OPEN

Reliability, validity and relevance of needs assessment instruments for informal dementia caregivers: a psychometric systematic review

Stephanie Kipfer[1,2] • Sandrine Pihet[1]

[1]School of Health Sciences Fribourg, HES-SO University of Applied Sciences and Arts Western Switzerland, [2]Bureau d'Echange des Savoirs pour des praTiques exemplaires de soins (BEST): a Joanna Briggs Institute Centre of Excellence

ABSTRACT

Objective: The objective of this review is to identify reliable and/or valid needs assessment instruments for informal dementia caregivers which are relevant for clinical practice, research and informal caregivers.

Introduction: Informal dementia caregivers report important unmet needs at all stages of the disease. In addition, they often indicate that health care providers insufficiently attend and adapt to their multiple needs. A systematic and patient-centered assessment is needed to address this lack of knowledge and understanding. However, existing quantitative needs assessment questionnaires are limited in terms of psychometric testing. Qualitative measures are time intensive and difficult to conduct on a large scale, with growing economic pressure. Information about the methodological quality and the characteristics of needs assessment instruments are crucial for clinicians and researcher to make informed decisions about the most reliable and valid tool for their specific purpose.

Inclusion criteria: This review considered studies on multidimensional needs assessment instruments for assessing the needs of informal dementia caregivers, living at home. Psychometric studies or other type of studies with sufficient data to evaluate methodological quality were included if they considered at least one outcome for reliability or validity.

Methods: Studies in English, French or German and published until February 2019 were searched in four databases: Embase, MEDLINE, CINAHL and PsycINFO. After screening the titles, abstracts or full-texts for eligibility, the provisional included studies were assessed for methodological quality with a standardized tool for systematic reviews of measurement properties. After data extraction using a standardized tool, the quality of the measurement properties was rated and compared using predefined quality criteria.

Results: Eighteen articles covering 14 different needs assessment instruments were included in the review. Eleven publications focused on the development or the evaluation of an instrument. In addition, a development report, a manual and five studies, not aimed primarily at validation but containing sufficient information about the development or the evaluation of the used instruments, were included. The systematic evaluation of the instruments revealed that half of them had excellent content validity. In contrast, structural validity was rarely examined, and mostly with an insufficient sample size or a questionable analysis. None of the instruments had an optimally tested and good internal consistency. Regarding reliability, test-retest agreement was rarely tested and inter-rater agreement was evaluated using controversial procedures. Comparing the different instruments reviewed, the "Partnering for better health – living with chronic illness: dementia" had the best psychometric evidence, and the "Questionnaire of consultation expectations" was also partly supported, while most other instruments presently had limited psychometric soundness.

Conclusions: Despite the good evidence for some psychometric properties, further developments in the field of needs assessment for informal dementia caregivers are needed, particularly regarding structural and construct validity, as well as test-retest reliability and sensitivity to change. To enhance conceptual clarity, the development of an underlying theoretical model of needs should be prioritized.

Keywords Dementia; informal caregivers; instrument; needs assessment; psychometrics

JBI Database System Rev Implement Rep 2019; 17(0):1–38.

Summary of Findings

Reliability, validity and relevance of needs assessment instruments for informal dementia caregivers				
Bibliography: Kipfer S, Pihet S. Reliability, validity and relevance of needs assessment instruments for informal dementia caregivers: a psychometric systematic review. JBI Database Sys Rev Implement Rep. 2019; 17				
Psychometric property	**Summary or pooled results**	**Overall rating**	**Factors determining the quality of evidence**	**GRADE of the quality of evidence**
Content validity	Seven studies with excellent content validity, but the identified domains were not consistent.	+/- inconsistent	- 1 inconsistency	Moderate
Internal consistency	One study had an adequate dimensionality analysis but low alphas, all others had either their dimensionality evaluated with small samples, or no proper evaluation of their dimensionality – although some had high alphas.	- insufficient	- 1 risk of bias	Moderate
Structural validity	Four studies evaluated the factor structure of their instrument, one with an excellent factor analysis meeting all criteria, two with an adequate factor analysis but samples were of too limited sizes, and a last one with a factor analysis inadequately performed separately for each dimension. Identified factor structures varied from five to eight dimensions.	- insufficient	- 1 inconsistency	Moderate
Reliability – test-retest	Two studies used a satisfactory procedure to evaluate test-retest agreement, one obtained a good correlation, and the other had two correlations above .70 and three below .70. Results for measurement error were rated as good for one study.	+/- inconsistent		High
Reliability – inter-rater	Inter-rater agreement was evaluated with questionable procedures.	? indeterminate	- 2 risk of bias	Low

Construct validity	In four studies, precise *a priori* hypotheses were formulated and at least 75% of the results were in accordance with them. These instruments seemed to contain items overlapping with the variables tested for association (e.g. burden, depression). Two other studies had less than 75% of hypotheses confirmed. The associations tested included very diverse outcomes and were rarely based on a theoretical model.	+/- inconsistent	- 1 risk of bias - 1 inconsistency	Moderate

GRADE Working group grades of evidence:

High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

The Summary of Findings follows an approach proposed by COSMIN on how to use GRADE for reviews of psychometric properties (Mokkink LB, Prinsen C, Patrick DL, Alonso J, Bouter LM, de Vet H, Terwee CB. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures [PROMs]. Amsterdam: University Medical Center, Department of Epidemiology and Biostatistics, Amsterdam Public Health research institute; 2018)

Introduction

Dementia is characterized by a progressive decline of cognitive and social functions. This limits the autonomy of those affected by dementia, and makes it difficult for them to cope with daily life. They become increasingly dependent on the care of others, particularly informal caregivers. Informal caregivers are individuals who regularly provide unpaid care, assistance and/or supervision to a close person with reduced autonomy.[1] Studies show that informal dementia caregivers provide care of a higher intensity (more hours per day) and longer duration (over more years), compared to caregivers of people without dementia. The caregiving time increases even more when the person with dementia shows behavioral symptoms.[2] In a survey of Alzheimer Europe, almost half of informal dementia caregivers spent more than 10 hours per day providing care.[3] This is comparable to data from the United States (US), where informal dementia caregivers spent on average nine hours per day caregiving,[4] and 31.1% of informal dementia caregivers provide care for two to three years, 18.5% for four to five years and 38.4% for six or more years.[5] For 2015, the estimated economic value of unpaid care provided by informal dementia caregivers in the US was US$221.3 billion.[5] In Switzerland, an example of a Western European country, informal caregivers contributed 80 million of unpaid hours in 2016, with a substantial increase compared to the 64 million of hours invested in 2012 and the 52 million in 2010.[6,7] The contribution of informal caregivers is expected to further increase due to the rising care needs of an aging population and the growing prevalence of (multiple) chronic conditions, in particular dementia.[6] In addition, the number of available formal carers is not expected to increase in line with this increase.[8] Informal caregivers thus play a key role, not only for the people with dementia but also for society in the sustainability of the health care system.[9-11] Recognizing the valuable contribution of informal

caregivers and providing them adequate support is therefore a core public health issue.

Caring for a person with dementia is a challenging experience and the burden of informal dementia caregivers is higher compared to informal caregivers of persons with other chronic conditions.[12,13] Straining continuous care, an unpredictable course, and neuropsychiatric symptoms of the person with dementia can cause high levels of stress which often leads to physical, psychological, emotional, social and financial problems.[2,13-15] In addition, family caregivers often have no experience in performing care, feel unprepared, and are lacking the required knowledge and support from health care providers to deliver appropriate care.[5,16] Informal caregivers report feelings of tiredness, stress, helplessness, and loneliness, and show high prevalence of depression and anxiety.[17] Due to the nature of dementia, informal caregivers also struggle with feelings of guilt, ambivalence, grief and loss. Identified physical problems can create an increased risk for vascular disease, impaired wound healing, decreased immunity, and reduced likelihood to engage in preventive health behavior.[17,18] Poor physical and psychological health conditions not only impair the quality of life of informal dementia caregivers but also affect their ability to provide care to the person with dementia and to sustain their own social support network, which leads to social isolation.[2,18-20] Burden and health deterioration of informal dementia caregivers are core predictors of early institutionalization and mistreatment of their care recipient.[18,21]

Due to the challenges of caregiving and the associated burden, informal dementia caregivers report important unmet needs at all stages of the disease. Their needs cover very diverse areas, such as information about the illness and support resources; support for their own emotional concerns; support on how to communicate with the care recipient, the family or the service providers; practical support in daily care and respite, and financial support.[2,22-26]

Informal dementia caregivers often report that health care providers insufficiently attend and adapt to their multiple needs, and complain about care fragmentation and poor coordination, which ultimately increases their stress and underutilization of support services despite their needs.[27-29] Underutilization of health care and other support resources contributes to the exhaustion of the informal dementia caregivers and precipitate institutionalization of

their care recipient, and thereby increases health care costs.[21,30,31] Informal caregivers do not always express spontaneously and directly how their needs can be met.[32] Therefore evaluating their needs in a systematic manner is crucial to supporting them in fulfilling their needs in a person-centered way in order to promote the quality of life of the caregiver and the affected person, as well as to maintain the caring situation at home.

Most studies on the needs of informal dementia caregivers have used qualitative study designs. Existing quantitative questionnaires have limitations: very few items for caregivers,[17,33] poor validation,[22,23,34,35] or lack of empirical evidence about needs dimensions (factor structure). This limits their use in both research and clinical practice. In addition, many of the assessment instruments, particularly semi-structured interviews, are time-intensive (e.g. assessment alone takes on average two hours;[26] or 90 minutes[23]). Furthermore, most of the collected information in interviews is qualitative, as such it is usually extensive, and more time is needed to prepare the information to make it available for the caregiver or other service providers (e.g. transcriptions). In view of the growing economic pressure on the social system, and of the rising support needs associated with population aging, such resources are impossible to secure on a large scale.

A search in MEDLINE, CINAHL, the *JBI Database of Systematic Reviews and Implementation Reports* and the Cochrane Database of Systematic Reviews was performed in January 2017, and again in January and July 2018, to identify completed and in-progress systematic reviews on needs assessment instruments for informal dementia caregivers. Three systematic reviews examining dementia needs assessment instruments were identified.[36,37,38] Two reviews focused on documenting the diverse instruments available, focusing on their content and methodological approach to measure the needs, with no specific interest in their psychometric properties: Novais *et al.*[36] included all types of studies which conducted a needs assessment, and Bangerter *et al.*[37] concentrated on quantitative findings. The third review (Mansfield *et al.*[38]) critically examined the psychometric properties of a very limited number of instruments (N = 4) due to highly restrictive inclusion criteria; only peer-reviewed studies in English published until August 2015 were included. In addition, they targeted instruments where all items

focused on the caregiver, and where the caregivers were directly asked about their needs.

Information about the methodological quality of measurement instruments are crucial for clinicians and researcher to make informed decisions about the best tool for their specific purpose. Therefore, the current review aimed to expand on the previous three reviews by focusing specifically on (at least partly) validated instruments and documenting in detail their psychometric properties. For this purpose, we followed the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) guidelines for systematic reviews of measurement properties which are recommended for psychometric reviews.[39] In addition, the current review expanded on that by Mansfield et al.[38] by including: i) instruments with diverse application methods (i.e. also professionally-assessed) or with items for both caregivers and persons with dementia, ii) studies published until February 2019 and in English, French and German, and iii) the CINHAL database as a relevant source for caregiver literature. The current review also includes recommendations for conducting further psychometric validations in the field of needs assessment among informal dementia caregivers. Therefore this review provides comprehensive and systematic information to guide further development of well-validated needs assessment instruments for use in research and clinical settings.

This review was conducted according to an a priori published protocol.[40]

Review questions/objectives

The objective of this psychometric review was to critically appraise, compare and summarize the measurement properties of needs assessment instruments for informal dementia caregivers. More specifically, the review questions were as follows:

- What are the measurement properties of needs assessment instruments for informal dementia caregivers (primary outcome)?
- What is the relevance of these instruments for clinical practice, research and informal caregivers according to their purpose, application method, administration burden, number of items and domain structure (secondary outcome)?

The objectives section has been revised, compared to the a priori protocol, to provide more clarity without changing the overall objectives of the review.

Inclusion criteria

The inclusion criteria were developed following COSMIN guidelines and JBI guidance. The COSMIN guidelines for systematic reviews of measurement properties recommend the following inclusion criteria: i) the instrument should aim to measure the construct of interest (= types of intervention(s)/phenomena of interest), ii) the study sample should comprise the target population of interest (= types of participants), iii) the study should concern the type of measurement instrument of interest (= self-reported or professionally interviewed), iv) the aim of the study should be the development of a measurement instrument or the evaluation of one or more of its measurement properties (= types of studies).[39]

Participants

This review considered studies that included informal caregivers of persons with dementia living at home as the study sample or as a part of it. Where possible, only the corresponding data was included in the review. Informal caregivers were defined as individuals who regularly provide unpaid care, assistance and/or supervision to a close person with reduced autonomy, in this context with dementia.[1] In regards to the person with dementia, there was no restriction regarding the type of dementia.

Instruments/construct

This review considered studies that report on needs assessment instruments for informal dementia caregivers. Needs was defined as "a condition that is important to the subject and that is not being satisfied in the subject's present environment".[41(p.772)] The application method of the instruments was either self-reported or professionally interviewed. Further inclusion criteria were added while identifying and screening the literature to complement those of the a priori protocol: Instruments needed to be multidimensional, for example, include more dimensions than only information needs. Instruments with items for informal caregivers and people with dementia were included if they contained at least two dimensions for the caregivers. Measuring the needs of informal dementia caregivers had to be an explicit objective of the instrument or of specific dimensions.

Outcomes

This review considered studies that included the following outcomes for psychometric properties:

- Reliability (test-retest reliability, inter-rater reliability, internal consistency)
- Validity (content validity, construct validity, structural validity, sensitivity to change)

In order to characterize the instruments, we additionally documented the following instrument characteristics:

- Purpose (original intended use)
- Application method (self-reported, professionally interviewed)
- Administration burden (training for clinicians, time for completion)
- Number of items and domain structure.

As not every article provided data for all of the psychometric outcomes, articles which reported at least one outcome regarding reliability or validity were considered. Criterion-related validity was not considered in the results as there was no reasonable gold-standard available for the included instruments (in accordance with COSMIN guidelines[39]). No data was found for the outcome sensivity to change.

Types of studies

This psychometric review considered psychometric studies, namely, instrument development or instrument evaluation studies. Other types of studies (in which needs assessment instruments were used) were included to identify eligible instruments and their responsible authors. If no published or unpublished psychometric study was available, other types of studies (example.g. a survey) were only included if they provided sufficient information to evaluate the methodological validity of at least one psychometric property according to the COSMIN checklist.[42]

Methods

This review was conducted according to an *a priori* published protocol[40] and has been registered with PROSPERO (CRD42018090611).

Search strategy

The search strategy aimed to find both published and unpublished studies. A three-step search strategy was utilized in this review. An initial limited search of MEDLINE and CINAHL was undertaken followed by analysis of the text words contained in the title and abstract, and of the index terms used to describe the articles. A second search was undertaken in February 2019 across all included databases. The search strategy considered all identified keywords and index terms as well as search blocks for dementia and patient reported outcome measures provided by the study group of the Biomedical Information of the Dutch Library Association (KNVI)[43] and a search block combined with a filter for measurement properties, as suggested by and available from the COSMIN website.[44] Thirdly, the reference lists of all selected full texts were searched for additional studies. A complementary search was performed in the included databases and gray literature using the names of the needs assessment instruments identified in the three foregoing steps, and authors were contacted to obtain possible additional gray literature relating to their instrument. Studies published in English, German and French were considered for inclusion in this review. There was no limitation regarding the publication time.

The databases searched included: Embase, MEDLINE, CINAHL and PsycINFO. In contrast to the *a priori* protocol, Embase was searched as recommended by COSMIN, and CINAHL was searched instead of OVID Nursing.[39]

The search for unpublished studies included: Google Scholar, ProQuest Dissertations and Theses, ResearchGate (contact with relevant researchers), homepages with information about needs assessment/outcome tools (e.g. www.decideproject.co.uk) and homepages of dementia or caregiver associations or organizations (e.g. www.alz.org, www.alzheimer-europe.org, www.caregiver.org, www.rosalynncarter.org).

Seven relevant researchers identified during the literature research were contacted by email to obtain unpublished psychometric studies or testing of the instrument, or to receive further or lacking data. Two of them provided the instrument itself to complete the data extraction. In two cases additional data was delivered to evaluate the methodological quality. One unpublished publication was found by contacting the responsible researcher. Three authors were not able to send additional information.

The full search strategy is provided in Appendix I.

Study selection

Following the search, all identified citations were collated and uploaded into EndNote X7.8/2016 (Clarivate Analytics, PA, USA) and duplicates removed. Titles and abstracts were then screened by two independent reviewers for assessment against the inclusion and exclusion criteria for the review. Potentially relevant publications were retrieved in full and their citation details imported into the standardized data extraction tool developed for this review. The full text of selected citations were assessed in detail against the inclusion and exclusion criteria by two independent reviewers. Reasons for exclusion of full text studies that did not meet the inclusion criteria are reported in Appendix II. Any disagreements that arose between the reviewers at each stage of the study selection process were either resolved through discussion, or with a third reviewer.

Assessment of methodological quality

Publications selected for retrieval were assessed by two independent reviewers for methodological validity prior to inclusion in the review using the COSMIN checklist.[42,45,46] The COSMIN checklist is a standardized tool which is recommended for use in systematic reviews of measurement properties.[42] This tool fulfils the specific requirements of a psychometric review and has already been successfully used in another JBI review protocol.[47] The original checklist consists of 12 boxes with five to 18 items per box, and is a modular tool. We therefore used only the seven boxes evaluating relevant psychometric properties for our review, namely, internal consistency, reliability, measurement error, content validity, structural validity, hypotheses testing and responsiveness. The box for criterion-related validity was not included as there is no reasonable gold-standard available. The three boxes for additional methodological standards (item response theory models, interpretability and cross-cultural validity) were not used as they focus on more advanced properties. The last box on generalizability was used for data extraction of study characteristics. Each item of the checklist is assessed on a four-point response scale: excellent, good, fair and poor. However, some items have only two or three response options (e.g. only excellent, fair or poor). The methodological quality scores per box were obtained by taking the lowest rating across all item in the box

("worst score count"). The lowest score of any box presented the overall score of the reviewed study. Studies with poor scores in all boxes would have been excluded from the review as this would indicate an inadequate methodological quality. In this review none of the included studies had to be excluded due to poor scores in all boxes. The checklist used for assessing the methodological quality can be found in the *a priori* published protocol.[40]

All disagreements that arose between the reviewers were resolved through discussion, or with a third reviewer.

Data extraction

Data were extracted from papers included in the review using the standardized data extraction tool developed for this review.[40] This tool was inspired by the standardized data extraction tools in Joanna Briggs Institute System for the Unified Management, Assessment and Review of Information (JBI SUMARI), by selecting the relevant parts and adapting them to the specificity of a psychometric review. JBI SUMARI is a web application designed to support researchers and practitioners in the entire process of conducting a systematic review. The data extracted included specific details about: i) study characteristics, ii) instrument characteristics and iii) outcomes of significance for the review question and specific objectives. Although the comprehensivness and accuracy of the provided data varied across studies, overall we could collect data on:

i)   Study characteristics: citation details, aim of the study, study design and method, setting, population characteristics, definition of informal caregivers and needs.

ii) Instrument characteristics: name of the instrument, purpose, target population, application method, administrative burden, number of items and domain structure, range of scores, response options/format.

iii) Psychometric outcomes: reliability (test-retest reliability, inter-rater reliability, internal consistency), validity (content validity, construct validity, structural validity).

Data were extracted and double checked independently by two reviewers (first and second author). To minimize errors, during the data extraction process, the data extraction form was first pilot tested and a standardized form was used. Disagreement between

the reviewers were resolved by discussion, or with a third reviewer.

Data synthesis

The main aim of the data synthesis was to compare outcomes to provide recommendations on the most suitable instrument for research, clinical use and informal caregivers. The findings on instrument characteristics, reliability and validity were compared and described in narrative form, including tables to aid data presentation. The domain structure of the instruments and a summary of their content is presented in a table and in a narrative form (see Appendix III and Table 1).

The quality criteria from Terwee *et al.* were used to judge the psychometric outcomes of the different instruments, namely, their content validity, internal consistency, construct validity, test-retest reliability (agreement), and inter-rater reliability (reliability).[48] The quality of the instruments was assessed as positive, indeterminate, or negative, with a fourth category for "no information available and doubtful design or method".[40] The results of this appraisal are presented in a narrative form and in Appendix IV.

Results

Study inclusion

A total of 4909 records were identified through the systematic search in the four databases. Searches in gray literature and requests to relevant researchers for unpublished literature or publications about specific instruments, identified *a priori*, revealed seven additional publications. After removing duplicates, 3468 records remained. Another 3404 records were excluded for not meeting the inclusion criteria after screening their titles or abstracts, resulting in 64 full texts assessed for eligibility. After reading the full texts, 19 articles met the inclusion criteria and were assessed for quality. One publication was excluded due to insufficient data to evaluate the methodological quality. Eighteen articles[49-66] covering 14 different needs assessment instruments were finally included in the review, with, in four cases, two different publications describing the same instrument, as they provided complementary information.[50-55,62,63] A flow chart of the study selection is presented in Figure 1. The excluded full-texts and the reasons for exclusion according to the inclusion criteria or the critical appraisal are listed in

Table 1: Summary of topics

Need for information and education
- Information about dementia and treatment
- Nursing skills/care tasks/dementia specific caring skills
- Information resources
- Formal help/services in the region/community resources
- Characteristics, access and availability of services
- Organizing care/type of provider

Needs related to emotional support
- Respite support
- Family time/shared activities
- Mental health counseling/psychiatric care
- Informal network/support from family and friends
- Support from society
- Relationship to person with dementia
- Counselling negative emotions
- Being a caregiver/assume caregiver role
- Sleep

Need for other accessible and appropriate services
- Medical care for caregiver
- General assistance or household chores
- Financial and legal support

Appendix II. Table 2 presents an overview of the included instruments, their acronyms and the authors of the included publications.

Characteristics of included studies

Eleven of the 18 included publications focused on the development or the evaluation of an instrument.[49,51-55,58,59,61,65,64] In addition, a development report and a manual were included to assess the content validity of an instrument or to provide additional results for the psychometric testing.[63,50] Five other studies, not primarily aiming at validation but containing sufficient information about the development or the evaluation of the used instruments, were also integrated in the review.[56,57,60,62,66] The publication dates ranged from 1996 to 2019.

Nine of the development or evaluation samples included only caregivers of persons with dementia
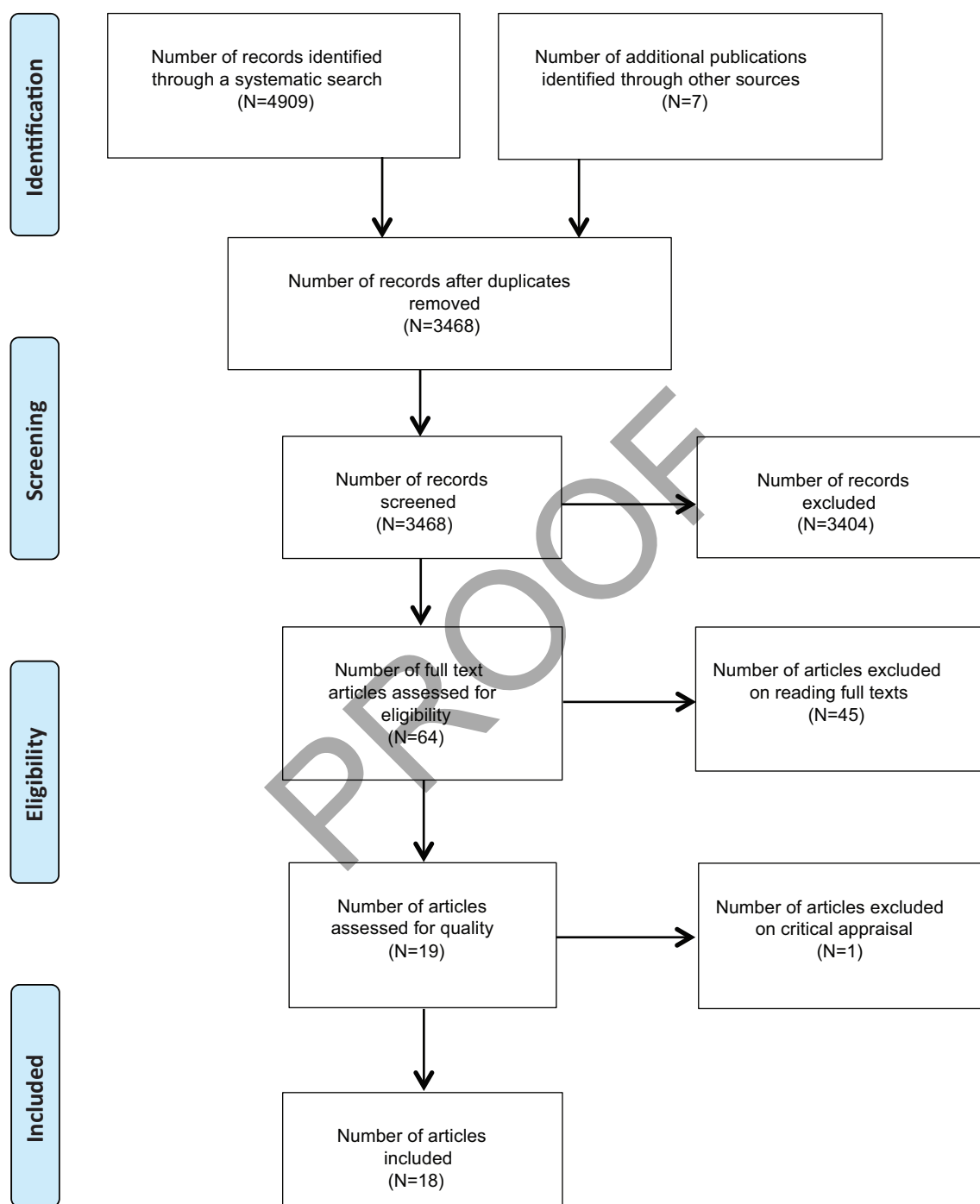
Figure 1: PRISMA flow diagram of search and study selection process[67]

living in the community.[49,54-60,64] The remaining nine samples were mixed, including caregivers of persons with dementia residing in the community or living in institutions.[50-53,61-63,65,66] Study samples differed in the relationship status, but caregivers were mostly spouses or offsprings of the persons with dementia. In all studies, the majority of caregivers were female. The mean age of the caregivers

Table 2: Overview of instruments and authors

Acronym	Instrument name	Authors of included studies
CADI	Carers Assessment of Difficulties Index	Charlesworth et al. (2007)[49]
CARENAP	The Care Needs Assessment Pack for Dementia	McWalter et al. (1996; 1998)[50,51]
CNA-D	The Carers' Needs Assessment for Dementia	Wancata et al. (2005)[52] Kaiser et al. (2005)[53]
CNCD	Caregivers' Needs Checklist for Dementia	Vaingankar et al. (2013; 2018)[54,55]
JHDCNA	The Johns Hopkins Dementia Care Needs Assessment	Hughes et al. (2014)[56]
NAS	Needs Assessment Survey	Wackerbarth et al. (2002)[57]
PBH-LCI:D	Partnering for Better Health – Living with Chronic Illness: Dementia	Sadak et al. (2015)[58]
RAM	Risk Appraisal Measure	Czaja et al. (2009)[59]
QCNE	Questionnaire of Carers Needs Evaluation	Dimakopoulou et al. (2015)[60]
EAC	Questionnaire onsultation expectations [Echelle d'attentes de consultation (EAC)]	Laprise et al. (2001)[61]
QNP	Questionnaire National Dementia Programme Survey Needs and problems of informal caregivers of persons with dementia	Peeters et al. (2010)[62] Van der Poel and van Beek (2006)[63]
SIDECAR	Scales measuring the Impact of DEmentia on CARers	Oyebode et al. (2019)[64]
Tayside	Tayside Profile for Dementia Planning	Gordon et al. (1997)[65]
UNM	Unmet Need Measure	Gaugler et al. (2004)[66]

ranged from 51 to 68 years (with a standard deviation of 10 to 14.5 years), and that of people with dementia ranged from 76 to 85 years (with a standard deviation of 5.9 to 9.8 years). An overview of the characteristics of the different samples is presented in Appendix V.

Description of instruments

Five of the instruments were developed or tested in the US,[56-59,66] four in the United Kingdom (UK),[49,50,51,64,65] and one each in Austria,[52,53] Singapore,[54,55] Greece,[60] Canada[61] and the Netherlands.[62,63] Eleven instruments contained only items for caregivers (CADI, RAM, QCNE, UNM, EAC, QNP, PBH-LCI:D, CNCD, NAS, CNA-D, SIDECAR) while three also included items for persons with dementia (CARENAP, JHDCNA, Tayside). Three instruments were recommended for use in both clinical and research settings (CADI,

PBH-LCI:D, RAM) while one was recommended only for research purposes (CNA-D), and four others only for clinical assessments (CARENAP, CNCD, EAC, Tayside). For the last six instruments, the intended context of use was not specified (JHDCNA, NAS, QCNE, QNP, SIDECAR, UNM). With the application method, seven instruments were self-administered (CNCD, NAS, PBH-LCI:D, QCNE, EAC, SIDECAR, UNM), three were used in professional interviews (CARENAP, CNA-D, JHDCNA), and two could be either self-administered or professionally interviewed (QNP, Tayside). For two instruments, the application method was not clearly stated, but the descriptions suggested for both a self-reported application method (CADI, RAM). For seven instruments, the administration time was described, ranging from five to 50 minutes (CADI, CARENAP, CNA-D, NAS, RAM, QCNE, QNP). Information about the required training for

clinicians was mentioned for two instruments. For one of them, no specific abilities or prior knowledge were necessary (QNP), whereas the other instrument could only be administered by professionals with experience in assessments and interviewing (CARENAP). Response options for all instruments were either nominally or ordinally scaled. For nine instruments, a total (CADI, CNA-D, CNCD, JHDCNA, PBH-LCI:D, RAM, EAC, UNM) or a mean score (NAS) could be obtained, with higher scores indicating higher unmet needs in most of the instruments. Four instruments did not use a scoring system (CARENAP, QCNE, QNP, Tayside). No concrete information was provided for SIDECAR on this aspect. The instruments differed regarding the domain structure and number of items. The number of items for caregivers ranged from 12 to 70 items. Appendix III provides an overview of the domain structure, number of items and response options of the different instruments. Further detailed information about the characteristics of the instruments is presented in Appendix VI.

Methodological quality

For each study, we evaluated the methodological quality of their assessment of six different psychometric properties, namely, content validity, structural validity (i.e. the factor structure of the instrument), internal consistency, reliability (including both test-retest and inter-rater), measurement error and construct validity. Criterion validity was not considered as there is currently no gold standard, and responsiveness (i.e. sensitivity to change) was excluded as no study assessed it. Table 3 provides an overview of the quality (excellent, good, fair or poor) of the assessment for each study and each specific standardized question (Q1 and following) regarding the psychometric properties evaluated in this study. For CNA-D, two studies testing construct validity with different variables were available and therefore evaluated separately, thus generating in total 15 studies for methodological quality. In the other three cases where two publications described the same instrument, as they contained complementary information, we treated them as one study when assessing methodological quality.

Content validity was documented in 13 of the 15 studies. Seven of these 13 studies had excellent ratings for each of the five specific criteria considered (QCNE, CARENAP, QNP, PBH-LCI:D, CNCD,

CNA-D, SIDECAR), and another study had good or excellent ratings (NAS). The five other studies all had at least one poor rating. Three of them failed to include informal caregivers in the item development process (i.e. Q2: RAM, UNM, EAC). Three did not assess if all items together comprehensively reflected the construct to be measured (i.e. Q4: CADI, Tayside, EAC). One did not assess if all items referred to relevant aspects of the construct to be measured (i.e. Q1: CADI). In summary, various examples of excellent methodological quality were available regarding content validity.

Structural validity is relevant for instruments aiming at measuring different domains of needs, which was clear for all studies. However, only four studies evaluated the factor structure of their instrument, one with a factor analysis meeting all the criteria (CADI), two with adequate factor analysis but samples of too limited size (PBH-LCI:D, CNCD), and a last one with a factor analysis inadequately performed separately for each dimension (QCNE). In summary, sample size issues and inappropriate statistical analysis limited the strength of the limited evidence regarding structural validity.

Internal consistency was reported in 10 of the 15 studies. Information on the percentage of missing data (Q2) and how they were handled (Q3) was provided in only two of these 10 studies (RAM, EAC). The sample size used to assess internal consistency was optimal (Q4, N = 100 or more) for seven studies (CADI, RAM, QCNE, UNM, QNP, PBH-LCI:D, CNCD), good for the EAC (50 < N < 99), moderate for the CNA-D (30 < N < 49), and poor for CARENAP (N < 30). As presented above, four studies tested the unidimensionality of their subscales before computing the Cronbach alphas (Q5), in two cases with an appropriate sample size (Q6: CADI, QCNE) and in two with far too small samples (PBH-LCI:D, CNCD), while six studies did not assess unidimensionality. Seven studies computed the internal consistency statistic for each subscale separately (Q7: CADI, QCNE, UNM, EAC, PBH-LCI:D, CNCD, CNA-D), while three computed a single alpha for all items irrespective of possible dimensions (RAM, CARENAP, QNP). Eight studies calculated an appropriate statistic (Q9 and/or Q10: CADI, RAM, QCNE, UNM, EAC, PBH-LCI:D, CNCD, CNA-D), while two provided only item-total correlations (CARENAP, QNP). In summary, most studies met a

Table 3: Methodological quality of included studies

Citation	Box	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14
JHDCNA Hughes *et al.* (2014)[56]	Construct validity	exc	good	exc	good	exc	good	good	exc	fair	good	-	-	-	-
CADI Charlesworth *et al.* (2007)[49]	Content validity	poor	exc	fair	poor	exc	-	-	-	-	-	-	-	-	-
	Structural validity	yes	exc	exc	exc	exc	exc	na	-	-	-	-	-	-	-
	Internal consistency	yes	exc	exc	exc	exc	exc	exc	exc	exc	na	na	-	-	-
RAM Cjaza *et al.* (2009)[59]	Content validity	exc	poor	exc	exc	exc	-	-	-	-	-	-	-	-	-
	Internal consistency	yes	good	fair	exc	poor	na	poor	exc	exc	exc	na	-	-	-
	Construct validity	good	fair	exc	fair	exc	good	good	fair	exc	good	-	-	-	-
QCNE Dimakopoulou *et al.* (2015)[60]	Content validity	exc	exc	exc	exc	exc	-	-	-	-	-	-	-	-	-
	Structural validity	yes	good	fair	exc	poor	poor	na	-	-	-	-	-	-	-
	Internal consistency	yes	good	fair	exc	exc	exc	exc	poor	na	exc	na	-	-	-
UNM Gaugler *et al.* (2004)[66]	Content validity	exc	poor	exc	exc	fair	-	-	-	-	-	-	-	-	-
	Internal consistency	yes	good	fair	exc	poor	na	exc	fair	na	exc	exc	-	-	-
	Construct validity	good	fair	exc	fair	exc	good	fair	fair	fair	good	-	-	-	-
Tayside Gordon *et al.* (1997)[65]	Content validity	exc	exc	exc	poor	fair	-	-	-	-	-	-	-	-	-
	Reliability	good	fair	good	exc	fair	fair	poor	poor	poor	poor	na	exc	na	na
EAC Laprise *et al.* (2001)[61]	Content validity	exc	poor	exc	poor	fair	-	-	-	-	-	-	-	-	-
	Internal consistency	yes	exc	exc	good	poor	na	exc	exc	exc	na	na	-	-	-
	Reliability	good	exc	fair	exc	good	exc	fair	fair	good	fair	fair	na	na	na
	Construct validity	good	fair	good	good	exc	exc	exc	fair	fair	exc	-	-	-	-
CARENAP McWalter (1996, 1998)[50,51]	Content validity	exc	exc	exc	exc	exc	-	-	-	-	-	-	-	-	-
	Internal consistency	yes	good	fair	poor	poor	na	poor	poor	na	good	na	-	-	-
	Reliability	exc	fair	poor	exc	fair	exc	exc	exc	poor	fair	na	exc	na	na
QNP – Peeters *et al.* (2010)[62] / Van der Poel and van Beek (2006)[63]	Content validity	exc	exc	exc	exc	exc	-	-	-	-	-	-	-	-	-
	Internal consistency	yes	good	fair	exc	poor	na	poor	poor	good	na	na	-	-	-
PBH-LCI:D Sadak *et al.* (2015)[58]	Content validity	exc	exc	exc	exc	exc	-	-	-	-	-	-	-	-	-
	Structural validity	yes	good	fair	poor	fair	exc	exc	-	-	-	-	-	-	-
	Internal consistency	yes	good	fair	exc	exc	poor	exc	fair	exc	na	poor	-	-	-
	Reliability	exc	exc	good	exc	good	exc	good	exc	good	exc	fair	na	na	na
	Measuremt error	good	fair	good	exc	good	exc	good	exc	good	fair	exc	-	-	-
	Construct validity	exc	exc	good	exc	exc	good	exc	fair	exc	fair	-	-	-	-
CNCD Vaingankar *et al.* (2013, 2017)[54,55]	Content validity	exc	exc	exc	exc	exc	-	-	-	-	-	-	-	-	-
	Structural validity	yes	good	fair	poor	fair	exc	na	-	-	-	-	-	-	-
	Internal consistency	yes	good	fair	exc	exc	poor	exc	fair	exc	na	na	-	-	-
	Construct validity	good	fair	exc	exc	exc	good	exc	fair	exc	exc	-	-	-	-
NAS - Wackerbarth *et al.* (2002)[57]	Content validity	exc	good	exc	exc	exc	-	-	-	-	-	-	-	-	-
CNA-D Wanacata *et al.* (2005)[52]	Content validity	exc	exc	exc	exc	exc	-	-	-	-	-	-	-	-	-
	Internal consistency	yes	good	fair	fair	poor	na	exc	fair	na	exc	na	-	-	-
	Reliability	good	fair	fair	exc	exc	exc	good	exc	good	exc	na	exc	na	na
	Construct validity	good	good	fair	good	exc	good	exc	good	exc	exc	-	-	-	-

Table 3. (*Continued*)

Citation	Box	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14
CNA-D Kaiser *et al.* (2005)[53]	Construct validity	good	good	fair	exc	exc	good	exc	fair	exc	exc	-	-	-	-
SIDECAR Oyebode *et al.* (2019)[64]	Content validity	exc	exc	exc	exc	exc	-	-	-	-	-	-	-	-	-

CADI: Carers Assessment of Difficulties Index; CARENAP: The Care Needs Assessment Pack for Dementia; CNA-D: The Carers' Needs Assessment for Dementia; CNCD: Caregivers' Needs Checklist for Dementia; EAC: Questionnaire consultation expectations; exc: excellent; JHDCNA: The Johns Hopkins Dementia Care Needs Assessment; na: not applicable; NAS: Needs assessment survey; PBH-LCI:D: Partnering for Better Health e Living with Chronic Illness: Dementia; QCNE: Questionnaire of carers needs evaluation; QNP: Questionnaire National Dementia Programme Survey Needs and problems of informal caregivers of persons with dementia; RAM: Risk Appraisal Measure; Reliability: test-retest and/or inter-rater; SIDECAR: Scales measuring the Impact of DEmentia on CARers; Tayside: Tayside Profile for Dementia Planning; UNM: Unmet need measure.

Content validity: Q1. Was there an assessment of whether all items refer to relevant aspects of the construct to be measured? Q2. Was there an assessment of whether all items are relevant for the study population? Q3. Was there an assessment of whether all items are relevant for the purpose of the measurement instrument? Q4. Was there an assessment of whether all items together comprehensively reflect the construct to be measured? Q5. Were there any important flaws in the design or methods of the study?

Structural validity: Q1. Does the scale consist of effect indicators, i.e. is it based on a reflective model? Q2. Was the percentage of missing items given? Q3. Was there a description of how missing items were handled? Q4. Was the sample size included in the analysis adequate? Q5. Were there any important flaws in the design or methods of the study? Q6. for CTT: Was exploratory or confirmatory factor analysis performed? Q7. for IRT: Were IRT tests for determining the (uni-) dimensionality of the items performed?

Internal consistency: Q1. Does the scale consist of effect indicators, i.e. is it based on a reflective model? Q2. Was the percentage of missing items given? Q3. Was there a description of how missing items were handled? Q4. Was the sample size included in the internal consistency analysis adequate? Q5. Was the unidimensionality of the scale checked? Q6. Was the sample size included in the unidimensionality analysis adequate? Q7. Was an internal consistency statistic calculated for each (unidimensional) (sub) scale separately? Q8. Were there any important flaws in the design or methods of the study? Q9. for classical test theory (CTT), continuous scores: Was Cronbach's alpha calculated? Q10. for CTT, dichotomous scores: Was Cronbach's alpha or KR-20 calculated? Q11. for IRT: Was a goodness of fit statistic at a global level calculated? E.g. χ^2, reliability coefficient of estimated latent trait value (index of (subject or item) separation)

Reliability: Q1. Was the percentage of missing items given? Q2. Was there a description of how missing items were handled? Q3. Was the sample size included in the analysis adequate? Q4. Were at least two measurements available? Q5. Were the administrations independent? Q6. Was the time interval stated? Q7. Were patients stable in the interim period on the construct to be measured? Q8. Was the time interval appropriate? Q9. Were the test conditions similar for both measurements? Q10. Were there any important flaws in the design or methods of the study? Q11. for continuous scores: Was an intraclass correlation coefficient (ICC) calculated? Q12. for dichotomous/nominal/ ordinal scores: Was kappa calculated? Q13. for ordinal scores: Was a weighted kappa calculated? Q14. for ordinal scores: Was the weighting scheme described?

Measurement error: Q1. Was the percentage of missing items given? Q2. Was there a description of how missing items were handled? Q3. Was the sample size included in the analysis adequate? Q4. Were at least two measurements available? Q5. Were the administrations independent? Q6. Was the time interval stated? Q7. Were patients stable in the interim period on the construct to be measured? Q8. Was the time interval appropriate? Q9. Were the test conditions similar for both measurements? Q10. Were there any important flaws in the design or methods of the study? Q11. for CTT: Was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?

Construct validity (i.e. hypotheses testing in COSMIN): Q1. Was the percentage of missing items given? Q2. Was there a description of how missing items were handled? Q3. Was the sample size included in the analysis adequate? Q4. Were hypotheses regarding correlations or mean differences formulated *a priori*? Q5. Was the expected direction of correlations or mean differences included in the hypotheses? Q6. Was the expected absolute or relative magnitude of correlations or mean differences included in the hypotheses? Q7. for convergent validity: Was an adequate description provided of the comparator instrument(s)? Q8. for convergent validity: Were the measurement properties of the comparator instrument(s) adequately described? Q9. Were there any important flaws in the design or methods of the study? Q10. Were design and statistical methods adequate for the hypotheses to be tested?

majority of criteria for methodological quality, with the exception of unidimensionality testing, which was often neglected.

Reliability includes both test-retest and inter-rater agreement, and was documented in only 5 of the 15 studies. Information on the percentage of missing data (Q1) and how they were handled (Q2) was provided in only one of these five studies (PBH-LCI:D). The sample size used to assess reliability was good for PBH-LCI:D and Tayside (Q3, 50<N<99), moderate for EAC and CNA-D (30<N<49), and poor for CARENAP (N<30). The measurement occasions were clearly or assumably independent (Q5) and made in similar conditions (Q9) for three studies (EAC, PBH-LCI:D, CNA-D), while this was not the case for CNA-D and CARENAP. For CNA-D, test-retest and inter-rater agreement were evaluated in a combined way by having two interviews conducted by two different persons two weeks apart. For the CARENAP, only

the inter-rater agreement was tested, based on a simultaneous evaluation made by an interviewer and an observer. The time interval between the two measurements (two or three weeks) was clearly stated (Q6) in four of the five studies. The exception was Tayside, where the time interval was highly variable, with an average of 21 days. The time interval was appropriate (Q8) and it was assumable or evidenced that informal dementia caregivers were stable in the interim period in terms of needs (Q7) for PBH-LCI:D, CARENAP and CNA-D. As mentioned, Tayside used a questionable time interval. EAC had a rather long interval of three weeks without any caution to ensure that informal dementia caregivers were stable within this period. Finally, three studies calculated adequate statistics for the assessment of reliability (Tayside, CARENAP, CNA-D), while EAC and PBH-LCI:D used less appropriate indices, namely, a Pearson correlation without evidence that no systematic change had

occurred. In summary, procedures mixing test-retest and inter-rater assessment, as well as as irregular or too long time intervals, and modest sample sizes, undermined the limited efforts to evaluate reliability.

Measurement error was only reported for one study (PBH-LCI:D), with no information on the percentage of missing data (Q1) and how they were handled (Q2). However the study had a good sample size of N = 79 (Q3), and used a clear (Q6) and appropriate (Q8) time interval between the two measures (i.e. two weeks), suggesting that the informal dementia caregivers were stable over that period (Q7). It was also assumable that the measurements were independent (Q5) and test conditions were similar (Q9), and appropriate statistics were used (Q11). The very limited evidence about measurement error was thus of good quality.

Construct validity was documented in eight of the 15 studies. Only JHDCNA and PBH-LCI:D provided information on the percentage of missing data (Q1) and how they were handled (Q2). Six studies had optimal or good sample sizes (Q3, N>50: JHDCNA, RAM, UNM, EAC, PBH-LCI:D, NAS), while CNA-D had a moderate sample in both studies with N = 45. In six studies, multiple hypotheses (PBH-LCI:D, CNCD, CNA-D) or a minimal number of hypotheses (JHDCNA, EAC, CNA-D) were formulated *a priori* (Q4), while RAM and UNM formulated only vague hypotheses. EAC was the sole study to specify the expected magnitude of the association (Q6). Seven studies described adequately all or most constructs measured by the comparator instruments (Q7), with only UNM delivering a poor description. In contrast, only JHDCNA and CNA-D provided comprehensive evidence of the measurement properties of the comparator instruments (Q8), with all of the others giving only partial information. Finally, all studies used clearly or assumably appropriate statistics, with the exception of PBH-LCI:D which used parametric statistics while the standard deviations for some variables suggested non-normal distributions. In summary, there was some quality empirical evidence on construct validity, although the psychometric properties of the comparator instruments were often insufficiently described.

Review findings
Quality of psychometric properties
We will first summarize the evidence available for each type of psychometric feature, with the exclusion of criterion validity, which is irrelevant in the absence of a gold standard, and of responsiveness or sensitivity to change, which was never evaluated for the reviewed instruments. We will then briefly discuss the psychometric properties of each instrument. The relevant information is presented in Appendix IV.

Content validity was documented for 13 of the 14 reviewed instruments. As mentioned above, the procedure used to optimize content validity was evaluated satisfactory in seven of these 13 instruments (CARENAP, CNA-D, CNCD, NAS, PBH-LCI:D, SIDECAR, QNP). Most studies generated items based on a literature review and/or an expert consultation, and then reviewed these items in collaboration with experts and at least five informal dementia caregivers. However the initial pool of the SIDECAR instrument was inductively developed on the basis of 42 interviews with caregivers. Caregivers, researchers and carer consultants were included in further steps of questionnaire development and testing of content validity. For the QNP, the final set of items was further submitted to additional informal dementia caregivers to optimize understandability. For three other instruments, a doubtful design was used (CADI, QCNE, Tayside; see Appendix IV for more details), and the three last instruments failed to include the target population in the process of item development (RAM, EAC, UNM).

Internal consistency was assessed for 10 of the 14 reviewed instruments. For three of these 10 instruments, the Cronbach alphas were computed for dimensions based on the results of a factor analysis. The latter supported a structure with five to eight dimensions (five for CNCD, with three to seven items; seven for PBH-LCI:D, with three to eight items; eight for the CADI, with two to seven items per dimension). However, for CADI, most alphas were below .70 despite an adequate factor analysis, and for CNCD and PBH-LCI:D, the factor analysis had largely insufficient sample sizes but all alphas were above .70. These results are therefore all partly problematic. For four other instruments, Cronbach alphas were provided for each dimension and comprised between .70 and .95, but for three of them no factor analysis was conducted (CNA-D, EAC and UNM), and for QCNE the factor analysis was inadequate, herefore this evidence should be considered inconclusive. Three other instruments reported only

the alpha for the full scale without considering the dimensions, with low values in for CARENAP and RAM, and a high value for QNP.

Reproducibility was evaluated for four of the 14 reviewed instruments in terms of test-retest agreement (CNA-D, PBH-LCI:D, EAC, Tayside), for three in terms of inter-rater reliability (CARENAP, CNA-D, Tayside), and for PBH-LCI:D in terms of measurement error. CNA-D and PBH-LCI:D showed satisfactory test-retest agreement, as evaluated with a proper procedure, with correlations in the .70 range. For two other instruments where the procedure was questionable, the average Kappa was excellent for EAC but varied between poor and excellent for the different subscales of Tayside. As presented above, all the procedures used to evaluate the inter-rater reliability were problematic. For CARENAP, which compared the simultaneous evaluation of an interviewer and an observer, thereby increasing the likelihood of agreement, the Kappas were high. For Tayside, which compared the ratings of a professional to the self-report of the informal caregiver, thereby reducing the chances of agreement, the Kappas were very low. For CNA-D, which evaluated the test-retest and inter-rater reliability in combination, based on two interviews conducted by different persons two weeks apart, the mean Kappa was high. The results for measurement error for PBH-LCI:D were good. Sound evidence regarding reproducibility is therefore still scarce, and limited to test-retest agreement and measurement error.

Validity was consistently assessed based on *construct validity* in seven of the 14 reviewed instruments. For CNA-D, PBH-LCI:D, RAM and EAC, precise *a priori* hypotheses were formulated and at least 75% of the results were in accordance with them. In contrast, CNCD had less than 75% of its well-formulated hypotheses confirmed. Most hypotheses focused on associations between unmet needs and the caregiver's objective (e.g. number of hours of care, problem behaviors or functional dependency of the care recipient) or subjective burden, depression or anxiety symptoms or psychological distress, amount of formal or social support received, self-care or quality of life. These postulated associations were either based on plausible links with other common relevant outcomes for informal dementia caregivers (e.g. depression and subjective burden for RAM; subjective burden for CNA-D; burden and psychological morbidity for CNCD),

or on different theoretical models, namely, the patient activation model for PBH-LCI:D, and Caplan's model of mental health consultation for EAC. For JHDCNA and UNM, the formulated hypotheses were too vague and numerous, seeming to pre-empt associations between all needs domains and all outcomes, which resulted in low rates of confirmation. They were based on plausible links with objective and subjective burden for JHDCNA, and Pearlin's stress process model for UNM. There is thus some evidence for construct validity.

Best validated instruments
Regarding individual instruments, the best-validated one is currently PBH-LCI:D by Sadak *et al.*[58] PBH-LCI:D was developed using an appropriate procedure regarding content validity, included six domains confirmed in a factor analysis and with good internal consistency, demonstrated adequate test-retest stability after two weeks, and showed the expected correlations with other variables indicating construct validity. EAC (in French) by Laprise *et al.*[61] had appropriate evidence of test-retest reliability and construct validity, although informal caregivers were not involved in the item development process and Cronbach alphas were computed without a dimensionality analysis. Four other instruments had adequate support for content validity, but insufficient evidence on all other psychometric properties (CARENAP, CNCD, QNP, NAS). The excellent content validity of SIDECAR provided a valid basis for further psychometric testing of this instrument, which seems to be currently underway. The CNA-D and RAM had good evidence for construct validity, but inconclusive evidence for all other properties. The last five instruments currently provide no convincing evidence on any psychometric property (CADI, JHDCNA, QCNE, Tayside, UNM).

Content and structure of dimensions
The topics assessed in the instruments can be divided into three thematic groups of needs: i) need for information and education, ii) needs related to emotional support, and iii) need for other accessible and appropriate services. All instruments contained at least one item in the first and second group. Ten instruments comprised items in the third group. The first group of needs was typically assessed with items about the need for education regarding care tasks, especially dementia specific caring skills (included in

all instruments), items about information on local services or community resources for persons with dementia or caregivers (included in 13 instruments). Needs for information about dementia and its treatment were present in ten instruments and needs in relation to the characteristics, accessibility and availability of services were included in nine instruments. In the second thematic group, the most considered topics were counselling for negative emotions (nine instruments), support from the informal network such as family, friends or other caregivers (nine instruments) or the society (eight instruments), and respite (eight instruments). Financial and legal support (seven items) was the most common topic in the third group of needs. Table 1 provides a summary of the topics assessed in the different instruments.

Discussion

This psychometric literature review identified 14 needs assessment instruments for informal dementia caregivers with empirically evaluated measurement properties. Their systematic evaluation, based on the COSMIN criteria,[42,45,46] revealed that half of them had excellent content validity. In contrast, the structure validity was rarely examined, and factor analyses were in most cases of low quality because of insufficient sample sizes or questionable procedures. None of the instruments had an optimally tested and good internal consistency, as the sole one with an adequate dimensionality analysis had low alphas, and all others had either their dimensionality evaluated with small samples, or no proper evaluation of their dimentionality – although some had high alphas. Regarding reliability, test-retest agreement was rarely tested, and only two instruments used a satisfactory procedure and obtained good correlations. Inter-rater agreement was relevant only for professionally interviewed instruments, and evaluated using inadequate procedures. Regarding validity, in the absence of a gold standard, no criterion validity could be assessed, but construct validity was evaluated in more than half of the instruments, with satisfactory procedures and results on four of them. Comparing the different instruments reviewed, PBH-LCI:D was rated as having the best psychometric evidence, and EAC was also partly supported, while most other instruments had limited or no proof of their psychometric soundness. SIDECAR showed very promising results regarding content validity.

However further psychometric testing is needed and seems to be currently underway.

This overview highlighted the importance of guidelines such as COSMIN[42,45,46,48] in guiding the efforts in the development of instruments with optimal psychometric properties in a specific field. As we noted for needs assessment in informal dementia caregivers, despite international investments in instrument development, optimal standards were on many occasions not achieved due to problematic procedures and analyses, as well as largely insufficient sample sizes. Our review therefore provides essential information to inform future efforts in the development of such measures in order to achieve more robust psychometric results.

Regarding needs assessment in informal dementia caregivers, our review showed that there are presently several instruments with adequate content validity, developed in diverse countries (Singapore, Austria, the Netherlands, the UK and the US). This information provides an excellent starting point for further development. The priority should now be to identify the structure, in terms of the number of different domains required to cover the diverse needs of informal dementia caregivers. The preliminary factor-analytic evidence that we reviewed identified five to eight domains of needs, but overall the instruments reviewed comprised two to 18 different need domains. Currently, there is no established theoretical model to organize the very diverse and complex needs of our population of interest. Pini et al.[68] made a first attempt in this direction by developing a needs-led framework based on qualitative interviews to conceptualize the impact of caring on the lives of family caregivers. A robust theoretical framework might enhance the conceptual clarity around the assessment of these needs, and could be further informed by explorations of the factorial structure with sufficient sample sizes. Nevertheless, cultural differences in the experience of dementia caregiving could affect the content of needs assessment instruments.[69] Such differences were however difficult to evaluate at the stage of content validity. They should be examined at later stages of instrument validation by testing the measurement invariance.

Our review also identified different approaches that were applied to assess the construct validity of some of the reviewed instruments. These approaches involved testing the associations between needs and other related constructs, predominantly the informal

caregiver's objective or subjective burden, psychological symptoms, the amount of support received, self-care or quality of life. The postulated associations were mostly based on plausible links with common outcomes for informal dementia caregivers, and sometimes on a theoretical model, with different models being used. The diversity of these procedures indicated that further conceptualisation of the place of needs within relevant theoretical models could strengthen the nomological net and thereby support more solid examination of the construct validity. Another challenge in this area is to ensure that the content of the items of the different instruments are not overlapping, to prevent spurious correlations. This would typically be the case for needs assessment including items about burden or depression for example, when their construct validity is tested with instruments measuring these constructs.

Regarding reliability, test-retest stability was scarcely assessed. And inter-rater agreement was always tested with questionable procedures, despite its importance as multiple professionals are normally involved with informal dementia caregivers. Substantial efforts need to be invested in these aspects, but we also have to keep in mind that their assessment is challenging in a fragile population such as informal dementia caregivers. Indeed, a short time interval is required as the situation of the person with dementia and the larger context can change quickly, thereby modifying the needs of the caregiver. Yet it is certainly difficult to obtain two assessments within one or two weeks from chronically stressed and often exhausted caregivers.

Finally, needs assessment were mainly used for evaluations at one specific point in time. Increasingly, the assessment of needs might be considered as an outcome to analyze the impact of interventions, or used to document the evolution of needs over time. The use of needs assessment as an outcome or longitudinal measure requires a satisfactory sensitivity to change, which was not assessed in the reviewed instruments, and warrants scientific attention in the future.

This review has both strength and limitations. First, our use of a highly structured procedure based on the COSMIN criteria certainly is a strength. However, COSMIN uses a very stringent evaluation of the reviewed instruments, as the final appreciation for each psychometric property is based on the lowest grade across all specific criteria for this property. Although the resulting synthesis could give the impression of a globally poor quality of evidence, we believe it is very helpful in the process of achieving high psychometric standards. Secondly, we were able to include instruments and articles in more diverse languages than previously published reviews,[36-38] although our limitation on our language skills did not allow us to examine publications in any Asian or Arabic languages. Thirdly, despite our efforts, we could not always access gray literature, as some authors did not respond, persons in charge of the project were absent, or authors were in a commercial process preventing them to provide access to the manual. Our conclusions are therefore limited to available information.

Conclusion
Recommendations for practice
This review revealed several instruments for measuring the needs of informal dementia caregivers. However the evidence for their use in the clinical or research setting is often limited. The two best validated instruments are PBH-LCI:D by Sadak et al.[58] and EAC by Laprise et al.[61] While PBH-LCI:D is in English and intended for use in clinical and research settings, EAC is in French and recommended only for clinical use. Both of them include a scoring system which allows for comparison of results at different time points and between different informal caregivers. PBH-LCI:D and EAC contain items covering the most common topics found across all instruments, namely, the need for information and education, needs related to emotional support and the need for other accessible and appropriate services. Their self-administered application method suggests less effort in administration time for professionals. Nontheless there is no information regarding administration burden for both, which would be important in terms of usability, especially in the clinical setting. In regards to well-documented cultural differences in the experience of dementia caregiving, and the absence of empirical evidence on the measurement invariance of the available instruments across diverse cultural groups, caution should be taken in using them in cultural contexts different from those in which the instruments were developed.[69]

Recommendations for research
Although we identified moderate to high evidence of strong psychometric properties for PBH-LCI:D and

EAC, this review highlights the need for further developments in the field of needs assessment in informal dementia caregivers, particularly in structural validity and construct validity, as well as test-retest reliability and sensitivity to change. The evaluation of both forms of validity would certainly benefit from a more robust theoretical framework about the core dimensions of needs in informal dementia caregivers, and the relationship between needs and other relevant outcomes for this population. We also need to identify appropriate procedures to assess test-retest reliability with minimal additional burden for informal dementia caregivers, and to evaluate sensitivity to change appropriately despite this involving a demanding procedure.

References

1. Swiss Confederation. Support for family caregivers: An analysis of the situation and the required measures in Switzerland [Soutien aux proches aidants: Analyse de la situation et mesures requises pour la Suisse] Bern: Federal Council; 2014; French.
2. Thompson GN, Roger K. Understanding the needs of family caregivers of older adults dying with dementia. Palliat Support Care 2014;123:223–31.
3. Alzheimer Europe,. Who cares? The state of dementia care in Europe. Luxembourg: Alzheimer Europe; 2006.
4. Fisher GG, Franks MM, Plassman BL, Brown SL, Potter GG, Llewellyn D, et al. Caring for individuals with dementia and cognitive impairment, not dementia: findings from the aging, demographics, and memory study. J Am Geriatr Soc 2011;593:488–94.
5. Alzheimer's Association. 2016 Alzheimer's Disease Facts and Figures. Alzheimer's & Dementia 2016;124:32–40.
6. De Pietro C, Camenzind P, Sturny I, Crivelli L, Edwards-Garavoglia S, Spranger A, et al. Switzerland - Health system review. Brussels: European Observatory on Health Systems and Policies; 2015.
7. Swiss Confederation. Explanatory report on the preliminary project of the federal law on improving the reconciliation between professional activities and the care of relatives. [Rapport explicatif concernant l'avant-projet de la loi fédérale sur l'amélioration de la conciliation entre activité professionnelle et prise en charge de proches]. Bern: Federal Council; 2018; French.
8. Zwaanswijk M, Peeters JM, van Beek AP, Meerveld JH, Francke AL. Informal caregivers of people with dementia: problems, needs and support in the initial stage and in subsequent stages of dementia: a questionnaire survey. Open Nurs J 2013;7:6–13.
9. Kesselring A. Caring for family members at home [Angehörige zu Hause pflegen: Anatomie einer Arbeit]. Swiss Medical Journal [Schweizerische Ärztezeitung] 2004;85(10):504–6; German.
10. Rosa E, Lussignoli G, Sabbatini F, Chiappa A, Di Cesare S, Lamanna L, et al. Needs of caregivers of the patients with dementia. Arch Gerontol Geriatr 2010;51(1):54–8.
11. Kraft E, Marti M, Werner S, Sommer H. Cost of dementia in Switzerland. Swiss Med Wkly 2010;140:w13093.
12. Butcher HK, Holkup PA, Buckwalter KC. The experience of caring for a family member with Alzheimer's disease. West J Nursing Res 2001;231:33–55.
13. de la Cuesta C. The craft of care: family care of relatives with advanced dementia. Qual Health Res 2005;157:881–96.
14. Vellone E, Sansoni J, Cohen MZ. The experience of Italians caring for family members with Alzheimer's disease. J Nurs Scholarsh 2002;344:323–9.
15. Knapp M, Prince M, Albanese E, Banerjee S, Dhanasiri S, Fernandez J-L, et al. Dementia UK. London: Alzheimer's Society; 2007.
16. Reinhard SC, Given B, Petlick NH, Bemis A. Supporting family caregivers in providing care. In: Hughes RG, editor. Patient safety and quality: An evidence-based handbook for nurses.. Rockville (MD): Agency for Healthcare Research and Quality (US), 2008;341–404.
17. EUROFAMCARE. EUROFAMCARE – Services for Supporting Family Carers of Older Dependent People in Europe: Characteristics, Coverage and Usage. Brussels: EUROFAMCARE Consortium; 2006.
18. Schulz R, Martire LM. Family caregiving of persons with dementia: prevalence, health effects, and support strategies. Am J Geriatr psychiatry 2004;123:240–9.
19. Perrig-Chiello P, Höpflinger F, Schnegg B. SwissAgeCare: Nursing relatives of older people in Switzerland. Bern: University of Bern; 2010; German.
20. Brodaty H, Donkin M. Family caregivers of people with dementia. Dialogues Clin Neurosci 2009;112:217–28.

21. Gaugler JE, Kane RL, Kane RA, Clay T, Newcomer R. Caregiving and institutionalization of cognitively impaired older people: utilizing dynamic predictors of change. Gerontologist 2003;432:219–29.

22. Black BS, Johnston D, Rabins PV, Morrison A, Lyketsos C, Samus QM. Unmet needs of community-residing persons with dementia and their informal caregivers: findings from the maximizing independence at home study. J Am Geriatr Soc 2013;6112:2087–95.

23. Ducharme F, Kergoat MJ, Coulombe R, Levesque L, Antoine P, Pasquier F. Unmet support needs of early-onset dementia family caregivers: a mixed-design study. BMC Nurs 2014;131:49.

24. Bass DM, Judge KS, Snow AL, Wilson NL, Morgan R, Looman WJ, et al. Caregiver outcomes of partners in dementia care: effect of a care coordination program for veterans with dementia and their family members and friends. J Am Geriatr Soc 2013;618:1377–86.

25. Afram B, Verbeek H, Bleijlevens MH, Hamers JP. Needs of informal caregivers during transition from home towards institutional care in dementia: a systematic review of qualitative studies. Int psychogeriatr 2015;276:891–902.

26. Freudiger Pittet S, Jordan A. Evaluation of caregiver burden and needs. [Evaluation de la charge et des besoins des proches aidants]. Lausanne: Association vaudoise d'aide et de soins à domicile (AVASD); 2012; French.

27. Brodaty H, Thomson C, Thompson C, Fine M. Why caregivers of people with dementia and memory loss don't use services. Int J Geriatr Psychiatry 2005;206:537–46.

28. Bass DM, Judge KS, Snow AL, Wilson NL, Looman WJ, McCarthy C, et al. Negative caregiving effects among caregivers of veterans with dementia. Am J geriatr psychiatry 2012;203:239–47.

29. Karlsson S, Bleijlevens M, Roe B, Saks K, Martin MS, Stephan A, et al. Dementia care in European countries, from the perspective of people with dementia and their caregivers. J Adv Nurs 2015;716:1405–16.

30. Gaugler JE, Kane RL, Kane RA, Newcomer R. Unmet care needs and key outcomes in dementia. J Am Geriatr Soc 2005;5312:2098–105.

31. Gaugler JE, Kane RL, Kane RA, Newcomer R. Early community-based service utilization and its effects on institutionalization in dementia caregiving. Gerontologist 2005;452:177–85.

32. van der Roest HG, Meiland FJ, Maroccini R, Comijs HC, Jonker C, Droes RM. Subjective needs of people with dementia: a review of the literature. Int psychogeriatr 2007;193:559–92.

33. Reynolds T, Thornicroft G, Abas M, Woods B, Hoe J, Leese M, et al. Camberwell Assessment of Need for the Elderly (CANE): Development, validity and reliability. BJPsych 2000;1765:444–52.

34. Wancata J, Krautgartner M, Berner J, Alexandrowicz R, Unger A, Kaiser G, et al. The Carers' Needs Assessment for Dementia (CNA-D): development, validity and reliability. Int psychogeriatr 2005;173:393–406.

35. McWalter G, Toner H, McWalter A, Eastwood J, Marshall M, Turvey T. A community needs assessment: the care needs assessment pack for dementia (CarenapD)–its development, reliability and validity. Int J Geriatr Psychiatry 1998;131:16–22.

36. Novais T, Dauphinot V, Krolak-Salmon P, Mouchoux C. How to explore the needs of informal caregivers of individuals with cognitive impairment in Alzheimer's disease or related diseases? A systematic review of quantitative and qualitative studies. BMC Geriatr 2017;171:86.

37. Bangerter LR, Griffin JM, Zarit SH, Havyer R. Measuring the Needs of Family Caregivers of People With Dementia: An Assessment of Current Methodological Strategies and Key Recommendations. J Appl Gerontol 2017;1:1–15.

38. Mansfield E, Boyes AW, Bryant J, Sanson-Fisher R. Quantifying the unmet needs of caregivers of people with dementia: a critical review of the quality of measures. Int J Geriatr Psychiatry 2017;323:274–87.

39. Terwee CB. Protocol for systematic reviews of measurement properties. Amsterdam: VU University Medical Center Knowledgecenter Measurement Instruments; 2011.

40. Kipfer S, Eicher M, Oulevey Bachmann A, Pihet S. Reliability, validity and relevance of needs assessment instruments for informal dementia caregivers: a psychometric systematic review protocol. JBI Database System Rev Implement Rep 2018;162:269–86.

41. Hileman JW, Lackey NR, Hassanein RS. Identifying the needs of home caregivers of patients with cancer. Oncol nurs forum 1992;195:771–7.

42. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012;214:651–7.

43. Biomedical Information of the Dutch Library Association. Search blocks. Nijkerk: Biomedical Information of the Dutch Library Association. [Internet]. 2019 Feb [cited 2019 Apr 18]. Available from: https://blocks.bmi-online.nl/

44. COSMIN. Search block and filter. COSMIN. [Internet]. 2019 Feb [cited 2019 Apr 18]. Available from: https://www.cosmin.nl/tools/pubmed-search-filters/

45. Terwee CB. COSMIN checklist with 4-point scale. Amsterdam: University Medical Center, Department of Epidemiology and Biostatics, EMGO Institute for Health and Care Research; 2011.

46. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. COSMIN checklist manual. Amsterdam: University Medical Center, Department of Epidemiology and Biostatics, EMGO Institute for Health and Care Research; 2012

47. Simpelaere I, White A, Bekkering GE, Geurden B, Van Nuffelen G, De Bodt M. Patient-reported and proxy-reported outcome measures for the assessment of health-related quality of life among patients receiving enteral feeding: a systematic review protocol. JBI database System Rev Implement Rep 2016;147:45–75.

48. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;601:34–42.

49. Charlesworth GM, Tzimoula XM, Newman SP. Carers Assessment of Difficulties Index (CADI): Psychometric properties for use with carers of people with dementia. Aging Ment Health 2007;112:218–25.

50. McWalter GJ, Toner HL, Eastwood J, Corser AS, Marshall MT, Turvey AA, Howie C. User Manual for the Care Needs Assessment Pack for Dementia (CarenapD). Stirling: University of Stirling; 1996.

51. McWalter G, Toner H, McWalter A, Eastwood J, Marshall M, Turvey T. A community needs assessment: the care needs assessment pack for dementia (CarenapD)–its development, reliability and validity. Int J Geriatric Psychiatry 1998;131:16–22.

52. Wancata J, Krautgartner M, Berner J, Alexandrowicz R, Unger A, Kaiser G, et al. The Carers' Needs Assessment for Dementia (CNA-D): development, validity and reliability. Int Psychogeriatr 2005;173:393–406.

53. Kaiser G, Krautgartner M, Alexandrowicz R, Unger A, Marquart B, Weiss M, et al. Die Übereinstimmungsvalidität des "Carers' Needs Assessment for Dementia" (CNA-D). Neuropsychiatrie 2005;194:134–40.

54. Vaingankar JA, Subramaniam M, Picco L, Eng GK, Shafie S, Sambasivam R, et al. Perceived unmet needs of informal caregivers of people with dementia in Singapore. Int Psychogeriatr 2013;2510:1605–19.

55. Vaingankar JA, Abdin E, Chong SA, Sambasivam R, Shafie S, Jeyagurunathan A, et al. Validity and reliability of the Caregivers' Needs Checklist for Dementia. Arch Psychol 2018;21:1–16.

56. Hughes TB, Black BS, Albert M, Gitlin LN, Johnson DM, Lyketsos CG, et al. Correlates of objective and subjective measures of caregiver burden among dementia caregivers: influence of unmet patient and caregiver dementia-related care needs. Int Psychogeriatr 2014;2611:1875–83.

57. Wackerbarth SB, Johnson MMS. Essential information and support needs of family caregivers. Patient Educ Couns 2002;472:95–100.

58. Sadak T, Korpak A, Borson S. Measuring caregiver activation for health care: Validation of PBH-LCI:D. Geriatr Nurs 2015;364:284–92.

59. Czaja SJ, Gitlin LN, Schulz R, Zhang S, Burgio LD, Stevens AB, et al. Development of the risk appraisal measure: a brief screen to identify risk areas and guide interventions for dementia caregivers. J Am Geriatr Soc 2009;576:1064–72.

60. Dimakopoulou E, Sakka P, Efthymiou A, Karpathiou N, Karydaki M. Evaluating the Needs of Dementia Patients' Caregivers in Greece: A Questionnaire Survey. Int J Caring Sci 2015;82:274–80.

61. Laprise R, Dufort F, Lavoie F. Construction and validation of a scale about consultation expectations of caregivers of older people. [Construction et validation d'une echelle d'attentes en matiere de consultation aupres d'aidant(e)s de personnes agees]. Can J Aging 2001;20(2):211–32; French.

62. Peeters JM, Van Beek A, Meerveld J, Spreeuwenberg P, Francke A. Informal caregivers of persons with dementia, their use of and needs for specific professional support: a survey of the National Dementia Programme. BMC Nurs 2010;99:1–8.

63. Van der Poel K, van Beek A. Development of the questionnaire 'Wishes and problems of caregivers of people with dementia'. [Ontwikkeling vragenlijst 'Wensen en problemen van mantelzorgers van mensen met dementie']. Utrecht: Nivel; 2006; Dutch.

64. Oyebode JR, Pini S, Ingleson E, Megson M, Horton M, Clare L, et al. Development of an item pool for a needs-based measure of quality of life of carers of a family member with dementia. PATIENT 2019;12:125–36.

65. Gordon DS, Spicker P, Ballinger BR, Gillies B, McWilliam N, Mutch WJ. A population needs assessment profile for dementia. Int J Geriatr Psychiatry 1997;126:642–7.

66. Gaugler JE, Anderson KA, Leach MS, Smith CD, Schmitt FA, Mendiondo M. The emotional ramifications of unmet need in dementia caregiving. Am J Alzheimers Dis Other Demen 2004;196:369–80.

67. Moher D, Liberati A, Tetzlaff J, Altman DG. The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 2009;66:e1000097.

68. Pini S, Ingleson E, Megson M, Clare L, Wright P, Oyebode JR. A needs-led framework for understanding the impact of caring for a family member with dementia. Gerontologist 2018;582:e68–77.

69. Janevic MR, Connell CM. Racial, ethnic, and cultural differences in the dementia caregiving experience: Recent findings. Gerontologist 2001;413:334–47.

Appendix I: Search strategy
Embase
Searched on 21 Feb 2019

Search	Query	Records retrieved
#1	('needs'/exp OR needs:ab,ti)	1340
#2	((carer* OR 'informal carer*' OR famil* OR caregiv* OR relative* OR spouse* OR 'spouse'/exp OR child* OR 'child'/exp OR child OR 'caregiver'/exp OR caregiver OR 'family'/exp OR family:ab,ti) AND ('dementia'/de OR 'alzheimer disease'/exp OR 'senile dementia'/de OR 'presenile dementia'/exp OR 'multiinfarct dementia'/exp OR 'diffuse lewy body disease'/exp OR 'frontotemporal dementia'/exp OR alzheimer* OR dementi* OR amenti*:ti,ab))	
#3	(assess* OR evaluate* OR 'outcome'/exp OR outcome OR 'outcomes'/exp OR outcomes OR identif* OR 'index'/exp OR index OR indices OR 'instrument'/exp OR instrument OR 'instruments'/exp OR instruments OR interview* OR 'inventory'/exp OR inventory OR measure OR measures OR 'questionnaire'/exp OR questionnaire OR 'questionnaires'/exp OR questionnaires OR 'profile'/exp OR profile OR profiles OR 'scale'/exp OR scale OR scales OR 'score'/exp OR score OR scores OR status OR 'survey'/exp OR survey OR 'surveys'/exp OR surveys OR test*:ab,ti)	
#4	#1 AND #2 AND #3 AND ('intermethod comparison'/exp OR 'intermethod comparison' OR 'data collection method'/exp OR 'data collection method' OR 'validation study'/exp OR 'validation study' OR 'feasibility study'/exp OR 'feasibility study' OR 'pilot study'/exp OR 'pilot study' OR 'psychometry'/exp OR 'psychometry' OR 'reproducibility'/exp OR 'reproducibility' OR reproducib*:ab,ti OR 'audit':ab,ti OR psychometr*:ab,ti OR clinimetr*:ab,ti OR clinometr*:ab,ti OR 'observer variation'/exp OR 'observer variation' OR 'observer variation':ab,ti OR 'discriminant analysis'/exp OR 'discriminant analysis' OR 'validity'/exp OR 'validity' OR reliab*:ab,ti OR valid*:ab,ti OR 'coefficient':ab,ti OR 'internal consistency':ab,ti OR (cronbach*:ab,ti AND ('alpha':ab,ti OR 'alphas':ab,ti)) OR 'item correlation':ab,ti OR 'item correlations':ab,ti OR 'item selection':ab,ti OR 'item selections':ab,ti OR 'item reduction':ab,ti OR 'item reductions':ab,ti OR 'agreement':ab,ti OR 'precision':ab,ti OR 'imprecision':ab,ti OR 'precise values':ab,ti OR 'test-retest':ab,ti OR ('test':ab,ti AND 'retest':ab,ti) OR (reliab*:ab,ti AND ('test':ab,ti OR 'retest':ab,ti)) OR 'stability':ab,ti OR 'interrater':ab,ti OR 'inter-rater':ab,ti OR 'intrarater':ab,ti OR 'intra-rater':ab,ti OR 'intertester':ab,ti OR 'inter-tester':ab,ti OR 'intratester':ab,ti OR 'interobeserver':ab,ti OR 'inter-observer':ab,ti OR 'intraobserver':ab,ti OR 'intertechnician':ab,ti OR 'inter-technician':ab,ti OR 'intratechnician':ab,ti OR 'interexaminer':ab,ti OR 'inter-examiner':ab,ti OR 'intraexaminer':ab,ti OR 'interassay':ab,ti OR 'inter-assay':ab,ti OR 'intraassay':ab,ti OR 'intra-assay':ab,ti OR 'interindividual':ab,ti OR 'inter-individual':ab,ti OR 'intraindividual':ab,ti OR 'intra-individual':ab,ti OR 'interparticipant':ab,ti OR 'inter-participant':ab,ti OR 'intraparticipant':ab,ti OR 'kappa':ab,ti OR 'kappas':ab,ti OR 'coefficient of variation':ab,ti OR repeatab*:ab,ti OR ((replicab*:ab,ti OR 'repeated':ab,ti) AND ('measure':ab,ti OR 'measures':ab,ti OR 'findings':ab,ti OR 'result':ab,ti OR 'results':ab,ti OR 'test':ab,ti OR 'tests':ab,ti)) OR generaliza*:ab,ti OR generalisa*:ab,ti OR 'concordance':ab,ti OR ('intraclass':ab,ti AND correlation*:ab,ti) OR 'discriminative':ab,ti OR 'known group':ab,ti OR 'factor analysis':ab,ti OR 'factor analyses':ab,ti OR 'factor structure':ab,ti OR 'factor structures':ab,ti OR 'dimensionality':ab,ti OR subscale*:ab,ti OR 'multitrait scaling analysis':ab,ti OR 'multitrait scaling analyses':ab,ti OR 'item discriminant':ab,ti OR 'interscale correlation':ab,ti OR 'interscale correlations':ab,ti OR (('error':ab,ti OR 'errors':ab,ti) AND (measure*:ab,ti OR correlat*:ab,ti OR evaluat*:ab,ti OR 'accuracy':ab,ti OR 'accurate':ab,ti OR 'precision':ab,ti OR 'mean':ab,ti)) OR 'individual variability':ab,ti OR 'interval variability':ab,ti OR 'rate variability':ab,ti OR 'variability analysis':ab,ti OR ('uncertainty':ab,ti AND ('measurement':ab,ti OR 'measuring':ab,ti)) OR 'standard error of measurement':ab,ti OR sensitiv*:ab,ti OR responsive*:ab,ti OR ('limit':ab,ti AND 'detection':ab,ti) OR 'minimal detectable concentration':ab,ti OR interpretab*:ab,ti OR (small*:ab,ti AND ('real':ab,ti OR 'detectable':ab,ti) AND ('change':ab,ti OR 'difference':ab,ti)) OR 'meaningful change':ab,ti OR 'minimal important change':ab,ti OR 'minimal important difference':ab,ti OR 'minimally important change':ab,ti OR 'minimally important difference':ab,ti OR 'minimal detectable change':ab,ti OR 'minimal detectable difference':ab,ti OR 'minimally detectable change':ab,ti OR 'minimally detectable difference':ab,ti OR 'minimal real change':ab,ti OR 'minimal real difference':ab,ti OR 'minimally real change':ab,ti OR 'minimally real difference':ab,ti OR 'ceiling effect':ab,ti OR 'floor effect':ab,ti OR 'item response model':ab,ti OR 'irt':ab,ti OR 'rasch':ab,ti OR 'differential item functioning':ab,ti OR 'dif':ab,ti OR 'computer adaptive testing':ab,ti OR 'item bank':ab,ti OR 'cross-cultural equivalence':ab,ti)	
#5	#4 NOT ('addresses':it OR 'biography':it OR 'case reports':it OR 'comment':it OR 'directory':it OR 'editorial':it OR 'festschrift':it OR 'interview':it OR 'lectures':it OR 'legal cases':it OR 'legislation':it OR 'letter':it OR 'news':it OR 'newspaper article':it OR 'patient education handout':it OR 'popular works':it OR 'congresses':it OR 'consensus development conference':it OR 'consensus development conference, nih':it OR 'practice guideline':it) NOT ('animals'/exp NOT 'humans'/exp)	
Publication time: no limitation / Language: no limitation		

MEDLINE via Pubmed

Searched on 21 Feb 2019

Search	Query	Records retrieved
#1	(need*[Title/Abstract])	
#2	((carer*[tiab] OR informal carer*[tiab] OR Family[Mesh] OR famil*[tiab] OR caregiv*[tiab] OR caregivers[Mesh] OR relative*[tiab] OR spouse*[tiab] OR child*[tiab]) AND (Dementia[Mesh] OR Alzheimer Disease[Mesh] AND alzheimer*[tiab] OR dementi*[tiab] OR amenti*[tiab]))	1620
#3	(assess*[tiab] OR evaluate*[tiab] OR outcome*[tiab] OR identif*[tiab] OR index[tiab] OR indices[tiab] OR instrument*[tiab] OR interview*[tiab] OR inventory[tiab] OR measure*[tiab] OR questionnaire*[tiab] OR profile*[tiab] OR scale*[tiab] OR score*[tiab] OR status[tiab] OR survey*[tiab] OR test*[tiab])	
#4	#1 AND #2 AND #3 AND ((instrumentation[sh] OR methods[sh] OR Validation Studies[pt] OR Comparative Study[pt] OR psychometrics[MeSH] OR psychometr*[tiab] OR clinimetr*[tw] OR clinometr*[tw] OR outcome assessment AND (health care) AND MeSH OR outcome assessment[tiab] OR outcome measure*[tw] OR observer variation[MeSH] OR observer variation[tiab] OR Health Status Indicators[Mesh] OR reproducibility of results[MeSH] OR reproducib*[tiab] OR discriminant analysis[-MeSH] OR reliab*[tiab] OR unreliab*[tiab] OR valid*[tiab] OR coefficient of variation[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR internal consistency[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement[tw] OR precision[tw] OR imprecision[tw] OR precise values[tw] OR test-retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intra-observer[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa[tiab] OR kappas[tiab] OR repeatab*[tw] OR ((replicab*[tw] OR repeated[tw]) AND (measure[tw] OR measures[tw] OR findings[tw] OR result[tw] OR results[tw] OR test[tw] OR tests[tw])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR known group[tiab] OR factor analysis[tiab] OR factor analyses[tiab] OR factor structure[tiab] OR factor structures[tiab] OR dimension*[tiab] OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR item discriminant[tiab] OR interscale correlation*[tiab] OR error[tiab] OR errors[tiab] OR individual variability[tiab] OR interval variability[-tiab] OR rate variability[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR standard error of measurement[-tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR (limit[tiab] AND detection[tiab]) OR minimal detectable concentration[tiab] OR interpretab*[tiab] OR ((minimal[tiab] OR minimally[tiab] OR clinical[tiab] OR clinically[tiab]) AND (important[tiab] OR significant[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR meaningful change[tiab] OR ceiling effect[tiab] OR floor effect[tiab] OR Item response model[tiab] OR IRT[tiab] OR Rasch[tiab] OR Differential item functioning[tiab] OR DIF[tiab] OR computer adaptive testing[tiab] OR item bank[tiab] OR cross-cultural equivalence[tiab])))	
#5	#4 NOT ((addresses[Publication Type] OR biography[Publication Type] OR case reports[Publication Type] OR comment[Publication Type] OR directory[Publication Type] OR editorial[Publication Type] OR festschrift[Publication Type] OR interview[Publication Type] OR lectures[Publication Type] OR legal cases[Publication Type] OR legislation[Publication Type] OR letter[Publication Type] OR news[Publication Type] OR newspaper article[Publication Type] OR patient education handout[Publication Type] OR popular works[Publication Type] OR congresses[Publication Type] OR consensus development conference[Publication Type] OR consensus development conference, nih[Publication Type] OR practice guideline[Publication Type]) NOT (animals[MeSH Terms] NOT humans[MeSH Terms])))	
Publication time: no limitation / Language: no limitation		

CINAHL via EBSCOhost
Searched on 21 Feb2019

Search	Query	Records retrieved
#1	(AB need*) OR (TI need*)	
#2	(TI (carer* OR informal carer* OR famil* OR caregiv* OR relative* OR spouse* OR child*) OR AB (carer* OR informal carer* OR famil* OR caregiv* OR relative* OR spouse* OR child*) OR (MH "family") OR (MH "caregivers")) AND ((MH "Dementia+") OR (MH "Dementia, Presenile+") OR (MH "Dementia, Senile+") OR (MH "Dementia, Vascular+") OR (MH "Lewy Body Disease") OR (MH "Alzheimer's Disease") OR (MH "Dementia, Multi-Infarct") OR (MH "Frontotemporal Lobar Degeneration+") OR (MH "Frontotemporal Dementia") OR TI (alzheimer* OR dementi* OR amenti*) OR AB (alzheimer* OR dementi* OR amenti*))	816
#3	TI (assess* OR evaluate* OR outcome OR outcomes OR identif* OR index OR indices OR instrument OR instruments OR interview* OR inventory OR measure OR measures OR questionnaire OR questionnaires OR profile OR profiles OR scale OR scales OR score OR scores OR status OR survey OR surveys OR test*) OR AB (assess* OR evaluate* OR outcome OR outcomes OR identif* OR index OR indices OR instrument OR instruments OR interview* OR inventory OR measure OR measures OR questionnaire OR questionnaires OR profile OR profiles OR scale OR scales OR score OR scores OR status OR survey OR surveys OR test*)	
#4	#1 AND #2 AND #3 AND (MH "Psychometrics") or (TI psychometr* or AB psychometr*) or (TI clinimetr* or AB clinimetr*) or (TI clinometr* OR AB clinometr*) or (MH "Outcome Assessment") or (TI outcome assessment or AB outcome assessment) or (TI outcome measure* or AB outcome measure*) or (MH "Health Status Indicators") or (MH "Reproducibility of Results") or (MH "Discriminant Analysis") or ((TI reproducib* or AB reproducib*) or (TI reliab* or AB reliab*) or (TI unreliab* or AB unreliab*)) or ((TI valid* or AB valid*) or (TI coefficient or AB coefficient) or (TI homogeneity or AB homogeneity)) or (TI homogeneous or AB homogeneous) or (TI "coefficient of variation" or AB "coefficient of variation") or (TI "internal consistency" or AB "internal consistency") or (MH "Internal Consistency+") or (MH "Reliability+") or (MH "Measurement Error+") or (MH "Content Validity+") or "hypothesis testing" or "structural validity" or "cross-cultural validity" or (MH "Criterion-Related Validity+") or "responsiveness" or "interpretability" or (TI reliab* or AB reliab*) and ((TI test or AB test) OR (TI retest or AB retest)) or (TI stability or AB stability) or (TI interrater or AB interrater) or (TI inter-rater or AB inter-rater) or (TI intrarater or AB intrarater) or (TI intra-rater or AB intrarater) or (TI intertester or AB intertester) or (TI inter-tester or AB inter-tester) or (TI intratester or AB intratester) or (TI intra-tester or AB intra-tester) or (TI interobserver or AB interobserver) or (TI inter-observer or AB inter-observer) or (TI intraobserver or AB intraobserver) or (TI intra-observer or AB intra-observer) or (TI intertechnician or AB intertechnician) or (TI inter-technician or AB inter-technician) or (TI intratechnician or AB intratechnician) or (TI intra-technician or AB intra-technician) or (TI interexaminer or AB interexaminer) or (TI inter-examiner or AB inter-examiner) or (TI intraexaminer or AB intraexaminer) OR (TI intra-examiner or AB intra-examiner) or (TI intra-examiner or AB intraexaminer) or (TI interassay or AB interassay) or (TI inter-assay or AB inter-assay) or (TI intraassay or AB intraassay) or (TI intra-assay or AB intra-assay) or (TI interindividual or AB interindividual) or (TI inter-individual or AB inter-individual) OR (TI intraindividual or AB intraindividual) or (TI intra-individual or AB intra-individual) or (TI interparticipant or AB interparticipant) or (TI inter-participant or AB inter-participant) or (TI intraparticipant or AB intraparticipant) or (TI intra-participant or AB intra-participant) or (TI kappa or AB kappa) or (TI kappa's or AB kappa's) or (TI kappas or AB kappas) or (TI repeatab* or AB repeatab*) or (TI responsive* or AB responsive*) or (TI interpretab* or AB interpretab*)	
#5	#4 NOT (PT "addresses" OR PT "biography" OR PT "case reports" OR PT "comment" OR PT "directory" OR PT "editorial" OR PT "festschrift" OR PT "interview" OR PT "lectures" OR PT "legal cases" OR PT "legislation" OR PT "letter" OR PT "news" OR PT "newspaper article" OR PT "patient education handout" OR PT "popular works" OR PT "congresses" OR PT "consensus development conference" OR PT "consensus development conference, nih" OR PT "practice guideline") NOT (MH "animals") NOT (MH "humans")	
Publication time: no limitation / Language: no limitation		

PsycINFO via OVID

Searched on 03.03.2019

Search	Query	Records retrieved
#1	((need*.ab. or need*.ti. or needs.sh.)	
#2	((carer* or informal carer* or famil* or caregiv* or relative* or spouse* or child*).ab. or (carer* or informal carer* or famil* or caregiv* or relative* or spouse* or child*).ti. or caregivers.sh. or family.sh.) AND (exp Dementia/ or exp Dementia with lewy bodies/ or exp Presenile Dementia/ or exp Semantic Dementia/ or exp Senile Dementia/ or exp Vascular Dementia/ or exp Alzheimer's Disease/ or (Alzheimer* or dementi* or amenti*).mp..ab. or (Alzheimer* or dementi* or amenti*).mp.).ti.)	1133
#3	(assess* OR evaluate* OR outcome OR outcomes OR identif* OR index OR indices OR instrument OR instruments OR interview* OR inventory OR measure OR measures OR questionnaire OR question-naires OR profile OR profiles OR scale OR scales OR score OR scores OR status OR survey OR surveys or test*).ab. OR (assess* OR evaluate* OR outcome OR outcomes OR identif* OR index OR indices OR instrument OR instruments OR interview* OR inventory OR measure OR measures OR questionnaire OR questionnaires OR profile OR profiles OR scale OR scales OR score OR scores OR status OR survey OR surveys or test*).ti. OR (exp outcome/ OR exp outcomes/ OR exp index/ OR exp instrument/ OR exp instruments/ OR exp inventory/ OR exp questionnaire/ OR exp questionnaires/ OR exp profile/ OR exp scale/ OR exp score/ OR exp survey/ OR exp surveys/)	
#4	#1 AND #2 AND #3 AND (Validation Studies.ti. or Comparative Study.af.) OR (exp Psychometrics/) OR (psychometr*).ti,ab. OR (clinimetr* or clinometr*).tw. OR (exp Psychometrics/ or exp Rating Scales/ or exp Test Construction/ or exp Measurement/ or exp Test Validity/ OR outcome assessment).ti,ab. OR (outcome measure*).tw. OR (observer variation).ti,ab. OR ('Health Status Indicators').ti,ab. OR (reproducib*).ti,ab. OR ('Discriminant Analysis').ti,ab. OR (reliab* or unreliab* or valid* or coefficient or homogeneity or homogeneous or 'internal consistency').ti,ab. OR (cronbach* and (alpha or alphas)).ti,ab. OR (item and (correlation* or selection* or reduction*)).ti,ab. OR (agreement or precision or imprecision or 'precise values' or 'test-retest').ti,ab. OR (test and retest).ti,ab. OR (reliab* and (test or retest)).ti,ab. OR (stability or interrater or inter-rater or intrarater or intra-rater or intertester or inter-tester or intratester or intra-tester or interobserver or inter-observer or intraobserver or intraobserver or intertechnician or inter-technician or intratechnician or intra-technician or interexaminer or inter-examiner or intraexaminer or intra-examiner or interassay or interassay or intraassay or intra-assay or interindividual or inter-individual or intraindividual or intra-individual or interparticipant or inter-participant or intraparticipant or intra-participant or kappa or kappas or repeatab*).ti,ab. OR ((replicab* or repeated) and (measure or measures or findings or result or results or test or tests)).ti,ab. OR (generaliza* or generalisa* or concordance).ti,ab. OR (intraclass and correlation*).ti,ab. OR (discrimina-tive or known group or factor analysis or factor analyses or dimension* or subscale*).ti,ab. OR (multitrait and scaling and (analysis or analyses)).ti,ab. OR (item discriminant or interscale correlation* or error or errors or individual variability).ti,ab. OR (variability and (analysis or values)).ti,ab. OR (uncertainty and (measurement or measuring)).ti,ab. OR (standard error of measurement or sensitiv* or responsive*).ti,ab. OR ((minimal or minimally or clinical or clinically) and (important or significant or detectable)and (change or difference)).ti,ab. OR (small* and (real or detectable) and (change or difference)).ti,ab. OR ('meaningful change').ti,ab. OR ('ceiling effect').ti,ab. OR (floor effect).ti,ab. OR ('item response model').ti,ab. OR (IRT).ti,ab. OR (Rasch).ti,ab. OR ('Differential item functioning').ti,ab. OR (DIF).ti,ab. OR ('computer testing').ti,ab. OR ('item bank').ti,ab. OR ('cross-cultural equivalence').-ti,ab.	
#5	#4 NOT ('addresses':it or 'biography':it or 'case reports':it or 'comment':it or 'directory':it or 'editorial':it or 'festschrift':it or 'interview':it or 'lectures':it or 'legal cases':it or 'legislation':it or 'letter':it or 'news':it or 'newspaper article':it or 'patient education handout':it or 'popular works':it or 'congresses':it or 'consensus development conference':it or 'consensus development conference, nih':it or 'practice guide-line':it).mp. not ('animals'/exp not 'humans'/exp) [mp = title, abstract, heading word, table of contents, key concepts, original title, tests & measures]	
Publication time: no limitation / Language: no limitation		

Appendix II: Excluded studies

Articles excluded on reading full-text

Reference	Reason for exclusion
Reynolds T, Thornicroft G, Abas M, Woods B, Hoe J, Leese M, *et al*. Camberwell Assessment of Need for the Elderly (CANE). Development, validity and reliability	Criteria 1
Nicolaou PL, Egan SJ, Gasson N, Kane RT. Identifying needs, burden, and distress of carers of people with Frontotemporal dementia compared to Alzheimer's disease	Criteria 1, 4
van der Roest HG, Meiland FJM, Comijs HC, Derksen E, Jansen APD, van Hout HPJ, *et al*. What do community-dwelling people with dementia need? A survey of those who are known to care and welfare services	Criteria 1, 4
Orrell M, Hancock GA, Liyanage KCG, Woods B, Challis D, Hoe J. The needs of people with dementia in care homes: the perspectives of users, staff and family caregivers	Criteria 1, 2, 4
Miranda-Castillo C, Woods B, Orrell M. The needs of people with dementia living at home from user, caregiver and professional perspectives: a crosssectional survey	Criteria 1, 4
Fernandes L, Goncalves-Pereira M, Leuschner A, Martins S, Sobral M, Azevedo LF *et al*. Validation study of the Camberwell Assessment of Need for the Elderly (CANE) in Portugal	Criteria 1
Ducharme F, Kergoat M-J, Coulombe R, Lévesque L, Antoine P, Pasquier F. Unmet support needs of early-onset dementia family caregivers: a mixed design study	Criteria 2,4
Ducharme F, Lévesque L, Caron C, Hanson E, Magnusson L, Nolan J *et al*. Mise à l'essai d'un outil d'évaluation des besoins de soutien des proches-aidants d'un parent âgé à domicile: Un outil ayant un potentiel d'application en Europe francophone	Criteria 2
Hanson E, Nolan J, Magnusson L, Sennemark E, Johansson L, Nolan M. COAT: The Carers Outcome Agreement Tool: A New Approach to Working with Family Carers	Criteria 2,4
Toye C, Lester L, Popescu A, McInerney F, Andrews S, Robinson AL. Dementia Knowledge Assessment Tool Version Two: development of a tool to inform preparation for care planning and delivery in families and care staff	Criteria 1
Scholzel-Dorenbos CJ, Arons AM, Wammes JJ, Rikkert MG, Krabbe PF. Validation study of the prototype of a disease-specific index measure for health-related quality of life in dementia	Criteria 1
Smith SC, Lamping DL, Banerjee S, Harwood R Foley B, Smith P *et al*. Measurement of health-related quality of life for people with dementia: development of a new instrument (DEMQOL) and an evaluation of current methodology	Criteria 1
Cooper C, Katona C, Livingston G. Validity and reliability of the brief COPE in carers of people with dementia: the LASER-AD Study	Criteria 1
Stephan A, Mayer H, Renom Guiteras A, Meyer G. Validity, reliability, and feasibility of the German version of the Caregiver Reaction Assessment scale (G-CRA): a validation study	Criteria 1
Chow TW, Pio FJ, Rockwood K. An international needs assessment of caregivers for frontotemporal dementia	Criteria 1,4

(*Continued*)	
Reference	**Reason for exclusion**
Ewing G, Brundle C, Payne S, Grande G. The Carer Support Needs Assessment Tool (CSNAT) for Use in Palliative and End-of-life Care at Home: A Validation Study	Criteria 2
Leong J, Madjar I, Fiveash B. Needs of Family Carers of Elderly People with Dementia Living in the Community	Criteria 1,4
George LK, Fillenbaum GG. OARS methodology. A decade of experience in geriatric assessment	Criteria 1
Francis GM, Munjas BA. Needs of family caregivers and persons with Alzheimer's disease	Criteria 1
Jennings LA, Reuben DB, Evertson LC, Serrano KS, Ercoli L, Grill J *et al.* Unmet needs of caregivers of individuals referred to a dementia care program	Criteria 1, 4
Fortinsky RH, Kercher K, Burant, CJ. Measurement and correlates of family caregiver self-efficacy for managing dementia	Criteria 1
Borson S, Scanlan JM, Sadak T, Lessig M, Vitaliano P. Dementia Services Mini-Screen: a simple method to identify patients and caregivers in need of enhanced dementia care services	Criteria 1
Roud H, Keeling S, Sainsbury R. Using the COPE assessment tool with informal carers of people with dementia in New Zealand	Criteria 1
Mizoguchi T, Iijima S, Niino N, Orimo H. Reliability and validity of a Japanese version of the cost of care index.	Criteria 1, 2
Pérez-Fuentesa M, Gázquez Linaresa JJ, Ruiz Fernández DM, Molero Juradoa M. Inventory of Overburden in Alzheimer's Patient Family Caregivers with no Specialized Training	Criteria 1
Givens JL, Jones RN, Mazor KM, Prigerson HG, Mitchell SL. Development and psychometric properties of the family distress in advanced dementia scale	Criteria 2
Rosness TA, Haugen PK, Gausdal M, Gjora L, Engedal K. Carers of patients with early-onset dementia, their burden and needs: A pilot study using a new questionnaire - Care-EOD	Criteria 4
Keady J, Nolan M. Behavioural and Instrumental Stressors in Dementia (BISID): refocussing the assessment of caregiver need in dementia	Criteria 1
Sato S, Kazui H, Shimizu Y, Yoshida T, Yoshiyama K *et al.* Usefulness of carer-held records to support informal caregivers of patients with dementia who live at home	Criteria 1, 4
Cova I, Travi N, Maggiore L, Cucumo V, Mariani C *et al.* What are the caregivers' needs on dementia care? An integrated qualitative and quantitative assessment	Criteria 4
Aoun, SM, Toye C, Slatyer S, Robinson A, Beattie E. A person-centred approach to family carer needs assessment and support in dementia community care in Western Australia	Criteria 4
Bass DM, Judge KS, Snow AL, Wilson NL, Morgan R, Looman WJ *et al.* Caregiver Outcomes of Partners in Dementia Care: Effect of a Care Coordination Program for Veterans with Dementia and Their Family Members and Friends	Criteria 4
Bass DM, Judge KS, Snow AL, Wilson NL, Looman WJ, McCarthy C *et al.* Negative caregiving effects among caregivers of veterans with dementia	Criteria 4

Reference	Reason for exclusion
(Continued)	
Bass DM, Judge KS, Snow AL, Wilson NL, Morgan RO, Maslow K *et al*. A controlled trial of Partners in Dementia Care: veteran outcomes after six and twelve month	Criteria 4
Amieva H, Rullier L, Bouisson J, Dartigues J-F, Dubois O, Salamon R. Needs and expectations of Alzheimer's disease family caregivers	Criteria 4
Zegelin A, Schnepp W, Riesner C, Herder K. Kurzbericht: Überprüfung der Anwendbarkeit, Praktikabilität und Modalitäten der Implementierung des Assessments CarenapD für den deutschen ambulanten Sektor bei Demenz	Criteria 4
Riesner C. Die Rolle pflegender Angehöriger von Menschen mit Demenz in der Bedarfsbestimmung am Beispiel der Carenap-Studie	Criteria 4
Zwaanswijk M, Peeters JM, van Beek AP, Meerveld JH, Francke AL. Informal Caregivers of People with Dementia: Problems, Needs and Support in the Initial Stage and in Subsequent Stages of Dementia: A Questionnaire Survey	Criteria 4
Sadak T, Wright J, Borson S. Managing Your Loved One's Health: Development of a New Care Management Measure for Dementia Family Caregivers	Criteria 4
Kuhn DR. Caring for relatives with early stage Alzheimer's disease: An exploratory study	Criteria 4
Edelman P, Kuhn D, Fulton BR, Kyrouac GA. Information and service needs of persons with Alzheimer's disease and their family caregivers living in rural communities	Criteria 4
Millenaar JK, Bakker C, Koopmans RTCM, Verhey FRJ. Care needs and experiences with the use of services of people with young onset dementia and their caregivers	Criteria 4
Millenaar JK, Van Vliet D, Bakker C, Vernooij-Dassen MJFJ, Koopmans RTCM, Verhey FRJ, *et al*. The experiences and needs of children living with a parent with young onset dementia: Results from the NeedYD study	Criteria 4
Rosa E, Lussignoli G, Sabbatini F, Chiappa A, Di Cesare S, Lamanna L, *et al*. Needs of caregivers of the patients with dementia	Criteria 4
Pini S, Ingleson E, Megson M, Clare L, Wright P, Oyebode JR. A needs-led framework for understanding the impact of caring for a family member with dementia	Criteria 4

Inclusion criteria recommended by the COSMIN guidelines for systematic reviews of measurement properties:[39] 1) the instrument should aim to measure the construct of interest (= types of intervention(s)/phenomena of interest), 2) the study sample should concern the target population of interest (= types of participants), 3) the study should concern the type of measurement instrument of interest (= self-reported or professionally interviewed), 4) the aim of the study should be the development of a measurement instrument or the evaluation of one or more of its measurement properties (= types of studies).

Articles excluded on critical appraisal

Reference	Reason for exclusion
Hodgson C, Higginson I, Jefferys P. Carers' Checklist: an outcome measure for people with dementia and their carers	Not enough information for critical appraisal, development report not available.

Appendix III: Domain structure, number of items and response options

Instrument acronym	Domain structure / sections (number of items)	response options/format	Range of scores
CADI Charlesworth et al. (2007)[49]	8 dimensions (30 items): 1. Carer's reaction to caring (7 items) 2. Degree of physical help (4 items) 3. Carer–PwD relationship (4 items) 4. Restrictions on social life (4 items) 5. Professional support (3 items) 6. Family support (3 items) 7. Interpersonal demands (3 items) 8. Financial consequences (2 items)	The frequency of experiences was rated on a threepoint scale 1 = 'never applies' 2 = 'sometimes applies' 3 = 'always applies'	Total range from 30 to 90
CARENAP McWalter (1996/1998)[50,51]	Instrument contains 4 sections: 1. Basic and referral information 2. Assessments for the PwD (7 dimensions): health and mobility, self-care and toiletting, social interaction, thinking and memory, behaviour and mental state, house- care, and community living 3. Assessments for the carer (5 dimensions): - Health - Daily difficulties - Support - Rest - Feelings 4. Personal history	**Response option for carers:** yes or no **Response options for PwD:** 'No Need' - they have no difficulty or they manage with no help at all; 'Met Need' - they have difficulties but they manage with the help they are getting; 'Unmet Need', - they need help or more help than they currently get **Types of help for unmet needs:** social stimulation/activity, prompting/supervision, doing tasks for the person, aids and adaptations, specialist assessment, counselling for the person, behaviour management, carer advice/training, don't know	No scoring
CNA-D Wancata et al. (2005)[52] / Kaiser et al. (2005)[53]	18 problem areas and 2 - 6 possible interventions per problem area 1. Lack of information about dementia 2. Lack of information about treatment 3. Lack of information about services 4. Financial burden 5. Legal issues 6. Disappointment caused by the illness, concerns about the patient's future 7. Communication problems and conflicts with the patient 8. Burdened by behavioral problems of the patients 9. Problems caused by crises 10. Not enough time for oneself (including caring for the patient when the relative becomes sick) 11. Social isolation, conflicts within the family 12. Burden caused by dangerous situations 13. Fear of stigmatization and discrimination 14. Feelings of guilt, being blamed 15. Missing nursing skills 16. Difficulties concerning household tasks 17. Burned out or overstrained by care 18. Physical or psychiatric illness of the carer	1. Problem areas (three-point scale): no, mild, moderate or serious problem. 2. Interventions (five-point scale): No need = intervention not needed and not received, Overprovision = intervention not needed, but received, Unmet need = intervention needed, but not received, Partially met need = intervention needed and received, but insufficiently, Met need = intervention needed and sufficiently received	3- and 5-point Likert scales Problems = 3x18 = 18–54 Met needs = 0–18 Unmet needs = 0–18
CNCD Vaingankar et al. (2013, 2017)[54,55]	5 dimensions (26 items): 1. Information (7 items) 2. Access to services (3 items) 3. Availability of services (6 items) 4. Other formal support services (4 items) 5. Informal support (6 items)	Caregivers agreement rated on a **5-point response scale**: I did not need 'X' / Strongly agree / Agree / Disagree / Strongly disagree	Response scale combined to a 3-point scale for analysis: I did not need 'X' = 0 // Strongly agree/ Agree = 1 // Disagree/ Strongly disagree = 2). **Range: 0 - 52** Total score obtained by adding the item scores, with higher scores indicative of **higher unmet need**

(*Continued*)			
Instrument acronym	**Domain structure / sections (number of items)**	**response options/format**	**Range of scores**
JHDCNA Hughes *et al.* (2014)[56]	**4 caregiver domains (12 items):** p. Caregiver Education (3) q. Resource Referrals (5) r. Caregiver Mental Health Care (4) s. Caregiver General Medical / Health Care (3) **15 care recipient domains (77 items):** a. Dementia Evaluation / Diagnosis (6), b. Treatment of Cognitive Symptoms (2), c. Treatment of Neuropsychiatric Symptoms (5), d. Behavior Management (3), e. Medication Management, f. Medication Administration (3), g. General Medical / Health Care (8), h. Allied Health & Nursing Care (5), i. Safety 9, j. Assistance with Daily Activities (10), k. Meaningful Activities (6), l. Legal Issues / Advance Care Planning (6), m. Assistance with Health Insurance (5), n. Patient Education (1), o. Caregiver Availability (3)	1) Clinicians rated items as being needed or not and if needed: 2) whether the need was "unmet" if: (a) = not addressed and potentially beneficial interventions are available or (b) = addressed but potential benefits of available interventions have not yet been achieved. A "fully met" need = addressed and potential benefits of available interventions achieved to the extent possible for the individual.	**Unmet caregiver needs** derived from the JHDCNA included: dementia education (1, 0), skills counseling (1, 0), community resource availability (1, 0), emotional support (1, 0), respite support (1, 0), mental health counseling/psychiatric care (1, 0), primary medical care (1, 0), specialist medical care (1, 0), and other professional medical care (1, 0). Unmet caregiver need for Community Resource Referrals services included: the Alzheimer's Association, eldercare attorney, Office on Aging/ social services, geriatric care management, and adult protective services. The total count of unmet needs was the sum of the five items of this domain

NAS
Wackerbarth *et al.* (2002)[57]
	Instrument contains **3 sections:**
	1. Information needs (**20 items**):
	    a. General information about dementia and providing care
	    b. Diagnosis and treatment
	    c. Legal and financial issues
	2. Support needs (**19 items**):
	    a. Supporting the care receiver
	    b. Interacting with and supporting others
	    c. Finding support for myself
	3. Information about the caregiving experience, socio-demographic information about the respondent and care receiver, and the caregiving duties (activities of daily living)
		Section 1 and 2: Respondents rate the need in terms of importance on a scale from 0 to 3: 0 = not important; 1 = slightly important, 2 = important; 3 = essential. Contains also the response option "doesn't apply to me".	No total score calculated (only means)

PBH-LCI:D Sadak *et al.* (2015)[58]	**6 domains (32 items)** 1. Ability to recognize and to anticipate day-to-day symptoms and challenges with multiple dimensions of patient health (8 items) 2. Ability to manage sudden changes in the dimensions of patient's health, to engage health services and to practice self-care (8 items) 3. Patient's medications (5 items) 4. Ability to manage day-to-day symptoms and challenges with the dimensions of patient's health (4 items) 5. Ability to recognize sudden/worrisome changes in the dimensions of patient's health (4 items) 6. Ability to advocate for patient in health care situations (3 items)	5-level Likert type response scale ranging from (1) 'Disagree Completely' to (4) 'Agree Completely,' with an additional response option (0), 'Not my responsibility.'	Possible total scores can range from 0 to 128, with higher scores indicating higher activation

(*Continued*)			
Instrument acronym	Domain structure / sections (number of items)	response options/format	Range of scores
RAM Czaja et al. (2009)[59]	**6 Domains (16 items) and suggested intervention strategies for each domain** 1. Self-care and healthy behaviours (2 items) 2. Patient Problem behaviours (2 items) 3. Burden (3 items) 4. Depression (1 item) 5. Social Support (2 Items) 6. Safety (6 items)	Different dichotomous or ordinal response options according to the question	Domains 1,2, 3, 4, 6: higher score indicates higher risk Domain 5: lower score indicates less satisfaction with support and higher risk **High risk:** summed all high risk answers for the upper limit of 38–40. 27 is sum of $^1/_2$ high risk and $^1/_2$ medium risk answers. **Moderate risk:** $^1/_2$ moderate answers and $^1/_2$ high-risk answers. **Low risk:** Upper limit of 11 is sum of all moderate risk answers
QCNE Dimakopoulou et al. (2015)[60]	Instrument contains **four sections (19 items):** 1. Socio-demographic parameters (4 items) 2. Caregiving details (5 items) 3. Problems deriving from care (5 items) a. with medical services b. with social services c. with benefits d. with financial, legal issues concerning care recipient e. any other problem, please describe **4. Extra services (5 items)** a. Psychosocial support b. Training/education c. Assistance with everyday life activities d. Information and assistance with financial, legal and other bureaucratic issues e. benefits f. Other	19 closed (yes or no response) and two open ended questions	No scoring, no summary score
EAC Laprise et al. (2001)[61]	**4 domains (28 items):** 1. Receiving information (7 items) 2. Developing skills (7 items) 3. Supporting self-confidence (7 items) 4. Overcoming relational difficulties (7 items)	5-level Likert type response scale ranging from (1) no expectations at all to (5) a lot of expecations (5): not at all, a little, moderately, many, a lot of expectations	Total score: 28–140
QNP Peeters et al. (2010)[62] / Van der Poel and van Beek (2006)[63]	Instrument contains **three sections:** 1. 14 problem areas, that are faced by persons with dementia and problems experienced by informal caregivers (30 items); 2. Importance of these problem areas, according to the informal caregiver; 3. Informal caregiver's needs for additional professional support (56 items). **14 problem areas (30 items):** 1. Feeling that 'something is wrong' 2. What is the matter and what might help? 3. Frightened, angry and confused 4. Having to cope all alone 5. Avoiding contacts 6. Physical care 7. Dangers 8. Health problems too 9. Loss 10. Feeling overwhelmed 11. Reduced say in matters, no say at all 12. In good times and bad 13. Miscommunication with care workers 14. Resistance to admission	Response option part 1 (perceived actual problems): 'yes', 'actually yes', 'actually no', 'no', or 'not applicable' Response option part "2 (importance of problems (2): 'of paramount importance', 'very important', 'important' or'not so important' Response option part 3 (needs for specific types of professional support): list of suggestions for support, directly linked to a particular problem area	No scoring

(*Continued*)			
Instrument acronym	**Domain structure / sections (number of items)**	**response options/format**	**Range of scores**
SIDECAR Oyebode *et al.* (2019)[64]	**70 items** (without explicit division into domains) See Oyebode *et al.* (2018)[64] for all items	Binary response format (Agree/Disagree)	No information found
Tayside Gordon *et al.* (1997)[65]	Instrument contains **three sections (31 items):** **1. Needs (14 items):** a. Mobility (1 item) b. Personal care (2 items) c. Domestic tasks (1 item) d. Behaviour (2 items) e. Solitude (2 items) f. Material needs (3 items) g. Informal carer problems (3 items: coping, practical problems & emotional upset) **2. Current care inputs (6 items):** Informal carer involvement (2 items), Service receipt (2 items), Respite (1 item) and Service refusal by sufferer (1 item) **3. Demographics and background (11 items):** Sufferer demographics (3 items), Evidence of cognitive impairment (2 items), Informal carer demographics (5 items), Informal support to carers (1 item)	Level of commitment for the domains: mobility, personal care, domestic tasks and behaviour needs rated as: * Independent = Attention needed less than once a week * Long interval = Attention needed not more than once a day * Short interval = Attention needed more than once a day at prearranged times * Critical interval = Attention needed at short (< 2h) and/or unpredictable intervals Other response options: frequencies, quantities, yes/no or other dichotomous response options.	Key concept is the use of a descriptive profile; no summative score or categorization
UNM Gaugler *et al.* (2004)[66]	**7 domains (34 items):** 1. Help with ADL tasks (5 items) 2. Help with IADL tasks (6 items) 3. Dementia symptoms (2 items) 4. Timing of care (3 items) 5. Formal support (8 items) 6. Information (4 items) 7. Confidante/family support (6 items)	Questions for eacht item: "Do you need more help with/help providing _____ ?" Response option: yes or no	The "yes" responses for each domain were summed to create unmet need scores

CG: Caregiver, PwD: Person with dementia, ADL: Activities of daily living, IADL: Instrumental activities of daily living, CADI: Carers Assessment of Difficulties Index, CARENAP: The Care Needs Assessment Pack for Dementia, CNA-D: The Carers' Needs Assessment for Dementia, CNCD: Caregivers' Needs Checklist for Dementia, JHDCNA: The Johns Hopkins Dementia Care Needs Assessment, NAS: Needs assessment survey, PBH-LCI:D: Partnering for Better Health e Living with Chronic Illness: Dementia, RAM: Risk Appraisal Measure, QCNE: Questionnaire of carers needs evaluation, EAC: Questionnaire consultation expectations, QNP: Questionnaire National Dementia Programme Survey Needs and problems of informal caregivers of persons with dementia, SIDECAR: Scales measuring the Impact of DEmentia on CARers, Tayside: Tayside Profile for Dementia Planning, UNM: Unmet need measure.

Appendix IV: Quality criteria for psychometric outcomes

Instrument acronym/ author (year)	Content validity	Internal consistency	Reproducibility		Construct validity
			Agreement (test-retest reliability)	Reliability (inter-rater reliability)	
CADI Charlesworth *et al.* (2007)[49]	? Target population was questioned for additional problems but, despite those mentioned, no changes were made to the items	- Despite adequate factor analysis, 5 out of 8 Cronbach's alphas are below 0.70	0	0	0
CARENAP McWalter (1996/ 1998)[50,51]	+	- Cronbach's alpha assessed for the full scale only AND is 0.56	0	? Simultaneous evaluation of an interviewer and an observer; 87% of Kappa calculated for each of the 30 items >.75	0
CNA-D Wancata *et al.* (2005)[52] / Kaiser *et al.* (2005)[53]	+	? Cronbach's alpha(s) calculated per dimension AND between 0.70 and 0.95; BUT not factor analysis performed	? Evaluated in combination with inter-rater: Two interviews conducted by different persons two weeks apart; Mean Kappa 0.84	? Evaluated in combination with test-retest: Two interviews conducted by different persons two weeks apart; Mean Kappa 0.84	+
CNCD Vaingankar *et al.* (2013, 2017)[54,55]	+	? Factor analysis performed but insufficient sample size (109 for 26 items); Cronbach's alpha(s) calculated per dimension AND between 0.70 and 0.95	0	0	- Only 67% of hypotheses confirmed and low correlations
JHDCNA Hughes *et al.* (2014)[56]	0 No information on procedure	0	0	0	? Hypotheses were vague and numerous, so that only 45% were confirmed
NAS Wackerbarth *et al.* (2002)[57]	+	0	0	0	0
PBH-LCI:D Sadak *et al.* (2015)[58]	+	? Factor analysis performed but insufficient sample size (130 for 32 items); Cronbach's alpha(s) calculated per dimension AND between 0.70 and 0.95;	+ Pearson correlations of 0.76	0	+
RAM Czaja *et al.* (2009)[59]	- Informal caregivers were not involved in item development	- Cronbach's alpha assessed for the full scale only AND is 0.65	0	0	+
QCNE Dimakopoulou *et al.* (2015)[60]	? Description of the concepts is too vague (e.g. wishes, problems with services)	? Factor analysis inadequately performed (separately for each dimension); Cronbach's alpha(s) calculated per dimension AND between 0.70 and 0.95	0	0	0

			Reproducibility		
Instrument acronym/ author (year)	Content validity	Internal consistency	Agreement (test-retest reliability)	Reliability (inter-rater reliability)	Construct validity
EAC Laprise et al. (2001)[61]	- Informal caregivers were not involved in item development	? Cronbach's alpha(s) calculated per dimension AND between 0.70 and 0.95; BUT not factor analysis performed	+ Pearson correlations of 0.63 for the total scale and 0.61 to 0.75 for the 4 subscales	0	+
QNP Peeters et al. (2010)[62]/ Van der Poel and van Beek (2006)[63]	+	? Cronbach's alpha assessed for the full scale only	0	0	0
SIDECAR Oyebode et al. (2019)[64]	+	0	0	0	0
Tayside Gordon et al. (1997)[65]	? Lack of details on the procedure	0	? Time interval is too variable; Weighted Kappa of 0.41 to 0.89 for the 4 subscales	? Time interval is too variable and ratings by worker and carer are not comparable; Weighted Kappa of -0.17 to 0.14 for the 4 subscales	0
UNM Gaugler et al. (2004)[66]	- Informal caregivers were not involved in item development	? 5 out of 6 Cronbach's alphas are adequate BUT factor analysis not performed	0	0	? Hypotheses are too vague and numerous, so that only 33% were confirmed

+: positive rating, ?: indeterminate rating due to doubtful design or method, -: negative rating, 0: no information available
CADI: Carers Assessment of Difficulties Index, CARENAP: The Care Needs Assessment Pack for Dementia, CNA-D: The Carers' Needs Assessment for Dementia, CNCD: Caregivers' Needs Checklist for Dementia, JHDCNA: The Johns Hopkins Dementia Care Needs Assessment, NAS: Needs assessment survey, PBH-LCI:D: Partnering for Better Health e Living with Chronic Illness: Dementia, RAM: Risk Appraisal Measure, QCNE: Questionnaire of carers needs evaluation, EAC: Questionnaire consultation expectations, QNP: Questionnaire National Dementia Programme Survey Needs and problems of informal caregivers of persons with dementia, SIDECAR: Scales measuring the Impact of DEmentia on CARers, Tayside: Tayside Profile for Dementia Planning, UNM: Unmet need measure

Appendix V: Characteristics of included studies

Acronym/instrument name	Authors/study design	Validation sample size	Mean age (SD or R)/ age groups (n)	Distribution of sex, relationship status	Supplementary characteristics	Setting(s)/country
CADI Carers Assessment of Difficulties Index	Charlesworth et al. (2007)[49] Psychometric study	**Internal consistency & structural validity:** CG: n = 232 **Content analysis:** 671 questionnaires	**Internal consistency & structural validity:** CG: M = 68 years (SD = 11.5) **Content analysis:** no information	**Internal consistency & structural validity:** CG: 65% female, 66% spouse, 26% children **Content analysis:** no information		Community setting UK
CARENAP The Care Needs Assessment Pack for Dementia	McWalter et al. (1996, 1998)[50,51] Same samples in both studies Psychometric study / Manual	**Inter-rater reliability:** CG: n = 27 PwD: n = 34 Professionals: n = 28 **Internal consistency:** same sample as for reliability but CG: n = 20 (due to missing data) **Content validity:** PwD: n = 157 Professionals: n = 40	**Internal consistency & inter-rater reliability:** CG: not reported PwD: M = 75.9 years (SD = 6.5) Professionals: not reported **Content validity:** PwD: M = 82.7 years (SD = 8.34) Professionals: not reported	**Internal consistency & inter-rater reliability:** CG: 63% female, 33% daughters, 30% wives, 26% husbands, 7% brothers, 4% son PwD: 70.6% female **Content validity:** PwD: 62% female, 25% male, 12% unidentified	**Internal consistency & inter-rater reliability:** CG: 81% cohabiting with PwD Professionals: 29% community based social work, 71% health agencies **Content validity:** PwD: 71% in NHS wards, 24% in private nursing homes, 6% in residential homes	**Internal consistency & inter-rater reliability:** Community setting **Content validity:** Community setting & Nursing homes Scotland
CNA-D The Carers' Needs Assessment for Dementia	Wancata et al. (2005)[52] Kaiser et al. (2005)[53] Same sample for concurrent validity in both studies Psychometric study	**Content validity:** CG: n = 40 Professionals: n = 40 **Concurrent validity, test-retest, inter-rater reliability & internal consistency:** CG: n = 45,	**Content validity:** CG: M = 51.1 years (SD 13.4) PwD: M = 77.5 years (SD 9.3) **Concurrent validity, test-retest, inter-rater reliability & internal consistency:** CG: M = 60.9 years (SD 11.9) PwD: M = 77.5 years (SD 9.3)	**Conten validity:** CG: 62.5% female, 59.5% offspring **Concurrent validity, test-retest, inter-rater reliability & Internal consistency:** CG: 73% female, 52% offspring and 46% partners of PwD PwD: 38% female	**Content validity:** Professionals: 28.2% medical doctors, 25.6% nurses, 12.8% psychologists, 12.8% social workers, 7.7% occupational therapists Duration prof. experience: M = 13.9 years (SD 9.2) **Concurrent validity, test-retest, inter-rater reliability & internal consistency:** PwD: 72% Alzheimer's, 13% vascular, 15% not diagnosed exactly, Housing situation PwD: 59% home-dwelling, 49% cohabiting with CG, 40% in nursing facility	Community setting, nursing facility Austria
CNCD Caregivers' Needs Checklist for Dementia	Vaingankar et al. (2013)[54] for content validity Vaingankar et al. (2018)[55] for internal consistency, structural validity & hypotheses testing Psychometric study	**Internal consistency, structural validity & hypothesis testing:** CG: n = 109 **Content validity:** GC: n = 63	**Internal consistency, structural validity & hypothesis testing:** CG: M = 55.9 years (SD = 10) PwD: M = 84.9 years (SD = 7.5) **Content validity:** GC: M = 52.9 years (SD not reported) PwD: R = 54–93 years (M, SD not reported)	**Internal consistency, structural validity & hypothesis testing:** CG: 62% female, 12% spouse, 75% child, 13% other relativ/friend PwD: 68.8% female **Content validity:** CG: 60% female 60% children, 20% spouses, 20% other relativ	**Internal consistency, structural validity & hypothesis testing:** CG Education: 7.4% None/minimal, 19.5% complete primary, 43.5% complete secondary, 29.6% completed tertirary CG Employment: 45.8% paid full-time, 12.1% paid part-time, 5.6% unemployed, 0.9% student, 15.9% housewife/husband, 19.6% retired = 19.6% **Content validity:** PwD duration of diagnosis: R = 1 - 23 years	Community setting Singapore

(*Continued*)						
Acronym/instrument name	Authors/study design	Validation sample size	Mean age (SD or R)/ age groups (n)	Distribution of sex, relationship status	Supplementary characteristics	Setting(s)/country
JHDCNA The Johns Hopkins Dementia Care Needs Assessment	Hughes *et al.* (2014)[56] Cross-sectional study	**Hypothesis testing:** CG: n = 246 PwD: n = 254	**Hypothesis testing:** CG: M = 66.13 years (SD 13.33) PwD: M = 83.6 years (SD 5.9)	**Hypothesis testing:** CG: 74.8% female 59.8% non-spouse PwD: 75.2% female	**Hypothesis testing:** PwD: Education M = 13 years, MMSE M = 17.7, PGDRS M = 10.4, NPI-Q M = 7.7, CSDD M = 6.1. CG: white race 68.7%, education M = 15.25 years, duration spent in CG-role M = 39.8 months, employed 48.8%, ZBI score M = 15.1 (SD = 8.5), total hours CGspent for and with participants in the past week M = 91.1 (SD = 77.5)	Community setting Baltimore, US
NAS Needs assessment survey	Wackerbarth *et al.* (2002)[57] Cross-sectional survey	**Content validity:** 15 interviews with CG	No information for sample of content validity (only survey)	No information for sample of content validity (only survey)	No information for sample of content validity only survey)	Community setting Kentucky, US
PBH-LCI:D Partnering for Better Health - Living with Chronic Illness: Dementia	Sadak *et al.* (2015)[58] Psychometric study	**Internal conistency, Test-retest reliabilty, Measurement error, Structural validity, hypothesis testing, Responsiveness:** CG: n = 130 / PwD: n = 130 **Content validity:** 16 clinical dementia experts, 35 informal dementia CG	**Reliabilty & validity testing:** CG: M = 65.9 years (SD = 11.1) PwD: M = 79.3 years (SD = 9.8) **Content validity:** no information	**Reliabilty & validity testing:** CG: 80% female, 63% spouse/domestic partner, 25% child, 12% other relative/ friend PwD: 20% female **Content validity:** no information	**Reliabilty & validity testing:** CG Race: 91% White, 4% Asian, 4%Black or African American, 4% American Indian, CG Education: 77% College graduate, 23% High school graduate, 6% Less then high school, Kingston caregiver stress measure: M = 24 (SD 8.3), CG anxiety (GAD-7): M = 4.6 (SD 3.9) CG depression (PHQ-9): M = 5.39 (SD 4.6) CG quality of life (SF-12): Mental health domain M = 46 (SD = 11), Physical health domain M = 45 (SD = 12) PwD Race: 90% White, 4% Asian, 5% Black or African American, PwD Education: 61% College graduate, 33% High school graduate, 6% Less then high school 6%, Patient's neuropsychiatric symptoms (NPI-Q): Mean = 7.34 (SD = .9)	Community setting Washington/ Minnesota, US

(Continued)

Acronym/instrument name	Authors/study design	Validation sample size	Mean age (SD or R)/ age groups (n)	Distribution of sex, relationship status	Supplementary characteristics	Setting(s)/country
RAM Risk Appraisal Measure	Czaja et al. (2009)[59] Psychometric study	**Internal consistency, hypothesis testing:** CG: n = 642 / PWD: n = 642 **Content validity:** no information	CG: M = 60.6 years (SD = 13.3) PWD: Hispanic: M = 78.2 years (13.9), White/Caucasian: M = 78.8 years (SD = 12.3), Black/ African American: M = 78 (SD = 13.2) **Content validity:** no information	CG: 80% female Hispanic: 39.2% spouse White/Caucasian: 57.5% spouse Black/African American: 29.9% spouse PWD: 58% female **Content validity:** no information	CG: 33% Hispanic, 34% White, 33% Black/African American PWD MMSE: severe cognitive deficits: Hispanic: M = 11.4, White/ Caucasian M = 14.4, Black/African American: M = 11.5 PwD ADL Limitations: substantial functional impairments: Hispanic: M = 3.5, White/Caucasian M = 3.3, Black/African American: M = 3.4. IADL Limitations: Hispanic: M = 6.8, White/ Caucasian M = 6.9, Black/African American: M = 6.8	Community setting 5 geographical areas in the US
QCNE Questionnaire of carers needs evaluation	Dimakopoulou et al. (2015)[60] Cross-sectional survey	**Internal consistency & structural validity:** n = 248 **Content analysis:** Focus groups: 10 informal carers & 10 health care professionals / Further evaluation: 6 informal carers	Internal consistency & structural validity: CG: M = 63.16 years (SD = 14.54)	Internal consistency & structural validity: CG: 78,2% female, 48.4% child, 41.1% spouse, 10.5% other	Years spent in caring: male CG M = 5.29 (SD = 2.92), female CG: M = 4.23 (SD = 2.4) PwD: 69.3% Cohabiting, 30.6% living alone CG Education: 16.9% 1–6 years, 35.5% 7–12 years, 41.1% 13–16 years, 6.5% +17 years	Community setting Greece
EAC Questionnaire consultation expectations [Echelle d'attentes de consultation (EAC)]	Laprise et al. (2001)[61] Psychometric study	**Content validity:** 6 psychologues ou travailleurs sociaux & 18 CG **Internal consistency, test-retest reliabilty, hypothesis testing:** CG: n = 78	CG: 56.4% between 28–58 years, 43.6% + 60 years	CG: 66.6% female, 23% spouse, 77% child	Care receiver: 59% dementia, 16.6% cardiopulmonary diseases, 6.4% cancer 37% living in an institution, 62.8% community setting	Hospital & Nursing home setting Canada (French-speaking part)
QNP Questionnaire National Dementia Programme Survey Needs and problems of informal caregivers of persons with dementia	Peeters et al.(2010)[62] / Van der Poel and van Beek (2006)[63] Using same sample for content validity Cross-sectional survey	**Internal consistency:** CG: n = 984 **Content validity:** Several dementia experts 3 focus groups with CG Comprehensibility and applicability: 15 CG	Internal consistency: CG: M = 62.8 years (SD = 13.3) PwD: 78.2 years (SD = 8.8) **Content validity:** no information	Internal consistency: CG: 71.6% female, 50.8% spouses, 38.2% children PwD: 58% female **Content validity:** no information	Internal consistency: CG: Frequency of care: 54.8% daily, 18.3% 3–6 times/week, 21.9% 1–2 times/week, Duration of caregiving: 23% more than 5years, 20% 2–3 years, 18% 1–2 years PwD: 69% homedwelling (with spouse or child), 31% living in institution, duration of dementia symptoms: 27% more than 5 years, 20% 2–3 years 16%, 1–2 years, 15% 4–5years **Content validity:** no information	Community setting, Nursing homes Netherlands

(*Continued*)						
Acronym/instrument name	Authors/study design	Validation sample size	Mean age (SD or R)/ age groups (n)	Distribution of sex, relationship status	Supplementary characteristics	Setting(s)/country
SIDECAR Scales measuring the Impact of DEmentia on CARers	Oyebode *et al.* (2019)[64]	Content validity: CG: n = 22 In addition six researchers, two CG and a charitable sector carers' organization were included in an initial check of content validity.	Content validity: CG (n = 22) age groups (years): 25−65: n = 12 65−79: n = 8 ≥ 80: n = 2	Content validity: CG (n = 22): 77.3% female, 63.6% spouses, 31.8% children, 4.5% grandchildren	Content validity: CG (n = 22): 72.7% white, 9% Asian, 13.6% black, 4.5% other CG Education: 50% up to 18 years of age or younger, 50% beyond 18 years of age CG Location of home: 95.5% urban, 4.5% rural	Community setting UK
Tayside Tayside Profile for Dementia Planning	Gordon *et al.* (1997)[65] Psychometric study	Inter-rater & test-retest validity: n = 1122 assessments of 805 individuals Content validity: CG: n = 25 semi-structured interviews & 5 exercise diary Professionals: n = not reported	No information	No information	No information	Community, residential and hospital setting, nursing home Scotland
UNM Unmet need measure	Gaugler *et al.* (2004)[66]	Internal conistency & hypothesis testing: CG: n = 344 (community setting) PwD: n = 344 (community setting) Total sample: n = 694 (Community sample = 344, Nursing home = 137, bereaved = 216) Content validity: consultations with clinical and research experts in dementia caregiving	Internal conistency & hypothesis testing: CG community setting: M = 62.23 years (SD = 13.48) PwD community setting: M = 76.36 years (SD = 9.09) Content validity: no information	Internal conistency & hypothesis testing: CG community setting: 69.7% female, 48.3% spouse PwD community setting: 63.6% female Content validity: no information	Internal conistency & hypothesis testing: CG community setting: 35.8% currently employed, 65.1% primary caregiver PwD community setting: 87.2% dementia diagnosis Duration of care: M = 48.09 months (SD = 31.17)	Community and institutional setting Kentucky, US

ADL: Activities of daily living; CADD: Cornell Scale for Depression in Dementia; CG: Caregiver; GAD: General Anxiety Disorder Assessment; IADL: Instrumental activities of daily living; M: Mean; MMSE: Mini-Mental State Examination; NPI-Q: Neuropsychiatric Interview-Questionnaire; PGDRS: Psychogeriatric Dependency Rating Scale; PHQ9: Patient Health Questionnaire; PwD: Person with dementia; SD: Standard deviation; SF-12: Short-Form Health Survey; UK: United Kingdom; US: United States of America; ZBI: Zarit Burden Inventory.

Appendix VI: Instrument characteristics

Instrument acronym	Purpose (research/ clinical)	Target population & setting	Application method	Responden /administrative burden	Languages/country in which the instrument was evaluated
CADI	Clinical & research use	CG of PwD living in the community	Not clearly stated, the authors repeatedly use the term interview, but they never say who conducted the interview, and the presentation format of the questionnaire is suitable for self-administration.	25 minutes to administer	English United Kingdom
CARENAP	Clinical use	CG and PwD living in the community	Professionally interviewed - information gathered from different sources & over a period of two weeks (2–3 visits)	PwD Assessment: 60 minutes to administer CG Assessment: 30 Minutes to administer. For professionals with experience in assessment/ interviewing and PwD.	English Scotland
CNA-D	Research use (service planning)	CG of PwD living in the community or in nursing facilities	Professionally interviewed	50 minutes to administer	German Austria
CNCD	Clinical use	CG of PwD living in the community	Self-administered	No information	English, Chinese, Malay, Tamil Singapore
EAC	Clinical use	CG of older persons living in the community or in an institution	Self-administered	No information	French Canada
JHDCNA	Not specified, was used in a cross-sectional study	CG and PwD living in the community	Clinician rated / professionally interviewed	no information found	EnglishUnited States
NAS	Not specified, was used in a survey	CG of PwD living in the community or in an institution	Self-adminstered	30 minutes	English United States
PBH-LCI:D	Clinical & research use	CG of PwD living in the Community	Self-administered	No information	English United States
RAM	Clinical & research use	CG of PwD living in the community	Not clearly stated, the authors often highlight the ease of administration, suggesting self-report.	5–7 minutes to assess key areas of potential risk	Spanish English United States
QCNE	Not specified (used in a cross-sectional survey)	Informal dementia caregivers	Self-administered	20 minutes to complete	Greek Greece
QNP	not specified (was used in a national survey)	CG of PwD living in the community or in an institution	Self-administered or professionally interviewed	30 minutes Interviewer does not need interview abilities or knowledge	Dutch Netherlands
SIDECAR	Not specified	CG of wliving in the community	Self-administered	No information found	English United Kingdom
Tayside	Clinical use	CG **and** PwD living in the community, residential or hospital setting	Self-administered or professionally interviewed	No information found	English Scotland
UNM	Not specified (was used in a cross-sectional survey)	CG of PwD living in the community, residential or hospital setting	Self-adminstered	No information found	English United States

CADI: Carers Assessment of Difficulties Index (Charlesworth et al., 2007)[49]; CARENAP: The Care Needs Assessment Pack for Dementia (McWalter et al.,1996;1998)[50,51]; CG: Caregiver; CNA-D: The Carers' Needs Assessment for Dementia (Wancata et al., 2005; Kaiser et al., 2005)[52,53]; CNCD: Caregivers' Needs Checklist for Dementia (Vaingankar 2013; 2017)[54,55]; EAC: Questionnaire consultation expectations (Laprise et al. (2001)[61]; JHDCNA: The Johns Hopkins Dementia Care Needs Assessment (Hughes et al., 2014)[56]; NAS: Needs assessment survey (Wackerbarth et al., 2002)[57]; PBH-LCI:D: Partnering for Better Health e Living with Chronic Illness: Dementia (Sadak et al., 2015)[58]; PwD: Person with dementia; QCNE: Questionnaire of carers needs evaluation (Dimakopoulou et al., 2015)[60]; QNP: Questionnaire National Dementia Programme Survey Needs and problems of informal caregivers of persons with dementia (Peeters et al., 2010; Van der Poel and van Beek, 2006)[62,63]; RAM: Risk Appraisal Measure (Czaja et al., 2009)[59]; SIDECAR: Scales measuring the Impact of DEmentia on CARers (Oyebode et al., 2019)[64]; Tayside: Tayside Profile for Dementia Planning (Gordon et al. 1997)[65]; UNM: Unmet need measure (Gaugler et al. (2004)[66].