mixing depression and cognitive impairment) also affects estimates based on cut-scores from other psychometric instruments like the MMPI and BSI, as well as the GDS (Harper, et al, 1990).

Applications:


Suttcliffe, et al (2000) have developed a new short form of the GDS, the GDS-12R, for residential care.

Nitcher, et al (1993) have developed a proxy version of the scale for those with mild to moderate cognitive impairment (though due the tendency of carers to endorse more symptoms, higher cut-scores are required). Brown & Schinka (2005) have also developed an informant version of the GDS-15.

The web-site suggests how the instrument can be used with aphasic patients. McDowell (2006) reports that a telephone format is also available.
In further developments:

- Arthur, et al. (1999) used the GDS-15 in an annual over 75 health check.
- Cully, et al. (2005b) have developed a 2-item screener for depression in rehabilitation inpatients based on the GDS.
- Recently, Segulin & Deponte (2007) have developed a modified version of the GDS for very old persons.

In terms of using the GDS with people with cognitive impairment, McDowell (2006) states that “In elderly people, depression commonly coexist with dementia; cognitive problems compromise the accuracy of self-reports just as depression may mask cognitive abilities.” This not surprising as the GDS is a recall task. McDowell (2006) recommends supplementing the measure by informant information.


Finally, the complex relationship between depression and cognitive impairment in the elderly has been studied in detail by Parmelee, et al. (1991b), Lichtenberg, et al. (1995) and Vinkers, et al. (2004a) using the GDS. From these studies, it seems that by examining depression as well as cognitive performance one can account for a greater amount of variance in cognitive test score results.

Carer and/or Patient Use of Instrument:

The GDS is a self-administered or interviewer administered questionnaire.

Psychometric Criteria

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td>Brink, et al. (1982)</td>
<td>X Alpha &gt;0.70</td>
<td>PROQOLID reports 0.94 internal consistency for healthy patients and those treated for depression. However, some investigators have found slightly lower reliability (Bowling (2005), see also Jefferson, et al. (2001) and Iglesias (2004) who found 0.83 – 0.84 and Friedman, et al. (2005)). Further details with different populations including younger age groups are provided by McDowell (2006). Abraham (1991) found reliabilities (KR-20) in the range of 0.69 – 0.88 (mean</td>
</tr>
<tr>
<td></td>
<td>Lyons, et al. (1989)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Abraham, (1991)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sutcliffe, et al. (2000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Jefferson, et al. (2001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Iglesias (2004)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Friedman, et al. (2005)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bowling (2005)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>McDowell (2006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PROQOLID</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
0.82) older people assessed 18 times over a 39 week period in a residential care setting.

However, a poor alpha of 0.46 was reported for the GDS-15 by Incalzi et al (2003) with older medical inpatients.

Sutcliffe et al (2000) improved the reliability of the GDS-15 by removing three items when used in a residential care setting.

<table>
<thead>
<tr>
<th>Test – retest</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>The extent to which the same results are obtained on repeated administrations of the same questionnaire when no change in physical functioning has occurred</td>
<td>Brink, et al. (1983) Parmelee, et al. (1989) Burns, et al. (2004) Bowling (2005)</td>
<td>ICC &gt;.70 Time intervals and confidence intervals reported</td>
<td>Brink, et al. (1983) found the following test-retest reliabilities in a residential care setting: 0.86 for one hour; 0.85 for one week; and 0.98 10 – 12 days. This is supported by Parmelee, et al. (1989) and Ingram (1996).</td>
</tr>
<tr>
<td>Calculation of an intraclass correlation coefficient (ICC); and an ICC &gt; 0.70 is desired</td>
<td></td>
<td>Marginal or inadequate test-retest reliability ICC&lt;.70 No information found on test-retest reliability</td>
<td></td>
</tr>
<tr>
<td>Preferred if time interval and confidence intervals were presented</td>
<td></td>
<td>No information</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inter – rater</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Inadequate inter-rater agreement No information provided</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VALIDITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>Weiss, et al. (1986) Adams (2001) Bowling (2005)</td>
<td>No patient involvement No information found on content validity</td>
<td>GDS was developed from items generated by clinicians and researchers (Bowling, 2005). It appears to miss some themes for the older person (Weiss, et al. 1986). Adams (2001) argues that some items of the GDS are measuring social withdrawal rather than depression. Specific data on readability of each item was not found. (As opposed to item format – multiple response vs. dichotomous response).</td>
</tr>
<tr>
<td>The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Construct

The extent to which scores on the questionnaire relate to other measures in a manner that is consistent with theoretically derived hypothesis concerning the domains that are measured

<table>
<thead>
<tr>
<th>Parmelee, et al. (1989)</th>
<th>X Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used □ Limited / inadequate construct validity reported □ No information provided</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cappeliez, et al. (1989)</td>
<td></td>
</tr>
<tr>
<td>Salamero &amp; Marcos (1992)</td>
<td></td>
</tr>
<tr>
<td>Cuijpers &amp; van Lammeren (1999)</td>
<td></td>
</tr>
<tr>
<td>Gazmarranian, et al. (2000)</td>
<td></td>
</tr>
<tr>
<td>Daaleman, et al. (2002)</td>
<td></td>
</tr>
<tr>
<td>Friedman, et al. (2005)</td>
<td></td>
</tr>
<tr>
<td>Heisel, et al. (2005)</td>
<td></td>
</tr>
<tr>
<td>Onishi, et al. (2006)</td>
<td></td>
</tr>
</tbody>
</table>

Low correlation with cognition scores (MMSE) (Onishi, 2006; McDowell, 2006).

Relationship to depressed mood, life satisfaction and suicidal ideation is reported by Friedman, et al. (2005). A paper by Heisel, et al. (2005) also supports the relationship suicidal ideation.

Scores on the GDS have a negative relationship with spirituality (Daaleman, et al. 2002). In accord with other literature in this area it correlates with HDRS - Melancholia Scale 0.77 (Salamero & Marcos, 1992).


There is evidence of immunity to social desirability – Cappeliez, et al. (1989).

Burns, et al. (2004) query its validity with people with dementia. See also the section on Construct: Correlation with other Measures.

### Construct: Internal Structure

Information provided on factor structure

| Parmelee, et al. (1989) | □ No evidence provided/failed a test of dimensionality |
| Sheikh, et al. (1991) | |
| Salamero & Marcos (1992) | |
| Adams (2001) | |
| Friedman, et al. (2005) | |
| Tang, et al. (2005) | |
| Onishi, et al. (2006) | |

are reported by Incalzi, et al. (2003) with older medical inpatients). Finally the work of Adams (Adams, 2001; Adams, et al. 2004) suggests the emergence of an Apathy (Withdrawal-Apathy-Vigor) sub-dimension, also known as depression without sadness. This sub-dimension may lead to an over-identification of the symptoms of depression.

Construct: Correlation with other measures

Comparisons made to other measures

|---------------------|----------------------|---------------------------|------------------------|---------------------------|------------------------|------------------------|------------------------|-------------------|-----------------|-----------------|-------------------|-----------------|-----------------|

X Correlations with other measures are reported □ Correlations not reported

Extensively researched. Here are some highlights:

- Correlation with HRSD = 0.62 – 0.81 (Bowling, 2005). Feher, et al. (1992) finds HDRS is a major predictor of GDS scores.
- Correlation with BDI = approximately 0.85 (Ferraro & Chelminski, 1996; Bowling, 2005). Correlation with BDI-II = 0.71 in a sample of older women (Jefferson, et al. 2001).
- Correlation with Montgomery-Asberg DRS = 0.82 (McDowell, 2006).
- Correlation with Zung Self-Rating Depression Scale = 0.76 for homebound elders (Iglesias, 2004). See also Dunn & Sacco (1989).
- GDS correlates with other well-being / self-esteem measures Philadelphia Center Morale Scale, Southampton Self-Esteem Scale and Bradburn Affect Balance Scale. Coleman, et al. (1995) questions the use of these measures in addition to GDS.
- Correlation with GHQ-12 = 0.5 – 0.6 (kappa) with elderly population in Brazil (Costa, et al. 2003).
- Clayton, et al. (1997) reports that GDS works better in picking up depression symptoms than the HRSD in elderly anxiety patients.

The GDS has higher correlation with measures of...
life satisfaction than CES-D in college students. Correlation CES-D and GDS = 0.66 (Brink and Niemeyer, 1992).

Finally, Korner, et al. (2006) in Denmark provides a direct comparison study with the Cornell Scale for Depression (CSDD) highlighting that the CSDD performs better in both dementia and non-dementia samples. While the GDS has diminished validity in dementia samples.

<table>
<thead>
<tr>
<th>Construct: Discriminant Validity</th>
<th>Litchenberg, et al. (1992)</th>
<th>X Scale differentiates between relevant categories of respondents</th>
<th>Distinguishes mild from moderate and severe depression (PROQOLID).</th>
</tr>
</thead>
<tbody>
<tr>
<td>D’Ath, et al. (1994)</td>
<td>No information on discriminant validity</td>
<td>Sensitivity high, though specificity is lower (Bowling, 2005).</td>
<td></td>
</tr>
<tr>
<td>Lyness, et al. (1997)</td>
<td></td>
<td>Similar or better when compared with other measures e.g. CES-D, BDI and diagnostic interviews / schedules. For further details see Lyness, et al. (1997) for CES-D as well as Litchenberg, et al. (1992), D’Ath, et al. (1994), Jongenelis, et al. (2005). There are also two systematic reviews by Watson &amp; Pigone (2003) and Wancata, et al. (2006).</td>
<td></td>
</tr>
<tr>
<td>Watson &amp; Pigone (2003)</td>
<td></td>
<td>Commenting on one study McDowell (2006) states that these discriminant properties apply “almost identically” to both the 30 and 15 item long versions of the GDS.</td>
<td></td>
</tr>
<tr>
<td>Bowling (2005)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jongenelis, et al. (2005)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McDowell (2006)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wancata, et al. (2006)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROQOLID</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Rapp, et al. (1988)</th>
<th>X Comparison made to criterion measures</th>
<th>Performs just as well or better than the Hamilton Rating Scale for Depression (HRSD) in discriminating elderly people at different severity levels when compared with Research Diagnostic Criteria, DSM-IV criteria, ICD-10 criteria, structured psychiatric interviews (Lichtenberg, et al. 1992; Jackson &amp; Baldwin, 1993; Almeida &amp; Almeida, 1999; Bowling, 2005).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parmelee, et al. (1989)</td>
<td>No comparison with criterion measures provided</td>
<td>Similar comparison were found with the CES-D by Wancata et al (2006) and Watson &amp; Pigone (2003).</td>
<td></td>
</tr>
<tr>
<td>Lichtenberg, et al. (1992)</td>
<td></td>
<td>This is also supported by Rapp, et al. (1988) using the BDI.</td>
<td></td>
</tr>
<tr>
<td>Jackson &amp; Baldwin (1993)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Almeida &amp; Almeida (1999)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friedman, et al. (2005)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bowling (2005)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wancata, et al. (2006)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
However, some problem in detecting minor depression is noted by Parmelee, et al. (1989). There is also a low correlation with suicide attempts in the community (Friedman, et al. 2005).

### Interpretability

The degree to which one can assign qualitative meaning to quantitative scores

Do authors provide the following:

- Presentation of means and SD of scores before and after treatment
- Comparative data on the distribution of scores in relevant subgroups
- Information on the relationship of scores to well-known functional measures or clinical diagnosis
- Information on the association between changes in scores and patients’ global ratings of the magnitude of change they have experienced

X Authors provide 2 or more types of information on interpretability
- Authors provide limited information to assist with interpretability
- No information provided

See the sections on Construct Validity and Sensitivity to Change.

Note: Mixed results with cognitively impaired / people with dementia.

### RESPONSIVENESS

<table>
<thead>
<tr>
<th>Floor and ceiling effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>The questionnaire fails to demonstrate a worse score in patients who clinically deteriorated and an improved score in patients who clinically improved</td>
</tr>
<tr>
<td>Authors should provide descriptive statistics of the distribution of scores</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Osborn, et al. (2002)</th>
</tr>
</thead>
<tbody>
<tr>
<td>X Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
</tr>
<tr>
<td>Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score</td>
</tr>
<tr>
<td>No information provided on floor and ceiling effects</td>
</tr>
</tbody>
</table>

In a large community sample of people 75 years and over (n = 14,545) in the UK, Osborn, et al. (2002) found that about 25% of the sample scored less than 1 on the GDS-15.
### Sensitivity to change
The ability to detect important change over time in the concept being measured

<table>
<thead>
<tr>
<th>Study References</th>
<th>Hypotheses Formulated</th>
<th>Results Agreement</th>
<th>Adequate Metric Used</th>
<th>Sensitivity to Change</th>
<th>MCID Information</th>
</tr>
</thead>
</table>
| Mossey, et al. (1996)  
McCurren, et al. (1999)  
Llewellyn-Jones, et al. (1999)  
Sumaya, et al. (2001)  
Vinkers, et al. (2004b)  
Sparks, et al. (2005) | X | □ | □ | □ | □ |

Clinical studies showing a change in scores on the GDS following treatment include:
- Interpersonal counselling therapy for subdysthymic depression in the medically ill (Mossey, et al. 1996).
- Cognitive exercises for one year in mild cognitive impairment / Alzheimer's Disease treated with cholinesterase inhibitor (ChEI) (Olazarab, et al, 2004).
- Treatment with Atorvastatin calcium for one year (Sparks, et al, 2005).
- Bright light treatment in residential care (Sumaya, et al. 2001).
- Geropsychiatric nurse and trained volunteers for depressed elders in nursing homes (McCurren, et al. 1999).

See also Vinkers, et al. (2004b) which showed a change in score in relation to the loss of a life partner (negative life event).

MCID information needs to be further investigated.

### Cultural Applicability and Cultural Adaptations:
There are numerous language versions including Chinese, Italian, Turkish, Vietnamese and Spanish. For a full list see the PROQOLID database. (Though users are advised to check for accuracy of translation - Bowling 2005; GDS web-site).

### Gender Appropriateness:
The GDS is appropriate for use with both genders. As is common with most depression questionnaires women score higher than men (Osborn, et al. 2002) and separate cut scores may be appropriate (Allen-Burge, et al. 1994).

### Age Appropriateness:
The GDS is appropriate for adults over 55 years (Source: PROQOLID). McDowell (2006) suggests that there are some issues using the instrument with those seventy five years and over, for example, the GDS does not look at the two week persistence of symptoms of depression.

### Summary:
A widely used and researched, self-report instrument for the assessment and screening of depression in elderly people. The GDS compares
favourably with other rating scales and self report measures of depression, for example, Hamilton Rating Scale for Depression (HRSD) and the CES-D; while being easier to administer and complete for elderly people (McDowell, 2006). The GDS has been used in hospital, community / primary care and residential settings (Bowling, 2005), and has good psychometric properties. However, care is needed when interpreting data from the GDS-15 obtained from community and hospital samples, as there is some evidence of lower reliability for this version of the scale outside of residential care settings.

In terms of psychometric development, further research work is needed in the following areas: (1) the issue of detecting minor depression with the GDS (Watson & Pigone, 2003); (2) the use of the GDS for those that are 75 years and older (McDowell, 2006); and (3) the applicability and suitability of the GDS for those with dementia / cognitive impairment. Here the evidence is mixed at best, and restricts the applicability of this instrument to those with milder forms of dementia - though it must be remembered that this scale was not specifically designed for people with dementia.

In terms of research design, future research studies should acknowledge the methodological issue of whether the blinding of research workers is operating when they use the GDS (Wacanta, et al. 2006).

Finally, McDowell (2006) provides an important clinical recommendation which is applicable to all psychiatric measures, namely that a psychiatric interview is required to confirm any classification.

Reporter: Nicholas Marosszeky

Date of report: August 2007

References


The PROQOLID Database (of the MAPI Research Trust in France) was used as an additional source of information for this review (Web-site: http://www.proqolid.org/).
### Appendix 10.3.7: Cornell Scale for Depression in Dementia

**AHOC INSTRUMENT REVIEW SHEET**

<table>
<thead>
<tr>
<th><strong>Title:</strong></th>
<th>Cornell Scale for Depression in Dementia or Cornell Scale for Depression.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Abbreviations:</strong></td>
<td>CSDD, CSD.</td>
</tr>
<tr>
<td><strong>Author(s) Name:</strong></td>
<td>George S. Alexopoulos, M.D.</td>
</tr>
</tbody>
</table>
| **Author(s) Address:** | George S. Alexopoulos, M.D.  
Professor of Psychiatry  
Weill-Cornell Institute of Geriatric Psychiatry  
Weill Medical College of Cornell University  
21 Bloomingdale Road  
White Plains, NY 10605. |
| **Supplied by:** | Copies of the CSDD are widely available in print and on the internet. See the following web-sites: |
| | [http://www.med.cornell.edu/research/galexopoulos/](http://www.med.cornell.edu/research/galexopoulos/) |
| **Cost:** | Need to seek permission to use from Dr Alexopoulos or Elsevier Science. |
| **Training requirements:** | Users need to be familiar with the administration and scoring guidelines.  
Web-site: |
| | [http://www.elsevier.com/wps/find/supportfaq.cws_home/permissionusematerial](http://www.elsevier.com/wps/find/supportfaq.cws_home/permissionusematerial) |
| **Purpose:** | Focuses on the Identification of depressive symptoms and signs in patients with Alzheimer’s Disease and other Dementias (McKeith, et al. 1999). |
| **Administration time:** | 20 minutes (Administration and scoring guidelines) - 30 minutes (20 minutes with carer and 10 minutes with the patient) (Burns, et al. 2004). |
| **Instrument Type:** | Clinical Rating Scale based on an interview.  
The clinical rating is based on semi structured interviews questions with informant and interview questions and signs from the patient. If discrepancies emerge you should re-interview (Administration and Scoring Guidelines). |
The timeframe for the symptom ratings is for the previous week.

The Informant can include nursing staff or a relative. However, they must have frequent contact with the patient (refer Administration and Scoring Guidelines).

**Structure:**
19 item scale. Items grouped under the following headings – Mood Related Signs; Behavioural Disturbance, Physical Signs, Cyclic Functions, Ideational Disturbance.

**Scoring:**
Items are added to give a total score. Item response format: absent, mild or intermittent and severe; plus unable to evaluate. (Scoring 0 – 2).

**Developed for:**
To develop a more suitable method to assess major depression symptoms in dementia patients by obtaining information from the patient and an informant (Alexopoulos, et al. 1988a).

**Normative Data:**
No normative information for the CSDD was found.

A score of 10 or more indicates a probable major depression. Scores above 18 indicate a definite major depression (Source: Administration and scoring guidelines). Papers by Watson, et al. (2003) and Watson, et al. (2006) set the cut score at 7 or more for residents in assisted living facilities.

**Clinical Data:**
 Numerous clinical studies were found, including treatment studies and clinical research into depression and dementia.

**Disease groups:**


**Treatment studies:**

- Bright light therapy in dementia: Lyketsos, et al. (1999a).
- Group psychotherapy for anxiety and depression in mild and moderate dementia; Cheston, et al. (2003).
- Maintaining social relationships in patients with AD via a day care centre: Vespa, et al. (2002).
- Sleep hygiene education, daily walking and increased light exposure treatment (NITE-AD): McCurry, et al. (2005)
- Snoezelen or controlled multisensory stimulation in residential care: van Weert, et al. (2005)

**Drug Treatments:**

- Estrogen skin patch for aggressive behaviour in male patients with AD: Hall, et al. (2005).
Rivastigmine treatment in VD: Moretti, et al. (2002)

Clinical Insights:


To set the context for the use of CSDD clinical data, a brief discussion about the importance of measuring depression in dementia follows:

Burns, et al. (2004) highlights the importance of depression in dementia as it is recognition of a potentially treatable condition (page 9). Cohen, et al. (2003) outlines how depression screening leads to increased rates of treatment (anti-depression medication) in residential care. Teresi, et al. (2002) have also used the CSDD and the GDS in depression recognition studies in residential care. Davidson, et al. (2006) shows how a single educational training session, which included using the CSDD, improved GPs ability to recognise depression in residential care.

A study by Purandare, et al. (2001) suggests that depression symptoms are common in both AD (without depression) and major depression in late life. Irritability, retardation and weight loss are common to both, while sadness, diurnal variation in mood, early or late insomnia differ.

Applications:

Like the GDS, the Cornell Depression Scale was used in the Challenge Depression Project conducted by the Hammond Care Group in 2004 (Website: http://www.health.gov.au/internet/wcms/Publishing.nsf/Content/ageing-chall-depress.htm/$FILE/challenge04.pdf).

The Cornell Depression Scale was used and recommended in the National Trial of the Aged Care Funding Instrument (ACFI) (Web-site: http://www.health.gov.au/internet/wcms/publishing.nsf/content/ageing-acfi-outcome.htm) and the National Framework for Documenting Care in Residential Aged Care Services (NATFRAME) (Web-site: http://www.health.gov.au/internet/wcms/publishing.nsf/content/ageing-rescare-natframe.htm~ageing-rescare-natframe01.htm). The ACFI modified the CSDD to streamline its administration in residential care. It also has provided severity grades.
Like the Geriatric Depression Scale, the CSDD is also in the Silver book of the Royal Australian College of General Practitioners (RACGP) http://www.racgp.org.au/silverbookonline/4-0.asp (Medical care of older persons in residential aged care facilities ('silver book') 4th Edition 2005.

NB: The Cornell Dysthymia Rating Scale (Cohen 1997 and Hellerstein et al 2002) for less severe but chronic depression contains some items and is related to the Hamilton Depression Rating Scale (HAM-D) not the Cornell Depression Scale in Dementia (see Cohen 1997 for further details).

An advantage of the CSDD is that it covers the entire range of severity of dementia. Cummings (2005) endorses this common view saying that the CSDD is “particularly useful because allows rating of depression across the entire range of severity” (page s20). This is supported by research papers by Alexopoulos, et al. (1988b), Ott & Fogel (1992), Muller-Thomsen, et al. (2005) and Korner, et al. (2006). However, a single paper was found by Allen, et al. (2000) which did not support this case. They found that CSDD could not discriminate between depressed and non depressed subject with cognitive impairment. The results of Kurlowicz, et al. (2002) supports the view that depression measurement methods less dependent on co-morbid medical illness, dementia and functional disability are to be preferred.

Detailed psychometric information comparing the CSDD and other noted instruments can be found in the following collections of papers.


- Burrows, et al. (2000) and Hendrix, et al. (2003) also provide data with the CSDD and Minimum Data Set; with the later suggesting that the MDS 2.0 requires more accurate assessment.

**Carer and/or Patient Use of Instrument:**

The CSDD has been used in studies with depressed carers (see Ballard et al 1995 and Nagatomo, et al. 1998). Logsdon & Teri, 1995 support the validity of the use of informant reports by carers (spouse or adult child). The use of informants produces higher cut scores for mild depression on all measures (BDI, HAM-D and CSDD), but the internal properties and correlations with other measures were found to be comparable to the self-report versions of these scales.
## Psychometric Criteria

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal consistency</strong></td>
<td>Alexopoulos, et al. (1988a)</td>
<td>X Alpha &gt;0.70</td>
<td>The original paper by Alexopoulos, et al. (1988a) outlines and alpha value of 0.84 for demented patients (Also reported in Burns, et al. 2004).</td>
</tr>
<tr>
<td></td>
<td>Alexopoulos, et al. (1988b)</td>
<td>□ Marginal or</td>
<td>Using Kuder-Richardson formula internal consistency = 0.98 for non demented patients (Alexopoulos, et al. 1988b).</td>
</tr>
<tr>
<td></td>
<td>Kurlowicz, et al. (2002)</td>
<td>inadequately internal consistency (&lt;0.70)</td>
<td>Additional internal consistency information can be found in Kurlowicz, et al. (2002) and Muller-Thomsen, et al. (2005).</td>
</tr>
<tr>
<td></td>
<td>Burns, et al. (2004)</td>
<td>□ No information found on internal consistency</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Muller-Thomsen, et al. (2005)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Alpha &gt;0.70</td>
<td>The original paper by Alexopoulos, et al. (1988a) outlines and alpha value of 0.84 for demented patients (Also reported in Burns, et al. 2004).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Marginal or</td>
<td>Using Kuder-Richardson formula internal consistency = 0.98 for non demented patients (Alexopoulos, et al. 1988b).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>inadequately internal consistency (&lt;0.70)</td>
<td>Additional internal consistency information can be found in Kurlowicz, et al. (2002) and Muller-Thomsen, et al. (2005).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information found on internal consistency</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Test – retest</strong></td>
<td>Alexopoulos, et al. (1988b)</td>
<td>□ ICC &gt;0.70</td>
<td>No information on test-retest reliability for English speaking samples was found.</td>
</tr>
<tr>
<td></td>
<td>Mack &amp; Patterson (1994)</td>
<td>□ Time intervals and confidence intervals reported</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maixer, et al. (1995)</td>
<td>□ Marginal or</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burns, et al. (2004)</td>
<td>inadequate test-retest reliability ICC&lt;.70</td>
<td>No information on test-retest reliability for English speaking samples was found.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information found on test-retest reliability</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Inter – rater</strong></td>
<td>Alexopoulos, et al. (1998a)</td>
<td>X Agreement</td>
<td>High inter-rater reliability with dementia patients kappa = 0.67 (Alexopoulos, et al. 1998a) This is also reported by Burns, et al. (2004).</td>
</tr>
<tr>
<td></td>
<td>Alexopoulos, et al. (1988b)</td>
<td>reported and</td>
<td>High inter-rater reliability with non demented patients 0.74 (Alexopoulos, et al. 1988b).</td>
</tr>
<tr>
<td></td>
<td>Mack &amp; Patterson (1994)</td>
<td>adequate</td>
<td>Little work has been undertaken in this area apart from the original studies by the authors.</td>
</tr>
<tr>
<td></td>
<td>Maixer, et al. (1995)</td>
<td>□ Inadequate inter-rater agreement</td>
<td>Problems with ratings have been identified by Mack &amp; Patterson (1994), namely, non-anchored scaling and confusing instructions.</td>
</tr>
<tr>
<td>VALIDITY</td>
<td>Studies Reported &amp; References</td>
<td>Adequacy Checks</td>
<td>Comment</td>
</tr>
<tr>
<td>----------</td>
<td>------------------------------</td>
<td>----------------</td>
<td>---------</td>
</tr>
<tr>
<td>Content</td>
<td>Ballard, et al. (1997b) Burns, et al. (2004)</td>
<td>□ Patients and experts were involved during item selection and/or item reduction&lt;br&gt;□ Patients were consulted for reading and comprehension&lt;br&gt;X No patient involvement&lt;br&gt;□ No information found on content validity&lt;br&gt;X There is an adequate coverage of relevant domains&lt;br&gt;□ There is limited coverage of relevant domains</td>
<td>Highly recommended measure (Burns, et al. 2004), especially with the method of administration (patient and informant). A critique by Ballard, et al. (1997b) of the CSDD is that it does not include the persistence of symptoms.</td>
</tr>
<tr>
<td>Construct: Correlation with other measures</td>
<td>Correlations with other measures are reported</td>
<td>There are a number of studies in this area:</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>---------------------------------------------</td>
<td>------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Comparisons made to other measures</td>
<td></td>
<td>Constructs: CSD correlates with:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>physically aggressive behaviour (Lyketsos, et al. 1999b); abnormal behaviour in residential care (Nagatomo, et al. 2001); QoL at one year follow-up (Selwood, et al. 2005); memory complaints (0.33) (Wong, et al. 2006); the need for ADL assistance (Watson, et al. 2006); pain related mood disturbance (Leong &amp; Nuo, 2007).</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>There are a number of studies in this area:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Correlations are reported</td>
<td>Correlations not reported</td>
</tr>
<tr>
<td></td>
<td></td>
<td>There are a number of studies in this area:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Constructs: CSD correlates with:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>physically aggressive behaviour (Lyketsos, et al. 1999b); abnormal behaviour in residential care (Nagatomo, et al. 2001); QoL at one year follow-up (Selwood, et al. 2005); memory complaints (0.33) (Wong, et al. 2006); the need for ADL assistance (Watson, et al. 2006); pain related mood disturbance (Leong &amp; Nuo, 2007).</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>There are a number of studies in this area:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Correlations are reported</td>
<td>Correlations not reported</td>
</tr>
<tr>
<td></td>
<td></td>
<td>There are a number of studies in this area:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Constructs: CSD correlates with:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>physically aggressive behaviour (Lyketsos, et al. 1999b); abnormal behaviour in residential care (Nagatomo, et al. 2001); QoL at one year follow-up (Selwood, et al. 2005); memory complaints (0.33) (Wong, et al. 2006); the need for ADL assistance (Watson, et al. 2006); pain related mood disturbance (Leong &amp; Nuo, 2007).</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>There are a number of studies in this area:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Correlations are reported</td>
<td>Correlations not reported</td>
</tr>
<tr>
<td></td>
<td></td>
<td>There are a number of studies in this area:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Constructs: CSD correlates with:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>physically aggressive behaviour (Lyketsos, et al. 1999b); abnormal behaviour in residential care (Nagatomo, et al. 2001); QoL at one year follow-up (Selwood, et al. 2005); memory complaints (0.33) (Wong, et al. 2006); the need for ADL assistance (Watson, et al. 2006); pain related mood disturbance (Leong &amp; Nuo, 2007).</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>There are a number of studies in this area:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Correlations are reported</td>
<td>Correlations not reported</td>
</tr>
<tr>
<td></td>
<td></td>
<td>There are a number of studies in this area:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Constructs: CSD correlates with:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>physically aggressive behaviour (Lyketsos, et al. 1999b); abnormal behaviour in residential care (Nagatomo, et al. 2001); QoL at one year follow-up (Selwood, et al. 2005); memory complaints (0.33) (Wong, et al. 2006); the need for ADL assistance (Watson, et al. 2006); pain related mood disturbance (Leong &amp; Nuo, 2007).</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>There are a number of studies in this area:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Correlations are reported</td>
<td>Correlations not reported</td>
</tr>
<tr>
<td></td>
<td></td>
<td>There are a number of studies in this area:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Constructs: CSD correlates with:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>physically aggressive behaviour (Lyketsos, et al. 1999b); abnormal behaviour in residential care (Nagatomo, et al. 2001); QoL at one year follow-up (Selwood, et al. 2005); memory complaints (0.33) (Wong, et al. 2006); the need for ADL assistance (Watson, et al. 2006); pain related mood disturbance (Leong &amp; Nuo, 2007).</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>depression across different types of dementia found that depression occurred more often and was more severe in VD patients rather than AD patients. In a follow-up study, Ballard, et al. (1996b) found that for 20% of all patients with depression their symptoms persisted for six months or longer.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finally, the HDRS / HAM-D seems to be better instrument in terms of scalability in cognitively intact and cognitively impaired populations than GDS and Cornell (using Rasch analysis and Mokken coefficient; Korner, et al. 2007).</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Criterion

<table>
<thead>
<tr>
<th>Information on the relationship of scores to gold standard measures or clinical diagnosis is provided</th>
<th>Alexopoulos et al 1988a</th>
<th>Alexopoulos et al 1988b</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ Comparison made to criterion measures</td>
<td>Vida et al 1994</td>
<td>Ballard et al 1997b</td>
</tr>
<tr>
<td>□ No comparison with criterion measures provided</td>
<td>Burns et al 2004</td>
<td>Greenberg et al 2004</td>
</tr>
<tr>
<td>□ No comparison with criterion measures provided</td>
<td>Lam et al 2004</td>
<td>Muller-Thomsen et al 2005</td>
</tr>
<tr>
<td>□ No comparison with criterion measures provided</td>
<td>Korner et al 2006</td>
<td></td>
</tr>
</tbody>
</table>

**X** In terms of criterion validity, the CSDD compares well with diagnostic criteria and HAM-D / HDRS (Burns, et al. 2004), and usually performs better than the GDS.

- The CSDD correlates 0.83 with depression severity levels in demented patients - according to Research Diagnostic Criteria (RDC) (Alexopoulos, et al. 1988a).
- The CSDD also correlates 0.81 (Spearman) with RDC for depression severity levels in demented and non demented patients (Alexopoulos, et al. 1988b).
- The CSDD is equivalent to HDS when using the RDC for major depression in mild to moderate AD (Vida, et al. 1994).
- The CSDD score is similar to DSM-III-R and RDC criteria for major depression – though the issue of the persistence of symptoms is missing (Ballard, et al. 1997b).
- The CSDD is better a screening tool than the GDS when compared to two independent clinicians using the ICD-10. Plus it is equally valid in demented and non-demented patients unlike the GDS (Korner, et al. 2006).
- There is lower agreement between the CSDD and DSM-IV in a palliative care population when compared to...
the GDS. But, as stated by the authors, the CSDD is the only tool which could be used with severe dementia patients (MMSE = 0) (Greenberg, et al. 2004).

The CSDD is the best at detecting depression when compared with the GDS and the Even Briefer Assessment Scale for Depression in Chinese Elderly. The authors recommended a single depression question followed by the CSDD if necessary (Lam, et al. 2004).

Muller-Thomsen, et al. (2005) also found the CSDD is better in detecting depression in mild and moderate to severe AD (using the MMSE as the criterion) than the GDS and NOSGER (Muller-Thomsen, et al. 2005).

See also the related papers by Mayer, et al. (2006) and Korner, et al. (2007) comparing the HAM-D / HDRS and CSDD.

**Interpretability**

The degree to which one can assign qualitative meaning to quantitative scores

Do authors provide the following:

- Presentation of means and SD of scores before and after treatment
- Comparative data on the distribution of scores in relevant subgroups
- Information on the relationship of scores to well-known functional measures or clinical diagnosis
- Information on the association between changes in scores and patients' global ratings of the magnitude of change they have experienced

<table>
<thead>
<tr>
<th>Authors</th>
<th>Information on interpretability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burns, et al. (2004) Challenge Depression Project (see above)</td>
<td></td>
</tr>
<tr>
<td>ACFI data (see above)</td>
<td></td>
</tr>
<tr>
<td>Watson, et al. (2006)</td>
<td></td>
</tr>
</tbody>
</table>

- X Authors provide 2 or more types of information on interpretability
- Authors provide limited information to assist with interpretability
- No information provided

A score of 10 or more indicates a probable major depression. Scores above 18 indicate a definite major depression. (Source: Administration and Scoring Guidelines).

Papers by Watson, et al. (2003) and Watson et al (2006) set the cut score at 7 or more for residents in assisted living facilities.

Burns, et al. (2004) repeats the common view that the cut-score is 8.

Australian data and guidelines are provided in the ACFI national trial data and the Challenge Depression Project.
<table>
<thead>
<tr>
<th>RESPONSIVENESS</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floor and ceiling effects</td>
<td>The questionnaire fails to demonstrate a worse score in patients who clinically deteriorated and an improved score in patients who clinically improved. Authors should provide descriptive statistics of the distribution of scores.</td>
<td>□ Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected. □ Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score. X No information provided on floor and ceiling effects.</td>
<td>No information for the CSDD was found.</td>
</tr>
<tr>
<td>Sensitivity to change</td>
<td>The ability to detect important change over time in the concept being measured.</td>
<td><strong>Mayer, et al. (2006)</strong>&lt;br&gt;See also the Drug treatment studies listed above.</td>
<td><strong>X</strong> Hypotheses were formulated and results were in agreement. □ An adequate metric was used (ES, SRM, comparison with external standard). □ No information on sensitivity to change was provided. □ MCID - Information was provided about the magnitude of score differences which would be clinically meaningful. X MCID – No information was provided.</td>
</tr>
</tbody>
</table>


**Gender Appropriateness:** Appropriate for use with both genders.<br>

**Age Appropriateness:** Although mainly used with elderly people it is appropriate for use with adults.<br>

**Summary:**<br>A widely used and highly respected measure. Burns, et al. (2004) state that the CSDD “sets the standard” in the area of depression measurement in severe dementia when measurement by an informant is required. However, since the original publications of the CSDD, little work has been published on scale’s inter-rater reliability. Up to date, reliability information is of vital importance.
importance if a clinical rating scale is going to be used in routine assessments by different practitioners, across different practice settings.

Report: Nicholas Marosszeky

Date of report: August 2007

References


Appendix 11: Reviews of Measures of Function
Appendix 11.1  Alzheimer's Disease Co-operative Study – Activities of Daily Living Inventory

AHOC INSTRUMENT REVIEW SHEET

Title: Alzheimer’s Disease Co-operative Study – Activities of Daily Living Inventory.

Abbreviations: ADCS-ADL.

Author(s) Name: Douglas Galasko, MD.

Author(s) Address: Department of Neurosciences
University of California - San Diego
9500 Gilman Drive
La Jolla CA 92093-0662
USA.

E-mail: dgalasko@ucsd.edu

Supplied by: A copy of the ADCS-ADL is available at medafile.com website:
http://www.medafile.com/cln/ADCSADL.htm

Cost: Need to seek permission for use from Dr. Galasko.

Training requirements: Minimal.

Little information is available in the scientific literature on the training of interviewers using the ADCS-ADL – though Galasko, et al. (1997) and Galasko, et al. (2005) refer to a training session and a procedures manual.

Purpose: An inventory of informant based items to assess activities of daily living (ADL) and instrumental activities of daily living (IADL) (also known as functional performance) of Alzheimer disease (AD) patients, with particular reference to outcomes in clinical trials (Burns, et al. 2004).

Administration time: 15 minutes (Schneider, 2001).

Instrument Type: Informant / proxy rating scale based on structured interview or questionnaire format. An informant is someone who spends at least 2 days per week in direct contact with the patient (Galasko, et al. 1997, Galasko, et al. 2005).

Structure: The inventory contains 24 items including basic ADL items (like grooming, dressing, walking, bathing, feeding and toileting), as instrumental items (like shopping, preparing meals, using household appliances, keeping appointments, reading) (Schneider, 2001). Informants are asked to rate the patients performance on the items in the inventory during the previous four weeks. For each item score, 0 reflects the inability to perform an activity or the need for extensive help, while the highest score represents complete independence (Feldman, et al. 2006). (Items have either specifically word response categories or general descriptors like “independently”, “with supervision” or “with physical help”).

NB: Some papers say that the ADCS-ADL is a 23 item instrument this is because the item on dressing has 2 parts.

Scoring: 0 – 78 points, with higher scores indicating better functioning.

Developed for: The informant assessment of activities of daily living for those with mild to moderate Alzheimer’s disease.
A group of clinicians created an item pool of activities based on existing scales and new items based on clinical experience concerning what normal elderly people would perform. These were relevant to patients with AD and at all severity levels. The pool included items on “personal care, communicating and interacting with other people, maintaining a household, conducting hobbies and interests, and making judgements and decisions” (page S34; Galasko, et al. 1997). After pre-testing, the inventory of 45 items was tested on 242 community dwelling AD patients (94 males and 148 females). All items were selected based on their performance in the following areas: applicability (as measured by attempting the item at baseline – allowing for the study of gender effects), test-retest reliability, relationship with MMSE defined levels of AD severity, correlation with MMSE scores, and measurement of functional decline at 12 months. 27 items remained from this analysis and a scale containing 23 items that were useful for “sensitively rating ADL abilities and detecting changes over time” for mild to moderate AD was developed (Galasko, et al. 2004, p.1071).

**Normative Data:**

In the original paper on the ADCS-ADL, Galasko et al. (1997) used a group of 64 healthy, high functioning elderly control subjects and found that all scored optimally for each item at baseline, 1 month, 2 months and at 12 months. However, they caution against taking these findings to reflect the expected performance on the scale of the oldest old or the elderly in general.

**Clinical Data:**

As a dementia specific instrument, the evidence base for the ADCS-ADL is derived from clinical trials for people with dementia.

It includes research papers into the latest pharmacological treatments like Cholinesterase (ChE) inhibitors (Donepezil, Galantamine) and N-methyl-D-aspartate receptor antagonists (NMDA) (Memantine) or in combination (Memantine + stable Donepezil):


Additional papers into other drugs like nonsteroidal anti-inflammatory drugs (NSAIDs) (Aisen, et al. 2003, Reines, et al. 2004), statins (Sparks, et al. 2005) and antioxidants (Thal, et al. 2003) have also used the ADCS-ADL but they did not find any treatment effects when compared to placebo. Aisen, et al. (2003) did use an interesting methodology though, of using a 15 point decline on the ADCS-ADL as a significant end-point for the study.

For further information there are a number of review papers in the scientific literature (see Potkin, 2002, Malouf & Birks, 2004, Craig & Birks, 2006, Loy & Schnieder, 2006, Kirby, et al. 2006).

**Studies in other related clinical areas using the ADCS-ADL:**

Applications:

Galasko and colleagues have adapted the scale for those patients with moderate to severe AD (MMSE scores 0 – 15). They have created a 19 item version containing a set of relevant ADL and IADL items for this group of patients (score range 0 – 54). It is known as the ADCS-ADL severe (ADCS-ADLsev) or ADCS-ADL19. This paper was published in 2005 and used similar methods and the same 45 item dataset in its development as in the original study – Galasko, et al. (1997). This version of the ADCS-ADL has been used in a number of clinical studies with moderate to severe AD patients, including: Reisberg, et al. (2003), Doody, et al. (2004), Tariot, et al. (2004), van Dyck, et al. (2006) and Winbald, et al. (2006). Test-retest reliability for the severe inventory is in the range of 0.89-0.94 (Spearman) (Galasko, et al. 2005). The mean baseline score for this group of patients (MMSE score 0 – 15) was 25.4, with declines of 5.6 and 10.3 points at 6 and 12 months respectively.

Using work from the work from the LASER-AD study in UK, Livingston and colleagues (Livingston, et al. 2004 and Livingston, et al. 2006) have developed a dependency model of AD from the ADCS-ADL and related it to the costs of care. They performed a factor analysis on the ADCS-ADL after removing the six ADL (or basic ADL) items, they found a two factor structure for the remaining 17 IADL items (domestic activities and communication activities). From this analysis, they found three groups non-dependent, non dependent but with instrumental functional disability, and dependent. These levels were significantly correlated with cognition (MMSE, ADAS-Cog, SIB), global measures (CIBIC-Plus), behavioural measures (NPI), quality of life (QOL-AD), health status (HSQ-12), residential setting and the costs of care. In terms of cost drivers, they found 43% of the variance could be determined by institutionalisation and dependency level. (In a cost-effectiveness study of the drug memantine, Jones et al. (2004), also used the ADCS-ADL and residential status as a measure of dependency).

In 2006, Livingston, et al. (2006) describe health status change for a group of AD patients over a 6 month period and found a significant decline of 4.4 points on ADCS-ADL score for the moderately severe – severe group (using MMSE groupings).

NB: The Alzheimer’s Disease Functional Assessment and Change Scale (ADFACS) contains 3 IADL items from the ADCS-ADL (Mohs et al. 2001).

Carer and/or Patient Use of Instrument: For use with carers of people with dementia.
### Psychometric Criteria

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td>Galasko, et al. (1997)</td>
<td>□ Alpha &gt;0.70 □ Marginal or inadequate internal consistency (&lt;0.70) X No information found on internal consistency</td>
<td>Information was not found on this aspect of the inventory.</td>
</tr>
<tr>
<td></td>
<td>Schneider (2001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test–retest</td>
<td>Galasko, et al. (1997)</td>
<td>X ICC &gt;.70 Time intervals and confidence intervals reported □ Marginal or inadequate test-retest reliability ICC&lt;.70 □ No information found on test-retest reliability</td>
<td>In their item work-up, Galasko, et al. (1997), found that the test-retest reliability for the selected items fell in the range of 0.4 – 0.75 (Kappa) measured at baseline and at one and two months. From this data, Schneider (2001) and Burns, et al. (2004) report that the ADCS-ADL has good test-retest reliability. Recently Galasko, et al. (2004) found test-retest reliability of 0.91 (spearman) for the total ADCS-ADL score, over 4 week period for placebo controls.</td>
</tr>
<tr>
<td></td>
<td>Schneider (2001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inter–rater</td>
<td>Galasko, et al. (1997)</td>
<td>□ Agreement reported and adequate □ Inadequate inter-rater agreement X No information provided</td>
<td>Information was not found on this aspect of the inventory.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VALIDITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>Galasko, et al. (1997) Data reanalysed in Galasko, et al. (2005)</td>
<td>X Patients and experts were involved during item selection and/or item reduction □ Patients were consulted for reading and comprehension □ No patient involvement □ No information found on content validity X There is an adequate coverage of relevant domains</td>
<td>The 45 items pool was selected by a panel of experts from existing and newly developed items. Informants / carers were questioned as to whether patients could perform (or attempted to perform) the tasks or whether they did so premorbidly (e.g. doing the laundry). The final items were selected if they were undertaken by 90% of the sample (AD and control) (Galasko, et al. 1997).</td>
</tr>
<tr>
<td>Construct</td>
<td>The extent to which scores on the questionnaire relate to other measures in a manner that is consistent with theoretically derived hypothesis concerning the domains that are measured</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Construct: Internal Structure</td>
<td>Information provided on factor structure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Construct: Correlation with other measures</td>
<td>Comparisons made to other measures</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Source | Results | Further validation work needs to be undertaken in comparing this informant measure with other dementia specific measures of function, especially clinical rating scales and direct measures of functional performance. |
| Doody, et al. (2004) and Feldman, et al. (2006) | □ Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used □ Limited /inadequate construct validity reported □ No information provided |
| Galasko, et al. (2005) and Feldman, et al. (2006) | For the ADCS-ADL: |
| Galasko, et al. (1997) | The original paper by Galasko, et al. (1997) reported correlations with MMSE scores for each selected item of between 0.4 and 0.7 (Spearman). |
| Reisberg, et al. (2003) | Hoe et al. (2005) found a correlation of 0.55 (Spearman) with the QOL-AD (for patients with people who have a MMSE score of between 3 and 11). |
| Hoe, et al. (2005) | |
| Galasko, et al. (2005) | |
| Livingston, et al. (2006) | |
| Walker, et al. (2006) | |
Walker, et al. (2006) found a correlation with functional impairment and a carers implementation of safety measures to prevent dangerous behaviour by person with dementia (linear regression).

See also the work of Livingston et al. (2004) and Livingston et al. (2006) described above.

For the ADCS-ADLsev:

Reisberg, et al. (2003) and Doody, et al. (2004) relate the severity inventory to the FAST.

Galasko, et al. (2005) also describes change score correlations with the MMSE, CDR, GDS and SIB.

No information on discriminant validity | The ADCS-ADL distinguishes between the stages of severity of AD patients from mild to severe (Schneider, 2001), as defined by the MMSE (Galasko, et al. 1997). |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Criterion</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Information on the relationship of scores to gold standard measures or clinical diagnosis is provided  
□ Comparison made to criterion measures  
X No comparison with criterion measures provided |  
Further work needs to be undertaken with other dementia specific measures of function (e.g. DAD), clinical rating scales and direct measures of functional performance. |  |
□ Authors provide limited information to assist with interpretability  
□ No information provided | The inventory is based on longitudinal data examining functional decline at 12 months (Galasko, et al. 1997). Galasko, et al. (2005) reports a mean baseline total score of 25.4 for the ADCS-ADLsev, and changes of 5.6 and 10.3 points at six and twelve months respectively.  
See also the item analyses by Doody, et al. (2004) and Feldman, et al. (2006) described above; and the papers in the Sensitivity to Change section below. |
### Association between changes in scores and patients’ global ratings of the magnitude of change they have experienced

### Responsiveness

<table>
<thead>
<tr>
<th>Floor and ceiling effects</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>The questionnaire fails to demonstrate a worse score in patients who clinically deteriorated and an improved score in patients who clinically improved</td>
<td>Galasko, et al. (1997) Galasko, et al. (2005)</td>
<td>X Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
<td>Galasko, et al. (1997) examined whether an item was performed (or attempted) by the patient, as rated by the informant/carers. They set the criteria for item selection at 90% performance to gain wide applicability. Items were also assessed across MMSE severity levels to examine suitability with regard to possible floor effects. One of the requirements for item selection for the ADCS-ADLsev required that at least 50% of group of severe AD patients performed the item meaningfully at baseline (Galasko, et al. 2005).</td>
</tr>
<tr>
<td>Authors should provide descriptive statistics of the distribution of scores</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sensitivity to change</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>□ An adequate metric was used (ES, SRM, comparison with external standard)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information on sensitivity to change was provided</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ MCID - Information was provided about the magnitude of score differences which would be clinically meaningful</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ MCID – No information was provided.</td>
<td></td>
</tr>
</tbody>
</table>
However, it should be noted that one recent study by Peskind, et al. (2006) into Memantine in mild to moderate AD did not find a significant treatment effect compared to placebo at 24 weeks. This was in contrast to significant results for measures of cognition, behaviour and global status.

The ADCS-ADL has also been shown to be sensitive to change:

Reisberg, et al. (2003) shows that moderate to severe AD patients on Memantine treatment have a better outcome than those on placebo.

Tariot, et al. (2004) also found significant results for Memantine and stable Donepezil treatment vs. placebo. This data was also subjected to responder analysis by van Dyck, et al. (2006).

Significant effects of Memantine vs placebo were shown in the item analysis of Doody, et al. (2004). This was also shown in the sophisticated item analysis of Feldman, et al. (2006) examining Memantine with stable Donepezil treatment.

Winblad, et al. (2006) in a placebo controlled trial showed less functional decline / preserved functioning for nursing home patients taking Donepezil for 6 months.

Cultural Applicability and Cultural Adaptations: Some clinical trials using the ADCS-ADL were noted to have occurred in Spain (Arrieta, 2006), Sweden (Winblad, et al. 2006), Latvia (Doody, et al. 2004) and Bosnia and Hercegovina (Rustembegovi, et al. 2003).

Gender Appropriateness: Appropriate for use with both genders. The original paper (Galasko, et al. 1997) designed detailed items (e.g. obtaining a snack or meal, not just using headings like cooking or preparing meals) and retained items for use with both men and women.

Age Appropriateness: Appropriate for use with older adults.

Summary: The ADCS-ADL is a psychometrically well designed instrument for measuring decline in functional performance in clinical trials. While shown
to be sensitive to change in a number of studies, further development work is needed in the areas of inter-rater reliability and construct / criterion validity. In particular, this informant / proxy rating inventory needs to be compared to other dementia specific measures of function, clinical rating scales and the direct assessments of function (eg. performance and timed tests). A wider application beyond clinical drug trials to other settings like hospitals and nursing homes is also required. These studies should examine the performance of the ADCS-ADL in dementia patients with other co-morbid conditions.

Reporter: Nicholas Marosszeky

Date of report: September 2007

References


Appendix 11.2 Disability Assessment for Dementia

AHOC INSTRUMENT REVIEW SHEET

Title: Disability Assessment for Dementia.

Abbreviations: DAD.

Author(s) Name: Louise Gauthier & Isabelle Gelinas.

Author(s) Address: Professor Gelinas
School of Physical & Occupational Therapy
McGill University
3654 Prom Sir William Osler
Montreal, Quebec
H3G 1Y5
Canada.

Email: isabelle.gelinas@mcgill.ca

Supplied by: A copy of the scale is available at the PROQOLID website.

Cost: Need to seek permission to use from Dr. Gelinas.

Training requirements: Minimal.

The PROQOLID website reports that there is a DAD Guide. This document and the original paper (Gelinas, et al. 1999) should be consulted.

Purpose: To develop an appropriate and psychometrically valid assessment of functional disability for caregivers of community-dwelling persons with AD (Gelinas, et al. 1999).

Administration time: Less than 15 minutes (Burns, et al. 2004).

Instrument Type: Informant / proxy rating scale based on interview.

Structure: The DAD consists of 40 items, 17 related to self-care and 23 involving instrumental activities of daily living, including leisure activities (Gelinas, et al. 1999, Burns, et al. 2004). The DAD covers the following domains or areas: hygiene, dressing, continence, eating, meal preparation, telephoning, going on an outing, finance and correspondence, medications and leisure and housework. A key feature of the DAD is that the activities items are broken up into sub-domains that assess the cognitive components involved, namely initiation (13 items), planning / organisation (10 items) and performance (17 items).

Scoring: Scoring each item is based on a Yes = 1 or No = 0, dichotomous format. A Yes score reflects the person's ability to do the activity without help or reminder during the past two weeks. Not applicable (NA) is also an acceptable response. To obtain the total score, as well as each domain and sub-domain, answers to applicable items are summed and then divided by the number of applicable items, and then multiplied by 100 to give a percentage score between 0 and 100. For all scores, a higher score indicates less disability in performing activities of daily living, while a lower score indicates more disability in performing activities of daily living. Scores can also be broken up into ADLs and IADLs. Feldman, et al. (2001b) and Feldman, et al. (2005b) provide examples of how to apply the DAD scores.

Developed for: The assessment of functional abilities and disabilities of people with dementia.
Normative Data:
No information on normative data was found.

Clinical Data:
Feldman, et al. (2001b) have published mild to moderate AD, placebo control data (n = 144) showing average rates of decline in DAD scores at 6 and 12 months. They found an average rate of decline of 12 points from the total score over 12 months. Other findings included:

- Regression analysis suggested that it is not possible to predict decline on an individual basis
- Variability in DAD subdomains – initiation, planning / organisation and performance - requiring caution when interpreting
- Men had a greater number of NA answers than women at baseline
- Greater decline noted in the cognitively impaired (MMSE) and those with preserved IADLs
- Roughly 25% of patients will remain unchanged or improved at 12 months

In a new study which added to this sample, Feldman et al. (2005b) found similar results over 12 months, however, it should be noted that both studies used a previous, 46 item version of the DAD. In this new study, some hierarchical patterns in functional decline were also found, and the relationship between DAD items and formal and informal care was explored.

As a dementia specific instrument, the evidence base for the DAD is based on clinical trials for people with dementia. These include:

- Rivastigmine: Karaman, et al. (2005)

One multinational trial that has compared the effects of Galantamine and Donepezil for mild to moderate AD by Jones, et al. (2004) found that Donepezil was superior - though it should be noted that this was an industry sponsored trial (Warner, 2004).

Finally Potkin (2002) comments “Early experience with this scale (DAD) in clinical trials indicates that it is both valid and reliable” (page 10).

Applications:
In terms of applications, the DAD has been used in immunization studies using human aggregated Abeta42 (Hock, et al. 2003, Bayer, et al. 2005) where DAD scores showed less decline in those with a positive antibody response versus controls (Bayer, et al. 2005). This instrument has also been used in studies examining the effects of APOE epsilon4 genotype (Wilcock, et al. 2000, Aerssens, et al. 2001, Suh, et al. 2006).

More generally, the relationship between functional performance as measured by DAD scores and depression and apathy in AD has been investigated by Lam, et al. (2007).

In terms of disease types: Erkinjuntti, et al. (2003) has applied the DAD to Vascular Dementia and cerebrovascular disease patients and Mioshi, et al. (2007) has applied the instrument to those with frontotemporal dementia. While Pantoni, et al. (2006) has compared DAD scores to increasing leukoaraiosis severity (white matter lesions) in the elderly population.
Finally, most studies using the DAD are about people with AD living in the community (see for example Feldman, et al. 2001b), though Lam et al. (2006) has used the DAD in nursing home residents.

**Carer and/or Patient Use of Instrument:**

A carer is interviewed about their care recipient who has dementia. Gelinas et al. (1999) recommend interviewing the caregiver away from the patient.

Importantly, a number of studies have used the DAD to examine carer burden or load. These include: Feldman, et al. (2003) which showed significantly less decline on ADLs and IADLs for people with moderate to severe AD on Donepezil versus placebo controls. The carers reported spending less time in ADL assistance, as well as being less stressed. Similar findings were found by Blesa (2000) with the drug Galantamine. Feldman, et al. (2005b) also found that a decreasing DAD score over 12 months was associated with increasing amount of paid and unpaid carer support.

**Psychometric Criteria**

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal consistency</strong></td>
<td>Gelinas, et al. (1999)</td>
<td>X Alpha &gt;0.70</td>
<td>Alpha = 0.956 (Gelinas, et al. 1999). Burns, et al. (2004) reports “a high degree of internal consistency and excellent inter-rater and test-retest reliability.”</td>
</tr>
<tr>
<td>The extent to which items in a (sub) scale are inter-correlated; a measure of the homogeneity of a (sub)scale Cronbach’s alpha should be between 0.70 and 0.90 for every dimension / sub-scale</td>
<td>Burns, et al. (2004)</td>
<td>□ Marginal or inadequate internal consistency (&lt;0.70) □ No information found on internal consistency</td>
<td></td>
</tr>
<tr>
<td><strong>Test – retest</strong></td>
<td>Gelinas, et al. (1999)</td>
<td>X ICC &gt;.70</td>
<td>ICC = 0.96 (n=45)(CI: 0.90 – 0.97) (1 week apart) (Gelinas, et al. 1999).</td>
</tr>
<tr>
<td>The extent to which the same results are obtained on repeated administrations of the same questionnaire when no change in physical functioning has occurred Calculation of an intraclass correlation coefficient (ICC); and an ICC &gt; 0.70 is desired Preferred if time interval and confidence intervals were presented</td>
<td></td>
<td>ICC = 0.96 (n=45)(CI: 0.90 – 0.97) (1 week apart) (Gelinas, et al. 1999).</td>
<td></td>
</tr>
<tr>
<td><strong>Inter – rater</strong></td>
<td>Gelinas, et al. (1999)</td>
<td>X Agreement reported and adequate □ Inadequate inter-rater agreement □ No information provided</td>
<td>ICC = 0.95 (CI: 0.90 – 0.97) (n=31, 2 raters) (Gelinas, et al. 1999).</td>
</tr>
<tr>
<td>Limits of agreement, Kappa, or standard error of measurement (SEM) were presented</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VALIDITY</td>
<td>Studies Reported &amp; References</td>
<td>Adequacy Checks</td>
<td>Comment</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------------------</td>
<td>----------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Content</td>
<td>Gelinas, et al. (1999)</td>
<td>X Patients and experts were involved during item selection and/or item reduction</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Patients were consulted for reading and comprehension</td>
<td>Instrument developed with the assistance of a panel of experts and caregivers (see above) (Gelinas, et al. 1999).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No patient involvement</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information found on content validity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>X There is an adequate coverage of relevant domains</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ There is limited coverage of relevant domains</td>
<td></td>
</tr>
<tr>
<td>Construct</td>
<td>Blesa (2000) Feldman, et al. (2003) Feldman, et al. (2005b)</td>
<td>X Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used</td>
<td>Most of the work here is in the area of carer burden or load.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Limited /inadequate construct validity reported</td>
<td>Feldman, et al. (2003) also showed that treatment changes in DAD scores were associated with carers spending less time in assisting patients with activities of daily living. This was similar to a finding in a study by Blesa (2000).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ No information provided</td>
<td>Feldman, et al. (2005b) found that decreasing DAD scores over a 12 month period for placebo controls with mild to moderate AD disease was associated with a increase in carer time (paid and unpaid) – using a previous, 46 item version of the DAD.</td>
</tr>
<tr>
<td>Construct: Internal Structure</td>
<td>No evidence provided/failed a test of dimensionality</td>
<td>No information was found.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Some evidence provided to support internal structure</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>□ Substantial evidence provided to support internal structure</td>
<td></td>
</tr>
</tbody>
</table>
### Construct: Correlation with other measures

Comparisons made to other measures

|-----------------------|------------------------|---------------------|-------------------|------------------------|-------------------|-------------------|

**X Correlations with other measures are reported  □ Correlations not reported**

- The DAD correlates - 0.70 with the GDS – Reisberg (Gelinas, et al. 1999).
- The DAD also correlates 0.54 with the MMSE (Gelinas, et al. 1999).
- The DAD correlates -0.52 or -0.57 with the ADASCog (Feldman, et al. 2001b; Feldman, et al. 2005b) – using a previous version of the DAD (46 items).
- Also change scores in DAD correlate with Clinical Global Impression change scores (CIBIC) and ADASCog change scores as noted by Feldman, et al. (2001b).
- The relationship between DAD scores and apathy and depression has been investigated by Lam, et al. (2007).
- Liu, et al. (2004) did not find a relationship between visuospatial neglect and performance on the DAD. This suggests further examination of this issue is required.
- Mok, et al. (2005) provides comparison data with the MMSE and Modified Barthel Index for the Chinese version of the scale.
- Caro, et al. (2002) found that cognitive score (ADASCog or MMSE) is a strong predictor of the DAD total score (transformed) in a group of mild to moderate AD patients (n= 1289) involved in clinical trials taken at the same time point.

### Construct: Discriminant Validity

The scale differentiates between relevant categories of respondents e.g. sick vs. well, varying degrees of severity

|-----------------------|------------------------|-------------------|

**X Scale differentiates between relevant categories of respondents  □ No information on discriminant validity**

- The DAD is valid across Global Deterioration Scale stages (Gelinas, et al. 1999).
- DAD scores (and severity) are related to living situation (living alone, living with spouse, living with children or other, living in a nursing home) (Sibley, et al. 2002).
- Caro, et al. (2002) also found group differences on DAD scores were also found to differ according to the amount of supervision per day (> or <
<table>
<thead>
<tr>
<th>Criterion</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information on the relationship of scores to gold standard measures or clinical diagnosis is provided</td>
<td>Feldman, et al. (2001b)</td>
<td>□ Comparison made to criterion measures</td>
<td>Further work needs to be undertaken with comparison to other dementia specific measures of function (e.g. ADCS-ADL), clinical rating scales and direct measures of functional performance.</td>
</tr>
<tr>
<td>Interpretability</td>
<td>Suh, et al. (2004a)</td>
<td>X Authors provide 2 or more types of information on interpretability</td>
<td>See the discussion of Feldman, et al. (2001b) and Feldman, et al. (2005b) in the Clinical Data section above.</td>
</tr>
<tr>
<td>The degree to which one can assign qualitative meaning to quantitative scores</td>
<td>Feldman, et al. (2005b)</td>
<td>□ Authors provide limited information to assist with interpretability</td>
<td>Suh, et al. (2004a) in their Korean sample (n = 107) found an average rate of decline of 15 points at 12 months, with linear declines in IADLs and planning / organisation and performance sub-domain scores (as the MMSE score declines) and a curvilinear relationship for ADLs and initiation scores.</td>
</tr>
<tr>
<td>Do authors provide the following:</td>
<td></td>
<td>□ No information provided</td>
<td></td>
</tr>
<tr>
<td>Presentation of means and SD of scores before and after treatment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparative data on the distribution of scores in relevant subgroups</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information on the relationship of scores to well-known functional measures or clinical diagnosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information on the association between changes in scores and patients’ global ratings of the magnitude of change they have experienced</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**RESPONSIVENESS**

**Floor and ceiling effects**

The questionnaire fails to demonstrate a worse score in patients who clinically deteriorated and an improved score in patients who clinically improved

Authors should provide descriptive statistics of the distribution of scores

<table>
<thead>
<tr>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gelas, et al. (1999)</td>
<td>X Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
<td>Gelas, et al. (1999) provides descriptive information on the distribution of DAD scores.</td>
</tr>
<tr>
<td>Feldman, et al. (2001b)</td>
<td>□ Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score □ No information provided on floor and ceiling effects</td>
<td>Feldman, et al. (2001b) provides baseline score distribution data for DAD scores for placebo control; mild to moderate AD patients – using a previous, 46 item version of the DAD.</td>
</tr>
</tbody>
</table>
### Sensitivity to change

<table>
<thead>
<tr>
<th>Study</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raskind, et al.</td>
<td>(1999)</td>
</tr>
<tr>
<td>Farlow &amp; Cyrus</td>
<td>(2000)</td>
</tr>
<tr>
<td>Blesa, (2000)</td>
<td></td>
</tr>
<tr>
<td>Gelinas, et al.</td>
<td>(2000)</td>
</tr>
<tr>
<td>Raskind, et al.</td>
<td>(2000)</td>
</tr>
<tr>
<td>Wilcock, et al.</td>
<td>(2000)</td>
</tr>
<tr>
<td>Aerssens, et al.</td>
<td>(2001)</td>
</tr>
<tr>
<td>Feldman, et al.</td>
<td>(2001a)</td>
</tr>
<tr>
<td>Gauthier, et al.</td>
<td>(2002)</td>
</tr>
<tr>
<td>Suh, et al. (2004b)</td>
<td></td>
</tr>
<tr>
<td>Feldman, et al.</td>
<td>(2005a)</td>
</tr>
<tr>
<td>Karaman, et al.</td>
<td>(2005)</td>
</tr>
<tr>
<td>Behl, et al. (2006)</td>
<td></td>
</tr>
<tr>
<td>Rockwood, et al.</td>
<td>(2006)</td>
</tr>
<tr>
<td>Johannsen, et al.</td>
<td>(2006)</td>
</tr>
</tbody>
</table>

X Hypotheses were formulated and results were in agreement
- An adequate metric was used (ES, SRM, comparison with external standard)
- No information on sensitivity to change was provided
- MCID - Information was provided about the magnitude of score differences which would be clinically meaningful

| Sensitivity to change was observed in DAD scores for the following drug treatment studies: |

Karaman, et al. (2005) found less functional decline in DAD scores for those moderate AD patients taking Rivastigmine versus controls.

DAD scores have shown the beneficial effects of the drug Metrifonate in mild to moderate AD (Raskind, et al. 1999, Farlow & Cyrus, 2000, Gelinas, et al. 2000).

In a group of studies, Feldman, et al. (2001a), Feldman, et al. (2003) and Feldman, et al. (2005a) used the DAD and reported significant treatment effects (i.e. slowing of the rate of decline) on the instrument for Donepezil medication versus placebo controls at 24 weeks, in moderate to severe AD patients. Gauthier, et al. (2002) also found similar results in a sub-group of moderate AD patients. In a separate study, Johannsen, et al. (2006) found a trend with mild to moderate AD patients.

Studies using Galantamine with mild to moderate AD patients showed significantly better outcomes than placebo controls on the DAD at 5 months (Blesa, 2000), and at 6 months (Raskind, et al. 2000, Wilcock, et al. 2000) and at 12 months (Blesa, 2000; Raskind et al. 2000; Blesa, et al. 2003). Similar results were found for patients with vascular dementia and cerebrovascular disease at 12 months by Erkinjuntti, et al. (2003). Suh, et al. (2004b) also found a dose response relationship for the drug using the DAD instrument. However, a recent study by Rockwood, et al. (2006) did not find any significant differences in DAD scores in comparison with a placebo group at 4 months for mild to moderate AD patients on Galantamine, in contrast to other measures like the ADASCog, CIBIC and the clinician rating of goal attainment.

In a study examining APOE, Aerssens, et al. (2001) found
an annual rate of decline of over 11 points over 12 months for placebo treated AD patients.

New research on second generation cholinesterase inhibitors found an effect size of 0.4-0.8 for the DAD scores between treated and non-treated mild to moderate AD patients at 12 months (Behl, et al. 2006).

Cultural Applicability and Cultural Adaptations: The DAD was developed for the English and French languages (Gelinas, et al. 1999). Korean (Suh, et al. 2004b) and Chinese (Mok, et al. 2005) versions are available; and papers have been noted to have been undertaken in Finland, Germany, Norway (Jones, et al. 2004), Denmark (Stokholm, et al. 2005) and Hong Kong (Lam, et al. 2006).

Gender Appropriateness: Appropriate for use with both genders. No sex differences were noted by Gelinas, et al. (1999).

Age Appropriateness: It is appropriate for use with adults of 18 years and over (PROQOLID).

Summary: The DAD is a logically developed and reliable measure for assessing functional disability in dementia patients and has been used in a number of clinical trials. While shown to be sensitive to change in a number of studies, further development work is needed in the areas of internal structure (i.e. factor analysis of the whole scale, and the value of the cognitive component sub-domains) and construct / criterion validity. In particular, this informant / proxy rating inventory needs to be compared to other dementia specific measures of function, clinical rating scales and the direct assessments of function (eg. performance and timed tests). This view is supported by comments made by Feldman, et al. (2003) about the DAD, and by Hancock & Charlesworth (2004) who state: “the validity of subjective ratings by caregivers has not been tested”.

Reporter: Nicholas Marosszeky

Date of report: September 2007

References


Assessment scales in old age psychiatry (2nd ed.). Taylor & Francis, London.


Craig D and Birks J (2006) 
Galantamine for vascular cognitive impairment. Cochrane Database of Systematic Reviews. No.1, pp.CD004746.


Farlow MR and Cyrus PA (2000) 


Feldman H, Gauthier S, Hecker J, et al. (2001a) 

Feldman H, Gauthier S, Hecker J, et al. (2005a) 

Feldman H, Sauter A, Donald A, et al. (2001b) 


*The PROQOLID Database (of the MAPI Research Trust in France) was used as an additional source of information for this review (Web-site: http://www.proqolid.org/).*
Appendix 11.3  Cleveland Scale for Activities of Daily Living

AHOC INSTRUMENT REVIEW SHEET

Title: Cleveland Scale for Activities of Daily Living.

Abbreviations: CSALD.

Author(s) Name: Marian B. Patterson, PhD & James L. Mack, PhD.

Author(s) Address: Dr. Patterson
2520 Fairmont Boulevard
Cleveland Heights, OH 44106
USA.

Dr. Mack
Case Western Reserve School of Medicine
29017 Lincoln Road
Bay Village, OH 44140-1942
USA.

Supplied by: A copy of the CSADL is in the paper by Patterson & Mack (2001).

Cost: Need to seek permission for use from Dr. Patterson.

Training requirements: Minimal. One needs to obtain the detailed training manual of Patterson & Mack (1998).

Purpose: A scale designed to measure in detail specific activities of daily living in individuals with dementia (Patterson & Mack, 2001).

Administration time: 10–20 minutes (Patterson & Mack, 2001).

Instrument Type: Informant / proxy rating scale based on interview.

Structure: The scale consists of 47 items covering 15 domains of everyday activity, including physical and instrumental ADL (Patterson & Mack, 2001) plus another item about other problem behaviours. Domains include; Bathing, Toileting, Personal Hygiene and Appearance, Dressing, Eating, Mobility, Medications, Shopping, Travel, Hobbies, Personal Interests, Employment, Housework / Home Maintenance, Telephone, Money Management, Communication Skills, Social Behaviour. In each domain “activities are broken down into several component behaviours” (Patterson & Mack, 2001). Some items reflect controlled or automatic processes.

For example the bathing item is broken down into its component parts: it includes four items: (1) initiates bath or shower with appropriate frequency and at appropriate times, (2) prepares bath / shower (draws water of proper temperature, ensures soap and towel are present, etc.), (3) gets in and out of tub or shower, and (4) cleans self.

“The scale was designed to evaluate dependency of patients in a variety of settings (including home, the community, and residential facilities) and to provide a comprehensive picture of dependency throughout the day, regardless of when it occurs. To accomplish these goals, dependency ratings are obtained from the patient’s primary caregiver, the individual most likely to have remained in proximity to the patient over an extended period of time.” (page 15) (Patterson & Mack, 2001).

Scoring: Primary carers / proxy informants are asked whether the patient performs the activity / item. If not, they are scored as dependent. For items like meal preparation, the informant is asked if the patient would be able to perform...
the activity if given the opportunity, as well as premorbid behaviour on the
task. (These items may reflect premorbid dependency and require special
questioning users are referred to in the CSADL manual).

Carers are not required to distinguish between types of dependency (i.e.
between the need for supervision or the need for help / assistance)
(Patterson, et al. 1992). Carers are asked to rate each activity or behaviour
based on the following standard response categories:

0 = Behaviour is carried out effectively, quite independently, without
direction or help
1 = Usually independent but sometimes or in some situation needs direction
or help
2 = Usually requires some direction or help, but sometimes or in some
situations is independent
3 = Always requires direction or help – behaviour is never carried out
independently
9 = Cannot rate because of insufficient information (Mack & Patterson
2001)

Three items are not scored: takes medications, works for pay and other
problem behaviours.

There are two types of total score, those weighted on dependency data
(Total Weighted Score) (TWS) (0 – 135), or those based on whether the
item is scored independent or dependent (Total Items Dependent) (TID) (0
– 45) (Mack & Patterson, 2001).

A truncated, 35 item version of the scale that does not use the items
requiring special questions of items about prior dependency can also be
applied. Also up to 5 items in the 45 item version can be missing or using
special questioning in order to derive the total score (Mack & Patterson,
2001).

Developed for:
The scale is designed to measure specific activities of daily living in
individuals with dementia.

Items were drawn from physical and instrumental items from the
OARSFAQ, the FAQ (Pfeffer) plus other more complex activities (hobbies,
communication skills, social behaviour) (Patterson, et al. 1992). Patterson,
et al. (1992) reports on the development and clinician review of the items
and pre-testing with AD patients (n = 113) examining 5 domains: eating,
bathing, dressing, toileting, hygiene using 22 ADL items. Patterson & Mack
(2001) report that the number of items in the original version of the scale
was reduce from 66 to 48 items with further wording refinement.

Normative Data:
Normative data on healthy elderly patients (n = 199) and physically
impaired patients with osteoarthritis / degenerative joint disease (n = 26) is
provided in the paper by Patterson & Mack (2001). These were based on a
direct interview with the person (i.e. not informant / proxy) and sometimes
they were conducted over the telephone.

Clinical Data:
Most clinical data on the CSADL is based on people with dementia living in
the community (see Patterson, et al. 1992, Patterson & Mack, 2001, Mack
& Patterson, 2006).

Fritsch, et al. (2002) in a large sample of possible / probable AD patients
found that find educational attainment slowed the rate of cognitive decline,
as measured by the MMSE. They found no effects of education attainment
or occupational status on the rate of functional decline using the CSADL
(analysis = multilevel analysis of repeated measures). Though Fritsch, et al.
(2002) did find an effect for ethnicity on functional decline.

Mack & Patterson (2006) now describe the scale as containing 16 instead of 15 activities of everyday living (ADLs + IADLs) with meal preparation being distinct from eating.

Patterson & Mack (2001) and Mack & Patterson (2006) have scored the CSADL based on telephone interviews.

Carer and/or Patient Use of Instrument: It is generally designed for use with an informant/carer although it has also been used with a direct interview with the person with dementia (see above). Spector (1997) notes some concerns with the wording of items for informants / carers, that they may not understand the meaning of terms like “initiates”, “appropriateness” and “acceptable”. Following these comments, Patterson & Mack (2001) responded by arguing that these terms are defined in the instruction manual.

Psychometric Criteria

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td>Patterson &amp; Mack (2001)</td>
<td>X Alpha &gt;0.70</td>
<td>Cronbach’s alpha = 0.97 (Patterson &amp; Mack, 2001).</td>
</tr>
<tr>
<td>The extent to which items in a (sub) scale are inter-correlated; a measure of the homogeneity of a (sub)scale</td>
<td></td>
<td>□ Marginal or inadequate internal consistency (&lt;0.70)</td>
<td></td>
</tr>
<tr>
<td>Cronbach’s alpha should be between 0.70 and 0.90 for every dimension / sub-scale</td>
<td></td>
<td>□ No information found on internal consistency</td>
<td></td>
</tr>
<tr>
<td>Test – retest</td>
<td>Patterson &amp; Mack (2001)</td>
<td>□ ICC &gt;.70</td>
<td>Information was not found on this aspect of the scale.</td>
</tr>
<tr>
<td>The extent to which the same results are obtained on repeated administrations of the same questionnaire when no change in physical functioning has occurred</td>
<td></td>
<td>Time intervals and confidence intervals reported</td>
<td></td>
</tr>
<tr>
<td>Calculation of an intraclass correlation coefficient (ICC); and an ICC &gt; 0.70 is desired</td>
<td></td>
<td>□ Marginal or inadequate test-retest reliability ICC&lt;.70</td>
<td></td>
</tr>
<tr>
<td>Preferred if time interval and confidence intervals were presented</td>
<td></td>
<td>X No information found on test-retest reliability</td>
<td></td>
</tr>
<tr>
<td>Inter – rater</td>
<td>Patterson &amp; Mack (2001)</td>
<td>X Agreement reported and adequate</td>
<td>Inter-rater reliability = 0.84 to 0.99 for the items, and 0.99 for the overall score (n = 31 AD patients with two raters) (Patterson &amp; Mack, 2001).</td>
</tr>
<tr>
<td>Limits of agreement, Kappa, or standard error of measurement (SEM) were presented</td>
<td></td>
<td>□ Inadequate inter-rater agreement</td>
<td></td>
</tr>
<tr>
<td>□ No information provided</td>
<td></td>
<td>□ No information found on inter-rater agreement</td>
<td></td>
</tr>
<tr>
<td>VALIDITY</td>
<td>Studies Reported &amp; References</td>
<td>Adequacy Checks</td>
<td>Comment</td>
</tr>
<tr>
<td>----------</td>
<td>------------------------------</td>
<td>----------------</td>
<td>---------</td>
</tr>
<tr>
<td><strong>Content</strong>&lt;br&gt;The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire</td>
<td>Patterson &amp; Mack (1992)</td>
<td>X Patients and experts were involved during item selection and/or item reduction □ Patients were consulted for reading and comprehension □ No patient involvement □ No information found on content validity</td>
<td>Scale development and testing on patients is described above. The early version of the CSALD was subject to clinician review.</td>
</tr>
</tbody>
</table>
| **Construct**<br>The extent to which scores on the questionnaire relate to other measures in a manner that is consistent with theoretically derived hypothesis concerning the domains that are measured | Patterson & Mack (2001) <br>McCleland, et al. (2004) | X Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used □ Limited/inadequate construct validity reported □ No information provided | The CSADL correlates with a modified version of the Dementia Scale (Blessed) = 0.82 (Patterson & Mack, 2001).

The CSADL correlates with the MMSE = -0.63 suggesting that the scale is sensitive to cognitive impairment (Patterson & Mack, 2001). McCleland, et al. (2004) also found a similar correlation (-0.81).

Mack & Patterson (2006) present item difficulty data in terms of dependency which shows greater difficulty with IADLs than ADLs for their group of AD patients the majority of whom live in the community. This is what one would expect. |
| **Construct: Internal Structure**<br>Information provided on factor structure | Neundorfer, et al. (2001) <br>Neundorfer, et al. (2006) | □ No evidence provided/failed a test of dimensionality □ Some evidence provided to support internal structure X Substantial evidence provided to support internal structure | High correlations between Basic ADL and Instrumental ADL items r = 0.71 - 0.80 were found by Neundorfer, et al. (2001) and Neundorfer, et al. (2006). However, low correlations were found when compared to the rate and acceleration of change scores (approx. 0.20) (Neundorfer, et al. 2001).

Exploratory and confirmatory factor analysis carried out by Mack & Patterson (2006) supports a two factor solution |
The two resultant factors were described as Basic and Instrumental ADLs, with some items found to load on both factors.

<table>
<thead>
<tr>
<th>Construct: Correlation with other measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparisons made to other measures</td>
</tr>
<tr>
<td>Patterson, et al. (1992)</td>
</tr>
<tr>
<td>Spector (1997)</td>
</tr>
<tr>
<td>Neundorfer, et al. (2001)</td>
</tr>
<tr>
<td>Patterson &amp; Mack (2001)</td>
</tr>
<tr>
<td>X Correlations with other measures are reported</td>
</tr>
<tr>
<td>□ Correlations not reported</td>
</tr>
</tbody>
</table>

When items were compared to overlapping items in the Dementia Scale (Blessed), the CSADL proved to be more sensitive with regard to dressing and eating (Patterson, et al. 1992). As Spector (1997) notes, by breaking ADLs into automatic and controlled processes, and using items that involve controlled processing (e.g., selecting clothes) the CSADL picks up more dysfunction than a version of the Blessed Scale.

Neundorfer, et al. (2001), a study into the relationship between carer depression and AD patient depression, also examined the relationship between patient IADLs and ADLs and the rate and acceleration of change in carer depression. They showed that changes in the functional variables predicted carer depression. Similar findings were reported by Neundorder et al. (2006) showing that greater dependency in IADLs produced greater depressive symptoms in carers (for those with no prior history or some non-clinical level of symptoms).

Neundorfer, et al. (2001) also found correlations with the duration of illness and BADL and IADL scores of the CSADL. They also found significant correlations with BADL scores and depressive symptoms (CERAD-BRSD depression subscale) and IADL scores with carer self-rated health (single item) and carer depression.

McClendon, et al. (2004) also found similar correlations with the duration of illness and age, as well as carer depression. Neundorfer, et al. (2006) also found a significant correlation between IADL score and carer self-rated health (single item).

Increases in ADL impairment (as measured by CSADL) and problematic behaviour were
<table>
<thead>
<tr>
<th>Construct: Discriminant Validity</th>
<th>Patterson, et al. (1992)</th>
<th>X Scale differentiates between relevant categories of respondents □ No information on discriminant validity</th>
<th>Patterson, et al. (1992) compared CSADL items across dementia severity groups (MMSE) and found that increasing mean scores for dependency were associated with dementia severity group membership. Patterson &amp; Mack (2001) showed significant group differences between mild, moderate and severe AD patients (defined by MMSE scores), as well as healthy and physically impaired controls. Though the authors did argue that they needed more patients with really severe impairment. The cross sectional design of Patterson &amp; Mack (2001) also supports arguments for the instrument’s sensitivity to change.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criterion</td>
<td>□ Comparison made to criterion measures X No comparison with criterion measures provided</td>
<td>Further comparisons need to be undertaken with other dementia specific measures of function (e.g. DAD), clinical rating scales and direct measures of functional performance.</td>
<td>Patterson &amp; Mack (2001) provide means and SDs for AD patients according to severity level (MMSE) and for a large group of controls (healthy and physically impaired). Item difficulty or distribution as defined by dependency (0 or 1) is also presented according group membership.</td>
</tr>
<tr>
<td>Interpretability</td>
<td>Patterson &amp; Mack (2001)</td>
<td>X Authors provide 2 or more types of information on interpretability □ Authors provide limited information to assist with interpretability □ No information provided</td>
<td>Patterson &amp; Mack (2001) provide means and SDs for AD patients according to severity level (MMSE) and for a large group of controls (healthy and physically impaired).</td>
</tr>
</tbody>
</table>

- McClendon, et al. (2004) also found a correlation of 0.22 with the CERAD-BRSD.
- Associated with an increased risk of death in AD patients (though a non linear relationship was found) (McClendon, et al. 2004).
<table>
<thead>
<tr>
<th>Responsiveness</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floor and ceiling effects</td>
<td>Patterson &amp; Mack (1992)</td>
<td>X Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
<td>For five basic ADL domains – bathing, dressing, eating, toileting and hygiene – Patterson &amp; Mack (1992) found only two items (initiates bath or shower and selects clothes) that were rated as dependent in more than 50% of patients and 10 of the 22 items were rated as dependent in less than 20% of patients in their sample. Mack &amp; Patterson (2006) also present item difficulty data for the full scale in their group of AD patients.</td>
</tr>
<tr>
<td>Sensitivity to change</td>
<td>Neundorfer, et al. (2001)</td>
<td>X Hypotheses were formulated and results were in agreement</td>
<td>The longitudinal analysis by Neundorfer, et al. (2001) over a number of data points suggests that the CSADL is highly sensitive to functional changes. Further work in this area is required.</td>
</tr>
</tbody>
</table>

Cultural Applicability and Cultural Adaptations: No information found on this aspect of the scale’s development. However, Fritsch, et al. (2002) did find an effect for ethnicity on functional decline when comparing caregivers of European American and African American patients.

Gender Appropriateness: Appropriate for use with both genders.

Age Appropriateness: Appropriate for use with older adults.
Summary: The Cleveland Scale for Activities of Daily Living (CSADL) is a promising new scale for the proxy measurement of function in people with dementia. The three papers, Patterson, et al. (1992), Patterson & Mack (2001), and Mack & Patterson (2006) chart the rigorous psychometric work-up of the scale, including information on internal consistency and inter-rater reliability, correlation with other well-known dementia measures (MMSE and Blessed Dementia Scale), factor structure, item difficulty and discriminant validity. However, further work is required in the areas of test – retest reliability and sensitivity to change. Additionally, like other measures reviewed in this section, the CSADL needs to be compared to other dementia specific measures of function (including performance and timed tests) in order to better gauge validity (especially criterion related validity). Finally, further application of the CSADL is required beyond the original development team.

Reporter: Nicholas Marosszeky

Date of report: September 2007

References


Appendix 12: Reviews of Patient and Carer Satisfaction Measures
### Appendix 12.1  Short Assessment of Patient Satisfaction

**AHOC INSTRUMENT REVIEW SHEET**

<table>
<thead>
<tr>
<th>Title:</th>
<th>Short Assessment of Patient Satisfaction.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbreviations:</td>
<td>SAPS.</td>
</tr>
<tr>
<td>Author(s) Name:</td>
<td>Hawthorne G, Sansoni J, Hayes L, Marosszeky N and Sansoni E.</td>
</tr>
<tr>
<td>Author(s) Address:</td>
<td>A/Professor Graeme Hawthorne, Department of Psychiatry, Level 1 North, Royal Melbourne Hospital, Grattan St, Parkville, VICTORIA, 3050. Email: <a href="mailto:graemeeh@unimelb.edu.au">graemeeh@unimelb.edu.au</a></td>
</tr>
<tr>
<td>Supplied by:</td>
<td>A/Professor Graeme Hawthorne.</td>
</tr>
<tr>
<td>Cost:</td>
<td>None.</td>
</tr>
<tr>
<td>Training requirements:</td>
<td>Nil.</td>
</tr>
<tr>
<td>Purpose:</td>
<td>To assess patient satisfaction with health care.</td>
</tr>
<tr>
<td>Administration time:</td>
<td>Less than 5 minutes.</td>
</tr>
<tr>
<td>Instrument Type:</td>
<td>Self-report survey.</td>
</tr>
<tr>
<td>Structure:</td>
<td>The SAPS consists of 7 items covering the 7 key dimensions of patient satisfaction. It assesses 7 key dimensions of patient satisfaction with health care: access and facilities, information, relationship with the health professional, participation in medical decision-making, the technical skill of the health professional, the effectiveness of treatment and a general satisfaction dimension (Hawthorne, 2006). There is one item in the SAPS for each of these 7 dimensions (Hawthorne, et al. 2007).</td>
</tr>
<tr>
<td>Scoring:</td>
<td>Following reversal of 4 items, scores on all items are summed.</td>
</tr>
<tr>
<td>Normative Data:</td>
<td>Not applicable.</td>
</tr>
<tr>
<td>Clinical Reference Data:</td>
<td>The construction sample was women who had received treatment for urinary incontinence.</td>
</tr>
<tr>
<td>Applications:</td>
<td>To assess patient satisfaction with health care. Because the items are all generic, the SAPS should have wide application.</td>
</tr>
<tr>
<td>Carer and/or Patient Use of Instrument:</td>
<td>As the items are generic, the SAPS should have wide application with people receiving treatment for a health condition. The questions will require minor rewording for carer/informant use.</td>
</tr>
</tbody>
</table>
## Psychometric Criteria

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal consistency</strong>&lt;br&gt;The extent to which items in a (sub) scale are inter-correlated; a measure of the homogeneity of a (sub)scale&lt;br&gt;Cronbach's alpha should be between 0.70 and 0.90 for every dimension / sub-scale</td>
<td>X Alpha &gt;0.70 □ Marginal or inadequate internal consistency (&lt;0.70) □ No information found on internal consistency</td>
<td>Cronbach α = 0.86 (Hawthorne, et al. 2006)</td>
<td></td>
</tr>
<tr>
<td><strong>Test – retest</strong>&lt;br&gt;The extent to which the same results are obtained on repeated administrations of the same questionnaire when no change in physical functioning has occurred&lt;br&gt;Calculation of an intraclass correlation coefficient (ICC); and an ICC &gt; 0.70 is desired&lt;br&gt;Preferred if time interval and confidence intervals were presented</td>
<td>□ ICC &gt;.70 Time intervals and confidence intervals reported □ Marginal or inadequate test-retest reliability ICC&lt;.70 X No information found on test-retest reliability</td>
<td>No information found on this aspect of the instrument.</td>
<td></td>
</tr>
<tr>
<td><strong>Inter – rater</strong>&lt;br&gt;Limits of agreement, Kappa, or standard error of measurement (SEM) were presented</td>
<td>□ Agreement reported and adequate □ Inadequate inter-rater agreement □ No information provided</td>
<td>Not applicable.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VALIDITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Content</strong>&lt;br&gt;The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire</td>
<td>□ Patients and experts were involved during item selection and/or item reduction □ Patients were consulted for reading and comprehension X No patient involvement □ No information found on content validity X There is an adequate coverage of relevant domains</td>
<td>The SAPS was based on a model of patient satisfaction drawn from an extensive review of the patient satisfaction literature. The review identified 7 key dimensions of patient satisfaction with health care: access and facilities, information, relationship with the health professional, participation in medical decision-making, the technical skill of the health professional, the effectiveness of treatment</td>
<td></td>
</tr>
</tbody>
</table>
There is limited coverage of relevant domains and a general satisfaction dimension (Hawthorne, 2006). There is one item in the SAPS for each of these 7 dimensions (Hawthorne, et al. 2007).

**Construct**
The extent to which scores on the questionnaire relate to other measures in a manner that is consistent with theoretically derived hypothesis concerning the domains that are measured

- X Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used
- □ Limited /inadequate construct validity reported
- □ No information provided

Spearman correlations with the ConsultSQ (Consultation Satisfaction Questionnaire), the CSQ (Client Satisfaction Questionnaire), the GUTSS (Genito-Urinary Treatment Satisfaction Scale) and the PSI (Patient Satisfaction Index) were 0.73, 0.78, 0.83 and 0.83, respectively (Hawthorne, et al. 2007).

**Construct: Internal Structure**
Information provided on factor structure

- □ No evidence provided/failed a test of dimensionality
- □ Some evidence provided to support internal structure
- X Substantial evidence provided to support internal structure

Mokken analysis and partial credit item response theory analysis were used to reduce the initial item pool (N=49) to the final 7 items of the SAPS (Hawthorne, et al. 2007). The Loevinger H for item fit ranged from 0.51 to 0.58, showing that each item exceeded the conventional cutpoint for inclusion in a homogenous scale. The only violations of Guttman monotony were due to sampling errors, and the scale Loevinger H was 0.55 suggesting a strong scale. The partial credit IRT analysis showed that there were no misfitting items. The point biserial correlations ranged from 0.61 to 0.81, suggesting very good consistency within the scale.

**Construct: Correlation with other measures**
Comparisons made to other measures

- X Correlations with other measures are reported
- □ Correlations not reported

Spearman correlations with the ConsultSQ (Consultation Satisfaction Questionnaire), the CSQ (Client Satisfaction Questionnaire), the GUTSS (Genito-Urinary Treatment Satisfaction Scale) and the PSI (Patient Satisfaction Index) were 0.73, 0.78, 0.83 and 0.83, respectively (Hawthorne, et al. 2007).

**Construct: Discriminant Validity**
The scale differentiates between relevant categories of respondent e.g. sick vs. well, varying degrees of severity

- X Scale differentiates between relevant categories of respondents
- □ No information on discriminant validity

Discriminant validity was examined by pooling data from four different patient satisfaction instruments, and then quartiling the resultant scores. The SAPS and each of the four original instruments were then examined against this pooled indicator. The SAPS was more sensitive than any of the other four measures (Hawthorne, et al. 2007).
<table>
<thead>
<tr>
<th>Criterion</th>
<th>Comparison made to criterion measures</th>
<th>The SAPS was examined by type of incontinence treatment, self-reported change in incontinence status, treatment success, and information given by the clinician to the patient. Scores on the SAPS significantly varied by each of these criteria (Hawthorne, et al. 2007).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information on the relationship of scores to gold standard measures or clinical diagnosis is provided</td>
<td>□ No comparison with criterion measures provided</td>
<td>□ Authors provide 2 or more types of information on interpretability</td>
</tr>
</tbody>
</table>

### Interpretability

- The degree to which one can assign qualitative meaning to quantitative scores
- Do authors provide the following:
  - Presentation of means and SD of scores before and after treatment
  - Comparative data on the distribution of scores in relevant subgroups
  - Information on the relationship of scores to well-known functional measures or clinical diagnosis
  - Information on the association between changes in scores and patients’ global ratings of the magnitude of change they have experienced

### RESPONSIVENESS

<table>
<thead>
<tr>
<th>RESPONSIVENESS</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floor and ceiling effects</td>
<td>Hawthorne, Sansoni, Hayes, Marosszeky and Sansoni (2006)</td>
<td>□ Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
<td>In a study evaluating measures of patient satisfaction with incontinence treatment, 20% of cases were in the top 5% of scale ranges indicating most cases were generally satisfied with their incontinence treatment.</td>
</tr>
</tbody>
</table>

The SAPS is easy to interpret: the higher the score the more satisfied the patient is with his/her health care.
Sensitivity to change

| The ability to detect important change over time in the concept being measured |
| □ Hypotheses were formulated and results were in agreement |
| □ An adequate metric was used (ES, SRM, comparison with external standard) |
| X No information on sensitivity to change was provided |
| □ MCID - Information was provided about the magnitude of score differences which would be clinically meaningful |
| X MCID – No information was provided. |

No information found on this aspect of the instrument.

Cultural Applicability and Cultural Adaptations: As this is a new instrument there are no translations available for other languages. There are no obvious cultural issues.

Gender Appropriateness: Appropriate for use with both genders.

Age Appropriateness: Appropriate for use with adults of all ages.

Summary: The SAPS appears to be a useful instrument for assessing patient satisfaction and it has good psychometric properties. Further research is needed as this assessment was based on a single study.

Reporter: A/Professor Graeme Hawthorne.

Date of report: September 2007.

References


Appendix 12.2  The Consultation Satisfaction Questionnaire

AHOC INSTRUMENT REVIEW SHEET

Title: Consultation Satisfaction Questionnaire.

Abbreviations: CSQ, or ConsultSQ.

Author(s) Name: Baker, R.

Author(s) Address: General Practice Unit, Department of Epidemiology and Community Medicine, University of Bristol, UK

Supplied by: A copy of the CSQ can be found in:

Cost: None identified.

Training requirements: Nil.

Purpose: The assessment of patient satisfaction with consultations in general practice.

Administration time: Less than 10 minutes.

Instrument Type: Self-report questionnaire.

Structure: The ConsultSQ comprises 18 items located in four scales: general satisfaction (3 items); professional care (factor 1: 7 items describing the patient's concerns, the provision of information, treatment by the doctor, agreement with the doctor's advice, and the doctor treating the person as a whole); depth of relationship (factor 2: 5 items measuring the doctor's intimate knowledge of the patient and the transmission of personal information to the doctor); and perceived time (factor 3: 3 items measuring the length of the consultation in relation to the patient's perceived needs). A limitation of the ConsultSQ is that there are no items assessing treatment effects.

Scoring: After recoding of negative variables, scoring is through simple summation.

Developed for: The assessment of patient satisfaction with consultations in general practice.

Normative Data: Not applicable.

Clinical Reference Data: The construction sample was patients attending general practitioners (Baker, 1990). It has since been used in studies of patients who changed doctor (Baker and Whitfield, 1992), for comparison between practice clinics (Baker and Whitfield, 1992), patient satisfaction with general practice (Kinnersley, et al. 1996) and medical students (McKinley, et al. 2004).

Applications: Assessment of patient satisfaction with consultations in general practice.

Carer and/or Patient Use of Instrument: For use with patients.
### Psychometric Criteria

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal consistency</strong>&lt;br&gt;The extent to which items in a (sub) scale are inter-correlated; a measure of the homogeneity of a (sub)scale Cronbach's alpha should be between 0.70 and 0.90 for every dimension / sub-scale</td>
<td>X Alpha &gt;0.70&lt;br&gt;□ Marginal or inadequate internal consistency (&lt;0.70)&lt;br&gt;□ No information found on internal consistency</td>
<td>Reliability of the ConsultSQ as assessed by Cronbach α has been reported in the range of 0.67 to 0.94 (Baker, 1990; Hawthorne, et al. 2006; Kinnersley, et al. 1996).</td>
<td></td>
</tr>
<tr>
<td><strong>Test – retest</strong>&lt;br&gt;The extent to which the same results are obtained on repeated administrations of the same questionnaire when no change in physical functioning has occurred Calculation of an intraclass correlation coefficient (ICC); and an ICC &gt; 0.70 is desired Preferred if time interval and confidence intervals were presented</td>
<td>X ICC &gt;.70&lt;br&gt;Time intervals and confidence intervals reported&lt;br&gt;□ Marginal or inadequate test-retest reliability ICC&lt;.70&lt;br&gt;□ No information found on test-retest reliability</td>
<td>Based on test-retest at 2-3 weeks using a mailed self-report version of the CSQ Baker &amp; Whitfield reported the correlation between 0.82 to 0.93 for the four different scales (Baker and Whitfield, 1992).</td>
<td></td>
</tr>
<tr>
<td><strong>Inter – rater</strong>&lt;br&gt;Limits of agreement, Kappa, or standard error of measurement (SEM) were presented</td>
<td>□ Agreement reported and adequate&lt;br&gt;□ Inadequate inter-rater agreement&lt;br&gt;□ No information provided</td>
<td>Not applicable.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VALIDITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Content</strong>&lt;br&gt;The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire</td>
<td>X Patients and experts were involved during item selection and/or item reduction&lt;br&gt;□ Patients were consulted for reading and comprehension&lt;br&gt;□ No patient involvement&lt;br&gt;□ No information found on content validity&lt;br&gt;□ There is an adequate coverage of relevant domains&lt;br&gt;□ There is limited</td>
<td>Based on a literature review and iterative consultation with clinicians and patients, the British ConsultSQ assesses patient’s satisfaction with a consultation with a general practitioner. From the review and consultation an item bank was developed and administered to patients in a surgery following a consultation. After further modification, it was re-issued. This procedure was iteratively followed and the bank progressively modified as</td>
<td></td>
</tr>
<tr>
<td>Construct</td>
<td>Coverage of relevant domains</td>
<td>More data about the performance of items within the bank became available. Following iteration, factor and correlation analyses were used to discard further items and refine the final form of the ConsultSQ (Baker, 1990).</td>
<td></td>
</tr>
</tbody>
</table>

| Construct | X Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used. Limited/inadequate construct validity reported. No information provided. |

| Construct: Internal Structure | X Some evidence provided to support internal structure. |

| Construct: Correlation with other measures | X Correlations with other measures are reported. |

## Construct

The extent to which scores on the questionnaire relate to other measures in a manner that is consistent with theoretically derived hypothesis concerning the domains that are measured.

### Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used.

- Limited/inadequate construct validity reported.
- No information provided.

- Kinnersley, et al. (1996) compared the ConsultSQ with the Medical Interview Satisfaction Scale (MISS, see below) in a sample of 198 patients attending GP surgeries. The findings were that there was very little difference in the psychometric properties of the two measures. The correlation between the ConsultSQ and MISS was 0.82, suggesting they were measuring the same latent concept.

- Hawthorne, et al. (2007) correlated the CSQ against the Client Satisfaction Questionnaire, the Genito-Urinary Treatment Satisfaction Scale and the Patient Satisfaction Index in a sample of women who had received treatment for urinary incontinence. The correlations were 0.67, 0.48 and 0.64 respectively.

### Construct: Internal Structure

Information provided on factor structure.

- No evidence provided/failed a test of dimensionality.
- Some evidence provided to support internal structure.
- Substantial evidence provided to support internal structure.

- Baker (1990) reported 3 factors within the CSQ, as described above. The general satisfaction scale did not form a factor per se, but comprised items that failed to load on other factors. Its validity, therefore, is in doubt.

- Hawthorne, et al. (2007) reported that the Loevinger H for the CSQ was 0.51, indicating a strong unidimensional scale. However, they also noted that this was achieved through item repetition within the CSQ (e.g. consider the items ‘I am totally satisfied with my visit to this doctor’ and ‘I am not completely satisfied with my visit to the doctor’. Altogether there are 6 pairs of such items in the CSQ (Hawthorne, 2006).

### Construct: Correlation with other measures

Comparisons made to other measures.

- Correlations with other measures are reported.
- Correlations not reported.

See construct validity section above.
<table>
<thead>
<tr>
<th>Construct: Discriminant Validity</th>
<th>X</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>The scale differentiates between relevant categories of respondent e.g. sick vs. well, varying degrees of severity</td>
<td>Scale differentiates between relevant categories of respondents □ No information on discriminant validity</td>
<td>Discriminant validity was examined by Hawthorne, et al. (Hawthorne, et al. 2007) by pooling data from four different patient satisfaction instruments, and then quartiling the resultant scores. The CSQ was sensitive across the four quartiles.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Criterion</th>
<th>X</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information on the relationship of scores to gold standard measures or clinical diagnosis is provided</td>
<td>Comparison made to criterion measures □ No comparison with criterion measures provided</td>
<td>Baker and Whitfield (1992) compared CSQ scores between stable patients (defined as those who had not moved between doctors) and patient who had moved, reporting that the CSQ was sensitive. Kinnersley, et al. (1996) reported it was sensitive to home completion versus surgery completion. Hawthorne, et al. (2007) reported the CSQ was sensitive to type of treatment, perceived treatment success and information given.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interpretability</th>
<th>□</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>The degree to which one can assign qualitative meaning to quantitative scores</td>
<td>Authors provide 2 or more types of information on interpretability □ Authors provide limited information to assist with interpretability □ No information provided</td>
<td>The CSQ is easy to interpret: the higher the score the more satisfied the patient is with his/her health care.</td>
</tr>
</tbody>
</table>

- **Presentation of means and SD of scores before and after treatment**
- **Comparative data on the distribution of scores in relevant subgroups**
- **Information on the relationship of scores to well-known functional measures or clinical diagnosis**
- **Information on the association between changes in scores and patients' global ratings of the magnitude of change they have experienced**
### RESPONSIVENESS

<table>
<thead>
<tr>
<th>Floor and ceiling effects</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hawthorne, Sansoni, Hayes, Marosszeky and Sansoni (2006)</td>
<td>X Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected □ Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score □ No information provided on floor and ceiling effects</td>
<td>In a study evaluating measures of patient satisfaction with incontinence treatment, 3% of scores were in the top 5% scale ranges indicating no ceiling effect. Due to the small numbers of cases endorsing the lowest category levels, the bottom two levels were collapsed for analysis and an average of 16% of cases selected the bottom two categories, indicating no floor effect.</td>
</tr>
</tbody>
</table>

| Sensitivity to change | | □ Hypotheses were formulated and results were in agreement □ An adequate metric was used (ES, SRM, comparison with external standard) X No information on sensitivity to change was provided □ MCID - Information was provided about the magnitude of score differences which would be clinically meaningful X MCID – No information was provided. | No information provided on this aspect of the instrument. |

### Cultural Applicability and Cultural Adaptations:
No information was provided.

### Gender Appropriateness:
Appropriate for use with both genders.

### Age Appropriateness:
Appropriate for use with adults.

### Summary:
The CSQ appears to be a good instrument, although there are concerns with the degree of repetition within the measure.

### Reporter:
A/Prof Graeme Hawthorne.

### Date of report:
September, 2007.
References


Appendix 12.3 Satisfaction with Care at the End of Life in Dementia Scale

AHOC INSTRUMENT REVIEW SHEET

Title: Satisfaction with Care at the End of Life in Dementia Scale.

Abbreviations: SWC-EOLD.

Author(s) Name: Volicer, L., Hurley, A. C. and Blasi, Z. V.

Author(s) Address: Geriatrics Research Education Clinical Center, E.N. Rogers Memorial Veterans Hospital, Bedford, Massachusetts 01730, USA.


Cost: No costs are mentioned in the seminal journal article.

Training requirements: Nil.

Purpose: Designed in response to the US MediCaring National Demonstration and Evaluation Project for those dying of chronic disease (e.g. congestive heart failure) (Skolnick, 1998), the SWC-EOLD scale assesses the quality of care during the last 90 days of life (Volicer, et al. 2001).

Administration time: Five minutes.

Instrument Type: Self-report Questionnaire for the carer of a person with dementia.

Structure: The questionnaire consists of 10 items assessing being fully involved in decision-making, the provision of information, keeping the care recipient comfortable, the health professionals being sensitive to the needs/feelings of the carer, understanding the care recipient's condition, knowing which doctor/nurse was in charge of providing care, that the care recipient received all necessary nursing assistance, that medication issues were clearly explained, that the care recipient received all treatments/interventions that he/she could have benefited from, and that the care recipient received the best treatment at the end of his/her life. Three of the items were negative, and needed to be reversed prior to scoring. The response set was a forced choice scale (strongly disagree/ disagree/ agree/ strongly agree); and a 'not applicable' option was included.

Scoring: After recoding of negative variables, scoring is through simple summation.

Developed for: Designed in response to the US MediCaring National Demonstration and Evaluation Project for those dying of chronic disease (e.g. congestive heart failure) (Skolnick, 1998), the SWC-EOLD scale assesses the quality of care during the last 90 days of life (Volicer, et al. 2001).

Normative Data: Not applicable.

Clinical Reference Data: Volicer, et al. (2001) constructed the SWC-EOLD in a sample of carers whose care recipient, suffering from Alzheimers disease, had died in the previous 12 months.

Applications: Assessment of carer satisfaction with the health care provided to the care recipient within the last year of life.
Carer and/or Patient Use of Instrument: The instrument is for use with carers.

Psychometric Criteria

<table>
<thead>
<tr>
<th>RELIABILITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal consistency</strong></td>
<td></td>
<td>X Alpha &gt;0.70</td>
<td>The reliability of the SWC-EOLD in the construction sample was Cronbach $\alpha = 0.90$ (Volicer, et al. 2001), elsewhere among dementia care dyads it was 0.83 (Kiely, et al. 2006).</td>
</tr>
<tr>
<td>The extent to which items in a (sub) scale are inter-correlated; a measure of the homogeneity of a (sub)scale</td>
<td></td>
<td>□ Marginal or inadequate internal consistency (&lt;0.70)</td>
<td></td>
</tr>
<tr>
<td>Cronbach's alpha should be between 0.70 and 0.90 for every dimension / sub-scale</td>
<td></td>
<td>□ No information found on internal consistency</td>
<td></td>
</tr>
<tr>
<td><strong>Test – retest</strong></td>
<td></td>
<td>□ ICC &gt;.70</td>
<td>No information was found on this aspect of the instrument.</td>
</tr>
<tr>
<td>The extent to which the same results are obtained on repeated administrations of the same questionnaire when no change in physical functioning has occurred</td>
<td></td>
<td>□ Marginal or inadequate test-retest reliability ICC&lt;.70</td>
<td></td>
</tr>
<tr>
<td>Calculation of an intraclass correlation coefficient (ICC); and an ICC &gt; 0.70 is desired</td>
<td></td>
<td>X No information found on test-retest reliability</td>
<td></td>
</tr>
<tr>
<td>Preferred if time interval and confidence intervals were presented</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Inter – rater</strong></td>
<td></td>
<td>□ Agreement reported and adequate</td>
<td>No information was found on this aspect of the instrument.</td>
</tr>
<tr>
<td>Limits of agreement, Kappa, or standard error of measurement (SEM) were presented</td>
<td></td>
<td>□ Inadequate inter-rater agreement</td>
<td></td>
</tr>
<tr>
<td>□ No information provided</td>
<td></td>
<td>X No information provided</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VALIDITY</th>
<th>Studies Reported &amp; References</th>
<th>Adequacy Checks</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Content</strong></td>
<td></td>
<td>□ Patients and experts were involved during item selection and/or item reduction</td>
<td>No evidence is presented in the seminal paper. The authors stated that most of the items were taken from other scales or the views of experts. No further information is given (Volicer, et al. 2001).</td>
</tr>
<tr>
<td>The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire</td>
<td></td>
<td>□ Patients were consulted for reading and comprehension</td>
<td></td>
</tr>
<tr>
<td>□ No patient involvement</td>
<td></td>
<td>□ No patient involvement</td>
<td></td>
</tr>
<tr>
<td>□ X No information found on content validity</td>
<td></td>
<td>□ X There is an</td>
<td></td>
</tr>
<tr>
<td>Construct</td>
<td>Adequate coverage of relevant domains □ There is limited coverage of relevant domains</td>
<td>□ Results were acceptable in accordance with the hypotheses and an adequate comparison measure was used □ Limited/inadequate construct validity reported</td>
<td>No evidence is available.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><strong>Construct</strong></td>
<td>The extent to which scores on the questionnaire relate to other measures in a manner that is consistent with theoretically derived hypothesis concerning the domains that are measured</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Construct: Internal Structure</strong></td>
<td>Information provided on factor structure</td>
<td>X No evidence provided/failed a test of dimensionality □ Some evidence provided to support internal structure □ Substantial evidence provided to support internal structure</td>
<td>No evidence is available.</td>
</tr>
<tr>
<td><strong>Construct: Correlation with other measures</strong></td>
<td>Comparisons made to other measures</td>
<td>X Correlations with other measures are reported □ Correlations not reported</td>
<td>The SWC-EOLD was correlated with two other scales, measuring symptom management at the end of life in dementia (SM-EOLD) and comfort assessment in dying with dementia (CAD-EOLD): the correlations were 0.28 and 0.30, respectively (Volicer, et al. 2001). Kiely, et al. (2006) in a study of dementia care dyads it correlated 0.81 with the Decision Satisfaction Inventory (a measure of satisfaction with medical decision-making).</td>
</tr>
<tr>
<td><strong>Construct: Discriminant Validity</strong></td>
<td>The scale differentiates between relevant categories of respondent e.g. sick vs. well, varying degrees of severity</td>
<td>□ Scale differentiates between relevant categories of respondents X No information on discriminant validity</td>
<td>No evidence is available.</td>
</tr>
<tr>
<td><strong>Criterion</strong></td>
<td>Information on the relationship of scores to gold standard measures or clinical diagnosis is provided</td>
<td>X Comparison made to criterion measures □ No comparison with criterion measures provided</td>
<td>In a study of end-of-life care for those with advanced dementia, scores on the SWC-EOLD systematically varied by care planning, symptom management, dementia ward status and tube feeding (Engel, et al. 2006).</td>
</tr>
<tr>
<td><strong>Interpretability</strong></td>
<td><strong>Studies Reported &amp; References</strong></td>
<td><strong>Adequacy Checks</strong></td>
<td><strong>Comment</strong></td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------------------</td>
<td>--------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>The degree to which one can assign qualitative meaning to quantitative scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do authors provide the following:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presentation of means and SD of scores before and after treatment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparative data on the distribution of scores in relevant subgroups</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information on the relationship of scores to well-known functional measures or clinical diagnosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information on the association between changes in scores and patients' global ratings of the magnitude of change they have experienced</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ Authors provide 2 or more types of information on interpretability</td>
<td>□ Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
<td>No significant floor or ceiling effects have been reported.</td>
<td></td>
</tr>
<tr>
<td>× Authors provide limited information to assist with interpretability</td>
<td>□ Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ No information provided</td>
<td>X No information provided on floor and ceiling effects</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The SWC-EOLD is easy to interpret: the higher the score the more satisfied the carer is with the health care provided to the care recipient.

<table>
<thead>
<tr>
<th><strong>RESPONSIVENESS</strong></th>
<th><strong>Studies Reported &amp; References</strong></th>
<th><strong>Adequacy Checks</strong></th>
<th><strong>Comment</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Floor and ceiling effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The questionnaire fails to demonstrate a worse score in patients who clinically deteriorated and an improved score in patients who clinically improved</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Authors should provide descriptive statistics of the distribution of scores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>□ Descriptive statistics of the distribution of scores were presented and no major floor or ceiling effects detected</td>
<td>□ Hypotheses were formulated and results were in agreement</td>
<td>No information was found on this aspect of the instrument.</td>
<td></td>
</tr>
<tr>
<td>□ Descriptive statistics of the distribution of scores were presented and more than 15% of respondents achieved the highest or lowest possible score</td>
<td>□ An adequate metric was used (ES, SRM, comparison with external standard)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X No information on sensitivity to change was provided</td>
<td>□ MCID - Information</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
was provided about the magnitude of score differences which would be clinically meaningful X MCID – No information was provided.

Cultural Applicability and Cultural Adaptations: There are no translations available. No information is provided on cultural adaptations.

Gender Appropriateness: Appropriate for use with both genders.

Age Appropriateness: Appropriate for use with adults.

Summary: At 10 items this is a short, unidimensional scale. There is limited evidence on its reliability, although this evidence appears to be satisfactory. There is almost no satisfactory validity evidence available for the scale. As acknowledged by the authors, much further work needs to be done on this scale.

Reporter: A/Prof Graeme Hawthorne.

Date of report: September, 2007.

References


Appendix 13: Some Resources Relevant for the Assessment of CALD populations

Following is a list of a few useful websites and references for dementia in CALD communities. The report by Alzheimer’s Australia – NCCDN also provides useful resources for practitioners (see Grypma, Mahajani & Tam, 2007). It should be noted, however, this is not a definitive list of resources.

General Information and Guidelines:

The Diversity Health Institute.

Centre for Cultural Diversity in Ageing.

Models:

The Purnell Model of Culturally Competent Health Care (Purnell & Paulanka, 2005); Person Centred Care in Dementia (O’Connor, et al. 2007).

Specific Tools:

Dementia and Culture Information Sheet – Migrant Information Centre East Melbourne.

Policy Checklist for service design and delivery - Centre for Cultural Diversity in Ageing.

Working with Interpreters: Guidelines for working effectively with interpreters in mental health settings – Victorian Transcultural Psychiatry Unit.

The Charter of Public Service in a Culturally Diverse Society – Australian Government, Department of Immigration and Citizenship.

Tools from the USA:

Online Teaching Modules – Stanford Geriatric Education Center, Stanford School Of Medicine, Palo Alto, CA.
(http://sgec.stanford.edu/training/)

(http://www.alz.org/professionals_and_researchers_caring_for_diverse_populations.asp)

Translating outcome measures:

Appendix 14: Appropriateness of Recommended Associated Symptoms Instruments for Indigenous people

Neuropsychiatric Inventory (NPI)

<table>
<thead>
<tr>
<th>Mode of administration</th>
<th>Clinical rating scale combined with informant interview.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concepts</td>
<td>Problems with terms such as ‘false beliefs’. Seeing things or seeing things that are not present may be perfectly appropriate in communities where there is a strong belief in spirits.</td>
</tr>
<tr>
<td>Language</td>
<td>Terms such as ‘resistive to help from others’ would have to be changed.</td>
</tr>
<tr>
<td>Time/value</td>
<td>Perhaps present the severity and carer distress ratings in a visual way which shows volume. Distinction between minimal and mild is difficult in any context. Difficulty in defining “the past month”.</td>
</tr>
<tr>
<td>Other</td>
<td>Brevity is an advantage.</td>
</tr>
<tr>
<td>Summary</td>
<td>Difficulty with this questionnaire in remote/traditional settings. The severity rating and carer distress values would have to be presented in another way. Possibility that this could be used in urban settings.</td>
</tr>
</tbody>
</table>
**BEHAVE-AD**

<table>
<thead>
<tr>
<th>Mode of administration</th>
<th>Clinical rating scale.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concepts</td>
<td>Some of the delusions may be very culturally specific, or not particularly meaningful. For example one's house is not one's home delusion may have little relevance for people who move between houses. Hallucinations may be very difficult to define and may not be regarded as aberrant behaviour.</td>
</tr>
<tr>
<td>Language</td>
<td>NA - clinical rating.</td>
</tr>
<tr>
<td>Time/value</td>
<td>Very difficult for a clinician to make value judgments in a cross cultural setting.</td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td>This is a very difficult instrument to use in a remote/traditional setting. May be possible to use in an urban setting.</td>
</tr>
</tbody>
</table>
### CERAD Behavioral Rating Scale

<table>
<thead>
<tr>
<th>Mode of administration</th>
<th>Clinical rating scale.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Concepts</strong></td>
<td>Difficult concepts are:</td>
</tr>
<tr>
<td></td>
<td>Sexual interest.</td>
</tr>
<tr>
<td></td>
<td>Believing that a dead person still alive- a very tricky question to ask.</td>
</tr>
<tr>
<td></td>
<td>Hallucinations as above.</td>
</tr>
<tr>
<td><strong>Language</strong></td>
<td>Sad or blue.</td>
</tr>
<tr>
<td><strong>Time/value</strong></td>
<td>Asks to rate symptoms over the last month.</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Summary</strong></td>
<td>This is a very difficult instrument to use in a remote/traditional setting. May be possible to use in an urban setting, but there are still problems with the questions about sexuality.</td>
</tr>
</tbody>
</table>
## Confusion Assessment Method

<table>
<thead>
<tr>
<th>Mode of administration</th>
<th>Clinical rating scale.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concepts</td>
<td>Very difficult to make ratings of inattention-when this may be simply be an appropriate response to an outsider asking questions. It is also difficult for an outsider to make assessments of disorganized thinking.</td>
</tr>
<tr>
<td>Language</td>
<td>NA.</td>
</tr>
<tr>
<td>Time/value</td>
<td>NA.</td>
</tr>
<tr>
<td>Other</td>
<td>Brevity is an advantage.</td>
</tr>
<tr>
<td>Summary</td>
<td>This appears to be an instrument which is used in a clinical setting. It may be very difficult for a clinician, who has not had extensive contact with the patient, to make some of these assessments. Language and cultural barriers may increase problems of understanding. May be possible to use in urban settings, validity with rural/remote people may be compromised.</td>
</tr>
</tbody>
</table>
### Delirium Rating Scale

<table>
<thead>
<tr>
<th>Mode of administration</th>
<th>Clinical rating scale.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concepts</td>
<td>Difficulties with defining hallucinations is problematic.</td>
</tr>
<tr>
<td></td>
<td>May be very difficult to assess cognitive status.</td>
</tr>
<tr>
<td>Language</td>
<td>NA.</td>
</tr>
<tr>
<td>Time/value</td>
<td>Change in symptoms over a six month period.</td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td>As with the previous instruments it may be difficult for a clinician to make assessments about cognitive status. Hallucinations may not be regarded as problematic. May be appropriate for use with urban Indigenous people but difficulties would be encountered with rural/remote people.</td>
</tr>
</tbody>
</table>
### Delirium Rating Scale – R98

<table>
<thead>
<tr>
<th>Mode of administration</th>
<th>Clinical rating scale.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concepts</td>
<td>Difficulties with defining hallucinations as problematic.</td>
</tr>
<tr>
<td></td>
<td>Difficulty in assessing abnormalities of language in a cross cultural setting.</td>
</tr>
<tr>
<td></td>
<td>Difficulty in assessing attention in a cross cultural setting.</td>
</tr>
<tr>
<td></td>
<td>Orientation questions in a multiple choice answer format are too difficult to understand.</td>
</tr>
<tr>
<td>Language</td>
<td>NA.</td>
</tr>
<tr>
<td>Time/value</td>
<td>Difficulties with time relating to long term memory.</td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td>As with the previous instruments it may be difficult for a clinician to make assessments about cognitive status and language abnormalities. Hallucinations may not be regarded as problematic. May be appropriate for use with urban Indigenous people but difficulties with rural/remote people would be encountered.</td>
</tr>
</tbody>
</table>
### Rating Scale for Aggressive Behaviour in the Elderly

<table>
<thead>
<tr>
<th>Mode of administration</th>
<th>Clinical rating scale.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concepts</td>
<td>Demanding or argumentative behaviour - this may be considered entirely appropriate behaviour for an elderly person. Many of these symptoms may be more related to distress at being in a clinical setting than indicative of a patient’s actual mental state.</td>
</tr>
<tr>
<td>Language</td>
<td>NA.</td>
</tr>
<tr>
<td>Time/value</td>
<td>NA.</td>
</tr>
<tr>
<td>Other</td>
<td>NA.</td>
</tr>
<tr>
<td>Summary</td>
<td>This instrument has very little meaning or relevance to Indigenous people. Symptoms may be present, but they may have little relationship to long term mental state.</td>
</tr>
</tbody>
</table>
## Cohen Mansfield Agitation Inventory (long form)

<table>
<thead>
<tr>
<th>Mode of administration</th>
<th>Carer rated instrument.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concepts</td>
<td>Requests for attention or help may be viewed as entirely appropriate for elderly people (and expected by their carers). Hoarding things is an effective strategy for preserving property in situations where there are many people in the house.</td>
</tr>
<tr>
<td>Language</td>
<td>NA.</td>
</tr>
<tr>
<td>Time/value</td>
<td>Times per week and day may be difficult concepts.</td>
</tr>
<tr>
<td>Other</td>
<td>Brevity.</td>
</tr>
<tr>
<td>Summary</td>
<td>Possible application, with recognition that one or two items may have little validity. Good to have the carer’s assessment.</td>
</tr>
</tbody>
</table>
## Rating Anxiety in Dementia

<table>
<thead>
<tr>
<th>Mode of administration</th>
<th>Clinical rating scale.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concepts</td>
<td>Items about worry may be difficult to assess for severity. Constant concern about family is normal and expected, especially when individual is removed from the home environment. Family problems/family health is in a very different category of worry to family finances. Irritability may be due to confusion and fear in the clinical setting. Phobias and panic attacks are concepts that are very difficult to describe in a cross cultural setting.</td>
</tr>
<tr>
<td>Language</td>
<td>NA.</td>
</tr>
<tr>
<td>Time/value</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>Brevity.</td>
</tr>
<tr>
<td>Summary</td>
<td>May have application in an urban setting. Serious problems with application with people from remote/ rural contexts.</td>
</tr>
</tbody>
</table>
## Apathy Evaluation Scale

<table>
<thead>
<tr>
<th>Mode of administration</th>
<th>Clinical rating scale.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concepts</td>
<td>Some judgments are very difficult to make in a cross cultural setting such as “she approaches life with intensity”. Values such as “seeing a job through” may have little meaning. Concepts such as initiative and motivation are difficult to assess in communities which are based on collective rather than individual decision making.</td>
</tr>
<tr>
<td>Language</td>
<td>NA.</td>
</tr>
<tr>
<td>Time/value</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>Brevity.</td>
</tr>
<tr>
<td>Summary</td>
<td>May have application in an urban setting. Serious problems with application with people from remote/ rural contexts.</td>
</tr>
</tbody>
</table>
### Cornell Scale for Depression in Dementia

<table>
<thead>
<tr>
<th>Mode of administration</th>
<th>Clinical rating scale.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concepts</td>
<td>Behavioural disturbance may be caused by circumstances and frustration at not being able to communicate effectively.</td>
</tr>
<tr>
<td>Language</td>
<td>NA.</td>
</tr>
<tr>
<td>Time/value</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>Brevity.</td>
</tr>
<tr>
<td>Summary</td>
<td>This has potential for application in Indigenous communities, but the person administering the instrument would have to be familiar to the individual and have a good knowledge of the person’s circumstances.</td>
</tr>
</tbody>
</table>
### Geriatric Depression Scale

<table>
<thead>
<tr>
<th>Mode of administration</th>
<th>Patient self-report.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concepts</td>
<td>Many of these concepts may have different meanings cross culturally (i.e. the desire to stay at home rather than try new things, getting started on new projects, making decisions, whether your mind as clear as it used to be etc.).</td>
</tr>
<tr>
<td>Language</td>
<td>Downhearted and blue.</td>
</tr>
<tr>
<td>Time/value</td>
<td>Past week.</td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td>May be possible to use in an urban setting, but this captures a very white middle class perception of depression. Possibly doesn't capture elements that may be important in the remote Indigenous context.</td>
</tr>
</tbody>
</table>
De Jong Gierveld Loneliness Scale

<table>
<thead>
<tr>
<th>Mode of administration</th>
<th>Patient self-report.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concepts</td>
<td>Many of these concepts may have different meanings cross culturally i.e. I find my circle of friends and acquaintances too limited. Loneliness may be a very difficult idea to explore in a community where dependence on family is the norm.</td>
</tr>
<tr>
<td>Language</td>
<td></td>
</tr>
<tr>
<td>Time/value</td>
<td>Past week.</td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td>People in remote communities may not have a concept of loneliness. Ideas about circle of friends may not be relevant in places where everyone is related.</td>
</tr>
</tbody>
</table>
Appendix 15: Kimberley Indigenous Cognitive Assessment Tool*

Orientation
Is this pension week? What time of year (season) is it now? What is the name of this community?

Recognition and naming
*Hold up three items in turn: comb; pannikin (cup); matches*
Then for each ask What do you call this? What is this one for?
*Then hide each object in turn; Don't forget where I put them.*

Registration
Tell me those things I showed you?
Verbal comprehension
Shut your eyes
First point to the sky then point to the ground

Verbal fluency
Tell me the names of all the animals that people hunt (*time for one minute*)
Recall
Where did I put the comb? The pannikin? The matches?

Visual naming
I'll show you some pictures. You tell me what they are. Remember the words for later on.
*(boy, emu, billy on fire, crocodile, bicycle)*

Graphomotor perseveration
Look at this. Now copy it. (*Alternating crosses and circles*)
Free recall
You remember the pictures I showed you before – What were they – tell me?
Cued recall
Which one did I show you before? (*show three pictures to choose from for each item shown in free recall*)

Praxis
Open this bottle and pour into this cup?
Show me how to use this comb?