

# Validity, Reliability, and Feasibility of Clinical Staging Scales in Dementia: A Systematic Review

American Journal of Alzheimer's Disease & Other Dementias®  
26(5) 357-365  
© The Author(s) 2011  
Reprints and permission:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/1533317511418954  
http://aja.sagepub.com



**Marcel G. M. Olde Rikkert, MD, PhD<sup>1</sup>,  
Klodiana Daphne Tona, BCS<sup>1</sup>, Lieneke Janssen, BCS<sup>1</sup>,  
Alistair Burns, MD, PhD<sup>2</sup>, Anatonio Lobo, MD, PhD<sup>3</sup>,  
Philippe Robert, MD, PhD<sup>4</sup>, Norman Sartorius, MD, PhD<sup>5</sup>,  
Gabriela Stoppe, MD, PhD<sup>5</sup>, and Gunhild Waldemar, MD, PhD<sup>6</sup>**

## Abstract

New staging systems of dementia require adaptation of disease management programs and adequate staging instruments. Therefore, we systematically reviewed the literature on validity and reliability of clinically applicable, multidomain, and dementia staging instruments. A total of 23 articles describing 12 staging instruments were identified (N = 6109 participants, age 65-87). Reliability was studied in most (91%) of the articles and was judged moderate to good. Approximately 78% of the articles evaluated concurrent validity, which was good to very good, while discriminant validity was assessed in only 25%. The scales can be applied in  $\pm 15$  minutes. Clinical Dementia Rating (CDR), Global Deterioration scale (GDS), and Functional Assessment Staging (FAST) have been monitored on reliability and validity, and the CDR currently is the best-evidenced scale, also studied in international perspective, and is available in 14 languages. Taking into account the increasing differentiation of Alzheimer's disease in preclinical and prodementia stages, there is an urgent need for global rating scales to be refined as well.

## Keywords

dementia, staging scale, validity, systematic review

## Introduction

Dementia is defined as an acquired syndrome of decline in memory and at least one other cognitive domain, such as language, visuospatial, or executive function, which is sufficiently severe to interfere with social or occupational function in an alert person (*Diagnostic and statistical manual of Mental disorder*, Fourth Edition, Text Revision [DSM-IV-TR]). Alzheimer's disease (AD) is the most common cause, which has a specific pathophysiology and clinical profile, beginning gradually and worsening over several years, thereby creating the notion of progressively passing several stages of severity.<sup>1</sup> In clinical practice, the diagnosis is based on behavioral assessments and cognitive tests that highlight quantitative and qualitative changes in cognitive functions and activities of daily living, which are characteristic of the dementia syndrome and its underlying diseases.

Following diagnosis, depending on the age of onset and comorbidities, the patients usually have a wide range of life expectancies. This range has a median of approximately 5 years, with men having a relatively shorter one than women.<sup>2</sup> Moreover, about 20% of patients with AD show a plateau phase of clinical stability in the course of the disease.<sup>3</sup> Because of this

natural course of the disease, dementia has sometimes been divided into stages (eg, prodementia, mild cognitive impairment, early dementia, moderate dementia, and advanced dementia), but the number and characteristics of these stages vary accordingly based on the assessment scales used. None of these staging systems has been generally accepted.

<sup>1</sup> Department of Geriatrics, Alzheimer Centre Nijmegen, Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands

<sup>2</sup> Department of Psychiatry, University of Manchester, Manchester, United Kingdom

<sup>3</sup> Department of Psychiatry, Hospital Clinico, Universitario and University of Zaragoza. CIBERSAM and Instituto Aragonès de Ciencias de la Salud, Zaragoza, Spain

<sup>4</sup> Department of Psychiatry, Nice University Hospital, Nice, France

<sup>5</sup> University Psychiatric Clinics (UPK), University of Basel, Basel, Switzerland

<sup>6</sup> Memory Disorder Research Unit, Department of Neurology, Neuroscience Centre, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark

## Corresponding Author:

Marcel G. M. Olde Rikkert, Department of Geriatrics, Alzheimer Centre Nijmegen, 925, Radboud University Nijmegen Medical Centre, PO Box, 9101, 6500 HB Nijmegen, Netherlands  
Email: m.olde-rikkert@ger.umcn.nl

Furthermore, intermittently over the years, new scales or alterations and expansions of existing ones are proposed.<sup>4</sup>

Implicit in the philosophy of staging is the gradual progression and deterioration of the clinical syndrome, particularly in AD. In relation to this, 2 distinct types of staging can be identified. First, the staging can aim at describing and monitoring the progress in the biological processes underlying the clinical deterioration. Therefore, a validation of such staging scales should hold this biological process as a gold standard or valid biomarkers as surrogate reference standard, as it has been done in other diseases (eg, chronic obstructive pulmonary disease and its staging by means of forced expiratory volumes). A second type of staging can focus on the clinical and health care needs of the person. Patients in different stages of such a scale have specific needs for care. For instance, at a certain stage, the recommended treatment may be focused toward alleviating symptomatic burdens, while at other stages it may be delaying the progression of the disease.

In order to attain progress in disease management regimens, there is a need for consensus in dementia staging, as both types of staging scales may be used in such programs. In this review, we focus on the clinical staging scales that monitor progress in symptoms and health care needs, while we do not focus on staging scales monitoring neuropathologic progress ("biological staging instruments," ie, staging by means of biomarkers which focus on the structural and biochemical disease process in the brain). Next, when we refer to dementia staging scales, we primarily refer to AD staging scales, which is analogous with the literature on staging instruments. If applicable, other specific underlying diseases will be mentioned.

The current uncertainty and disagreement on different stages of dementia affects existing research strategies, study results, service organization, education concerning dementia, and the way diagnoses and staging of dementia are delivered to individual patient. Confusion about staging may partly explain why health care services that provide care to people with dementia have not yet clearly defined the priority of interventions at different stages of the disease. Next, public and legal arrangements for people with dementia, such as eligibility for driving, voting, and autonomy in health care decision making are often formally determined by the stage in which a patient is judged to be in. Valid staging is becoming even more important as there is a strong trend to establish a diagnosis much earlier, by applying several diagnostic biomarkers. Because of this increasing relevance of dementia staging, we conducted a systematic review to describe the current state of the art and the evidence base of the clinically applied dementia staging instruments.

## Methods

The main objective of this systematic review is to investigate the reliability, validity, and feasibility of the currently available and clinically applicable staging scales for dementia as a syndrome and Alzheimer's dementia as a more specific disease entity. Methodological search filters were used as described

by the Cochrane Effective Practice and Organization of Care (EPOC) group. The EPOC group is a review group of the Cochrane, an international network of people helping health care providers, policy makers, patients, their advocates, and carers and makes well-informed decisions about human health care by preparing and publishing systematic reviews (SRs). The research focuses of the EPOC group are interventions designed to improve the delivery, practice, and organization of health care services (EPOC; [www.epoc.cochrane.org](http://www.epoc.cochrane.org)). Additionally, the snowball method was used to manually identify relevant references from the reference lists of included articles.

## Search Strategy

A search of the relevant literature was carried out in the databases of Medline, PsycInfo, Cinahl, and Cochrane library using general search terms and the following medical subject headings (1) "dementia," (2) "Alzheimer's disease," (3) "(staging) scales/inventories (of dementia)," (4) "geriatric/patient assessment," (5) "severity of illness/(disorders)," (6) "clinical assessment tools," (7) "(multidimensional) scaling (testing)," and (8) "(rating) scales." Bibliographies of identified articles were hand searched for further relevant references (snowball method).

## Selection

A selection was made from all titles and abstracts that were found in the databases using the following inclusion criteria: empirical studies (prospective data collection); English language (articles in other language were registered); and participants with cognitive impairment, dementia, or AD (Table 1 and Figure 1). Studies using only biomarkers and/or only neuroimaging techniques were out of the scope of this article and were excluded from further analysis.

A staging instrument was defined as an instrument developed to assess disease progression in dementia patients and to position them in one of several stages or somewhere along the timeline that characterizes the (natural) course of the disease. As dementia is affecting several domains (cognition, behavior, autonomy, functional performance, and health care need), staging systems or scales should cover at least more than one of these domains (characterized as global staging scales). This article examines global staging scales used in dementia. Scales covering only one domain are generally referred to as clinimetric rating scales for this specific domain and thus were not considered as dementia staging scales. Additionally, articles that examined the validity of global scales only in one domain (eg, only cognition) are out of the scope of this review and were excluded from further analysis. No other restriction was placed on the subjects regarding age, population, and whether the patients were institutionalized or not.

The publications that met the aforementioned criteria were studied independently by 2 authors (L.J. and K.D.T.) and reviewed by a third author (M.O.R.). Next, all authors were

**Table 1.** Inclusion and Exclusion Criteria for the Systematic Review

Selection Criteria	Inclusion Criteria	Exclusion Criteria
Language	English	Non-English
Study type	Empirical studies (prospective data collection)	Neuropathological studies that only use biomarkers
	Covering more than 1 domain of the subsequent domains, eg, cognition, behavior, autonomy, functional performance, and health care need	Studies that only use neuroimaging techniques
		Studies of specific subtypes of dementia
		Publication prior to 1980
Population	Human studies (all ages) Disease status: evidence of decline and fulfilling dementia syndrome criteria and/or Alzheimer's disease criteria	

consulted until consensus was reached. No publications prior to 1980 were taken into account, because the criteria used for diagnosing dementia prior to 1980 substantially differ from those currently in use.<sup>5</sup> Articles were selected until the end of 2009.

### Selection Procedure

The systematic search strategy resulted in 963 articles, of which 23 (2.4%) fulfilled all inclusion and exclusion criteria as summarized in Table 1 and Figure 1. A total of 809 were excluded because of duplicate searches in the 4 databases (Medline, PsycInfo, Cinahl, and Cochrane library) or because based on their abstract they did not fulfill the criteria mentioned above. This led to the retrieval of 154 potentially relevant articles that were screened in full-text from these articles: 133 studies did not meet all the inclusion and exclusion criteria (most were scales examining only one domain in dementia). Studies that did not fulfill or fulfilled only partially the above-mentioned criteria were not included in the analysis, leading to the inclusion of 21 articles. The reference list of these 21 articles were screened manually and led to 2 additional relevant articles (snowball method). In the end, 23 articles were included in the review.

### Important Properties of Staging Instruments

Important properties of the staging instruments included were first of all their discriminatory validity, that is the power to validly discriminate distinct groups of patients that are accepted to be differently impaired in cognition, behavior, and functional performance, and thus in need of help. Second, their predictive validity would have to be studied, which means that the power to estimate the time that the patient will remain in a specific stage is evidenced by empirical data, and that the scales are able to predict the next dementia phase.

### Methodological Quality Assessment

All articles that fulfilled the inclusion and exclusion criteria were reviewed. A quality assessment of articles per scale was carried out. Research quality was judged for risk of bias and a table of quality assessment was made according to the *Cochrane Handbook For Systematic Reviewing*,<sup>6</sup> as far as this was applicable to nontherapeutic research. These quality criteria were complemented by the evidence-based criteria for research quality in diagnostic and prognostic studies.<sup>7,8</sup>

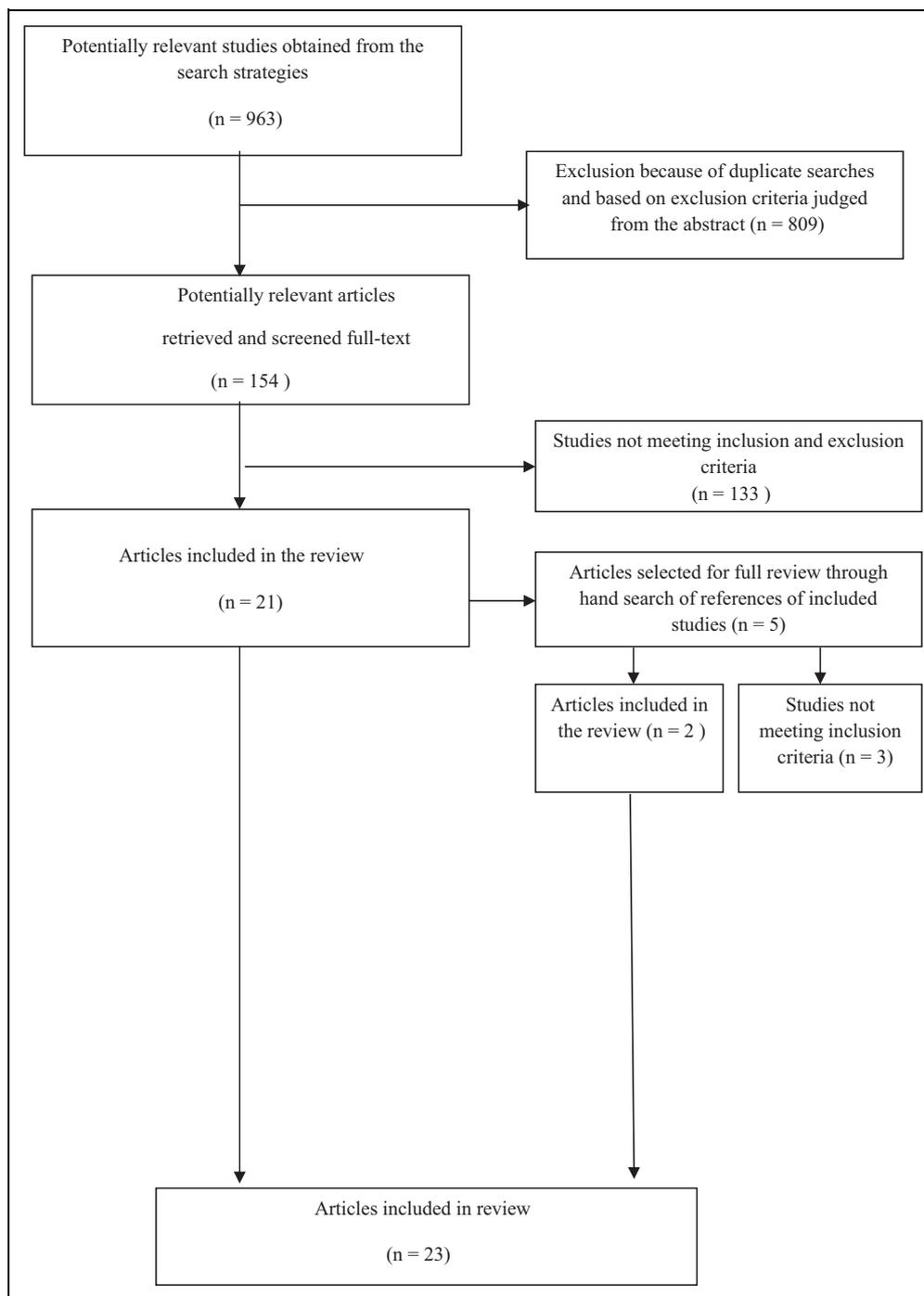
### Analysis

Due to the complexity and diversity of the review results, meta-analysis was considered inappropriate. Therefore, results of included studies were analyzed by making qualitative summaries.

### Results

The systematic search strategy resulted in 963 articles, of which 23 (2.4%) fulfilled all inclusion and exclusion criteria as summarized in Table 1. These articles<sup>9-31</sup> can be found on Supplement Table 1. The total number of participants included in 23 articles was 6109 and their mean age ranged from 65 to 87. In the selected articles, we could identify 12 different staging instruments (Table 2). Most of the scales have different versions for different languages, describe a different number of stages (eg, Clinical Dementia Rating [CDR] scale) often with extension of the original scale to more severe stages,<sup>14</sup> or have different ways of data acquisition (interview vs combining test scores; eg, CDR and the Machine Learning [ML] version).<sup>19</sup> All scales but one has English as a primary language.

The scales make different distinctions in domains of disease progress, ranging from 3 (Global Assessment of Dementia [GAD]) to 20 domains (Hierarchic Dementia scale [HDS]; see Supplement Table 1 for these scale details). The staging scales are to be considered as categorical (nominal) or ordinal scales. No staging scale fulfills the criteria of an interval or continuous scale, which probably is logically impossible, according to the definition of staging scales. Remarkably, the Bedford Alzheimer Nursing Severity scale (BANS-S) and the Direct Assessment of Functional Status (DAFS) do not use specifically named stages but refer to staging linked to scores ranging from 7 to 28 and 0 to 86, respectively.<sup>9,31</sup> Two other staging scales, namely Gottfries-Brane-Steen Scale (GBS) scale and Dementia Severity scale (DSS), do not present unequivocally defined dementia stages.<sup>23,28</sup> Only 6 of the 23 studies were carried out in the last 10 years, which resulted in 1 scale being developed in the last decade (DSS). An important difference between the Global Deterioration scale (GDS),<sup>29</sup> the CDR ML version,<sup>19</sup> and the DAFS on one hand versus the other scales on the other hand is that only these 3 count scores from formalized cognitive or functional performance testing. The other scales rely on history taken both from the patient



**Figure 1.** Flow diagram of selection of empirical research articles on clinical dementia staging.

and/or the caregiver and may be completed merely by information available by professional observations during an interview or during delivery of care.

### Reliability and Validity

Reliability and validity were only tested using observational clinical data in most of the articles included, while only 3 studies validated the clinical staging scales with autopsy or by comparing the staging data with results from biomarker studies

(Supplement Table 1). The first is the BANS-S, in which psychometric properties of BANS-S were related to neuropathological findings obtained from brain autopsies. BANS-S belongs to the category of staging instruments that are useful only for late stages of dementia. According to Volicer et al,<sup>11</sup> the BANS-S score that was determined within 3 months before death correlated significantly with the density of neurofibrillary tangles in the hippocampus ( $r = .443$ ,  $P = .038$ ). Second, the GDS, developed by Reisberg and colleagues,<sup>30</sup> was validated twice against biomarkers.<sup>32,33</sup> These studies indicated that GDS

**Table 2.** Staging Scales Overview

Staging Scales	References
Bedford Alzheimer Nursing Severity Scale (BNAS-S)	Appollonio et al <sup>9</sup>
Clinical Dementia Rating (CDR)	Burke et al <sup>12</sup>
Machine Learning (ML) analysis of CDR	Shankle et al <sup>19</sup>
CDR Sum of Boxes Scores	O'Bryant et al <sup>20</sup>
Clinical Global Impression (CGI) scale	Dahlke et al <sup>21</sup>
Dementia Severity Rating scale (DSRS)	Clark et al <sup>22</sup>
Dementia Severity scale (DSS)	Harvey et al <sup>23</sup>
Direct Assessment of Functional Status (DAFS)	Zanetti et al <sup>24</sup>
Functional Assessment Staging (FAST)	Scian and Reisberg <sup>25</sup>
Functional Rating scale (FRS)	Lanctot et al <sup>26</sup>
Global Assessment of Dementia (GAD)	Ashford et al <sup>27</sup>
Gottfries-Brane-Steen Scale (GBS)-scale	Gottfries et al <sup>28</sup>
Global Deterioration scale (GDS)	Eisdorfer et al <sup>29</sup>
Hierarchic Dementia scale (HDS)	Cole et al <sup>31</sup>

correlated significantly with computerized tomography (CT) scan rankings of ventricular dilation ( $r = .62$ ) and with CT scan-based assessments of sulcus enlargement ( $r = .53$ ). Third, positron-emission tomography (PET) showed that decreased (from baseline/when compared with “normal”) glucose utilization in the caudate nucleus, thalamus, and temporal lobe/cortex significantly correlated with CDR scores ( $r = .69-.83$ ,  $P < .05$ ). Nevertheless, the number of participants on which these findings were based was rather small ( $N = 43-77$ ).<sup>32,33</sup>

Reliability was defined as the repeatability or consistency of the results over time or across different raters: if a staging effort would be repeated (many times) a reliable staging instrument gives (nearly) identical results. Only the studies on the GDS and the Functional Rating scale (FRS)<sup>26,29,30</sup> did not test a single aspect of reliability (Supplement Table 2). The BANS-S, the Dementia Severity Rating scale (DSRS), the Functional Assessment Staging (FAST), and the HDS are best studied for reliability and show good to excellent results on intrarater and interrater reliability.

Validity expresses how close a staging score is to the actual (true) value, which in this context is the actual stage of the disease. As the disease stage was rarely studied from pathophysiological perspective, it remains uncertain whether the stages really reflect, qualitatively and quantitatively, different biological stages in the dementia disease course. Validity was mostly studied by expressing concurrent validity between different staging scales (in 18 of 23 studies, 78%) and by comparing the correlations of the staging scale with other cognitive and functional scales (see Supplement Table 2 for details on this validity testing). On average, concurrent validity was good, in comparison with the cognitive scales such as Mini-Mental State Examination (MMSE), Cambridge Cognitive impairment test (CAMCOG), and Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS-Cog) (Table 3). Discriminant validity was most often assessed by testing whether the staging results were statistically different across distinct groups of patients within the global study population. These distinct groups were most often defined by combining test scores on

cognitive or functional performance tests or by subdividing populations using another staging instrument. Discriminant validity was only studied in 25% of the articles, and on average resulted in good or very good discriminant validity (Table 3). The BANS-S and CDR scored highest on validity testing.

### Feasibility

Feasibility was primarily evaluated by measuring the time needed to complete the staging procedure and counting the number of languages per scale. This time varied from 2 to 40 minutes, with an average of 15 minutes per scale. Only the CDR was studied in a considerable number of languages ( $N = 14$ ). Some scales were specifically adapted to be valid and feasible for severe dementia (eg, BANS-S and CDR extension). There were also scales, in which the concurrent validity was tested in multiracial populations (eg, Chinese, Malays, and Indians) within a single language domain (ie, Chinese<sup>15</sup>), which may be regarded as important value for feasibility especially for application of a scale in an international perspective.

### Discussion

This systematic review showed that at least 12 clinical dementia staging instruments have been developed in the last 30 years, although no single-staging instrument is complete in the sense that it is excellently validated, shows good reliability, is applicable in the entire course of the disease, and is most widely applied in cross-cultural perspective. In this study, we did not focus on biomarker-based staging scales, because this requires a specific systematic review. We focused on functional phenotype-based staging scales, most directly meeting clinical questions. The current evidence for staging instruments is strong for intrarater and interrater reliability, which is moderate to good for most staging instrument, For concurrent validity with other scales for cognition and functional performance, reliability evidence is good to very good.

Formally, correlation testing is not allowed because the staging scales by definition do not fulfill criteria of interval or continuous scales. However, this concurrent validity testing is often used in clinimetric scales, though this theoretical precondition is similarly violated (eg, in studies on the MMSE, which also is not an interval scale).

The instruments that were retrieved are sufficiently feasible, as the time needed to administer is short (on average 15 min) and administration of the scale is easily taught to the personnel/administrator. Although only the CDR was validated in many languages, many other scales were also translated into other languages and could be found in different languages. However, as far as we could retrieve them in this systematic search, the majority of these translations are not empirically studied on reliability and validity.

### Validity of Staging Instruments

The evidence base for staging instruments is not yet sufficiently strong for discriminant validity, and responsiveness to

**Table 3.** Psychometric Characteristics of Dementia Staging Scales<sup>a,b</sup>

Scales	Reliability			Validity		Feasibility Time Required/ Number of Languages	Reference
	Interrater	Intrater	Internal Consistency	Concurrent	Discriminant		
1. Bedford Alzheimer Nursing Severity scale (BANS-S)	++++	+++	-	++	-	-/2	Appollonio et al <sup>9</sup> Bellelli et al <sup>10</sup> Volicer et al <sup>11</sup>
2. Clinical Dementia Rating (CDR)	+++	-	-	-	-	20 to 40 min/14	Burke et al <sup>12</sup>
	+++	-	-	+++	++++	-/1	Chaves et al <sup>13</sup> Dooneif et al <sup>14</sup>
	-	-	-	++	-	-/1	Lim et al <sup>15</sup>
	+++	-	-	+	-	-/1	McCulla et al <sup>16</sup>
	-	-	-	+++	-	-/1	Hughes et al <sup>17</sup>
	-	-	-	+++	-	-/1	Juva et al <sup>18</sup>
	+++	-	-	-	-	-/1	McCulla et al <sup>16</sup>
	-	-	-	+++	-	-/1	Hughes et al <sup>17</sup>
	-	-	-	+++	-	-/1	Juva et al <sup>18</sup>
CDR using Machine Learning (ML)	-	-	-	-	++++ (diagnostic validity)	-/1	Shankle et al <sup>19</sup>
CDR Sum of Boxes	-	-	-	-	++++ (diagnostic validity)	-/1	O'Bryant et al <sup>20</sup>
3. Clinical Global Impression (CGI) scale	+	++	-	-	-	-/1	Dahlke et al <sup>21</sup>
4. Dementia Severity Rating scale (DSRS)	++++	+++	++++	-	-	-/1	Clark et al <sup>22</sup>
5. Dementia Severity scale (DSS)	-	++++	++++	++	++	10 min/1	Harvey et al <sup>23</sup>
6. Direct Assessment of Functional Status (DAFS)	++++	-	++++	++	-	-/2	Zanetti et al <sup>24</sup>
7. Functional Assessment Staging (FAST)	++++	++++	-	++++	-	2 min/1	Scian and Reisberg <sup>25</sup>
8. Functional Rating scale (FRS)	-	-	-	+++	-	15 min/1	Lanctôt et al <sup>26</sup>
9. Global Assessment of Dementia (GAD)	-	-	-	++++	-	-/1	Ashford et al <sup>27</sup>
10. Gottfries-Brane-Steen Scale (GBS) scale	++++	-	-	+++	-	30 min/1	Gottfries et al <sup>28</sup>
11. Global Deterioration scale (GDS)	-	-	-	-	-	2 min/1	Eisdorfer et al <sup>29</sup>
	-	-	-	++	-	-/1	Reisberg et al <sup>30</sup>
12. Hierarchic Dementia scale (HDS)	+++	+++	++++	+++	-	15 to 30 min/1	Cole et al <sup>31</sup>

<sup>a</sup> Per study, according to Quality Assessment Criteria. Adapted from Streiner.<sup>8</sup>

<sup>b</sup> -, data not available. Interrater reliability: Intraclass correlation (ICC) tested but < 80%: +; ICC > 80%: +++; ICC > 90%: +++++. Intrater reliability: Kappa (K) < 0.20: -; K = 0.20-0.40: +/-; K = 0.40-0.60: ++; K = 0.60-0.80: +++; K ≥ .80: +++++. Internal consistency: Cronbach's  $\alpha$  < .70: -; Cronbach's  $\alpha$  > .70: +++++. Concurrent validity: correlation < 0.30: +; 0.30-0.50: ++; > 0.50: +++; > 0.70: +++++. Discriminant validity: weak: +; moderate: ++; strong: +++ (no uniform measure). Cutoffs: according to Streiner.<sup>8</sup>

therapeutic interventions is only rarely studied. From a statistical and methodological point of view, one may conclude that concurrent validity based on other (staging) scales of insufficiently documented validity should be interpreted as "circular reasoning." Additionally, a fundamental question arising is whether the current staging instruments, almost all based on pathophysiological data and disease models more than 10 years old, still have sufficient content validity in this era of modern diagnostic biomarker testing (eg, magnetic resonance imaging [MRI], cerebrospinal fluid [CSF], and PET) and earlier

detection of AD. Next, one can also debate the validity of the domain descriptions presented in the scales.

Currently, we know that dementia progression over time varies from dementia trajectories, with stability for years to rapidly progressive AD.<sup>3,34</sup> In addition, there is growing evidence that Alzheimer dementia (especially late onset AD) is a complex disease based on multicausality, with many contributing factors that cause the diversity in the clinical course of the disease and in the order and clustering of symptoms.<sup>35-37</sup> Furthermore, AD among the elderly persons may have different

etiological factors and thus different stages, when compared with AD among younger person.<sup>38-40</sup> This all contributes to the notion that at present, dementia and the different subtypes can be much better characterized in their course of disease, by addressing the increasing interindividual variability in pattern and rate of progression. Developing a system of staging that fits this knowledge base, but still is applicable to clinical practice, poses a serious challenge on capturing this variability in a prognostic system that still allows comparisons among individuals and valid predictions of future changes.<sup>41</sup> The need for increased precision and accuracy of staging is partly addressed already by the definition of a variety of syndromes together called "mild cognitive impairment," in which amnesic, vascular, and other subtypes are distinguished.<sup>42</sup> Correct staging depends on the outcome of the discussion on the definition of early (prodromal) AD based on biomarker positivity.

### Future Research

Based on the results of this review, one can draw some conclusions on what a future research agenda should include. First of all, future research should concentrate and describe in detail the diverse characteristics and needs of the patients. It is important mentioning that in a number of reliability and validity articles, the population was not well described. In some cases, even basic demographics such as age and education were not mentioned; therefore, the reader cannot tell whether different scales were validated in comparable cohorts. This is of crucial importance in future research, as the disease process can now be described with a growing number of biomarkers. Linked to the neurobiological staging a clinical staging is needed, as it is widely known and accepted that the disease severity of the brain does not parallel the functional and clinical performances of the individual patients.

A second point in the research agenda should be the validation of staging scales for the new concepts of AD. Recently, new techniques have come available and influence the way clinicians diagnose and approach first stages of AD. This fact urges for incorporation of these changes in the staging instruments and therefore requires extension and improvement of validity studies or even creation of new scales.

A last point to fuel the future research agenda is the use of biomarkers in staging scales. In a clinical and research environment, in which heterogeneity in dementia is increasingly disentangled, there seems less and less place for global staging scales that lump together this heterogeneity. On the contrary, it seems more in line with this scientific progress, to develop and validate a multidimensional staging instrument, in which progression of the disease can be assessed separately on all domains affected, with sufficient responsiveness to detect small changes in the course of the disease. Biomarkers and repeated imaging could constitute additional domains incorporated in such a complex staging instrument. Unfortunately, for the time being, biomarker candidates have not managed to reflect subtle changes in clinical phenotypes, still necessitating both biomarker-based and phenotype-based staging instruments.

### Applicability in Health Care Planning

In our aging societies, dementia and specifically AD is becoming one of the major threats for quality of life in old age, for societal stability in health care planning and financial sustainability of health care. To plan labor forces and health care services and to define dementia treatment programs, physicians, managers, policy makers, and governments still need global staging instruments to predict the need for professional support and medical care. To be able to fulfill this target, clinical staging instruments should group and order stages of different levels of disease burden and thus different care needs. However, the current staging scales are not validated on such discriminating properties. Moreover, both on group and individual level, comorbidity in dementia increasingly affects the disease burden of individuals; this is not covered at all by the currently available staging scales. In order to fill the gap of health care planning for older participants, staging scales will have to be redefined and revalidated. Premorbid characteristics such as socioeconomic level, educational background, stability of the social network, and overall biomedical frailty and comorbidity should be considered as possible items in such staging scales.

So far, the discussion on staging instruments mostly addresses the dementia care challenge from the perspective of the Western economies. However, aging is also entering the developing world, which poses largely different questions on predicting the course of dementias for these societies and their families. Quick, simple, but still valid dementia staging scales are highly needed in rural areas and may fit best to the priorities of such societies. No specific reference was made to these applications in the studies retrieved (eg, by validation in Africa or South-East Asia).

### Conclusions

Not only dementia itself and its underlying diseases are increasingly shown to be complex in pathophysiology but also dementia care is becoming more and more complex and heterogeneous, especially when considered from an international perspective. To be able to still fruitfully apply global staging instruments for dementia, these instruments should be very carefully selected and developed after assessing the exact purpose of staging and the populations aimed at.

Multidomain global dementia staging is sensible only for global clinical purposes, but for most other aims of staging, such as monitoring therapeutic response and supplying adequate care, more detailed biological staging instruments implementing modern medical techniques are required. Considering the evidence base for staging instruments, the scenery seems ready for the development of fundamentally different dementia staging instruments, with a dichotomy of clinically and biologically validated staging scales. The new clinical staging scales should include the staging of early cognitive, social, and/or occupational difficulties. Studies on new staging methods should directly focus on the evaluation of staging scales in disease management programs.

## Authors' Note

The authors M.O.R., A.B., A.L., P.R., N.S., G.S., and G.W. are members of the European Dementia Consensus Network (EDCON).

## Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

## References

1. Perneczky R, Wagenpfeil S, Komossa K, Grimmer T, Diehl J, Kurz A. Mapping scores onto stages: mini-mental state examination and clinical dementia rating. *Am J Geriatr Psychiatry*. 2006;14(2):139-144.
2. Larson EB, Shadlen MF, Wang L, et al. Survival after initial diagnosis of Alzheimer disease. *Ann Intern Med*. 2004;140(7):501-509.
3. Bozoki AC, An H, Bozoki ES, Little RJ. The existence of cognitive plateaus in Alzheimer's disease. *Alzheimers Dement*. 2009;5(6):470-478.
4. Burns A, Lawlor B, Craig S, eds. *Assessment Scales in Old Age Psychiatry*. London, UK: Martin Dunitz; 1999:30-34.
5. Zuidema SU, van Iersel MB, Koompan RT, Verhey FT, Olde Rikkert MG. Efficacy and adverse reactions of anti-psychotics for neuropsychiatric symptoms in dementia: a systematic review. *Ned Tijdschr Geneesk*. 2006;150(28):1565-1573.
6. Higgins JPT, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2 [updated July 2010]*. The Cochrane Collaboration; 2009. www.cochrane-handbook.org
7. Sackett DL, Strauss SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence Based Medicine*. 5th ed. New York: Churchill-Livingstone; 2008.
8. Streiner DL. Clinimetrics vs. psychometrics: an unnecessary distinction. *J Clin Epidemiol*. 2003;56(12):1142-1145.
9. Appollonio I, Gori C, Riva G, et al. Assessing early to late stage dementia: the TSI and BANS-S scales in the nursing-home. *Int J Geriatr Psychiatry*. 2005;20(12):1138-1145.
10. Bellelli G, Frisoni GB, Bianchetti A, Trabucchi M. The bedford Alzheimer nursing severity scale for the severely demented: validation study. *Alzheimer Dis Assoc Disord*. 1997;11(2):71-77.
11. Volicer L, Hurley AC, Lathi DC, Kowall NW. Measurement of severity in advanced Alzheimer's disease. *J Gerontol*. 1994;49(5):223-226.
12. Burke WJ, Houston MJ, Boust SJ, Roccarorte WH. Use of the Geriatric depression scale in dementia of the Alzheimer type. *J Am Geriatr Soc*. 1989;37(9):856-860.
13. Chaves MLF, Camozzato AL, Godinho C, et al. Validity of the clinical dementia rating scale for the detection and staging of dementia in Brazilian patients. *Alzheimer Dis Assoc Disord*. 2007;21(3):210-217.
14. Dooneief G, Marder K, Tang MX, Stern Y. The clinical dementia rating scale: community-based validation of "profound" and "terminal" stages. *Neurology*. 1996;46(6):1746-1749.
15. Lim WS, Chin JJ, Lam CK, Lim PP, Sahadevan S. Clinical dementia rating: experience of a multi-racial Asian population. *Alzheimer Dis Assoc Disord*. 2005;19(3):135-142.
16. McCulla MM, Coats M, Van Fleet N, Duckek J, Grant E, Morris JC. Reliability of clinical nurse specialists in the staging of dementia. *Arch Neurol*. 1989;46(11):1210-1211.
17. Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. *Br J Psychiatry*. 1982;140:566-572.
18. Juva K, Sulkava R, Erkinjuntti T, Ylikoski R, Valvanne J, Tilvis R. Staging the severity of dementia: comparison of clinical (CDR, DSM II-R), functional (ADL, IADL) and cognitive (MMSE) scales. *Acta Neurol Scand*. 1994;90(4):293-298.
19. Shankle WR, Mani S, Dick MB, Pazzani MJ. Simple models for estimating dementia severity using machine learning. *Stud Health Technol Inform*. 1998;52(pt 1):472-476.
20. O'Bryant SE, Waring SC, Cullum CM, et al. Staging dementia using clinical dementia rating scale sum of boxes scores: A Texas Alzheimer's research consortium study. *Arch Neurol*. 2008;65(8):1091-1095.
21. Dahlke F, Lohaus A, Gutzmann H. Reliability and clinical concepts underlying global judgments in dementia: implications for clinical research. *Psychopharmacol Bull*. 1992;28(4):425-432.
22. Clark CM, Ewbank DC. Performance of the dementia severity rating scale: a caregiver questionnaire for rating severity in Alzheimer disease. *Alzheimer Dis Assoc Disord*. 1996;10(1):31-39.
23. Harvey PD, Moriarty PJ, Kleinmann L, et al. The validation of a caregiver assessment of dementia: the dementia severity scale. *Alzheimer Dis Assoc Disord*. 2005;19(4):186-194.
24. Zanetti O, Frisoni GB, Rozzini L, Bianchetti A, Trabucchi M. Validity of direct assessment of functional status as a tool for measuring Alzheimer's disease severity. *Age Ageing*. 1998;27(5):615-622.
25. Sclan SG, Reisberg B. Functional assessment staging (FAST) in Alzheimer's disease: reliability, validity, and ordinality. *Int Psychogeriatr*. 1994;4(suppl 1):55-69.
26. Lanctôt KL, Hsiung GR, Feldman HH, Masoud ST, Sham L, Heemann N. Assessing the validity of deriving clinical dementia rating (CDR) global scores from independently-obtained functional rating scale (FRS) scores in vascular dementia with and without Alzheimer's disease. *Int J Geriatr Psychiatry*. 2009; 24(10):1174-1176.
27. Ashford J, Kumar V, Barringer M, et al. Assessing Alzheimer severity with a global clinical scale. *Int Psychogeriatr*. 1992;4(1):55-74.
28. Gottfries CG, Brane G, Gullberg B, Steen G. A new rating scale for dementia syndromes. *Arch Gerontol Geriatr*. 1982;1(4):311-330.
29. Eisdorfer C, Cohen D, Tang MX, et al. An empirical evaluation of the global deterioration scale for staging Alzheimer's disease: correction. *Am J Psychiatry*. 1992;149(2):1129.
30. Reisberg B, Ferris SH, de Leon MJ, Crook T. The global deterioration scale for assessment of primary degenerative dementia. *Am J Psychiatry*. 1982;139(9):1136-1139.

31. Cole MG, Dastoor DP. The Hierarchic dementia scale: conceptualization. *Int Psychogeriatr*. 1996;8(2):205-212.
32. Ferris SH, de Leon MJ, Wolf AP, et al. Positron emission tomography in the study of aging and senile dementia. *Neurobiol Aging*. 1980;1:127-131.
33. Schmidt C, Redyk K, Meissner B, et al. Clinical features of rapidly progressive Alzheimer's disease. *Dement Geriatr Cogn Disord*. 2010(4);29:371-378.
34. Olde Rikkert MG, Teunisse JP, Vernooij-Dassen M. One hundred years of Alzheimer's disease and the neglected second lesson of Alois Alzheimer on multicausality in dementia. *Am J Alzheimers Dis Other Demen*. 2005;20(5):269-272.
35. Olde Rikkert MG, van der Flier WM, de Leeuw FE, et al. Multiple diagnostic tests are needed to assess multiple causes of dementia. *Arch Neurol*. 2006;63(1):144-146.
36. Shineman DW, Fillit HM. A multi-targeted approach for a complex multifaceted disease. *Curr Alzheimer Res*. 2009;6(5):407-408.
37. Savva GM, Wharton SB, Ince PG, Forster G, Matthews FE, Brayne C. Medical research council cognitive function and ageing study. Age, neuropathology, and dementia. *N Engl J Med*. 2009;360(22):2302-2309.
38. Haroutunian V, Schnaider-Beeri M, Schmeidler J, et al. Role of the neuropathology of Alzheimer disease in dementia in the oldest-old. *Arch Neurol*. 2008;65(9):1211-1217.
39. Dolan D, Troncoso J, Resnick SM, Crain BJ, Zonderman AB, O'Brian RJ. Age, Alzheimer's disease and dementia in the baltimore longitudinal study of ageing. *Brain*. 2010;133(pt 8):2225-2231.
40. Cohen-Mansfield J, Reisberg B, Bonnema J, et al. Staging methods for the assessment of dementia: Perspectives. *J Clin Psychiatry*. 1996;57(5):190-198.
41. Petersen RC. Early diagnosis of Alzheimer's disease: is MCI too late? *Curr Alzheimer Res*. 2009;6(4):324-330.
42. Gauthier S, Dubois B, Feldman H, Scheltens P. Revised research diagnostic criteria for Alzheimer's disease. *Lancet Neurol*. 2008;7(8):668-670.