Title: Health-Related Quality of Life instruments for Alzheimer’s disease and mixed dementia

Author List:
Jaime Perales\textsuperscript{1, 2}, MPH; Theodore D Cosco\textsuperscript{2}, MSc; Blossom CM Stephan\textsuperscript{2}, PhD; Josep Maria Haro\textsuperscript{1, 3}, PhD; Carol Brayne\textsuperscript{2}, MD

Institution:
1. Parc Sanitari Sant Joan de Déu, Universitat de Barcelona, Sant Boi de Llobregat, Barcelona, Spain.
2. Department of Public Health & Primary Care, Cambridge Institute of Public Health, University of Cambridge.
3. Instituto de Salud Carlos III, Centro de Investigación Biomédica en Red de Salud Mental. CIBERSAM, Spain.

Proposed Journal: International Psychogeriatrics

Corresponding Author Contact Details:
Jaime Perales MPH, MBPsS
Department of Public Health & Primary Care, Cambridge Institute of Public Health. University Forvie Site, Robinson Way, Cambridge, CB20SR.
email: jaume.perales@pssjd.org

Sponsors of research: Instituto de Salud Carlos III

Characters: 4286
Abstract

Background: Over the last 20 years, a number of instruments developed for the assessment of Health-related Quality of Life (HRQL) in dementia have been introduced. The aim of this review is to synthesize evidence from published reviews on HRQL measures in dementia and any new literature in order to identify dementia specific HRQL instruments, the domains they measure and their operationalization.

Methods: An electronic search of PsycINFO and PubMed was conducted, from inception to December 2011 using a combination of key words that included quality of life and dementia.

Results: Fifteen dementia-specific HRQL instruments were identified. Instruments varied depending on their country of development/validation, dementia severity, data collection method, operationalization of HRQL in dementia, psychometric properties and the scoring. The most common domains assessed include mood, self-esteem, social interaction and enjoyment of activities.

Conclusions: A number of HRQL instruments for dementia are available. The suitability of the scales for different contexts is discussed. Many studies do not specifically set out to measure dementia-specific HRQL but do include related items. Determining how best to operationalize the many HRQL domains will be helpful for mapping measures of HRQL in such studies maximizing the value of existing resources.
Keywords: Systematic review, psychometric properties, dimensions, operationalization, scales, mood

Running title:
Health-Related Quality of Life instruments for dementia
Introduction

Dementia is one of the most common disorders of old age (Ferri et al., 2005; Marengoni et al., 2008) and a leading cause of mortality and disability in high income countries (Lopez et al., 2006). Medications temporarily reduce symptoms for some, but without modifying the underlying course of the disorder, which is related to many underlying causes. Narrow assessment of cognition and functional ability is insufficient for clinical decision-making and policy development as they only reflect a part of the impact of dementia (Whitehouse, 2000). Treatment is increasingly focused on improving or maintaining optimal quality of life (QoL) (Ettema et al., 2005) as this has become a key outcome for evaluating the effectiveness of dementia interventions (Small et al., 1997; Whitehouse, 2000; Moniz-Cook et al., 2008).

Health-related quality of life (HRQL) can be defined as the individual's perception of the impact of a health condition on everyday life (Bullinger et al., 1993). It differs from the concept of quality of life in that HRQL includes only aspects of quality of life that are affected by a health condition. Despite the lack of agreement of what domains constitute HRQL, most authors appear to agree upon the multidimensional and subjective nature of the concept and that the assessment should include measurement of positive and negative dimensions (Lawton, 1994; The WHOQOL Group, 1995; Brown et al., 2004; Ready, 2011).
Multiple measures have been developed specifically for assessing HRQL in patients with Alzheimer’s disease or related dementias. These instruments can be either generic HRQL questionnaires used in dementia populations such as the World Health Organization Quality of Life Assessment (WHOQOL) or Schedule for the Evaluation of Individual Quality of Life (SEIQoL) (The WHOQOL Group, 1998; Schölzel-Dorenbos, 2000), dimension specific scales such as the Progressive deteriorations Scale (PDS) (DeJong et al., 1989) or dementia specific such as the Quality of Life in Alzheimer’s Disease (QoL-AD) or the DEMQOL (Logsdon et al., 1999; Smith et al., 2005). A number of studies have reviewed the different instruments used to measure HRQL in dementia (e.g. Ready and Ott, 2003; Ettema et al., 2005). However, no attempt has been made as yet to synthesize the findings across the different reviews. Although one study has discussed the appropriateness of different scales (Schozel-Dorenbos et al., 2007), this did not use a systematic methodology. No review has assessed the features of the different scales in detail in order to assess their suitability in different contexts.

The aim of this paper is to synthesize the evidence from reviews on HRQL instruments for dementia and any new literature to bring together and provide an update on dementia-specific HRQL measures providing details of their features including: number of items, scoring, data collection method, severity in which the instrument may be administered, domains, time frame, countries where the scales were developed or validated and measures of reliability and validity. The suitability of the scales in different contexts is discussed.
Conceptualization and operationalization of dementia-specific HRQL domains by each dementia-specific instrument are analyzed and compared for consistency.

**Methods**

In order to synthesize the evidence from reviews on HRQL instruments in dementia, JP conducted an electronic search in PubMed from inception to December 2011 using a combination of key words that included quality of life and dementia as major topic Mesh terms and specifying reviews as type of publication. The publications could be in either English or Spanish language. In order to identify previous and new literature, a second electronic search was conducted of PsycINFO and PubMed, from inception to December 2011 using a combination of words that included quality of life and dementia as major topic Mesh terms or keywords. The criteria for inclusion were: (a) articles, abstracts or scale manuals that were published in English or Spanish language on dementia-specific HRQL measures developed for use with on patients with Alzheimer's disease (AD) or mixed dementia; and (b) reporting on the development, description of the dimensions and psychometric properties of an instrument. In order to report country of validation of the scales, those papers with abstracts in English irrespective of the rest of the text were also included. Generic, domain-specific QOL measures such as those measuring only activities of daily living (DeJong et al., 1989) and Parkinson's disease-specific HRQL measures are beyond the scope of this review and were excluded. This review excluded considerations of caregiver
quality of life except where this was related to measures of patient HRQL.

Literature selection was undertaken in two phases. In phase 1, titles and abstracts of all articles were independently reviewed (JP and TDC) to exclude non-related publications. Full-text articles accepted in phase 1 were reviewed in order to extract the appropriate information (phase 2). In case of discrepancy, if no agreement was met by both readers after discussion, a third author would decide (BS). In order to rate the psychometric quality assessment of each of the instruments, standardized criteria was used. These criteria are shown in Table 1.

---Table 1 about here---

Results

Reviews identified (Literature Search 1)

Of the 271 articles identified in the first literature search, eight were reviews on HRQL instruments in dementia (Salek et al., 1998; Walker et al., 1998; Ready and Ott, 2003; Ettema et al., 2005; Smith et al., 2005; Lucas-Carrasco, 2007; Schozel-Dorenbos et al., 2007; Ready, 2011). Only four (Salek et al., 1998; Walker et al., 1998; Ready and Ott, 2003; Ettema et al., 2005) used a systematic approach. Table 2 shows the dementia-specific HRQL instruments included in each of the eight reviews identified. Although the number of reviews increases by year of publication, there are some inconsistencies in the instruments included. The reviews highlight the wide variation in instruments, their items and methods of scoring, psychometric properties,
data collection method, severity of dementia and the populations used to assess validity and reliability.

---Table 2 about here---

Instruments identified (Literature Search 2)

Figure 1 shows the selection process for the review on dementia-specific HRQL. In total 848 articles were identified. After the title-abstract screen, 76 articles remained. Based on the full-text search, 68 publications covering 15 different dementia-specific HRQL scales were identified and included in this review. The dementia-specific HRQL measures reviewed included: the Alzheimer Disease Related Quality of Life (ADRQL), Bath Assessment of Subjective Quality of Life in Dementia (BASQUID), Cornell-Brown Scale for Quality of Life in Dementia (CBS), Dementia Care Mapping (DCM), Dementia Quality of Life Instrument (D-QoL), DEMQOL, Patient Activity Scale-AD plus the Modified Apparent Emotion Scale (PES-AD + AES), Quality of Life in Late-stage Dementia Scale (QUALID), Quality of Life-Alzheimer’s Disease (QOL-AD), Quality of Life Assessment Schedule (QOLAS), Quality of Life for Dementia (QOL-D), QUALIDEM, Quality of Life for Older People Experiencing Dementia (QLDJ), Vienna List and the Community Dementia Quality of Life Profile. As shown in table 2, none of the reviews included the QLDJ or identified all the instruments. Table 3 shows the publications reviewed for each of the instruments. Some publications have assessed the psychometric properties of more than one instrument at a time (Edelman et al., 2005; Sloane et al., 2005; Smith et al., 2005).
Table 4 summarizes the characteristics of the original version of each of the 15 different instruments and reports the existence of validations in other countries. There are two instruments where only self-report is possible, six proxy-report-only measures, four measures that include both self and proxy ratings and three observer-rated instruments. The number of items varies between 10 in the QOLAS and 47 in the ADRQL. Three scales can obtain both total and subscale scores whereas six can only obtain subscale scores and six only total scores. Response scales vary from being binary to 6-point Likert scales. Most response scales of observation-rated measures or those rated only by proxies ask about frequency of a certain behavior, whereas the other types of instruments vary between quality, how much of a problem, satisfaction or worry of a trait. The psychometric properties of self-report only measures (Brod et al., 1999; Trigg et al., 2007) have been tested in patients with mild-moderate dementia severity as have all measures using both types of report except for DEMQOL-proxy and QoL-AD. Two proxy-report only (Salek et al., 1997; Black et al., 1999) and two observer-rated instruments (Fossey et al., 2002; Ettema and Lange, 2007) have tested their psychometric properties in patients of all dementia severity whereas the psychometric properties of the rest have been tested in patients with moderate-severe dementia severity. The time frame of the questions is usually either in the present or in the previous week/2 weeks. The number of dimensions ranges
Most instruments have been developed in the USA (6) (Albert et al., 1996; Black et al., 1999; Brod et al., 1999; Logsdon et al., 1999; Weiner et al., 2000; Ready et al., 2002) or the UK (5) (Salek et al., 1997; Selai et al., 2001; Fossey et al., 2002; Smith et al., 2005; Trigg et al., 2007). However, other instruments have also been developed in Japan (2) (Terada et al., 2002; Yamamoto-Mitani et al., 2002), the Netherlands (1) (Ettema et al., 2007) and Austria (1) (Porzsolt et al., 2004). Five instruments (Brod et al., 1999; Logsdon et al., 1999; Weiner et al., 2000; Fossey et al., 2002; Smith et al., 2005) have been validated in other countries. None of the samples used to assess the psychometric properties of each of the instruments were population based.

**Conceptualization and operationalization**

The most common domains across different dementia-specific HRQL instruments are mood, social interaction, enjoyment of activities/ sense of aesthetics and self-esteem/self-concept. Other common domains are cognition, activities, health, living conditions and feelings of usefulness. Domains are operationalized differently across instruments. Most domains are operationalised either through the patient or proxy’s perception of the different domains or through the rating of observational behaviors or contexts. Most instruments measure mood as positive and negative affect although two exceptions are the QoL-AD, which asks for mood in general, and the QOLAS, which needs to be given 2 examples within the mood domain by the respondent. Common items on positive affect are cheerfulness, happiness, contentment and calm. The most common items indicating negative effect are
sadness, anger, worry and anxiety. Proxy and observation-based instruments usually measure either observational indicators of patient’s mood (such as crying) or proxy’s perception of the patient’s mood. A similar trend can be observed in other dimensions. For example, social interaction has been assessed as the perception by the patient or the proxy of the patient’s quality of interactions such as worries about not having enough company, satisfaction or problems with social relationships, or through observation-based ratings such as frequency of visiting friends, talking or seeking or rejecting contact with people. Enjoyment of activities has been measured through perception, satisfaction or worry about enjoyment of different activities, or observation-based assessments such as voluntarily participation in activities and talking about work or activities. A special case is the domain sense of aesthetics, which has been defined as the ability to enjoy sensory stimuli. Items on self-esteem involve perceptions of usefulness, self-confidence, and satisfaction with oneself. Observer-based ratings consist usually in noticing that patients mention they are worthless, hopeless or useless. Some questionnaires such as the ADRQL measure awareness of self with items such as responding to one’s own name. Some instruments also include activities of daily living in their assessment of HRQL. These items have been measured as the patient or proxy’s perception of the ability to do chores such as walking, cooking, eating and shopping or how much problem or worry these activities involve. The same applies to the assessment of cognitive functioning. Other domains that have not been included so often include self-rated health, financial and living context.

---Table 4 about here---
Psychometric properties

Table 5 shows the psychometric quality assessment using the criteria shown in table 1. Ten instruments (Albert et al., 1996; Salek et al., 1997; Black et al., 1999; Logsdon et al., 1999; Weiner et al., 2000; Selai et al., 2001; Fossey et al., 2002; Ready et al., 2002; Yamamoto-Mitani et al., 2002; Porzsolt et al., 2004) either were not tested for acceptability or did not meet the criteria.

Regarding reliability, all instruments but PES-AD+AES, DCM and DEMQOL-Proxy show good evidence of internal consistency. Eight instruments (Salek et al., 1997; Brod et al., 1999; Logsdon et al., 1999; Weiner et al., 2000; Yamamoto-Mitani et al., 2002; Smith et al., 2005; Ettema et al., 2007; Trigg et al., 2007) showed good evidence of test-retest reliability. Among those instruments assessed by a proxy or an observer, six instruments assessed inter-rater reliability, most of them showing good evidence. Regarding validity, 10 instruments showed good evidence of content validity (Salek et al., 1997; Black et al., 1999; Brod et al., 1999; Logsdon et al., 1999; Selai et al., 2001; Terada et al., 2002; Yamamoto-Mitani et al., 2002; Smith et al., 2005; Ettema et al., 2007; Trigg et al., 2007). Despite the fact that criterion-related validity cannot be tested due to the lack of a gold standard, two measures (Selai et al., 2001; Ready et al., 2002) claim to show some evidence of it, perhaps referring to convergent validity. Regarding analyses against external criteria, 14 instruments (Albert et al., 1996; Salek et al., 1997; Black et al., 1999; Brod et al., 1999; Logsdon et al., 1999; Weiner et al., 2000; Selai et al., 2001; Terada et al., 2002; Fossey et al., 2002; Ready et al., 2002; Porzsolt et al., 2004; Smith et al., 2005; Trigg et al., 2007; Ettema et al., 2007) show
evidence of convergent validity but not discriminant or known groups differences. Six measures have shown responsiveness (D-QoL, BASQID, QoL-AD, ADRQL, QUALID and DCM). Finally, the factorial structure of nine instruments has been analyzed (Brod et al., 1999; Salek et al., 1999; Weiner et al., 2000; Fossey et al., 2002; Terada et al., 2002; Yamamoto-Mitani et al., 2002; Smith et al., 2005; Trigg et al., 2007; Revell et al., 2009), most of them finding subscales.

---Table 5 about here---

**Discussion**

This review synthesizes the information on dementia specific HRQL instruments from reviews (Salek et al., 1998; Walker et al., 1998; Ready and Ott, 2003; Ettema et al., 2005; Smith et al., 2005; Lucas-Carrasco, 2007; Schozel-Dorenbos et al., 2007; Ready, 2011) and builds on these by gathering them all into a single study and identifying one new measure, the QLDJ.

While this is a systematic review, we did not include measures assessing only a specific domain of HRQL such as the PDS (DeJong et al., 1989), or generic HRQL scales that are not tailored to people with dementia (Rabins et al., 1999; Silberfeld et al., 2002). Their focus on health and function imply that HRQL will decrease automatically with disease progression (Ettema et al., 2005) and some of these lack evidence on validity or reliability in dementia.
populations (Ettema et al., 2005). We also did not include instruments that measure the QoL of caregivers of people with dementia as these are outside the scope of this review. The literature searches were conducted in English and in Anglophone databases and therefore instruments developed in non-English-speaking countries without a translation available would have been missed. However, regarding instrument development and validation, we accepted any publication as long as the abstract was in English or Spanish.

**Conceptualization and operationalization**

The concept of HRQL in dementia has been influenced by the broader concept of QoL that refers to “evaluation by subjective and social-normative criteria, of the behavioral and environmental situation of a person” (Lawton, 1994). There has been a movement in QoL towards the measurement of the experience of the person, including usually perceptions or satisfactions with psychological, physical and social domains (Ettema et al., 2005). In dementia-specific HRQL instruments, the measurement is also aimed at capturing the experience of the person with dementia. Domains commonly measured are mood, self-esteem, social interaction and enjoyment of activities. Similarly to the general concept of QoL, normative measures such as income or cognitive tests (different to perception or satisfaction with such domains) have been excluded and considered as a different outcome (Brod et al., 1999; Rabins et al., 1999). Conceptual frameworks do not differ in essence with Lawton’s model of HRQL in dementia (Lawton, 1994), although they limit the number of domains. Different perspectives on HRQL abound and are manifest in the
variety of domains represented in the different scales. For example, the measurement of HRQL using the ADRQL will widely differ to the one using the CDQLP since they have very few domains in common. Instruments also differ in their breadth of the assessment and whether subjective experience of the person with dementia is assessed or proxy measures (observational or not) are used instead. While self-reported measures are more appropriate for measuring the patient’s own experience, in severe stages, the use of instruments based on proxy ratings is inevitable. Further, we found that there is no single protocol for assessing the different domains across instruments. This limits cross-study comparability.

*Country of development and validation*

Most dementia-specific HRQL measures were developed in the USA and UK. A number of scales have been validated in other countries such as Spain and Japan (Suzuki *et al.*, 2005, 2006; Matsui *et al.*, 2006), increasing the availability. The QoL-AD has been validated in at least 10 different countries and the D-QoL has been validated in at least 5 different countries in the Americas, Europe and Asia. This allows for cross-county comparisons. However, the concept of HRQL may vary across different cultures and culturally specific scales may also be needed. Further, for validation studies is the issue of the representativeness of samples used in development and validations studies. None of the 15 instruments have been developed using a population-based sample raising issues of generalizability.
Dementia severity

Decisions for scale selection will depend on the dementia severity. Self-rated instruments would be more appropriate for mild-moderate stages and proxy and observer rated instruments more appropriate at more severe stages of the disease. Most of the extant dementia-specific HRQL measures are proxy rated. This is a consequence of the idea that people with dementia are not able to rate their own HRQL due to cognitive impairment. However, there is a strong movement in measurement of HRQL in dementia to obtain self-reports from the individual with dementia where possible given recent findings that suggest that people with mild to moderate dementia are aware and able to assess their HRQL (Brod et al., 1999; Mozley et al., 1999; Selai et al., 2001; Logsdon et al., 2002; Ready et al., 2002; Smith et al., 2005; Trigg et al., 2007). Proxy informants have a different point of view on the patient’s HRQL and tend to give lower rating than people with dementia (Logsdon et al., 2002; Thorgrimsen et al., 2003). Even with careful training of observers it is uncertain whether the observed behaviors represent the most important and relevant aspects of HRQL, as these measures have been developed to assess the subjective perceptions of QoL. However, it has been shown that proxies are as good as patients in detecting changes in HRQL over time (Sneeuw et al., 1997). Studies suggest that these differences in reporting can be explained by the disability paradox, caregiver states such as depression or burden and the lack of patient’s insight (Carr and Higginson, 2001; Logsdon et al., 2002; Novella et al., 2012), but also by methodological issues such as precision bias or response bias (Smith et al., 2005). Obtaining information from both sources (patients and caregivers) is important since they reflect
different and imperfect measurements of “true” state and may be contrasted, providing richer information. A different approach is assessing HRQL using different instruments covering all dementia severities (Edelman et al., 2005), however this method can be confusing since different conceptual frameworks are assumed when using different measures.

Purpose of assessment

Instruments may be used for clinical practice or for research. For clinical practice, since the main aim is to help the individual to continue living in the optimal fashion, the optimal scale will be the one that lets the persons with dementia assess a personal perception of HRQL allowing a more individualized and therefore effective treatment plan. To date, there is only one scale that allows this type of assessment, the QOLAS (Selai et al., 2001). This is a double edged sword since a drawback of this measure in research is that, for this same reason, scores of the different participants or the same participant in different stages might not reflect the same concept of HRQL and therefore comparisons do not seem to be completely pertinent. This raises the question of whether people can learn to consider certain aspects when assessing their HRQL and dismiss others. This would be useful for improving HRQL in dementia.

For cross-sectional studies, operationalization, country of development/validation and data collection method will guide the selection of
the scale. However, when it comes to longitudinal designs, the use of proxy and observer rated scales seems more appropriate since severity of dementia is likely to change over time. For both longitudinal and randomized control trials, responsiveness is an important factor. To date, at least six Instruments have proved to have a certain degree of responsiveness or sensitivity to change. These measures are the BASQID, QoL-AD, ADRQL, QUALID, PES-AD+AES and DCM. (Black et al., 1999; Albert et al., 2001; Fossey et al., 2002; Thorgrimsen et al., 2003; Martin-Cook et al., 2005; Trigg et al., 2007).

Psychometric properties

Almost every measure showed good evidence of internal consistency, and most measures have at least tested reliability in two different ways. All self-rated instruments and one scale of each of the other data collection methods showed at least some good evidence of acceptability. These scales are the D-QoL and BASQID for self-rated, DEMQOL for both self and proxy rated, and QoL-D and QUALIDEM for proxy-rated. Content validity has been mostly reached by qualitative evidence from pre-testing with patients, expert opinion and literature review that items in the scale are representative of the construct. In-depth qualitative interviews with people with dementia and/or their carers were conducted for the D-QoL, DEMQOL, BASQID, ADRQL, QUALIDEM, and QoL-D in order to generate the scale items. A special case is the QOLAS, in which the items are tailored to each respondent. Although 2 measures (Selai et al., 2001; Ready et al., 2002) claim to measure it, criterion-related validity cannot be assessed due to the lack of a gold standard with
which to compare it. Convergent validity has mainly been assessed by means of correlations between the scales and measures of dementia severity (Albert et al., 1996; Salek et al., 1996; Brod et al., 1999; Rabins et al., 1999; Weiner et al., 2000; Selai et al., 2001; Yamamoto-Mitani et al., 2002; Ready et al., 2002; Porzsolt et al., 2004; Smith et al., 2005; Ettema et al., 2007; Trigg et al., 2007), depression (Rabins et al., 1999; Selai et al., 2001; Logsdon et al., 2002; Smith et al., 2005; Ettema et al., 2007), activities of daily living (Rabins et al., 1999; Weiner et al., 2000; Selai et al., 2001; Logsdon et al., 2002; Terada et al., 2002; Yamamoto-Mitani et al., 2002; Porzsolt et al., 2004; Smith et al., 2005), behavioral and psychological symptoms (Weiner et al., 2000; Ettema et al., 2007) and other measures of QoL (Fossey et al., 2002; Smith et al., 2005; Trigg et al., 2007). Discriminant validity has not often been assessed and usually includes associations with gender, age or caregiver characteristics.

Scoring

A profile score or subscales of nine out of 15 instruments can be obtained. Within these nine, three (ADRQL, QLDJ and BASQID) (Rabins et al., 1999; Yamamoto-Mitani et al., 2002; Trigg et al., 2007) can also calculate a total HRQL score. The possibility of calculating subscale scores is vital for the assessment of HRQL. Firstly, because HRQL is by definition a multi-dimensional concept and therefore, the scores should represent each domain. Secondly, if HRQL is to be used in order to assess treatment benefits, subscale scores may shape those treatments.

Conclusion
There has been much development on the measurement of dementia-specific HRQL in the last two decades. This is perhaps due to the lack of consensus on the concept of HRQL in dementia. It is also a reflection of the importance of assessing HRQL in this group of people (Karlawish et al., 2000; Whitehouse, 2000; Moniz-Cook et al., 2008). The fact that dementia may be expressed in many different ways makes the concept of HRQL in dementia even more complex. The suitability of the scale depends on several factors; namely countries (development/validation), dementia severity (mild, moderate or severe), data collection method (patient, proxy, both or observer rating), purpose of the assessment (clinical practice, research), conceptualization and operationalization (domains and items), psychometric properties (validity and reliability) and scoring (total scores vs. subscales). Many studies do not specifically set out to measure dementia-specific-HRQL but do include related items as part of their surveys that together may function as a HRQL scale (e.g. items on depression, social interaction or enjoyment of activities). This paper may also be useful for mapping HRQL dimensions in such studies maximizing the value of existing resources. In order to do so, item analyses could determine the performance of frequently used questions in dementia HRQL instruments. These items could also be mapped onto already existing conceptual frameworks. The items to be used will depend on how severe the dementia is and how the information was administered. However, validity and reliability will have to be analyzed if this is to be conducted. Future studies validating each scale across samples (community vs. clinical), populations and cohorts are needed to be able to generalize results.
Conflict of interest declaration

None

Description of authors’ roles

J. Perales conducted literature review, the data collection and wrote the paper
T. D. Cosco conducted the literature review and assisted with writing the paper
B. C.M. Stephan supervised the data collection and assisted with writing the paper
J.M. Haro assisted with writing the paper
C. Brayne supervised the data collection, development of the paper and assisted with writing the paper

Acknowledgments

J.P. is grateful to the Instituto de Salud Carlos III for a predoctoral grant (PFIS) and to Prof Betty E Black and Prof Sam Salek for providing additional literature.
References


Quality of Life in Alzheimer’s Disease (QOL-AD) scale in Mexican patients with Alzheimer, vascular and mixed-type dementia. *Revista de Neurologia.* 51, 804–808.


Walker, M.D., Salek, S. and Bayer, A. (2001). The Relationship Between the Quality of Life (QQL) of Dementia Patients and Their Carers: Validation of the Community Dementia Quality of Life Profile. *Age and ageing*. 30, 50.


<table>
<thead>
<tr>
<th>Psychometric property</th>
<th>Definition/ test</th>
<th>Criteria for acceptability</th>
</tr>
</thead>
</table>
| **Item analysis/ reduction** | Identification of items for possible elimination owing to weak psychometric performance; assessed on the basis of (1) unrotated principal component factor analysis to determine whether all items are measuring a single factor; and (2) item analyses for all items | **Acceptability**<br>Principal component factor analysis:<br>All items should load on the first unrotated factor >0.30<br>**Item analyses (applied to all items):**<br>Missing data < 5%<br>No item redundancy (inter-item correlations ≤ 0.75)<br>Maximum endorsement frequencies ≤80% (i.e. the proportion of respondents who endorse each response category), including floor/ceiling effects <80% (i.e. response categories with high endorsement rates at the bottom/top ends of the scale, respectively)<br>Aggregate adjacent endorsement frequencies ≥10%<br>Missing data for summary scores <5%<br>Even distribution of endorsement frequencies across response categories<br>Floor ceiling effects for summary scores <10%<br>**Reliability**<br>**Internal consistency**<br>The extent to which items comprising a scale measure the same construct (e.g. homogeneity of the scale); assessed by Cronbach’s alpha and item-total correlations<br>Cronbach’s alphas for summary scores ≥.70<br>**Test-retest**<br>The stability of a measuring instrument; assessed by administering the instrument to respondents on two different occasions and examining the correlation between test and retest scores<br>Test–retest reliability correlations for summary scores ≥.70<br>**Inter-rater**<br>Agreement between independent raters/observers; assessed by ICCs<br>ICC≥.70<br>**Parallel (alternative) forms**<br>Agreement between two or more parallel/ alternative forms or different versions of the same measure (e.g. form A/B, short/long form) that indicates that they can be used interchangeably; assessed on the basis of correlations between parallel/ alternative forms of a measure<br>High correlations between parallel/alternative forms of the measure (e.g. between long and short form)<br>**Validity**<br>**Content**<br>The extent to which the content of a scale is representative of the conceptual domain it is intended to cover; assessed qualitatively during the questionnaire development stage through pre-testing with patients, expert opinion and literature review<br>Qualitative evidence from pre-testing with patients, expert opinion and literature review that items in the scale are representative of the construct being measured<br>**Criterion related**
### Construct validity

<table>
<thead>
<tr>
<th>Evidence</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concurrent evidence</td>
<td>High correlation between the scale and the criterion measure.</td>
</tr>
<tr>
<td>Predictive evidence</td>
<td>High correlation between the scale and the criterion measure.</td>
</tr>
</tbody>
</table>

**Concurrent**
- Evidence that the scale predicts a gold-standard criterion that is measured at the same time; assessed on the basis of correlations between the scale and the criterion measure.
- Evidence that the scale predicts a gold-standard criterion that is measured in the future; assessed on the basis of correlations between the scale and the criterion measure.

**Predictive**
- Evidence that a single entity (construct) is being measured and that items can be combined to form a summary score; assessed on the basis of evidence of good internal consistency and correlations between scale scores (which purport to measure related aspects of the construct).
- High correlation between the scale and the criterion measure.

**Analyses against external criteria**

#### Convergent
- Evidence that the scale is correlated with other measures of the same or similar constructs; assessed on the basis of correlations between the measure and other similar measures.
- Moderate to high correlations between scale scores.

#### Discriminant
- Evidence that the scale is not correlated with measures of different constructs; assessed on the basis of correlations with measures of different constructs.
- Significant differences between known groups or difference of expected magnitude.

**Known groups differences**
- The ability of a scale to differentiate known groups; assessed by comparing scores for subgroups who are expected to differ on the construct being measured.
- Significant differences between known groups or difference of expected magnitude.

### Responsiveness

- The ability of a scale to detect clinically important change over time; assessed by comparing scores before and after an intervention of known efficacy (on the basis of various methods including t-tests, effect sizes, standardized response means, or responsiveness statistics).
- Significant differences between known groups or difference of expected magnitude.

Adapted from Smith *et al.*, 2005 (Smith *et al.*, 2005)
Table 2. Dementia-specific HRQL instruments included in each of the eight reviews identified in our search

<table>
<thead>
<tr>
<th>Scale/Review</th>
<th>Walker &amp; Salek 1998(^1), 2</th>
<th>Ready 2003(^3)</th>
<th>Ettema 2005(^4)</th>
<th>Smith 2005(^5)</th>
<th>Lucas-Carrasco 2007(^6)</th>
<th>Schozel-Dorenbos 2007(^7)</th>
<th>Ready 2011(^8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-QOL</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>QOL-AD</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCM</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>CBS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>QOLAS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADRQL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QUALID</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QOL-D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>DEMQOL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>QUALIDE M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>BASQID</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>QLDJ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PES-AD+AES</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDQLP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vienna List</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1: (Salek et al., 1998); 2: (Walker et al., 1998) 3: (Ready and Ott, 2003); 4: (Ettema et al., 2005); 5: (Smith et al., 2005); 6: (Lucas-Carrasco, 2007); 7: (Schozel-Dorenbos et al., 2007); 8: (Ready, 2011)
Table 3. Primary and secondary papers on dementia-specific HRQL measures identified in our search

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Primary papers (development)</th>
<th>Secondary papers (modifications &amp; validations in other countries)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzheimer Disease Related Quality of Life (ADRQL)</td>
<td>1, 2,</td>
<td>3, 4, 5, 6</td>
</tr>
<tr>
<td>Bath Assessment of Subjective Quality of Life in Dementia (BASQID)</td>
<td>7, 8</td>
<td></td>
</tr>
<tr>
<td>Cornell-Brown Scale for Quality of Life in Dementia (CBS)</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Community Dementia Quality of Life Profile (CDQLP)</td>
<td>10-14</td>
<td></td>
</tr>
<tr>
<td>Dementia Care Mapping (DCM)</td>
<td>15, 20</td>
<td>5, 16-18</td>
</tr>
<tr>
<td>Dementia Quality of Life Instrument (D-QoL)</td>
<td>23</td>
<td>5, 6, 19, 24-32, 33</td>
</tr>
<tr>
<td>Patient Activity Scale-AD plus the Modified Apparent Emotion Scale (PES-AD + AES)</td>
<td>34, 35</td>
<td>5</td>
</tr>
<tr>
<td>Quality of Life for Older People Experiencing Dementia (QLDJ)</td>
<td>36, 37</td>
<td></td>
</tr>
<tr>
<td>Quality of Life-Alzheimer's Disease (QOL-AD)</td>
<td>38, 39</td>
<td>6, 19, 40-56</td>
</tr>
<tr>
<td>Quality of Life Assessment Schedule (QOLAS)</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>Quality of Life for Dementia (QOL-D)</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>Quality of Life in Late-stage Dementia Scale (QUALID)</td>
<td>59</td>
<td>60-64</td>
</tr>
<tr>
<td>QUALIDEM</td>
<td>65-67</td>
<td></td>
</tr>
<tr>
<td>Vienna List</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>Measure</td>
<td>No. items</td>
<td>Scoring</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------</td>
<td>---------</td>
</tr>
<tr>
<td>Self-rated Only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BASQID</td>
<td>14</td>
<td>Total &amp; subscales</td>
</tr>
<tr>
<td>D-QoL</td>
<td>29</td>
<td>Subscales</td>
</tr>
<tr>
<td>Patient &amp; Proxy rated</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBS</td>
<td>19</td>
<td>Total</td>
</tr>
<tr>
<td>DEMQOL</td>
<td>28 (p), 31 (c)</td>
<td>Total</td>
</tr>
<tr>
<td>QoL-AD</td>
<td>12</td>
<td>Total</td>
</tr>
<tr>
<td>QOLAS</td>
<td>10</td>
<td>Total</td>
</tr>
</tbody>
</table>

p= patient version, c= caregiver version
<table>
<thead>
<tr>
<th>Measure</th>
<th>No. items</th>
<th>Scoring</th>
<th>Response Scale</th>
<th>Severity</th>
<th>Time frame</th>
<th>Dimensions</th>
<th>Validated in population studies</th>
<th>Country of validation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proxy rated Only</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADRQL</td>
<td>47</td>
<td>Total &amp; subscales</td>
<td>Dichotomous (agree/disagree)</td>
<td>All stages</td>
<td>Previous 2 weeks</td>
<td>Social interaction, awareness of self, feelings and mood, enjoyment of activities, response to surroundings</td>
<td>No</td>
<td>USA (original)</td>
</tr>
<tr>
<td>CDQLP</td>
<td>33</td>
<td>Total</td>
<td>4 point: (not at all-always)</td>
<td>All stages</td>
<td>Previous 2 weeks</td>
<td>Thinking &amp; behavior, family &amp; social life, physical activities, other aspects of daily living</td>
<td>No</td>
<td>UK (original)</td>
</tr>
<tr>
<td>PES-AD+AES</td>
<td>21</td>
<td>Subscales</td>
<td>Activity: 3 point (frequently-never), affect: 5 point (never- ≥3xday)</td>
<td>Mild</td>
<td>Affect: previous 2 weeks, activity: previous week</td>
<td>Previous 2 weeks</td>
<td>No</td>
<td>USA (original)</td>
</tr>
<tr>
<td>QLDJ</td>
<td>24</td>
<td>Total &amp; subscales</td>
<td>4 point (not at all applicable, very much applicable)</td>
<td>Moderate-severe</td>
<td>Previous 2 weeks</td>
<td>Interacting with surroundings, expressing self, experiencing minimum negative behaviors</td>
<td>No</td>
<td>Japan (original)</td>
</tr>
<tr>
<td>QoL-D</td>
<td>31</td>
<td>Subscales</td>
<td>4 point (none-frequent)</td>
<td>Moderate-severe</td>
<td>Not specified</td>
<td>Positive affect, negative affect and actions, ability of communication, restlessness, attachment with others, spontaneity and activity</td>
<td>No</td>
<td>Japan (original)</td>
</tr>
<tr>
<td>QUALID</td>
<td>11</td>
<td>Total</td>
<td>5 point (different options of frequency)</td>
<td>Moderate-Severe</td>
<td>Previous week</td>
<td>Affect, comfort and engagement with activities</td>
<td>No</td>
<td>USA (original), Sweden, Spain</td>
</tr>
<tr>
<td><strong>Observational assessment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCM</td>
<td>26</td>
<td>Subscales</td>
<td>Well-being: 6 point (extreme ill-being-extreme well-being)</td>
<td>All stages</td>
<td>Every 5 minutes for 6 hours</td>
<td>Well-being, activity</td>
<td>No</td>
<td>UK (original)</td>
</tr>
<tr>
<td>QUALIDEM</td>
<td>37</td>
<td>Subscales</td>
<td>4 point (never-often)</td>
<td>All stages</td>
<td>Previous 2 weeks</td>
<td>Care relationships, positive affect, negative affect, restless tense behavior, positive self-image, social relations, social isolation, feeling at home, having something to do</td>
<td>No</td>
<td>Netherlands (original)</td>
</tr>
<tr>
<td>Vienna List</td>
<td>40</td>
<td>Subscales</td>
<td>5 point: (never-always)</td>
<td>Severe</td>
<td>Present</td>
<td>Communication, negative affect, bodily contact, aggression, mobility</td>
<td>No</td>
<td>Austria (original)</td>
</tr>
<tr>
<td>Measure</td>
<td>Acceptability</td>
<td>Internal consistency</td>
<td>Test-retest</td>
<td>Inter-rater reliability</td>
<td>Content validity</td>
<td>Criterion-related validity</td>
<td>Convergent validity</td>
<td>Discriminant validity</td>
</tr>
<tr>
<td>---------</td>
<td>---------------</td>
<td>----------------------</td>
<td>-------------</td>
<td>------------------------</td>
<td>-----------------</td>
<td>--------------------------</td>
<td>-------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>BASQID</td>
<td>++</td>
<td>+++</td>
<td>+++</td>
<td>NA</td>
<td>+++</td>
<td>-</td>
<td>+++</td>
<td>++</td>
</tr>
<tr>
<td>D-QoL</td>
<td>++</td>
<td>+++</td>
<td>+++</td>
<td>NA</td>
<td>+++</td>
<td>-</td>
<td>++</td>
<td>-</td>
</tr>
<tr>
<td>CBS</td>
<td>-</td>
<td>+++</td>
<td>-</td>
<td>NA (c)/+++(p)</td>
<td>+</td>
<td>++</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DEMQOL</td>
<td>++ (c)/++++(p)</td>
<td>+++ (c)</td>
<td>+++</td>
<td>NA (c)/+++(p)</td>
<td>+++</td>
<td>-</td>
<td>++</td>
<td>- (c)/+++(p)</td>
</tr>
<tr>
<td>QoL-AD</td>
<td>-</td>
<td>+++</td>
<td>+++</td>
<td>NA (c)/-(p)</td>
<td>+++</td>
<td>-</td>
<td>+++</td>
<td>- (c)/+++(p)</td>
</tr>
<tr>
<td>QOLAS</td>
<td>-</td>
<td>+++</td>
<td>-</td>
<td>NA (c)/-(p)</td>
<td>+++</td>
<td>+</td>
<td>++</td>
<td>-</td>
</tr>
</tbody>
</table>

p= patient version, c= caregiver version
<table>
<thead>
<tr>
<th>Measure</th>
<th>Acceptability</th>
<th>Internal consistency</th>
<th>Test-retest</th>
<th>Inter-rater reliability</th>
<th>Content validity</th>
<th>Criterion-related validity</th>
<th>Convergent validity</th>
<th>Discriminant validity</th>
<th>Known groups differences</th>
<th>Responsiveness</th>
<th>Factor or principal components analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proxy rated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADRQL</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+++</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Yes</td>
<td>-</td>
</tr>
<tr>
<td>CDQLP</td>
<td>-</td>
<td>+++</td>
<td>-</td>
<td>-</td>
<td>+++</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Yes</td>
<td>-</td>
</tr>
<tr>
<td>PES-AD + AES</td>
<td>-</td>
<td>-</td>
<td>+++</td>
<td>-</td>
<td>+++</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>Yes</td>
<td>-</td>
</tr>
<tr>
<td>QLDJ</td>
<td>-</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>++</td>
<td>-</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>Yes</td>
<td>-</td>
</tr>
<tr>
<td>QoL-D</td>
<td>+++</td>
<td>+++</td>
<td>-</td>
<td>+++</td>
<td>+++</td>
<td>-</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>Yes</td>
<td>-</td>
</tr>
<tr>
<td>QUALID</td>
<td>-</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>++</td>
<td>-</td>
<td>+++</td>
<td>-</td>
<td>-</td>
<td>Yes</td>
<td>-</td>
</tr>
<tr>
<td>QoL-D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OBSERVATIONAL ASSESSMENT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADRQL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDQLP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PES-AD + AES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QLDJ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QoL-D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QUALID</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QoL-D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QUALIDEM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vienna List</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>

Criteria
- No evidence or not tested.
+ some limited evidence
++ some good evidence but some aspect do not meet criteria or some aspects not tested/reported
+++ good evidence

Example
ADRQL did not assess acceptability
CDQLP only reported the p value when examining the association between the total score and the MMSE score
DEMOLQ acceptability: missingness is always higher than 5% although there is no floor or ceiling effect.
DEMOLQ proxy’s acceptability: missingness is always lower than 5% and there are no floor or ceiling effects.
Inter-rater reliability in self-rated scales
Figure 1. Flow chart of article selection for the review on dementia-specific health related quality of life instruments