

S. Barth¹
 P. Schönknecht¹
 J. Pantel²
 J. Schröder¹

Neuropsychologische Profile in der Demenzdiagnostik: Eine Untersuchung mit der CERAD-NP-Testbatterie

*Mild Cognitive Impairment and Alzheimer's Disease:
 An Investigation of the CERAD-NP Test Battery*

Zusammenfassung

Die vorliegende Arbeit analysiert die Diskriminationsfähigkeit der deutschen Version der neuropsychologischen Testbatterie CERAD-NP. Dazu wurden 49 Patienten mit leichter kognitiver Beeinträchtigung (LKB), 80 Patienten mit Alzheimer-Demenz (AD), 36 Patienten mit Major-Depression (MD) und 26 gesunde ältere Kontrollen eingeschlossen. Alle Probanden waren ambulante Patienten der Sektion Gerontopsychiatrie der Universität Heidelberg. Die Diagnosen wurden anhand der klinischen und apparativen Untersuchungsbefunde sowie der neuropsychologischen Testung unter Ausschluss der CERAD-NP-Testbatterie nach den Kriterien für aging-associated cognitive decline (AACD) für LKB, NINCDS-ADRDA für AD und DSM-IV für MD gestellt. Alle CERAD-NP-Subtests unterschieden signifikant ($p \leq 0,05$) zwischen gesunden Kontrollen und leichter AD. Zwischen einer leichten und einer mittelschweren AD trennten die Untertests Verbale Flüssigkeit und Konstruktive Praxis, während in den Gedächtnisparametern schon bei leichter AD Bodeneffekte auftraten. LKB-Patienten zeigten im Vergleich zu gesunden Kontrollen eine geringere Verbale Flüssigkeit und Defizite des Episodischen Gedächtnisses; eine Abgrenzung zur leichten AD gelang mit Ausnahme der Konstruktiven Praxis in allen Subtests. Affektiv bedingte kognitive Leistungseinbußen wurden hinsichtlich des Episodischen Gedächtnisses beobachtet. Dagegen konnte die Testbatterie nicht zwischen Depressiven und LKB-Patienten unterscheiden. Die CERAD-NP-Testbatterie kann als ein Verfahren bewertet werden, das sensitiv zwischen leichter und mittelschwerer AD, LKB, Depression und gesunden Kontrollen trennen kann. In Abhängigkeit vom Erkrankungsverlauf verändert sich der Schwierigkeitsgrad der einzelnen Subskalen. Während Auf-

Abstract

To investigate the psychometric properties of the German version of the CERAD-NP, neuropsychological deficits were compared between 49 patients with mild cognitive impairment (LKB), 80 patients with Alzheimer's disease (AD), 36 with major depression (MD), and 26 elderly controls. All participants were outpatients of the memory clinic of the Section of Geriatric Psychiatry, Heidelberg University. Diagnoses were established based on clinical examination, laboratory testing, neuroimaging, and routine neuropsychological testing according to the criteria of aging-associated cognitive decline (AACD) for LKB, NINCDS-ADRDA for AD, and DSM-IV for MD, respectively. All CERAD-NP subtests discriminated between controls and AD patients with the latter showing significantly ($p \leq 0.05$) lower test scores. The subtests verbal fluency and constructive apraxia differed significantly between mildly and moderately AD, while the subtests assessing declarative (episodic) memory performance showed only minor, non-significant differences between the respective groups. The LKB patients took an intermediate position between controls and AD patients with significantly lower scores in verbal fluency and declarative memory performance than the controls. When compared with the AD patients, LKB patients were significantly impaired in all subtests except constructive apraxia. Relative to the controls, the patients with MD showed a decreased episodic memory performance but no evidence suggesting an impairment in other neuropsychological domains. Our results indicate that the CERAD-NP is a psychometric instrument which allows a sensitive discrimination between mild and moderate AD, LKB, MD and healthy controls. However, sensitivity of discrimination between different stages of dementia varies with respect to the

Institutsangaben

¹Sektion Gerontopsychiatrie der Psychiatrischen Klinik der Universität Heidelberg
²Klinik für Psychiatrie und Psychotherapie der Universität Frankfurt

Korrespondenzadresse

Dr. phil. Sonja Barth · Sektion für Gerontopsychiatrie · Psychiatrische Klinik der Universität Heidelberg ·
 Voßstraße 4 · 69115 Heidelberg · E-mail: sonja_barth@med.uni-heidelberg.de

Bibliografie

Fortschr Neurol Psychiatr 2005; 73: 1–9 © Georg Thieme Verlag KG Stuttgart · New York
 DOI 10.1055/s-2004-830249
 ISSN 0720-4299

gaben zum Episodischen Gedächtnis schon bei leichter AD Bodeffekte aufweisen, können Aufgaben zur Sprache und Konstruktiven Praxis auch noch zwischen weiter fortgeschrittenen Stadien trennen.

different subtest. While the subtest for episodic memory showed floor effects already for mild dementia, subtests for verbal fluency and constructive apraxia were able to discriminate even between more advanced stages of the disease.

Einleitung

Neuropsychologische Untersuchungsverfahren haben in der klinischen Demenzdiagnostik im Hinblick auf Früherkennung demenzieller Erkrankungen und Therapieinterventionen eine hohe Bedeutung. Über ein differenziertes neuropsychologisches Leistungsprofil kann zwischen normwertigem kognitivem Leistungsniveau, einer leichten kognitiven Beeinträchtigung (LKB) und einer Demenzerkrankung unterschieden werden. Weiterhin ermöglichen neuropsychologische Verfahren Aussagen zum Schweregrad und Erkrankungsverlauf sowie zur Therapieevaluation. Während das Erkennen manifester Demenzen in der Regel keine Schwierigkeiten bereitet, ist dagegen die Beurteilung einer LKB schwierig. Von erheblicher Bedeutung ist auch die Erfassung depressiver Störungen, insbesondere wenn sie in Form so genannter Spätdepressionen erstmals im Alter oder aber vor dem Hintergrund körperlicher Erkrankungen auftreten [1,2].

Um eine hohe Validität in der Prüfung der kognitiven Leistungsfähigkeit zu gewährleisten, sind standardisierte Testverfahren zur Demenzdiagnostik unabdingbar. Seit 1998 hat sich im deutschsprachigen Raum die Neuropsychologische Testbatterie CERAD-NP etabliert, die 1986 von dem Consortium to Establish a Registry for Alzheimer's Disease des National Institute on Aging in den USA entwickelt bzw. aus bereits bestehenden Verfahren erstellt wurde. Das CERAD-NP sollte als ein standardisiertes Instrument zur Erfassung von klinischen, neuropathologischen und neuropsychologischen Symptomen einer Alzheimer-Demenz (AD) eine Vereinheitlichung in der Demenzdiagnostik bringen [3] und wurde ursprünglich zur Erfassung derjenigen kognitiven Parametern entwickelt, die auch schon in frühen Stadien der AD betroffen sind [4,5]. Das Verfahren gliedert sich in acht Untertests, welche die Bereiche Sprache, Orientierung, Episodisches Gedächtnis sowie Konstruktive Praxis abbilden (vgl. Tab. 1). Die Sensitivität der einzelnen Untertests unterscheidet sich in Abhängigkeit vom Erkrankungsstadium. Als besonders sensitiv zur Unterscheidung zwischen einer LKB und einer leichten AD haben sich Aufgaben zum verzögerten Abruf erwiesen, während die Wortfindung erst zwischen mittelschwerer und schwerer Demenz valide trennt [6–9]. Morris u. Mitarb. [4] und Welsh u. Mitarb. [9] überprüften die Testgütekriterien der Objektivität (Auswertungsobjektivität), der Reliabilität (Retest-Reliabilität) und der Validität, die insgesamt als gut bewertet wurden.

Die vorliegende Studie untersucht die Sensitivität der CERAD-NP-Batterie in der Differenzialdiagnose zwischen zwei Schweregraden der AD (leicht und mittelschwer), LKB, Depression und psychisch gesunden älteren Kontrollpersonen. Sie geht dabei der Frage nach, welche Untertests am besten zwischen den Gruppen trennen. Während die Sensitivität des Verfahrens für Demenzpatienten in Abgrenzung zu Gesunden als ausreichend

Tab. 1 Darstellung der CERAD-NP-Testbatterie und der damit erhobenen kognitiven Parameter

| | |
|--|---|
| 1. verbale Flüssigkeit, Kategorie „Tiere“ | verbale Produktionsfähigkeit, semantisches Gedächtnis, kognitive Flexibilität |
| 2. modifizierter Boston Naming Test | Wortfindung und -benennung, visuelle Wahrnehmung |
| 3. mini Mental State Examination (MMSE) | Orientierung, Konzentrationsfähigkeit, Merkfähigkeit, Sprache, Ideatorische und Konstruktive Praxis |
| 4. Wortliste Gedächtnis unmittelbar (3 Lerndurchgänge) | unmittelbare Merkfähigkeit und Lernvermögen von nicht assoziiertem verbalem Material |
| 5. konstruktive Praxis | visuokonstruktive Fähigkeiten |
| 6. Wortliste Abrufen | verzögerte verbale Merkfähigkeit, freie Reproduktion |
| 7. Wortliste Wiedererkennen – Treffer – korrekte Zurückweisung | verzögerte verbale Merkfähigkeit, Rekognition; Abruf- vs. Speicherdefizite |
| 8. konstruktive Praxis Abrufen | verzögerte figurale Merkfähigkeit, freie Reproduktion |

gesichert angesehen werden kann, kommt der Diskriminationsfähigkeit von Personen mit einer LKB gegenüber gesunden Kontrollen und Patienten mit leichter Demenz eine besondere Bedeutung zu. Weiterhin hat die Differenzialdiagnose möglicher affektiv bedingter kognitiver Störungen zentralen Stellenwert.

Methode

In die Auswertung gingen 191 Patienten der Gedächtnisambulanz der Universität Heidelberg ein, die von August 2001 bis Juli 2003 neuropsychologisch mit der CERAD-NP-Testbatterie untersucht wurden. Die Patienten wurden von niedergelassenen Allgemeinmedizinern und Fachärzten überwiesen. Eingeschlossen wurden Patienten mit einer LKB (n = 49), einer leichten (n = 64) oder mittelschweren (n = 16) AD oder einer depressiven Störung (n = 36). Weiterhin eingeschlossen wurden Patienten, die sich aufgrund subjektiver Leistungseinbußen in unserer Gedächtnisambulanz vorstellten, ohne dass diese in der ausführlichen neuropsychologischen Untersuchung bestätigt werden konnten. Neben einem normwertigen Leistungsprofil durften auch im Anamnesegespräch und den somatischen Untersuchungsbefunden keine Hinweise auf psychiatrische und neurologische Erkrankungen bestehen. Insgesamt traf dies für 26 Patienten zu, die als gesunde Kontrollpersonen in die Analysen gingen. Patienten mit einer schweren AD wurden ausgeschlossen, da diese aufgrund des weit fortgeschrittenen kognitiven Leistungsabbaus nicht mehr an der ausführlichen neuropsychologischen Untersu-

chung teilnehmen konnten. Ebenfalls ausgeschlossen wurden Patienten mit anderen Demenzformen.

Die Diagnosen wurden anhand der Befunde der psychiatrisch-neurologischen Untersuchung, der Exploration, der Fremdanamnese, laborchemischer Parameter, bildgebender Untersuchungen des Gehirns (CCT oder MRT) und der neuropsychologischen Untersuchung unter Ausschluss der CERAD-NP-Testbatterie gestellt. Auf das Mini Mental State Examination (MMSE) als ein Bestandteil des CERAD-NP konnte zur Diagnosestellung und Schweregradeinteilung allerdings nicht verzichtet werden, so dass dieses in den weiteren Analysen nicht berücksichtigt werden konnte. Weitere neuropsychologische Untersuchungsverfahren waren die Wechsler Memory Scale – Revidierte Fassung (WMS-R) [10], der Uhrentest [11] sowie der Trail Making Test (TMT) [12]. Die Global Deterioration Scale [13] diente der Einschätzung des Schweregrades der funktionellen Einbußen auf einer siebenstufigen Skala. Eine klinisch diagnostizierte depressive Symptomatik wurde nach den Kriterien des ICD-10 diagnostiziert und bei Patienten über 60 Jahren mittels der Kurzform der Geriatric Depression Scale (GDS) [14] bzw. bei jüngeren Patienten mittels des Beck-Depressions-Inventars (BDI) [15] standardisiert erfasst.

Nach den Kriterien der NINCDS-ADRDA [16] wurden die eingeschlossenen AD-Diagnosen als wahrscheinliche Demenzen klassifiziert, wenn Defizite in mindestens zwei kognitiven Funktionen vorlagen, ein progredienter Verlauf von Gedächtnisstörungen und anderen kognitiven Beeinträchtigungen bestand und das Alter der Patienten zwischen 40 und 90 Jahren lag. Bewusstseinsstörungen, andere psychische oder somatische Erkrankungen, welche die kognitiven Defizite erklären könnten, galten als Ausschlusskriterium. Die Diagnose LKB richtete sich nach den Kriterien für aging-associated cognitive decline (AACD) nach Levy [17] mit eigen- oder fremdanamnestic kognitiven Leistungseinbußen und psychometrisch objektivierbaren Defiziten (Testleistung unterhalb einer Standardabweichung des Erwartungswertes der Normstichprobe) in mindestens einem der fünf Bereiche: Gedächtnis und Lernen, Aufmerksamkeit und Konzentration, Sprache, visuell-räumliches Vorstellungsvermögen und Denken. Hinweise auf internistische, neurologische oder psychiatrische Erkrankungen, welche die kognitiven Defizite erklären können, gelten als Ausschlusskriterien. Das AACD-Konzept ist damit in Anlehnung an die ICD-10-Kriterien einer leichten kognitiven Störung nicht auf das Vorliegen von Gedächtnisstörungen begrenzt.

Die Rohwerte wurden z-transformiert und somit anhand der publizierten Originalstichprobe [18] alters- und geschlechtsnormiert. Bisher liegen noch keine einheitlichen Regeln zur Bewertung der z-Werte vor, um zu beurteilen, ob das Ergebnis im unterdurchschnittlichen Bereich liegt oder den Erwartungswerten der Normstichprobe entspricht. Fisseni [19] klassifiziert eine Leistung als durchschnittlich, wenn sie im Bereich \pm einer Standardabweichung liegt. Ein Ergebnis wird dann als unterdurchschnittlich bewertet, wenn der z-Wert < -1 beträgt. Ein weniger strenger Schwellenwert für ein unterdurchschnittliches Ergebnis ist ein z-Wert von $\leq -1,28$, der einem Prozentrang von 10 entspricht. Je kleiner der Schwellenwert, desto höher ist die Sensitivität, allerdings auf Kosten einer geringeren Spezifität. In der vor-

liegenden Arbeit wurde der Schwellenwert für eine unterdurchschnittliche Leistung nach Fisseni [19] bei einem z-Wert < -1 festgelegt.

Die statistische Auswertung erfolgte über das Programm SPSS 11.5 mittels bivariater Korrelationen, Varianzanalysen und Scheffé-Tests als post-hoc-t-Test-Vergleiche. Weiterhin wurden Receiver Operating Characteristic (ROC)-Kurven berechnet. Für die nichtparametrischen Variablen kamen Kontingenztafelanalysen (Chi-Quadrat-Tests) und der Mann-Whitney-U-Test zur Anwendung.

Ergebnisse

Die Stichprobenbeschreibung ist in Tab. 2 zusammengefasst. Der Anteil der männlichen Probanden war in den Diagnosegruppen Depression, leichte AD und mittelschwere AD signifikant niedriger. Das statistisch bedeutsam höhere Durchschnittsalter dementer Patienten im Vergleich zu den gesunden Kontrollen und depressiven Patienten ist dadurch zu erklären, dass in der vorliegenden Studie eine Inanspruchnahmepopulation untersucht wurde und das Demenzrisiko bekanntermaßen mit höherem Lebensalter ansteigt. Die Gruppenunterschiede bezüglich der Alters- und Geschlechtsverteilung wurden über die Umrechnung der Rohwerte in z-transformierte Werte kontrolliert, so dass diese in den Analyseverfahren nicht von Bedeutung sind und eine Konfundierung über Alters- und Geschlechtsunterschiede ausgeschlossen ist. Bei 116 Patienten wurde die GDS eingesetzt, bei 16 das BDI; bei Patienten mit einer mittelschweren Ausprägung der Demenz wurde aufgrund der zu erwartenden Defiziten des Sprachverständnisses auf eine standardisierte Schweregradeinschätzung einer möglichen depressiven Symptomatik verzichtet. Insgesamt waren Patienten mit einer LKB, einer leichten AD und gesunde Kontrollen in ihrer Ausprägung einer gegenwärtigen depressiven Symptomatik vergleichbar. Die Werte depressiver Patienten waren erwartungsgemäß signifikant höher und lagen im Bereich einer klinisch relevanten aktuellen depressiven Störung.

Über die Gesamtgruppe berechnet korrelierte der MMSE hochsignifikant ($0,29 \leq r \leq 0,67$; $p < 0,000$) mit allen CERAD-NP-Untertests. Auch die einzelnen Untertests waren durchweg interkorreliert. Der MMSE erwies sich als sensitiv zur Differenzierung gesunder Kontrollen, Patienten mit LKB, leichter und mittelschwerer AD, während Depressive nicht von Gesunden und LKB-Patienten über das Verfahren unterscheidbar waren. In der Gruppe der Depressiven war ein höherer Wert in der GDS und damit eine stärkere Ausprägung der gegenwärtigen depressiven Symptomatik nicht mit einem höheren MMSE-Score korreliert. Von den CERAD-NP-Maßen korrelierte nur der Untertest Konstruktive Praxis Abrufen, also das verzögerte Erinnern zuvor gezeichneter Figuren, signifikant mit der Ausprägung der depressiven Symptomatik ($r = -0,45$; $p \leq 0,05$). Der BDI-Score, der lediglich bei 12 Probanden der Depressionsgruppe zur Anwendung kam, war weder mit dem MMSE noch mit den CERAD-NP-Maßen korreliert.

Betrachtet man die Leistungsprofile der CERAD-NP-Testbatterie in Abb. 1, liegen die Testergebnisse der gesunden Kontrollperso-

Tab. 2 Stichprobenbeschreibung

| | gesunde Kontrollen | depressive | LKB | AD, leicht | AD, mittelschwer | gesamt | Signifikanz χ^2 -Quadrat ¹ F-Wert ² |
|--|-------------------------|-------------------------|-------------------------|-------------------------|--------------------------|-------------------------|--|
| Stichprobenverteilung n (%) | 26 (13,6) | 36 (18,8) | 49 (25,7) | 64 (33,5) | 16 (8,4) | 191 (100) | |
| Geschlecht männlich/weiblich n (%) | 15/11 (58/42) | 10/26 (27/73) | 29/20 (59/41) | 24/40 (37/63) | 6/10 (38/62) | 84/107 (44/56) | 11,781 ^{*1} |
| Alter mean \pm sd range | 64,9 \pm 8,4 50–78 | 61,4 \pm 7,6 46–77 | 69,8 \pm 8,9 53–89 | 72,3 \pm 8,0 54–87 | 72,4 \pm 11,3 52–85 | 68,6 \pm 9,4 46–89 | 11,545 ^{***2} |
| MMSE mean \pm sd range | 28,9 \pm 1,1 26–30 | 28,1 \pm 1,5 24–30 | 27,0 \pm 1,3 25–30 | 23,3 \pm 2,9 17–30 | 15,0 \pm 2,9 10–20 | 25,2 \pm 4,3 10–30 | 149,677 ^{***2} |
| Global Deterioration Scale mean \pm sd range | 1,2 \pm 0,4 1–2 | 2,1 \pm 0,8 1–4 | 2,7 \pm 0,5 1–4 | 3,6 \pm 0,8 1–5 | 4,7 \pm 0,7 2–5 | 2,9 \pm 1,2 1–5 | 90,69 ^{***2} |
| Geriatric Depression Scale mean \pm sd range | 3,2 \pm 2,9 0–11 | 8,0 \pm 3,6 2–15 | 2,7 \pm 1,9 0–7 | 2,9 \pm 3,0 0–12 | nicht durchführbar | 3,7 \pm 3,4 0–15 | 15,682 ^{***2} |
| Beck Depressionsinventar mean \pm sd range | 8 \pm 0 0–8 | 20,2 \pm 7,7 12–32 | 8 \pm 5,7 4–12 | nicht durchführbar | nicht durchführbar | 17,1 \pm 8,7 4–32 | 4,198 ^{*2} |

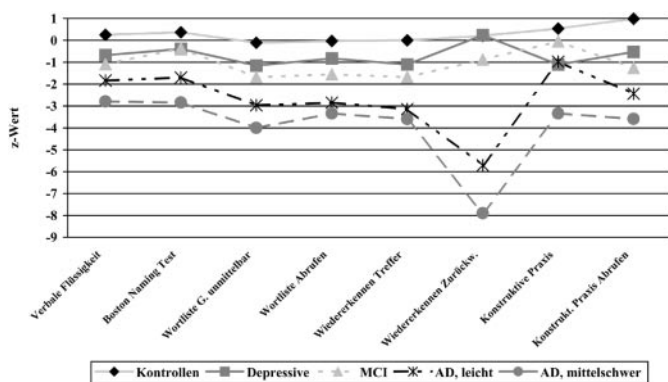
*p \leq 0,05; **p \leq 0,01; ***p \leq 0,001

Abb 1 Neuropsychologische Profile der CERAD-NP-Testbatterie im Vergleich zur Normstichprobe (z-Werte).

nen erwartungsgemäß im Bereich von \pm einer Standardabweichung in Bezug zu den alters- und geschlechtsabhängigen Normwerten. Dagegen zeigen depressive Patienten Normabweichungen im Episodischen Gedächtnis und der Konstruktiven Praxis, LKB-Patienten im Episodischen Gedächtnis sowie in der Verbalen Flüssigkeit. Mit Ausnahme der Konstruktiven Praxis (z-Wert = $-0,98$) liegen alle Leistungen der Patienten mit leichter AD in dem Bereich von minus einer bis minus zwei Standardabweichungen, Patienten mit einer mittelschweren Demenz liegen in allen Untertests mindestens zwei Standardabweichungen unterhalb der Erwartungswerte. Die Leistungsprofile der Diagnosegruppen verlaufen weitestgehend parallel, allerdings je nach Schweregrad der Beeinträchtigung auf einem niedrigeren Niveau. Die Werte depressiver Patienten lagen überwiegend zwischen gesunden Kontrollen und LKB-Patienten. Auffällig ist der starke Leistungsabfall in den Demenzgruppen bezogen auf die verzögerte Wiedererkennensleistung der Wortliste. Die Mittelwerte und Standardabweichungen in den einzelnen Subtests sind für die Diagnosegruppen in Tab. 3 aufgelistet.

In einer einfaktoriellem multivariaten Varianzanalyse konnten alle Subtests hochsignifikant ($p \leq 0,001$; df 4, 186) zwischen den Diagnosegruppen unterscheiden (Verbale Flüssigkeit: F-Wert

28,4; Boston Naming Test: F-Wert 10,7; Wortliste Gedächtnis unmittlbar: F-Wert 39,6; Wortliste Abrufen: F-Wert 48,0; Wortliste Wiedererkennen Treffer: F-Wert 5,6; Wortliste Wiedererkennen korrekte Zurückweisung: F-Wert 15,4; Konstruktive Praxis: F-Wert 10,6; Konstruktive Praxis Abrufen: F-Wert 37,1). . Mit einem anschließenden Scheffé-Test als post-hoc-Vergleich wurde geprüft, welche der acht Untertests zwischen welchen Diagnosegruppen signifikant trennen (vgl. Tab. 3). Jeweils im Vergleich zu gesunden Kontrollen zeigen Depressive signifikant geringere Leistungen in den Subtests Wortliste Gedächtnis unmittlbar und Konstruktive Praxis Abrufen, LKB-Patienten in den Subtests Verbale Flüssigkeit, Wortliste Gedächtnis unmittlbar, Wortliste Abrufen und Konstruktive Praxis Abrufen und Demenzpatienten in allen acht Untertests. Verglichen mit LKB-Patienten lagen die Leistungen der Patienten mit leichter AD in allen Untertests signifikant niedriger mit Ausnahme der Konstruktiven Praxis. Depressive Patienten und LKB-Patienten unterschieden sich in keinem Subtest signifikant voneinander. Eine leichte von einer mittelschweren AD zu trennen, gelang über die Untertests Verbale Flüssigkeit und Konstruktive Praxis, während beide Demenzgruppen in den Gedächtnisleistungen vergleichbar waren.

In einem weiteren Analyseschritt wurde die Diskriminationsfähigkeit der CERAD-NP-Subtests für die jeweiligen Diagnosegruppen berechnet. Dazu wurden die z-Werte von jeweils zwei Gruppen vergleichend gegenübergestellt und ROC-Kurven errechnet. Je größer die Fläche unter der jeweiligen Kurve ist (AUC: Area Under the Curve; vgl. Tab. 4), desto besser kann die Variable zwischen den Gruppen trennen, d.h. desto größer ist die diagnostische Diskriminationsfähigkeit. Ein AUC-Wert von 1 bedeutet, dass die Variable perfekt zwischen zwei Gruppen trennt, ein AUC-Wert von 0,5, dass die Variable nur zufällig zwischen den Gruppen variiert und damit keine diagnostische Aussage ermöglicht. Alle Gedächtnisparameter sowie der Untertest Verbale Flüssigkeit konnten zwischen gesunden Kontrollen und depressiven Patienten trennen, zwischen gesunden Kontrollen und LKB-Patienten war eine Unterscheidung in allen Subskalen mit Ausnahme der Konstruktiven Praxis möglich. Alle CERAD-NP-

Tab. 3 Ergebnisse der CERAD-NP-Testbatterie in den Untersuchungsgruppen. Mittelwerte und Standardabweichungen mit den Ergebnissen eines Scheffé-Tests auf dem 95 % Signifikanzniveau

| | gesunde Kontrollen | depressive | LKB | AD, leicht | AD, mittel-schwer | Scheffé-Test |
|---|--------------------|--------------|--------------|---------------|-------------------|---|
| verbale Flüssigkeit | 0,25 ± 1,0 | - 0,68 ± 0,9 | - 1,09 ± 1,0 | - 1,84 ± 1,0 | - 2,79 ± 0,6 | Kon > LKB > AD(l) > AD(m) Kon,MD,LKB |
| modifizierter Boston Naming Test | 0,37 ± 0,7 | - 0,39 ± 1,8 | - 0,36 ± 1,2 | - 1,7 ± 2,5 | - 2,84 ± 3,1 | Kon > AD(l), AD(m) Kon,MD,LKB > AD(l) |
| Wortliste Gedächtnis unmittelbar: Summe über drei Lerndurchgänge | - 0,11 ± 1,0 | - 1,15 ± 1,4 | - 1,69 ± 1,1 | - 2,97 ± 1,4 | - 4,0 ± 0,9 | Kon > MD, LKB > AD(l), AD(m) |
| Wortliste Abrufen | - 0,03 ± 0,9 | - 0,83 ± 1,2 | - 1,55 ± 1,2 | - 2,85 ± 1,1 | - 3,34 ± 0,5 | Kon > LKB > AD(l), AD(m) Kon,MD,LKB > AD |
| Wortliste Wiedererkennen | | | | | | |
| - Treffer | 0 ± 1,6 | - 1,1 ± 2,1 | - 1,68 ± 2,8 | - 3,14 ± 4,6 | - 3,6 ± 4,2 | Kon > AD(l), AD(m) Kon,MD,LKB,AD(l),AD(m) |
| - korrekte Zurückweisung | 0,21 ± 0,5 | 0,24 ± 1,9 | - 0,88 ± 2,8 | - 5,71 ± 8,11 | - 7,9 ± 6,48 | Kon > AD(l), AD(m) Kon,MD,LKB > AD(l) |
| konstruktive Praxis | 0,54 ± 0,4 | - 1,11 ± 1,3 | - 0,05 ± 1,2 | - 0,98 ± 2,7 | - 3,34 ± 3,8 | Kon > AD(l) > AD(m) Kon,MD,LKB,AD(l) > AD(m) |
| konstruktive Praxis Abrufen | 0,99 ± 1,0 | - 0,53 ± 1,4 | - 1,26 ± 1,8 | - 2,44 ± 1,5 | - 3,59 ± 1,2 | Kon > Dep, LKB > AD(l), AD(m) |

Legende: gesunde Probanden (Kon), depressive Störung (MD), leichte kognitive Beeinträchtigung (LKB), leichte Alzheimer Demenz (AD(l)), mittelgradige Alzheimer Demenz (AD(m))

Tab. 4 Diskriminationsfähigkeit der CERAD-NP-Subtests im Vergleich zu gesunden Kontrollen bei einem Schwellenwert von $z < -1$

| Diagnose | | verbale Flüssigkeit | BNT | Wortliste unmittelbar | Wortliste Abrufen | Wiedererkennen Treffer | Wiedererkennen Zurückw. | konstr. Praxis | konstr. Praxis Abrufen |
|---------------|--------------|---------------------|----------|-----------------------|-------------------|------------------------|-------------------------|----------------|------------------------|
| MD | AUC | 0,645* | 0,614 | 0,729** | 0,695** | 0,71** | 0,548 | 0,558 | 0,81*** |
| | Sensitivität | 0,808 | 0,923 | 0,808 | 0,885 | 0,923 | 0,962 | 1,0 | 1,0 |
| | Spezifität | 0,389 | 0,194 | 0,556 | 0,444 | 0,306 | 1,0 | 0,194 | 0,306 |
| | Genauigkeit | 0,599 | 0,559 | 0,682 | 0,665 | 0,615 | 0,981 | 0,597 | 0,653 |
| LKB | AUC | 0,752*** | 0,633*** | 0,862*** | 0,841*** | 0,729*** | 0,542 | 0,588 | 0,865*** |
| | Sensitivität | 0,808 | 0,731 | 0,846 | 0,885 | 0,923 | 0,962 | 1,0 | 1,0 |
| | Spezifität | 0,569 | 0,569 | 0,765 | 0,608 | 0,51 | 0,235 | 0,137 | 0,51 |
| | Genauigkeit | 0,689 | 0,65 | 0,805 | 0,747 | 0,717 | 0,599 | 0,569 | 0,755 |
| AD (l) | AUC | 0,903*** | 0,844*** | 0,96*** | 0,967*** | 0,8*** | 0,726*** | 0,693** | 0,969*** |
| | Sensitivität | 0,808 | 0,923 | 0,846 | 0,885 | 0,923 | 0,962 | 1,0 | 1,0 |
| | Spezifität | 0,855 | 0,468 | 0,968 | 0,935 | 0,565 | 0,548 | 0,29 | 0,955 |
| | Genauigkeit | 0,832 | 0,696 | 0,907 | 0,91 | 0,744 | 0,755 | 0,645 | 0,978 |
| AD (m) | AUC | 0,966*** | 0,958*** | 1,0*** | 1,0*** | 0,75** | 0,977*** | 0,883*** | 1,0*** |
| | Sensitivität | 0,808 | 0,923 | 0,846 | 0,885 | 0,923 | 0,926 | 1,0 | 1,0 |
| | Spezifität | 1,0 | 0,75 | 1,0 | 1,0 | 0,687 | 0,837 | 0,562 | 1,0 |
| | Genauigkeit | 0,904 | 0,837 | 0,923 | 0,923 | 0,805 | 0,882 | 0,781 | 1,0 |

* $p \leq 0,05$; ** $p \leq 0,01$; *** $p \leq 0,001$

Untertests konnten signifikant zwischen gesunden Kontrollen und leichter AD trennen, allerdings mit unterschiedlicher Diskriminationsfähigkeit. Die höchste Diskriminationsfähigkeit trat erwartungsgemäß bei dem Vergleich zwischen Gesunden und mittelschwer Dementen auf: alle Variablen trennten hochsignifikant.

Mittels der ROC-Kurven wurden die Maße für Sensitivität und Spezifität der CERAD-Subtests in den Diagnosegruppen jeweils im Vergleich zu den gesunden Kontrollen errechnet, wobei ein Schwellenwert von $z = -1$ bestimmt wurde. Die zusätzlich ermit-

telte diagnostische Genauigkeit ergibt sich aus dem Mittelwert zwischen Sensitivität und Spezifität und damit einer Gleichgewichtung beider Parameter. Eine diagnostische Genauigkeit von 75% kann als gut, eine von 80% als sehr gut beurteilt werden. Entsprechende Werte sind in Tab. 4 hervorgehoben. Im Vergleich von Patienten mit MD zu gesunden Kontrollen wurde eine sehr hohe Differenzierungsfähigkeit von 98,1% für die Leistung in der Wortliste Wiedererkennen korrekte Zurückweisung beobachtet. In den übrigen Parametern konnten zwar hohe Maße für Sensitivität ($> 80\%$) ermittelt werden, jedoch mit geringer Spezifität und damit einer hohen Fehlerwahrscheinlichkeit, Gesunde als

Tab. 5 Gruppenunterschiede im Trail Making Test (TMT) und Uhrentest

| Angaben in % (n) | gesunde Kontrollen | depressive | LKB | AD, leicht | AD, mittelschwer | χ^2 -Quadrat |
|--|--------------------|------------|-----------|------------|------------------|-------------------|
| TMT, Teil A | | | | | | 63,27*** |
| durchschnittlich bis überdurchschnittlich (Prozentrang ≥ 25) | 96,2 (25) | 69,5 (25) | 66,7 (32) | 51,6 (32) | 6,3 (1) | |
| niedrig durchschnittlich (Prozentrang 10–25) | 3,8 (1) | 16,7 (6) | 12,5 (6) | 17,7 (11) | 0 | |
| unterdurchschnittlich (Prozentrang < 10) | 0 | 13,9 (5) | 20,8 (10) | 30,6 (19) | 93,8 (15) | |
| TMT, Teil B | | | | | | 72,30*** |
| durchschnittlich bis überdurchschnittlich (Prozentrang ≥ 25) | 84,6 (22) | 36,2 (13) | 41,7 (20) | 9,6 (6) | 0 | |
| niedrig durchschnittlich (Prozentrang 10–25) | 11,5 (3) | 25,0 (9) | 10,4 (5) | 12,9 (8) | 0 | |
| unterdurchschnittlich (Prozentrang < 10) | 3,8 (1) | 38,9 (14) | 47,9 (23) | 77,4 (48) | 100 (16) | |
| Uhrentest | | | | | | 128,43*** |
| 1: perfekte Darstellung der Uhr | 88,0 (22) | 55,6 (20) | 49,0 (24) | 23,4 (15) | 0 | |
| 2: leichte visuell-räumliche Fehler | 4,0 (1) | 13,9 (5) | 18,4 (9) | 6,3 (4) | 6,3 (1) | |
| 3: fehlerhafte Uhrzeit bei erhaltener visuell-räumlicher Darstellung der Uhr | 8,0 (2) | 27,8 (10) | 30,6 (15) | 50,0 (32) | 6,3 (1) | |
| 4: mittelgradige visuell-räumliche Desorganisation | 0 | 2,8 (1) | 2,0 (1) | 18,8 (12) | 56,3 (9) | |
| 5: schweregradige visuell-räumliche Desorganisation | 0 | 0 | 0 | 1,6 (1) | 31,3 (5) | |
| 6: keinerlei Darstellung einer Uhr | 0 | 0 | 0 | 0 | 0 | |

* $p \leq 0,05$; ** $p \leq 0,01$; *** $p \leq 0,001$

krank zu klassifizieren. Bei dem vorgegebenen Schwellenwert konnten Patienten mit LKB in den Subtests Wortliste unmittelbar (80,5%), Konstruktive Praxis Abrufen (75,5%) und Wortliste Abrufen (74,7%) mit hoher diagnostischer Genauigkeit von Gesunden differenziert werden, während Aufgaben zur Sprachproduktion und zur Konstruktiven Praxis nicht geeignet waren, die beiden Gruppen zu trennen. Neben einer sehr hohen diagnostischen Genauigkeit in den Parametern zum Episodischen Gedächtnis gelang eine gute Differenzierung von Patienten mit leichter AD und Gesunden zusätzlich in dem Untertest Verbale Flüssigkeit mit 83,2%, dagegen erreichten der Boston Naming Test und Konstruktive Praxis bei niedriger Spezifität keine ausreichende Differenzierungsfähigkeit. Patienten mit mittelschwerer AD waren über alle Parameter mit guter bis sehr guter diagnostischer Genauigkeit von Gesunden zu trennen.

Tab. 5 gibt einen Überblick über die Ergebnisse im TMT und Uhrentest. Da die genannten Verfahren neben der WMS und dem MMSE zur Diagnosestellung berücksichtigt wurden, ist eine Interpretation der Ergebnisse nur bedingt zulässig.

Über Kontingenztafelanalysen wurde geprüft, inwieweit sich die Diagnosegruppen in den ordinal skalierten Ergebnissen des TMT unterschieden, indem der Anteil unterdurchschnittlicher Werte (Prozentrang < 10) vergleichend gegenübergestellt wurde. Im TMT A, der die Informationsverarbeitungsgeschwindigkeit misst, war der Anteil unterdurchschnittlicher Leistungen in allen Diagnosegruppen im Vergleich zu kognitiv Gesunden erhöht (MD: $\chi^2 = 3,9$; $p \leq 0,05$; LKB: $\chi^2 = 6,3$; $p \leq 0,01$; leichte AD: $\chi^2 = 10,2$; $p \leq 0,01$; mittelschwere AD: $\chi^2 = 37,9$; $p \leq 0,001$). Die Leistungen

von Patienten mit mittelschwerer AD waren im Vergleich zu den Patienten mit leichter AD deutlich reduziert ($\chi^2 = 20,6$; $p \leq 0,001$). Dagegen war eine Differenzierung zwischen den Diagnosegruppen LKB, MD und leichter AD nicht möglich.

Im TMT B, einem Verfahren zur Erfassung der kognitiven Komplexität und Umstellungsfähigkeit als Maß exekutiver Funktionen, konnten ebenfalls gesunde Kontrollen von allen vier Diagnosegruppen abgegrenzt werden (MD: $\chi^2 = 10,1$; $p \leq 0,001$; LKB: $\chi^2 = 14,9$; $p \leq 0,001$; leichte AD: $\chi^2 = 40,2$; $p \leq 0,001$; mittelschwere AD: $\chi^2 = 38,0$; $p \leq 0,001$). Patienten mit einer leichten AD hatten einen signifikant höheren Anteil unterdurchschnittlicher Ergebnisse im Vergleich zu Patienten mit LKB ($\chi^2 = 10,3$; $p \leq 0,001$) und Depression ($\chi^2 = 14,5$; $p \leq 0,001$), Patienten mit mittelschwerer AD und leichter AD unterschieden sich ebenfalls signifikant ($\chi^2 = 4,4$; $p \leq 0,05$). Dagegen waren die Leistungen von Depressiven und LKB-Patienten vergleichbar.

Der Uhrentest ist gerade zur Bestimmung des Schweregrades einer Demenz von hoher differenzialdiagnostischer Bedeutung. In dem Mann-Whitney-U-Test unterschieden sich die mittleren Rangsummen kognitiv Gesunder von allen Diagnosegruppen signifikant voneinander (MD: $Z = -2,7$; $p \leq 0,01$; LKB: $Z = -3,2$; $p \leq 0,01$; leichte AD: $Z = -5,4$; $p \leq 0,001$; mittelschwere AD: $Z = -5,8$; $p \leq 0,001$). Eine Abgrenzung einer leichten AD von einer LKB ($Z = -4,0$; $p \leq 0,001$) und MD ($Z = -3,9$; $p \leq 0,001$) war gegeben, ebenso die Abgrenzung einer leichten von einer mittelschweren AD ($Z = -4,7$; $p \leq 0,001$). Eine Differenzierung zwischen Patienten mit einer LKB und einer Depression war dagegen über den Uhrentest nicht möglich.

Diskussion

Das Ziel der vorliegenden Studie war eine Analyse der Diskriminationsfähigkeit der deutschsprachigen CERAD-NP-Testbatterie zur Unterscheidung zwischen gesunden Kontrollen, Depressiven, Patienten mit einer LKB, einer leichten sowie einer mittelschweren AD. Untersucht werden sollte dabei, welche der acht Untertests die höchste Trennschärfe in Abhängigkeit von der Diagnose bzw. dem Erkrankungsverlauf aufweisen.

Alle CERAD-NP-Untertests unterschieden signifikant zwischen kognitiv gesunden Probanden und Patienten mit einer leichten bzw. mittelschweren AD. Die Diskriminationsfähigkeit, die über die Berechnung von ROC-Kurven ermittelt wurde, war für alle Subtests bedeutsam. Die höchste Diskriminationsfähigkeit wurde jeweils für die Untertests zum Episodischen Gedächtnis (verzögerte und unmittelbare Merkfähigkeit) als Teil expliziter Gedächtnisfunktionen erzielt. Im Vergleich zu den alters- und geschlechtsabhängigen Normen lagen Patienten mit leichter AD mit Ausnahme der visuokonstruktiven Fähigkeiten mindestens um eine Standardabweichung, mittelschwer Demente in allen Subskalen mindestens um zwei Standardabweichungen unterhalb der Mittelwerte der Normstichprobe. Tatsächlich bestätigt die gute Differenzierungsfähigkeit zwischen gesunden Kontrollen und Demenzpatienten die Ergebnisse früherer Studien mit der deutschsprachigen [6, 7] und der englischsprachigen [4] Version der CERAD-NP.

Die differenzialdiagnostisch wichtige Differenzierung zwischen LKB und altersgemäßer kognitiver Leistungsfähigkeit war über den Untertest Verbale Flüssigkeit und die Aufgaben zum Episodischen Gedächtnis (Wortliste Gedächtnis, Wortliste Abrufen, Konstruktive Praxis Abrufen) möglich, und zwar mit einer diagnostischen Genauigkeit von mindestens 75%. Die standardisierten z-Werte waren in diesen Funktionen < -1 und lagen damit unterhalb des Schwellenwertes. Der Untersuchung des Episodischen Gedächtnisses kommt gerade in der Früherkennung einer kognitiven Störung eine hohe Bedeutung zu [20]. Elias u. Mitarb. [21] konnten in einer prospektiven Studie den Nachweis erbringen, dass bereits 10 Jahre vor Erstmanifestation einer AD Leistungseinbußen im Episodischen Gedächtnis (verzögerte freie Reproduktion) und des abstrakten Verständnisses bestanden; innerhalb dieser langen präklinischen Phase blieben die Leistungseinbußen allerdings relativ stabil. Störungen der auditiven Kurzzeitgedächtnisspanne wurden dagegen in der präklinischen Phase nicht beobachtet [22]. Die Beeinträchtigung des Episodischen Gedächtnisses korrespondiert mit frühen Veränderungen im Hippokampuskomplex und anderer Strukturen im medialen Temporallappen, wie sie mit der Magnetresonanztomographie (MRT) bei der LKB und der AD nachgewiesen wurden [23–25].

Mit Ausnahme der Konstruktiven Praxis, in der auch Patienten mit leichter Demenz noch unauffällige Ergebnisse erreichten, war eine Differenzierung zwischen LKB und leichter AD in allen Untertests gegeben. Eventuell ist der Schwierigkeitsgrad der Aufgaben zur Konstruktiven Praxis in der CERAD-NP zu niedrig, so dass AD-Patienten trotz bekannter früher parietaler Schädigungen noch altersgemäße Ergebnisse erreichen können und die Aufgaben erst in einem weiter fortgeschrittenen Stadium valide trennen. Hierfür spricht, dass eine eigene Untersuchung des

Vierjahresverlaufs der LKB Störungen der Wortflüssigkeit und des räumlichen Vorstellungsvermögens bei einem erheblichen Anteil der Betroffenen erbrachte [26].

Depressive Störungen können mit kognitiven Defiziten assoziiert sein, die im Sinne einer depressiven Pseudodemenz nur schwer von kognitiven Defiziten bei leichter Demenz unterscheidbar sind [27, 28]. In der vorliegenden Studie waren depressive Patienten im Vergleich zu gesunden Kontrollen in den Subtests Wortliste Gedächtnis unmittelbar und Konstruktive Praxis Abrufen beeinträchtigt. Dies steht teilweise in Übereinstimmung mit Ergebnissen von Lachner und Engel [29], die bei depressiven Patienten überwiegend Beeinträchtigungen in der verzögerten Merkfähigkeit und der Wortflüssigkeit finden. Weiterhin beschreiben die Autoren Defizite in der Aufmerksamkeit und in den Exekutivfunktionen sowie in den tempoabhängigen Verfahren, die sich in unserer Studie in einer geringeren Leistung im TMT Teil A und B abbildeten. Paradiso u. Mitarb. [30] vermuten, dass die Störungen der Merkfähigkeit und der Exekutivfunktionen durch Aufmerksamkeitsdefizite und psychomotorische Verlangsamung ausreichend erklärbar sind.

Eine Unterscheidung zwischen LKB und Depression war dagegen in der einfaktoriellen multivariaten Varianzanalyse nicht möglich. Die Ergänzung um den Uhrentest sowie den TMT A und B erbrachte keinen weiteren Vorteil. Dagegen erhielten LKB-Patienten signifikant negativere Einschätzungen in der Global Deterioration Scale (Chi-Quadrat 19,9; df 3; $p \leq 0,000$), die trotz gleichbarer Testergebnisse in den neuropsychologischen Untersuchungsverfahren auf einen höheren Grad der kognitiv bedingten Alltagsbeeinträchtigungen im Vergleich zu den depressiven Patienten schließen lassen.

Eine Trennung zwischen depressiven Patienten und Patienten mit leichter AD war mit Ausnahme der Wortliste Wiedererkennen Treffer und der Konstruktiven Praxis in allen Parametern möglich. In der verzögerten Wiedererkennenaufgabe lagen Depressive im Erkennen bereits dargebotener Begriffe mit einem z-Wert von $-1,1$ im unterdurchschnittlichen Bereich, während die korrekte Zurückweisung noch nicht dargebotener Begriffe altersentsprechend war. Gainotti und Marra [31] bestätigen, dass Demenzpatienten in einer Wiedererkennenaufgabe mehr falsch positive Fehler im Sinne eines „liberal response bias“ machen, während Depressive eher zu falsch negativen Aussagen tendieren und damit bereits dargebotene Begriffe als unbekannt klassifizieren. Dies konnte in der vorliegenden Studie bestätigt werden. Die Beeinträchtigungen der Konstruktiven Praxis und der verzögerten Wiedererkennensleistung stehen in Übereinstimmung mit den Ergebnissen von Satzger u. Mitarb. [6], die vorschlagen, diese Parameter bei depressiven Patienten geringer zu gewichten.

Eine Unterscheidung zwischen leichter und mittelschwerer AD war über die Subtests Verbale Flüssigkeit und Konstruktive Praxis möglich, während in den Verfahren zum Episodischen Gedächtnis und zur Wortfindung bereits im frühen Erkrankungsstadium Bodeneffekte auftraten und somit eine Differenzierung zwischen den Demenzstadien nicht möglich war. Dies korrespondiert mit den Ergebnissen früherer Studien [8, 32], die auf die hohe Bedeutung der Gedächtnisparameter in der Früherken-

nung einer Demenz hinweisen, in denen andere Defizite noch nicht auftraten, während die Bedeutung in der Verlaufsdiagnostik bei bekannter Demenz abnimmt. Im späteren Verlauf sind nach Chen u. Mitarb. [32] und in Übereinstimmung mit unseren Ergebnissen visuokonstruktive und sprachliche Fähigkeiten zusätzlich beeinträchtigt, die eine Differenzierung zwischen beginnender und mittelschwerer Demenz erlauben. Die Ergänzung der CERAD-NP-Testbatterie um zusätzliche Verfahren, die auch fortgeschrittene Erkrankungsstadien valide trennen, sollte hier bedacht werden. Der Uhrentest sowie die Global Deterioration Scale sind gerade zur Bestimmung des Schweregrades einer Demenz von hoher differenzialdiagnostischer Bedeutung; In der vorliegenden Untersuchung zeigten Patienten mit mittelschwerer AD in beiden Skalen signifikant geringere Leistungen verglichen mit leicht Dementen. Der TMT Teil A als Maß für die Informationsverarbeitungsgeschwindigkeit sowie Teil B zur Erhebung der kognitiven Komplexität und Umstellungsvermögen waren ebenfalls von hoher diskriminativer Bedeutung.

Zusammenfassend zeigt die vorliegende Studie, dass die CERAD-NP-Testbatterie valide zwischen gesunden Kontrollen, Patienten mit leichter kognitiver Beeinträchtigung, beginnender und mittelschwerer AD trennen kann und somit sowohl zur Frühdiagnose einer AD als auch zur Evaluation des Erkrankungsverlaufs geeignet ist. Eine Unterscheidung zwischen einer LKB und affektiv bedingten kognitiven Beeinträchtigung kann dagegen nicht sicher getroffen werden. Allerdings gelingt die Abgrenzung depressiver Patienten zu gesunden Kontrollen und Patienten mit beginnender Demenz, d. h. eine Unterscheidung zwischen affektiv bedingten kognitiven Leistungseinbußen und einer demenziellen Entwicklung ist möglich. Die Auswertungen ergaben weiterhin, dass auf keinen Untertest des CERAD-NP verzichtet werden sollte. Die Aufgaben zum Episodischen Gedächtnis trennen zwar schon für sich genommen zwischen gesunden Kontrollen, LKB und leichter AD, zeigen allerdings früh Bodeneffekte und sind daher nicht sensitiv für Veränderungen im Verlauf [7,33]. Eine differenzierte Abklärung des weiteren Erkrankungsverlaufs ist mit den Untertests Verbale Flüssigkeit, Boston Naming Test und Konstruktive Praxis möglich. Eine Kritik an der Testbatterie betrifft das Fehlen eines Verfahrens zur Erfassung der Exekutivfunktionen und der Informationsverarbeitungsgeschwindigkeit, so dass eine Ergänzung um den TMT als sinnvoll erachtet werden kann. Zur Differenzierung zwischen leichter und mittelschwerer Demenz eignet sich der Uhrentest.

Literatur

- Schröder J, Pantel J, Förstl H. Demenzielle Erkrankungen – Ein Überblick. In: Kruse A, Martin M (Hrsg). Enzyklopädie der Gerontologie. Zürich: Huber, im Druck,
- Seidl U, Pantel J, Re S, Schröder J. Depressive Störung und Spätdepression. In: Kruse A, Martin M (Hrsg). Enzyklopädie der Gerontologie. Zürich: Huber, im Druck,
- Heyman A, Fillenbaum GG. Overview: clinical sites, case material, and special studies. *Neurology* 1997; 49: 2–6
- Morris JC, Heyman A, Mohs RC, Hughes JP, Belle G van, Fillenbaum G et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD-NP). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* 1989; 39: 1159–1165
- Welsh KA, Butters N, Mohs RC, Beekly D, Edland S, Fillenbaum G, Heyman A. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD-NP). Part V. A normative study of the neuropsychologic battery. *Neurology* 1994; 44: 609–614
- Satzger W, Hampel H, Padberg F, Bürger K, Nolde T, Ingrassia G, Engel RR. Zur praktischen Anwendung der CERAD-NP-Testbatterie als neuropsychologisches Demenzscreening. *Nervenarzt* 2001; 72: 196–203
- Aebi C. Validierung der neuropsychologischen Testbatterie CERAD-NP: eine Multi-Center-Studie. Basel: Dissertation, Universität Basel, 2002
- Welsh K, Butters N, Hughes J, Mohs R, Heyman A. Detection of abnormal memory decline in mild cases of Alzheimer's disease using CERAD-NP neuropsychological measures. *Arch Neurol* 1991; 48: 278–281
- Welsh KA, Butters N, Hughes JP, Mohs RC, Heyman A. Detection and staging of dementia in Alzheimer's Disease: Use of the neuropsychological measures developed for the Consortium to Establish a Registry for Alzheimer's Disease. *Arch Neurol* 1992; 49: 448–452
- Härting C, Markowitsch H-J, Neufeld H, Calabrese P, Deisinger K, Kessler J. Wechsler Gedächtnistest – Revidierte Fassung. Göttingen: Huber, 2000
- Shulman K, Sheldetsky R, Silver IL. The challenge of time: Clock-drawing and cognitive function in the elderly. *Int J Geriatr Psychiatry* 1986; 1: 135–140
- Spree O, Strauss E. A compendium of neuropsychological tests. New York: Oxford University Press, 1991
- Reisberg B, Ferris SH, Leon MJ, Crook T de. The Global Deterioration Scale (GDS): an instrument for the assessment of Primary Degenerative Dementia (PDD). *Am Psychiat* 1982; 139: 1135–1139
- Yesavage JA, Sheikh JL. Geriatric depression scale (GDS) short form. In: Shulz SC. (1998) Assessment in diagnosis and treatment of gerontopsychiatric patients. *Psychopharmacol Bull*, 1986; 24: 709–711
- Hautzinger M, Bailer M, Worall H, Keller F. Beck-Depressions-Inventar. Testhandbuch. Göttingen: Huber, 1994
- McKhann G, Drachmann D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 1984; 34: 939–944
- Levy R. Aging-associated cognitive decline. *International Psychogeriatrics* 1994; 6: 63–68
- Thalmann B, Monsch AU. CERAD-NP – Neuropsychologische Testbatterie: Vorläufige Normen. Basel: Memory Clinic, 1997
- Fisseni H-J. Lehrbuch der psychologischen Diagnostik. Göttingen: Verlag für Psychologie, 1990
- Small BJ, Mobly JL, Laukka EJ, Jones S, Bäckman L. Cognitive deficits in preclinical Alzheimer's disease. *Acta Neurol Scand* 2003; 107: 29–33
- Elias MF, Beiser A, Wolf PA, Au R, White RF, D'Agostino RB. The preclinical phase of Alzheimer disease. A 22-year prospective study of the Framingham Cohort. *Arch Neurol* 2000; 57: 808–813
- Bäckman L, Small BJ, Fratiglioni L. Stability of the preclinical episodic memory deficit in Alzheimer's disease. *Brain* 2001; 124: 96–102
- Schröder J, Pantel J. Morphologische und funktionelle Bildgebung. In: Förstl H. (Hrsg). Alzheimer Demenz. Berlin, Heidelberg, New York: Springer, 1999
- Pantel J, Kratz B, Essig M, Schröder J. Parahippocampal volume deficits in subjects with aging-associated cognitive decline. *Am J Psychiatry* 2003; 160: 379–382
- Schröder J, Buchsbaum MS, Shihabuddin L, Tang C, Wei T, Spiegel-Cohen J, Hazlett EA, Abel L, Luu-Hsia C, Ciaravolo TM, Marin D, Davis KL.

- Patterns of cortical activity and memory performance in Alzheimer's disease. *Biol Psychiatry* 2001; 49: 426–436
- ²⁶ Schönknecht P, Pantel J, Kruse A, Schröder J. Prevalence and natural course of aging-associated cognitive decline in a population-based sample of "young-old" subjects. *American Journal of Psychiatry*, in Druck
- ²⁷ La-Rue A. *Aging and Neuropsychological Assessment*. New York: Plenum Press, 1992
- ²⁸ Caine ED. The neuropsychology of depression: The pseudodementia syndrome. In: Grant I, Adams KM (Hrsg). *Neuropsychological assessment of neuropsychiatric disorders*. New York: Oxford University Press, 1986: 221–243
- ²⁹ Lachner G, Engel RR. Differentiation of dementia and depression by memory tests: a meta-analysis. *J Nerv Ment Dis* 1994; 182: 34–39
- ³⁰ Paradiso S, Lamberty GJ, Garvey MJ, Robinson RG. Cognitive impairment in the euthymic phase of chronic unipolar depression. *J Nerv Ment Dis* 1997; 185: 748–754
- ³¹ Gainotti G, Marra C. Some aspects of memory disorders clearly distinguish dementia of the Alzheimer's type from depressive pseudodementia. *J Clin Exp Neuropsychol* 1994; 16: 65–78
- ³² Chen P, Ratcliff G, Belle SH, Cauley JA, DeKosky ST, Ganguli M. Cognitive tests that best discriminate between presymptomatic AD and those who remain nondemented. *Neurology* 2000; 55: 1847–1853
- ³³ Calabrese P. Neuropsychologie der Alzheimer-Demenz. In: Calabrese P, Förstl H (Hrsg). *Psychopathologie und Neuropsychologie der Demenzen*. Lengerich: Pabst Science Publishers Lengerich, 2000: 31–50