

Postprint Version	1.0
Journal website	http://onlinelibrary.wiley.com/doi/10.1002/gps.2585/abstract
Pubmed link	http://www.ncbi.nlm.nih.gov/pubmed/20845397
DOI	10.1002/gps.2585

This is a NIVEL certified Post Print, more info at <http://www.nivel.eu>

Evaluation of Qualidem: a dementia-specific quality of life instrument for persons with dementia in residential settings; scalability and reliability of subscales in four Dutch field surveys

I. E. BOUMAN^{1,2}, T. P. ETTEMA³, R. B. WETZELS⁴, A. P. A. VAN BEEK⁵, J. DE LANGE⁶ AND R. M. DRÖES^{1,2}

¹Department of Psychiatry, VU University Medical Center, Valeriusplein 9, 1075 BG Amsterdam, The Netherlands

²Institute for Health and Care Research (EMGO), Alzheimer Center, Amsterdam, The Netherlands

³Stichting SHDH, Haarlem, The Netherlands

⁴Primary and Community Care, Centre for Family Medicine, Elderly Care Medicine and Public Health, Radboud University Nijmegen, Medical Centre, Nijmegen, The Netherlands

⁵NIVEL-Institute, Netherlands Institute for Health Services Research, Utrecht, The Netherlands

⁶Trimbos-Institute, Netherlands Institute of Mental Health and Addiction, Utrecht, The Netherlands

Correspondence to: Prof. Dr. R. M. Dröes, E-mail: rm.droes@vumc.nl

Objective: To evaluate the scalability and reliability of Qualidem, a quality of life observation instrument rated by professional caregivers of persons with mild to very severe dementia living in residential settings.

Method: Data from four field surveys in the Netherlands were used. The instrument consists of nine subscales for mild to severe dementia; of which six can be applied in very severe dementia. The Mokken scaling model was used to compute scalability and reliability coefficients for each subscale and dementia group.

Results: Seven hundred fifty nine persons with mild to severe dementia and 214 persons with very severe dementia residing in 36 nursing homes and 4 homes for the elderly were included. In general, the subscales for the mild to severely demented group were scalable and (moderate) sufficiently reliable; the results confirmed the results of previous research to develop Qualidem. For the very severe demented group, four out of six subscales were scalable (Care relationship, $H=0.56$; Positive affect, $H=0.55$; Restless tense behavior, $H=0.42$; and Social isolation, $H=0.34$); they were also sufficiently reliable to measure quality of life ($0.57 \leq p \leq 0.82$). From the other two measured subscales, Social relations was considered not scalable ($H=0.26$) and Negative affect was scalable ($H=0.36$), but insufficiently reliable ($p=0.40$).

Conclusion: Qualidem is an easy to administer and overall moderately sufficient reliable rating scale that provides a quality of life profile of persons with mild to even very severe dementia living in residential settings.

INTRODUCTION

Quality of life (QoL) has become a major concept within the care for people with dementia (Lawton, 1997; Whitehouse et al., 1997; Logsdon and Albert, 1999). Some authors even consider it the primary purpose of health care for those persons (Gwyther, 1997). It is, therefore, an essential outcome of treatment effectiveness. New methods of care, or the improvement of existing programs, aiming to improve QoL of people suffering from dementia need to be assessed with adequate QoL measures. Several dementia-specific QoL measures have been developed over the last decade, mostly in the English language (Ettema et al., 2005). In Dutch dementia research and in the residential care of people with dementia the need for a QoL measure was also felt. The Discomfort Scale- Dementia of Alzheimer Type (DS-DAT) was specially developed for non-communicative patients with advanced Alzheimer's Disease (Hurley et al., 1992). A Dutch version of the DS-DAT showed good interobserver reliability and to constitute one concept in a sample with moderate to severe dementia (Van der Steen et al., 2002). In an recent overview of dementia-specific QoL measures by Schölzel-Dorenbos et al. (2007a) for patients with advanced dementia in residential care, the DS-DAT and the Qualidem were recommended. The use of these observational instruments can be seen as complementary to each other for people with very severe dementia. In addition to emotional and social domains of QoL, the Qualidem also focuses on the care relationship and coping with the nursing home environment. The Qualidem instrument was recently developed in the Netherlands for people suffering from dementia living in residential settings (Ettema et al., 2007a, 2007b). In developing Qualidem, it became possible to assess QoL of all residents, including the very severely demented. The Qualidem instrument was tested on psychometric properties in persons with dementia and proved reliable and valid (Ettema, 2007). However, the instrument should be regarded as a first version. Replication of the analysis on new data is a good practice in research and can be carried out on data gathered in several new research projects. If the results are similar any doubts about chance results can then be eliminated. The results from the subscales developed for the very severely demented group were also based on a small sample size (n=36). The focus of this paper is therefore to investigate whether the scalability of the subscales of the Qualidem instrument can be replicated, especially in persons with very severe dementia, using data from four Dutch field surveys. Additionally, we aimed to provide insight into the distribution of scores on the subscales in the groups with mild to severe and very severe dementia to make comparative information available.

METHOD

Four field tests have been carried out using the Qualidem, a dementia-specific QoL observation questionnaire rated by professional caregivers. We combined data from: (1) the original Qualidem study for the development and evaluation of this instrument (Ettema, 2007); (2) a pilot project into the QoL of residents of psychogeriatric nursing homes carried out by the Netherlands Institute of Health Services Research (NIVEL-a) and initiated by ActiZ, branch organization for care entrepreneurs (Poortvliet et al., 2006); (3) another field test carried out by NIVEL (NIVEL-b; Merten et al., 2007); (4) and the 'Waalbed-II' study, a longitudinal prospective cohort study of neuropsychiatric symptoms and QoL in dementia patients in nursing homes (De Jonghe et al., 2009). The Qualidem data were anonymously handed over for further analysis. All data were collected between 2003 and 2008.

Participants

Residents of nursing homes in the Netherlands were included when diagnosed with dementia. Patients suffering from Parkinson's disease, from a diminution of consciousness, or receiving terminal care were excluded. The maximum number of residents per nursing home was set at 40 to keep the required effort of nursing home staff at an acceptable level. A representative group in each nursing home was approached for the research projects. The Qualidem and Nivel studies were performed in several Dutch nursing homes spread over the country. The nursing homes volunteered in reaction to a call for participation. The major requirement for testing a questionnaire is a heterogeneous sample with respect to severity of dementia drawn from the population in which the questionnaire is to be used. Such a sample is not necessarily random. Although the method of selecting the sample was aimed at a broad range of nursing homes to increase representation, we are not sure how representative the sample is of the larger population of people with dementia. In the prospective cohort Waalbed-II-study, residents were enrolled from fourteen dementia

special care units (SCU) from nine nursing homes in the Netherlands who are part of the Nijmegen University Nursing Home Network; they represent an average sample of the Dutch nursing home residents.

Procedures

In the Qualidem study, certified nursing assistants (CNAs) scored the instrument, applying an observation period of 2 weeks. The CNAs received an oral instruction during a planned meeting of 1 h in the nursing home at the start of the observation period. The rating scales were provided with a clearly written instruction. In each nursing home a random sample of seven residents from the participating residents was drawn for investigation of inter-observer reliability. Two CNAs were instructed to observe the same resident independently from each other, and not to compare their scores. In the other three field surveys, two CNAs independently scored the Qualidem and consensus was reached in case of discrepancies. The questions were answered about the past week, in which they had observed the resident. More details of the participants and followed procedures can be found in the original publications (Poortvliet et al., 2006; Ettema, 2007; Merten et al., 2007; De Jonghe et al., 2009).

Instruments

The Qualidem consists of 37 items describing observable behavior. The list contains 16 indicative and 21 contra-indicative items that make up nine homogeneous subscales: Care relationship, Positive affect, Negative affect, Restless tense behavior, Positive self image, Social relations, Social isolation, Feeling at home, and Having something to do. The four response options are never, rarely, sometimes, and frequently. The items are printed in random order, so that items supposed to reflect a subscale are spread around the list. Administration is easy and scoring by the professional caregiver takes approximately 15 min. Six of the nine subscales can be applied in very severe dementia, using approximately half the items (18/37 items; see Table 4). Three items are retained for future research: 'Enjoys meals' was not scalable, but was considered important in establishing QoL, and 'Does not want to eat' and 'Likes to lie down (in bed)' may supplement the Qualidem instrument, because these are observable in very severely demented persons. The validity of the Qualidem instrument was supported by the method of construction in combination with a validation study. First, QoL in dementia was defined and the QoL domains were established based on theory. Second, the results of a performed study on convergent and discriminant validity further supported construct validity. The inter-rater reliability was acceptable for most subscales. The scores of single observers of the same person were used to calculate the intra-class correlation (ICC) based on the two-way random effects model, following the definition of absolute agreement (McGraw and Wong, 1996). Calculation of the average score ICC provided an estimate of inter-rater reliability when two observers cooperated in scoring the Qualidem. The error variance substantially reduced in this way and it was therefore strongly recommended to use two raters who score the Qualidem independently and then reach consensus on differences in scores (Ettema, 2007). Indeed in the other three field tests included in this manuscript, two observers rated the Qualidem according to the instructions accompanying this test. Qualidem does not measure disease severity. To assess the severity of dementia of the participants, the Dutch version of the Global Deterioration Scale (GDS; Reisberg et al., 1982; Muskens, 1993) was used. This scale is a staging instrument indicating cognitive deterioration in dementia. The GDS provides an overview of seven different stages: from pre-dementia stages 1–3 to stage 7 for persons with very severe cognitive decline. Based on observable behavior matching stage 7 of the GDS, the NIVEL studies determined whether participants either belonged to the mild to severe dementia group (GDS<7) or to the very severe dementia group (GDS=7). A detailed description of the assessment is available in the manual of the observation questionnaire 'Zicht op eigen leven' (Poortvliet et al., 2006).

Analysis

To examine whether the subscales of the Qualidem are represented in the data, Mokken's non-parametric latent trait model of one-dimensional scaling (Sijtsma and Molenaar, 2002) was used, by means of the MSP software (version 5.0) (Molenaar and Sijtsma, 2000). In common factor analysis, one would test whether a set of items fits the linear common one-factor model. However, this method requires item scores to be normally distributed and measured at interval level. It is very unlikely that this assumption is met with fourpoint Likert type items. With the technique it is possible to test whether a set of ordinal polytomous items form a homogeneous unidimensional scale, that allows for reliable ranking of the subjects. Central in

the analysis is Loevinger's H-coefficient of homogeneity or scalability. The scalability of a single item in relation to the other items in the scale is defined by H_i and the scalability of the total scale is measured by H . High coefficients indicate that the assumptions of a Mokken scale are very likely met. An important assumption of the Mokken scaling model is unidimensionality of the scales, meaning that all items of a subscale measure the same latent trait. It is easy to understand the logic behind this assumption: suppose a scale is multidimensional and measures two latent traits. It would be almost impossible to interpret the measurement results: which of the latent traits accounts for the data structure? A thorough treatment of the subject can be found in Sijtsma and Molenaar (2002). All H_i should be non-negative for the Mokken model to hold. However, items with positive values below 0.30 have weak discrimination power and are considered not scalable. The total scale should have a H -value of at least 0.30 to form a weak scale. Values of H between 0.40 and 0.50 indicate moderate scalability and values of 0.50 and above indicate strong scalability. The scores of contra-indicative items were reversed so that high scores on any item reflect a high score on the latent trait, reflecting a better QoL. The scores were linearly transformed to range from 0 to 100. Scalability and reliability coefficients are presented. For the distribution of scores within each subscale, percentile scores were calculated. Score range, mean and standard deviation per subscale were calculated. The missing value analysis was carried out using SPSS version 15.0.

RESULTS

Sample for analyses

The total sample size we used for the analysis ($n=973$) consisted of 238 cases from the Qualidem study, 179 cases from the NIVEL-a study, 268 cases from the NIVEL-b study and 288 cases from the Waalbed-II study. From these study samples, the number of cases reported suffering from mild to severe dementia (total $n=759$) was 202/238, 161/179, 189/268 and 207/288, respectively; and the number of cases reported with very severe dementia (total $n=214$) was 36/238, 18/179, 79/268 and 81/288, respectively. In total, 36 nursing homes and 4 homes for the elderly participated, and between 9 and 36 persons with dementia were included per ward. A detailed overview of the characteristics from the included surveys is presented in Table 1.

[TABLE 1]

Characteristics of the participants with mild to severe dementia and very severe dementia are presented in Tables 2 and 3, respectively. The average age of the participants with mild to severe dementia varied between 83 and 85 years in the four studies; between 71% and 76% was female (Table 2). The persons with very severe dementia were, on average, between 84 and 85 years of age; between 81% and 94% was female; a higher percentage of participants was diagnosed with Alzheimer type and/or vascular dementia in the Qualidem study (87%) compared to the Waalbed-II study (60%) (diagnoses in the NIVEL studies were not available, Table 3). Within the 'mild to severe' and 'very severe' group most characteristics of the participants were comparable between the studies.

Missing value analysis

From the original Qualidem data, 29 responses out of the maximum of 11 662 responses (238 demented persons times 49 originally tested Qualidem items) were missing (0.2%, see Table 1). From the other three studies, the percentages of missing items varied between 0.1% and 0.3% in the data from participants with mild to severe dementia, and between 0.8% and 0.9% in the data from participants with very severe dementia. This was considered sufficient support for these responses to be missing completely at random. The missing responses were imputed using the EM algorithm (Schafer and Graham, 2002).

Mokken scale analysis

For replication purposes, we examined the scalability and reliability of the nine subscales of Qualidem for the group of persons with mild to severe dementia. The results are presented in Table 4. For each study, and the combined studies, the item- and scale- H coefficients for each subscale are presented (Loevinger's scalability coefficients). The reliability coefficients (r) are also presented per subscale. The total scores ($n=759$) showed that the scalability of the subscales Care relationship, Positive affect, Negative affect and

Restless tense behavior is strong ($0.50 \leq H \leq 0.64$); for Positive self image, Social relations, Feeling at home and Having something to do, the scalability was moderate ($0.40 \leq H \leq 0.45$); and for Social isolation it was very weak ($H=0.29$). The reliability coefficient r varied from 0.55 for Social isolation/Having something to do to 0.90 for Positive affect. All subscales are considered (moderate) sufficiently reliable in this group. Overall, the scale- H values of the four studies combined were comparable to the original Qualidem data. Only minor variations occurred, for example, the scale- H value for the subscale Negative affect increased from 0.50 (Qualidem) to 0.55 (total score), while the subscales for Social relations, Social isolation and Having something to do decreased somewhat (from 0.50 to 0.45, from 0.34 to 0.29 and from 0.46 to 0.40, respectively). The reliability coefficients were comparable between the four studies. The analysis was repeated in the group of residents with a GDS score of 7. Only the items from each subscale which are observable in this stage of the disease were used (see Method Section) (Table 5, the underlined item numbers). The scalability and reliability coefficients are presented again for each study separately and for the entire group of very severely demented persons. The total scores ($n=214$) showed again that the scalability of the subscale Care relationship and Positive affect proved strong ($H=0.56$ and $H=0.55$, respectively); for Restless tense behavior, however, scalability was moderate ($H=0.42$); and for Negative affect and Social isolation it was weak ($H=0.36$ and $H=0.34$, respectively). The subscale Social relations was considered not scalable ($H=0.26$). Except for the subscales Negative affect and Social relations (r 0.46 and 0.40), the reliability coefficient r was satisfactory to very satisfactory (0.57 for Social isolation to 0.82 for Positive affect). Most subscales (total scores) became stronger compared to the original Qualidem data for this group of very severely demented persons, for example for Care relationship the scale- H increased from 0.41 to 0.56, for Positive affect from 0.33 to 0.55, and for Restless tense behavior from 0.30 to 0.42. The scale- H value of the subscale Social isolation remained the same (0.34), while the scale- H value for Negative affect decreased somewhat (from 0.40 to 0.36). All H -values were positive.

[TABLE 2, 3, 4, 5]

Percentile scores

Percentile scores of 10%, 25%, 50%, 75%, and 90% were calculated for the subscales of the Qualidem for both dementia groups ($GDS < 7$ and $GDS = 7$). The results are shown in Table 6. The table shows the percentages of persons that received a score above or below a certain score on the particular subscale. For instance, on the subscale Care relationship, only 10% of the participants with mild to severe dementia had a score 43 or lower, while 50% scored between 76 and 100 (range 0–100, transformed scale). For all scales, a higher score means more-favorable behavior, reflecting a better quality of life; score 100 on the Care relationship subscale means that the person, for example, had never rejected help nor had been in conflict with the nursing assistants or had been angry in the week prior to filling in the questionnaire; score 0 means that this behavior had occurred frequently. These three items may also be observed in persons with very severe dementia: 50% had a score above 78 (range 0–100); score 100 means these negative traits had never occurred and score 0 means all three traits had frequently occurred in the previous week. For Restless tense behavior, containing the same three items for both groups, the persons with mild to severe dementia had a higher, more-favorable, score on the 50% percentile than the persons with very severe dementia (56 vs. 44, respectively); for the subscale Social isolation, however, 50% of persons with mild to severe dementia had a score of 67 or higher, while in the group of persons with very severe dementia 50% even had a score of 78 or higher (Table 6). For two subscales the scores were lower compared to the rest of the subscales. On the subscales Restless tense behavior and Social relations for persons with very severe dementia, the 50% percentiles were on the left side of the mid-range score. (The latter subscale was, however, considered not scalable and insufficiently reliable.) The 50% percentiles of the other subscales, for both groups, were on the right side of the midrange score, indicating that for those 50% of persons with dementia, the subscale scores are relatively high, reflecting a better quality of life in that particular domain.

[TABLE 6]

DISCUSSION

The overall data for persons with mild to severe dementia showed that the original Qualidem results can be replicated in other residential settings. The instrument is a sufficiently reliable rating scale that provides

a QoL profile of persons with dementia. The overall data from four field studies showed furthermore that four out of six subscales of the Qualidem questionnaire for people with very severe dementia living in residential settings were well scalable and also sufficiently reliable to measure QoL. From the other two measured subscales, one was considered not scalable (Social relations), and the other one (Negative affect) was scalable (weak), but insufficiently reliable. It can therefore be concluded that four, out of total nine, subscales of the Qualidem, that is, Care relationship, Positive affect, Restless tense behavior and Social isolation, can be applied in very severe dementia as well, using approximately half the items (Table 5). The subscales Positive self image, Feeling at home, and Having something to do cannot be observed in this patient group, due to loss of function (Ettema, 2007). This patient group in the final stage can be characterized by loss of all verbal abilities, the need for full assistance in toileting and feeding, and loss of basic psychomotor skills, like the ability to walk. The changes in the H-values of the subscales for the group with mild to severe dementia, comparing the original Qualidem data (n=202) with the combined data of the four surveys (n=759, including Qualidem), showed only minor variations. The sample of 202 nursing home residents with dementia proved large enough to study the scalability of the Qualidem subscales. For the very severe dementia group, the H-value improved for three out of four subscales and remained the same for one other; two subscales were not useable. Because these results are based on a larger sample size used (originally n=36; combined n=214), they will provide a better picture of the coefficients found (H-value and r). The Qualidem questionnaire can therefore continued to be used in the field, with only a restricted number of subscales (4/9) useable for the very severely demented persons. Although the scalability and reliability of the subscales improved for the larger sample of persons with very severe dementia, further testing of the remaining items may improve the structure of the Qualidem for this group. The two retained items (see Method Section) ‘Does not want to eat’ and ‘Likes to lie down (in bed)’, that are observable in the very severely demented, may be fitted to supplement the Qualidem instrument. It was suggested in a recent study by Koopmans et al. (2009), also using the Qualidem in a population of very severe demented persons (n=39) in two nursing homes, that the item ‘Likes to lie down’ is not suited in this population, because the nurses were not able to assess the patients on this item (25/39). In our surveys (not including the original Qualidem data; 30/36 missing values), we could not confirm this finding of Koopmans et al. (2009): the caregivers in our study reported that the patient showed this behavior either sometimes or frequently in 67% of the cases (120/178). The reliability coefficient p varied from just acceptable (0.55) to very good (0.90) for the mild to severe dementia subscales, and from 0.57 to 0.82 for very severe dementia. In an overview of dementiaspecific QoL measures by Schölzel-Dorenbos et al. (2007a), similar results were reported for the internal consistency of eight, mainly self-report, instruments, with internal consistency varying from 0.60 to 0.94. The instruments differed, however, not only in their content, but also in their applicability, for example, for which stage of dementia it was used, how the data were collected (by interview or observation), who rated it (patient, proxy or professional), and for which setting it was used (community and/or institution). Although the psychometric properties of the dementia-specific instruments have been extensively reviewed, only recently the use of these instruments has been reviewed in relation to clinical variables of interest, such as cognition, behavior, depression and activity limitation (Banerjee et al., 2009). The findings of the review showed that, among others, there is no convincing evidence that lower cognition is associated with lower health related quality of life (HRQL); and also that currently little is known about how HRQL in dementia changes over longer time periods or which interventions may promote HRQL for people with dementia. Considering that many identified needs of people with dementia in residential care remain unmet (27%), as reported by Hancock et al. (2005), mostly regarding daytime activities, psychological distress and memory, improvements in the daily life and QoL of persons with dementia can be expected from interventions aiming to address these needs. In a first (cluster) randomized controlled trial, aiming to reduce unmet needs of patients with dementia in residential care, the results did not indicate that the intervention reduced unmet needs or improved QoL at follow-up (Orrell et al., 2007). The study, however, may have been underpowered with a low-intensity intervention of short duration. The available valid and reliable QoL measures for dementia should be applied more often in clinical trials to provide information about treatment benefits. With this regard, however, more information on the responsiveness of dementia-specific instruments, including the Qualidem, is indispensable. So far, information on responsiveness indeed is scarce (Schölzel-Dorenbos et al., 2007b).

Limitations of this study

First, because not all items are applicable for persons in the last stage of dementia, some subscale scores of the Qualidem are based on a very limited number of items. This made it difficult to investigate the internal validity of those scales. In the Netherlands, approximately one in seven nursing home residents live up to this stage. More work is needed to assess the QoL of this group. Second, additional measures, such as the DS-DAT, may be used to extend the covered domains (e.g., physical health). Dröes et al. (2006), in an exploratory study, also indicated that there are differences in perspectives on QoL between persons with dementia, their caregivers and theoretical models. Several domains indeed are not captured by the subscales of the Qualidem, for example, religion, spirituality, financial security, and a sense of being useful. Third, even though we had a large sample of 759 persons with mild to severe dementia and 214 persons with very severe dementia, large enough to study the psychometric properties of the Qualidem, this was not sufficient to establish norms for both groups. Van den Brink and Engels-Freeke (1998) indicated that a sample of 1000–1500 respondents would be needed to set norms for such a measuring instrument. Fourth, we do not know how much the subscale scores are influenced by differences in dementia diagnoses (as reported in 2/4 included field studies) for the group very severe dementia. There is a suggestion from a French study using the Alzheimer's disease health-related QoL (ADRQL) that people with Lewy Body dementia may have lower HRQL scores than people with Alzheimer's disease or mixed dementia (Tomas et al., 2006). In general, little information is available on the influence of dementia subtypes on QoL. With the subscales of the Qualidem it is possible to rank residents of nursing homes and homes for the elderly on the respective QoL domains and the instrument can thus well be used for general (longitudinal) care and intervention research in groups of residents with dementia, including the group with very severe dementia. Although the use for clinical application in psychogeriatric care (to make decisions about individual residents) requires reliability coefficients of at least 0.90 (Streiner and Norman, 2002), the scores on subscales may be of assistance to the clinician's decision making process and care evaluation. This will certainly be the case when norms for the subscale score are available. Although for this purpose a larger sample is preferred, the percentile scores presented for each subscale in this study already give some insight in the distribution of scores in people with mild to severe and very severe dementia living in nursing homes and homes for the elderly. Because the Qualidem instrument measures the QoL of persons with dementia, the items and the domains actually steer CNAs away from the behavioral and psychological problems of dementia, which most behavioral observation instruments focus on, and towards other behaviors that reflect quality of life. This helps them to think about how QoL of the residents can be improved and makes them aware that quality of life is more than the absence of behavior problems (Ettema, 2007). The emotion-oriented approaches in dementia care all focus on improving QoL. At present this is the main focus, as no cure for dementia is foreseen in the near future. There is a great need for a dementia-specific QoL measure to evaluate these approaches or other new interventions in dementia care that can be used in different stages of the disease. With the Qualidem this need in the Netherlands has been fulfilled. The Qualidem may be helpful by providing useful information to care professionals, relatives and healthcare inspectors in judging the quality of care of people with dementia in residential settings.

Conflict of interest

None declared.

Acknowledgements

We would like to thank the nursing homes, homes for the elderly, professional caregivers and the residents that participated in this study for their cooperation in the data collection. This research was supported by grants from ActiZ and VUmc. None of the sponsors have played a role in the study. There is no financial relationship between the authors and the sponsors.

REFERENCES

Banerjee S, Samsi K, Petrie CD, et al. 2009. What do we know about quality of life in dementia? A review of the emerging evidence on the predictive and explanatory value of disease specific measures of health related quality of life in people with dementia. *Int J Geriatr Psychiatry* 24: 15–24.

- De Jonghe JF, Wetzels RB, Mulders A. 2009. Validity of the severely impairment battery short version. *J Neurol Neurosurg Psychiatry*. 80: 954–959.
- Dröes R-M, Boelens E, Bos J, et al. 2006. Quality of life in dementia in perspective: an explorative study of variations in opinions among people with dementia and their professional caregivers, and in literature. *Dementia Int J Soc Res Pract* 5: 533–558.
- Ettema TP. 2007. The Construction of a dementia-specific Quality of life instrument rated by professional caregivers: the QUALIDEM. Amsterdam: Free University. PhD Thesis.
- Ettema TP, Dröes R-M, de Lange J, Mellenbergh GJ, Ribbe MW. 2005. A review of quality of life instruments used in dementia. *Qual Life Res* 14: 675–686.
- Ettema TP, Dröes R-M, de Lange J, Mellenbergh GJ, Ribbe MW. 2007a. QUALIDEM: development and evaluation of a Dementia-specific quality of life instrument: validation. *Int J Geriatr Psychiatry* 22: 424–430.
- Ettema TP, Dröes R-M, de Lange J, Mellenbergh GJ, Ribbe MW. 2007b. QUALIDEM: development and evaluation of a Dementia-specific quality of life instrument: scalability, reliability and internal structure. *Int J Geriatr Psychiatry* 22: 549–556.
- Gwyther LP. 1997. The perspective of the person with Alzheimer disease: which outcomes matter in early to middle stages of dementia? *Alzheimer Dis Assoc Disord* 11 (Suppl 6): 18–24.
- Hancock G, Woods RT, Challis D, Orrell M. 2005. The needs of older people with dementia in residential care. *Int J Geriatr Psychiatry* 21: 43–49.
- Hurley AC, Volicer BJ, Hanrahan PA, Houde S, Volicer L. 1992. Assessment of discomfort in advanced Alzheimer patients. *Research in Nursing & Health* 15: 369–77.
- Koopmans RT, van der Molen M, Raats M, Ettema TP. 2009. Neuropsychiatric symptoms and quality of life in patients in the final phase of dementia. *Int J Geriatr Psychiatry* 24: 25–32.
- Lawton MP. 1997. Assessing quality of life in Alzheimer disease research. *Alzheimer Dis Assoc Disord* 11 (Suppl 6): 91–99.
- Logsdon RG, Albert SM. 1999. Assessing quality of life in Alzheimer's disease: conceptual and methodological issues. *J Ment Health Aging* 5(1): 3–6
- McGraw KO, Wong SP. 1996. Forming Inferences about some Intraclass Correlation Coefficients. *Psychological Methods* 1: 30–46.
- Merten H, van Beek APA, Gerritsen DL, et al. 2007. Dagelijkse bezetting van personeel en de kwaliteit van leven van bewoners met psychogeriatrische problemen. NIVEL, Netherlands Institute for Health Services Research. Utrecht, the Netherlands.
- Molenaar IW, Sijtsma K. 2000. User's Manual MSP5 for Windows. Groningen: iecProGAMMA.
- Muskens JB. 1993. Het beloop van dementia: een exploratief longitudinal onderzoek in de huisartsenpraktijk. Nijmegen: Katholieke Universiteit Nijmegen. PhD Thesis.
- Orrell M, Hancock G, Hoe J, et al. 2007. A cluster randomized controlled trial to reduce the unmet needs of people with dementia living in residential care. *Int J Geriatr Psychiatry* 22: 1127–1134.
- Poortvliet MC, van Beek APA, de Boer ME, Gerritsen DL, Wagner C. 2006. Het vaststellen van kwaliteit van leven bij cliënten in de ouderenzorg. NIVEL, Netherlands Institute for Health Services Research. Utrecht, the Netherlands.
- Reisberg B, Ferris SH, De Leon MJ, Crook T. 1982. The Global Deterioration Scale for assessment of primary degenerative dementia. *Am J Psychiatry* 139: 1136–1139.
- Schafer JL, Graham JW. 2002. Missing data: our view of the state of the art. *Psychol Methods* 7(2): 147–177.
- Schölzel-Dorenbos CJM, Ettema TP, Bos J, et al. 2007a. Evaluating the outcome of interventions on quality of life in dementia: selection of the appropriate scale. *Int J Geriatr Psychiatry* 22: 511–519.
- Schölzel-Dorenbos CJM, van der Steen MJMM, Engels LK, Olde Rikkert MGM. 2007b. Assessment of quality of life as outcome in dementia and MCI intervention trials. A systematic review. *Alzheimer Dis Assoc Disord* 21: 172–178.
- Sijtsma K, Molenaar IW. 2002. Introduction to Nonparametric Item Response Theory. Thousand Oaks: Sage Publications Ltd.
- Streiner DL, Norman GR. 2002. *Health Measurement Scales: A Practical Guide to Their Development and Use*, 3rd edn. Oxford University Press Inc.: New York.
- Tomas P, Lalloue F, Preux P-M. 2006. Dementia patients caregivers quality of life: the PIXEL study. *Int J Geriatr Psychiatry* 21: 50–56.
- Van den Brink WP, Engels-Freeke 1998. Beschrijvend testgebruik. In Van den Brink WP, Mellenbergh GJ (eds). *Testleer en testconstructie*. Amsterdam: Boom.
- Van der Steen JT, Ooms ME, Van der Wal G, Ribbe MW. 2002. Measuring discomfort in patients with dementia. Validity of a Dutch version of the Discomfort Scale – dementia of Alzheimer type (DS-DAT). *Tijdschrift voor Gerontologie en Geriatrie* 33: 257–63.

Whitehouse PJ, Orgogozo JM, Becker RE, et al. 1997. Quality-of-life assessment in dementia drug development. Position paper from the International Working Group on Harmonization of Dementia Drug Guidelines. *Alzheimer Dis Assoc Disord* 11 (Suppl 3): 56–60.

[TABLES]

Table 1 Details of the used field surveys

	Qualidem (Ettema <i>et al.</i> , 2007)	NIVEL-a (Poortvliet <i>et al.</i> , 2006)	NIVEL-b (Merten <i>et al.</i> , 2007)	Waalbed-II (De Jonghe <i>et al.</i> , 2009)
Participants ^a , <i>n</i>				
Mild to severe dementia	202	161	189	207
Very severe dementia	36	18	79	81
Residential care, <i>n</i>				
Nursing home	10	5	12	9
Home for the elderly	0	4	0	0
Psycho geriatric wards, <i>n</i>	25	9	12	14
Participants per ward, <i>n</i>	14–32	13–30	13–36	9–31
Assessment severity dementia	GDS ^b	GDS ^c	GDS ^c	GDS ^b
Assessor Qualidem	CNA	CNA	CNA	CNA
Data collection, year	2003/2004	2005	2006/2007	2007/2008
Missing item-response, <i>n</i> (%)	$\left\{ \begin{array}{l} 29/11\ 662^d \\ (0.2) \end{array} \right.$	$\left\{ \begin{array}{l} 19/5957^d (0.3) \\ 3/324^d (0.9) \end{array} \right.$	$\left\{ \begin{array}{l} 11/6993 (0.2) \\ 12/1422 (0.8) \end{array} \right.$	$\left\{ \begin{array}{l} 4/7659 (0.1) \\ 11/1458 (0.8) \end{array} \right.$
Mild to severe dementia				
Very severe dementia				

GDS, Global Deterioration Scale; CNA, certified nursing assistant.

^aNumber of participants in item-analysis (in total, six cases were omitted due to incomplete data).

^bGDS scores <7 are the stages mild to severe dementia; score 7 is stage very severe dementia.

^cBased on observable behavior matching stage 7 of the GDS, the NIVEL studies determined whether participants either belonged to the mild to severe dementia group (GDS < 7) or to the very severe dementia group (GDS = 7). A detailed description of the assessment is available in the manual of the observation questionnaire 'Zicht op eigen leven' (Poortvliet *et al.*, 2006).

For the Qualidem study: 238 respondents times 49 items (number of items originally tested); other studies: number of respondents times 37 items (9 subscales); or number of respondents times 18 items (6 subscales).

Table 2 Characteristics of the participants with mild to severe dementia from the Qualidem, NIVEL (a and b), and Waalbed-II study

	Qualidem (<i>n</i> = 202)	NIVEL-a (<i>n</i> = 161)	NIVEL-b (<i>n</i> = 189)	Waalbed-II (<i>n</i> = 207)
Age, mean ± SD	84.6 (1.78)	84.9 (7.33)	83.2 (7.33)	82.9 (6.76)
Range	55–102	49–98	66–100	59–98
Sex, <i>n</i> (%)				
Male	49 (24)	44 (27)	54 (29)	57 (28)
Female	153 (76)	116 (73)	133 (71)	150 (72)
Marital status, <i>n</i> (%)				
Married	42 (21)	—	—	53 (28)
Widowed	124 (61)	—	—	106 (57)
Never married	28 (14)	—	—	20 (11)
Divorced	8 (4)	—	—	8 (4)
Duration of admission in years, mean ± SD	1.8 (1.64)			1.9 (1.96)
<i>n</i> (%), <1 year		55 (36)	61 (33)	
1–2 years		44 (29)	33 (18)	
>2 years		52 (34)	89 (49)	
Dementia, <i>n</i> (%)				
Alzheimer type (AD)	97 (48)	—	—	75 (36)
Vascular dementia	24 (12)	—	—	24 (12)
Mixed AD and vascular dementia	16 (8)	—	—	33 (16)
Other (e.g., Pick's disease, fronto-temporal dementia, Korsakov dementia)	19 (9)	—	—	10 (5)
Not otherwise specified	46 (23)	—	—	65 (31)

Table 3 Characteristics of the participants with very severe dementia from the Qualidem, NIVEL (a and b), and Waalbed-II study

	Qualidem (n = 36)	NIVEL (a) (n = 18)	NIVEL (b) (n = 79)	Waalbed-II (n = 81)
Age, mean \pm SD	85.4 (5.72)	83.6 (9.0)	83.7 (8.51)	84.0 (8.00)
Range	73–98	62–96	56–98	65–102
Sex, n (%)				
Male	7 (19)	1 (6)	15 (19)	11 (14)
Female	29 (81)	17 (94)	64 (81)	70 (86)
Marital status, n (%)				
Married	12 (33)	—	—	18 (26)
Widowed	17 (47)	—	—	44 (64)
Never married	6 (17)	—	—	4 (6)
Divorced	1 (0)	—	—	3 (4)
Duration of admission in years, mean \pm SD	4.1 (3.25)			3.8 (3.10)
n (%), <1 year		1 (6)	4 (6)	
1–2 years		1 (6)	15 (19)	
>2 years		14 (88)	53 (67)	
Dementia, n (%)				
Alzheimer type (AD)	20 (56)	—	—	37 (46)
Vascular dementia	5 (14)	—	—	6 (7)
Mixed AD and vascular dementia	6 (17)	—	—	6 (7)
Other (e.g., Pick's disease, fronto-temporal dementia, Korsakov dementia)	2 (6)	—	—	2 (3)
Not otherwise specified	3 (8)	—	—	30 (37)

Table 4 Scalability and reliability coefficients from nine subscales of the Qualidem instrument for persons with mild to severe dementia; data from Qualidem, NIVEL (a and b), and Waalbed-II study

Item nr in list	Subscale (items)	GDS < 7			Qualidem (n = 202)			NIVEL-a (n = 161)			NIVEL-b (n = 189)			Waalbed-II (n = 207)			Total (n = 759)		
		Scale-H ^a (item H _i)	ρ	Scale-H (item H _i)	ρ	Scale-H (item H _i)	ρ	Scale-H (item H _i)	ρ	Scale-H (item H _i)	ρ	Scale-H (item H _i)	ρ	Scale-H (item H _i)	ρ	Scale-H (item H _i)	ρ		
A. Care relationship																			
4	Rejects help from nursing assistants	0.48	0.85	0.51	0.85	0.53	0.87	0.50	0.85	0.50	0.85	0.50	0.85	0.50	0.86				
7	Is angry	0.45		0.48		0.50		0.62		0.62		0.62		0.61					
14	Has conflicts with nursing assistants	0.58		0.57		0.59		0.43		0.43		0.39		0.58					
17	Accuses others	0.44		0.48		0.43		0.53		0.41		0.50		0.43					
24	Appreciates help that he or she receives	0.43		0.55		0.59		0.47		0.51		0.48		0.45					
31	Accepts help	0.43		0.47		0.47		0.69		0.64		0.64		0.64					
33	Criticizes the daily routine	0.44		0.63		0.69		0.92		0.91		0.91		0.90					
B. Positive affect																			
1	Is cheerful	0.66	0.90	0.67	0.90	0.69	0.92	0.69	0.90	0.64	0.91	0.66	0.90	0.66	0.90				
5	Has a contented appearance	0.66		0.67		0.75		0.67		0.67		0.68		0.68					
8	Is capable of enjoying things in daily life	0.66		0.61		0.62		0.64		0.64		0.63		0.63					
10	Is in a good mood	0.62		0.73		0.76		0.73		0.73		0.71		0.71					
21	Has a smile around the mouth	0.63		0.62		0.72		0.66		0.66		0.66		0.66					
40	Mood can be influenced in a positive sense	0.52		0.53		0.60		0.51		0.51		0.53		0.53					
C. Negative affect																			
6	Makes an anxious impression	0.41	0.72	0.39	0.72	0.57	0.77	0.49	0.72	0.61	0.78	0.51	0.78	0.55	0.75				
11	Is sad	0.59		0.61		0.63		0.63		0.71		0.64		0.64					
23	Cries	0.52		0.55		0.58		0.58		0.63		0.57		0.57					
D. Restless tense behavior																			
2	Makes restless movements	0.53	0.74	0.51	0.73	0.50	0.75	0.51	0.73	0.46	0.67	0.52	0.73	0.50	0.73				
19	Is restless	0.54		0.53		0.51		0.52		0.52		0.52		0.52					
22	Has tense body language	0.48		0.40		0.39		0.33		0.33		0.40		0.40					
E. Positive self image																			
27	Indicates he or she would like more help	0.42		0.41		0.46		0.46		0.32		0.40		0.40					
35	Indicates not being able to do anything	0.45		0.53		0.52		0.53		0.53		0.50		0.50					
37	Indicates feeling worthless	0.37		0.34		0.49		0.49		0.40		0.41		0.41					
F. Social relations																			
3	Has contact with other residents	0.60	0.82	0.44	0.78	0.45	0.78	0.50	0.78	0.42	0.74	0.53	0.77	0.45	0.77				
12	Responds positively when approached	0.41		0.27		0.41		0.52		0.52		0.53		0.53					
18	Takes care of other residents	0.51		0.55		0.49		0.37		0.37		0.36		0.36					
25	Cuts oneself off from environment	0.42		0.34		0.33		0.50		0.50		0.51		0.51					
29	Is on friendly terms with one or more residents	0.55		0.50		0.54		0.33		0.24		0.33		0.33					
34	Feels at ease in company of others	0.53		0.39		0.42		0.47		0.45		0.51		0.51					
G. Social isolation																			
16	Is rejected by other residents	0.34	0.60	0.25	0.50	0.21	0.50	0.42	0.50	0.29	0.53	0.45	0.55	0.29	0.55				
20	Openly rejects contact with others	0.38		0.30		0.28		0.30		0.30		0.33		0.33					
32	Calls out	0.32		0.23		0.24		0.24		0.31		0.29		0.29					
32	Calls out	0.31		0.22		0.11		0.22		0.26		0.24		0.24					

(Continues)

Bouman, A.I.E., Ettema, T.P., Wetzels, R.B., Beek, A.P.A. van, Lange, J. de, Droës, R.M. Evaluation of Qualidem: a dementia-specific quality of life instrument for persons with dementia in residential settings. Scalability and reliability of subscales in four Dutch field surveys. *International Journal of Geriatric Psychiatry*: 2010



Table 4. (Continued)

Item nr in list	Subscale (items)	Qualidem (n = 202)		NIVEL-a (n = 161)		NIVEL-b (n = 189)		Waalbed-II (n = 207)		Total (n = 759)	
		Scale-H ^a (item H _i)	ρ	Scale-H (item H _i)	ρ	Scale-H (item H _i)	ρ	Scale-H (item H _i)	ρ	Scale-H (item H _i)	ρ
GDS < 7											
H. Feeling at home		0.46	0.76	0.47	0.76	0.47	0.77	0.45	0.75	0.46	0.76
13	Indicates that he or she is bored	0.35		0.41		0.43		0.39		0.40	
28	Indicates feeling locked up	0.55		0.53		0.54		0.51		0.53	
36	Feels at home on the ward	0.41		0.46		0.37		0.37		0.40	
39	Wants to get off the ward	0.51		0.49		0.52		0.48		0.50	
I. Having something to do		0.46	0.63	0.42	0.58	0.33	0.47	0.39	0.53	0.40	0.55
26	Finds things to do without help from others	0.46		0.42		0.33		0.39		0.40	
38	Enjoys helping with chores on the ward	0.46		0.42		0.33		0.39		0.40	

Scale-H: Loewingers coefficient of homogeneity and scalability. Values between 0.30 and 0.40 indicate a weak scale, 0.40–0.50 indicate moderate scalability and 0.50 and above indicate strong scales. Thirty-seven out of 40 items are used for the analysis; it is recommended to retain the following 3/40 items for future research (applicable in all stages): enjoys meals, does not want to eat and likes to lie down (in bed).

All items were translated from Dutch to English by a native English speaker and translated back to Dutch by a second translator. Differences between the original and translated version were discussed before a definite English translation of the item was established.

Table 5 Scalability and reliability coefficients from six (out of nine) subscales of the Qualidem instrument for persons with very severe dementia; data from Qualidem, NIVEL (a and b), and Waalbed-II study

Item nr in list	Subscale (items)	GDS = 7 (underlined item numbers)					Total (n = 214)
		Qualidem (n = 36)	NIVEL-a (n = 18)	NIVEL-b (n = 79)	Waalbed-II (n = 81)		
		Scale-H ^a (item H _i)	Scale-H (item H _i)	Scale-H (item H _i)	Scale-H (item H _i)	Scale-H (item H _i)	ρ
A. Care relationship							
4	Rejects help from nursing assistants	0.41	0.50	0.51	0.63	0.56	0.80
7	Is angry	0.55	0.60	0.67	0.65	0.65	0.65
<u>14</u>	Has conflicts with nursing assistants	0.55	0.59	0.52	0.68	0.61	0.61
<u>17</u>	Accuses others						
24	Appreciates help that he or she receives	0.06	0.26	0.30	0.53	0.38	0.38
31	Accepts help						
<u>33</u>	Criticizes the daily routine	0.33	0.74	0.60	0.50	0.55	0.82
B. Positive affect							
1	Is cheerful	0.21	0.68	0.61	0.52	0.55	0.55
5	Has a contented appearance	0.48	0.73	0.62	0.45	0.54	0.54
8	Is capable of enjoying things in daily life						
<u>10</u>	Is in a good mood	0.38	0.80	0.58	0.55	0.58	0.58
21	Has a smile around the mouth	0.27	0.74	0.60	0.50	0.54	0.54
<u>40</u>	Mood can be influenced in a positive sense	0.40	0.29	0.23	0.41	0.36	0.46
C. Negative affect							
6	Makes an anxious impression	0.40	0.29	0.23	0.41	0.36	0.36
<u>11</u>	Is sad	0.36	0.29	0.23	0.41	0.36	0.36
23	Cries						
D. Restless tense behavior							
2	Makes restless movements	0.30	0.36	0.43	0.48	0.42	0.68
19	Is restless	0.35	0.57	0.49	0.56	0.51	0.51
<u>22</u>	Has tense body language	0.42	0.31	0.49	0.52	0.48	0.48
E. Positive self image							
27	Indicates he or she would like more help	0.12	0.17	0.32	0.34	0.29	0.29
35	Indicates not being able to do anything						
37	Indicates feeling worthless						
F. Social relations							
3	Has contact with other residents	0.10	0.43	0.17	0.40	0.26	0.40
<u>12</u>	Responds positively when approached	0.13	0.27	0.24	0.38	0.28	0.28
<u>18</u>	Takes care of other residents	0.14	0.58	0.19	0.47	0.32	0.32
25	Cuts oneself off from environment						
<u>29</u>	Is on friendly terms with one or more residents	0.04	0.40	0.08	0.36	0.21	0.21
34	Feels at ease in company of others						
G. Social isolation							
16	Is rejected by other residents	0.34	0.21	0.47	0.23	0.34	0.57
<u>20</u>	Openly rejects contact with others	0.32	0.37	0.53	0.24	0.38	0.38
<u>32</u>	Calls out	0.28	0.10	0.46	0.10	0.27	0.27
		0.41	0.15	0.42	0.35	0.37	0.37

(Continues)

Bouman, A.I.E., Ettema, T.P., Wetzels, R.B., Beek, A.P.A. van, Lange, J. de, Droës, R.M. Evaluation of Qualidem: a dementia-specific quality of life instrument for persons with dementia in residential settings. Scalability and reliability of subscales in four Dutch field surveys. *International Journal of Geriatric Psychiatry*: 2010



Table 5. (Continued)

Item nr in list	GDS = 7 (underlined item numbers)	Subscale (items)	Qualidem (n = 36)		NIVEL-a (n = 18)		NIVEL-b (n = 79)		Waalbed-II (n = 81)		Total (n = 214)	
			Scale-H ^a (item H _i)	ρ	Scale-H (item H _i)	ρ	Scale-H (item H _i)	ρ	Scale-H (item H _i)	ρ	Scale-H (item H _i)	ρ
H. Feeling at home												
13	<u>13</u>	Indicates that he or she is bored										
28	<u>28</u>	Indicates feeling locked up										
36	<u>36</u>	Feels at home on the ward										
39	<u>39</u>	Wants to get off the ward										
I. Having something to do												
26	<u>26</u>	Finds things to do without help from others										
38	<u>38</u>	Enjoys helping with chores on the ward										

GDS 7: the underlined items are suited for response when dementia is very severe.

Scale-H: Loevingers coefficient of homogeneity and scalability. Values between 0.30 and 0.40 indicate a weak scale, 0.40–0.50 indicate moderate scalability and 0.50 and above indicate strong scales. Thirty-seven out of 40 items are used for the analysis; it is recommended to retain the following 3/40 items for future research (applicable in all stages): enjoys meals, does not want to eat and likes to lie down (in bed).

All items were translated from Dutch to English by a native English speaker and translated back to Dutch by a second translator. Differences between the original and translated version were discussed before a definite English translation of the item was established.

^aIn bold.

Table 6 Percentiles of subscale scores for persons with mild to severe dementia ($n = 759$) and for persons with very severe dementia ($n = 214$) in residential care

Subscale, 0–100 ^a ; Dementia group	No. of items	Mean (SD)	Percentiles				
			10%	25%	50%	75%	90%
A. Care relationship							
Mild to severe	7	72 (22.5)	43	57	76	90	100
Very severe	3	71 (25.6)	33	56	78	100	
B. Positive affect							
Mild to severe	6	75 (22.0)	44	61	78	94	100
Very severe	4	60 (25.5)	25	42	67	75	92
C. Negative affect							
Mild to severe	3	68 (26.9)	33	44	67	89	100
Very severe ^b	2	69 (26.5)	33	50	67	100	
D. Restless tense behavior							
Mild to severe	3	56 (31.1)	11	33	56	78	100
Very severe	3	48 (31.0)	0	22	44	67	100
E. Positive self image							
Mild to severe	3	78 (25.4)	44	67	89	100	
F. Social relations							
Mild to severe	6	60 (22.1)	33	44	61	78	89
Very severe ^c	3	44 (22.3)	22	33	44	67	100
G. Social isolation							
Mild to severe	3	69 (25.4)	33	56	67	89	100
Very severe	3	72 (25.6)	33	56	78	100	
H. Feeling at home							
Mild to severe	4	79 (24.1)	42	67	83	100	
I. Having something to do							
Mild to severe	2	32 (32.6)	17	50	83	100	

^aHigher scores are more favorable, reflecting a better QoL.

^bUnreliable subscale.

^cNot scalable (and unreliable) subscale.