

The CERAD Neuropsychological Assessment Battery Is Sensitive to Alcohol-Related Cognitive Deficiencies in Elderly Patients: A Retrospective Matched Case-Control Study

Liane Kaufmann,¹ Stefan Huber,² Daniel Mayer,¹ Korbinian Moeller,^{2,3,4} AND Josef Marksteiner¹

¹Department of Psychiatry and Psychotherapy A, General Hospital Hall, Hall in Tyrol, Austria

²Junior Research Group Neuro-cognitive Plasticity, Leibniz Institut für Wissensmedien, Tuebingen, Germany

³Department of Psychology, University of Tuebingen, Tuebingen, Germany

⁴LEAD Graduate School and Research Network, University of Tuebingen, Tuebingen, Germany

(RECEIVED October 13, 2016; FINAL REVISION August 31, 2017; ACCEPTED September 2, 2017; FIRST PUBLISHED ONLINE November 6, 2017)

Abstract

Objectives: Adverse effects of heavy drinking on cognition have frequently been reported. In the present study, we systematically examined for the first time whether clinical neuropsychological assessments may be sensitive to alcohol abuse in elderly patients with suspected minor neurocognitive disorder. **Methods:** A total of 144 elderly with and without alcohol abuse (each group $n = 72$; mean age 66.7 years) were selected from a patient pool of $n = 738$ by applying propensity score matching (a statistical method allowing to match participants in experimental and control group by balancing various covariates to reduce selection bias). Accordingly, study groups were almost perfectly matched regarding age, education, gender, and Mini Mental State Examination score. Neuropsychological performance was measured using the CERAD (Consortium to Establish a Registry for Alzheimer's Disease). Classification analyses (i.e., decision tree and boosted trees models) were conducted to examine whether CERAD variables or total score contributed to group classification. **Results:** Decision tree models disclosed that groups could be reliably classified based on the CERAD variables "Word List Discriminability" (tapping verbal recognition memory, 64% classification accuracy) and "Trail Making Test A" (measuring visuo-motor speed, 59% classification accuracy). Boosted tree analyses further indicated the sensitivity of "Word List Recall" (measuring free verbal recall) for discriminating elderly with *versus* without a history of alcohol abuse. **Conclusions:** This indicates that specific CERAD variables seem to be sensitive to alcohol-related cognitive dysfunctions in elderly patients with suspected minor neurocognitive disorder. (*JINS*, 2018, 24, 360–371)

Keywords: Elderly, Minor neurocognitive disorder, Alcohol abuse, Neuropsychological performance, CERAD, Verbal recognition memory, Visuo-motor speed

INTRODUCTION

Chronic excessive alcohol consumption has been repeatedly reported to impair brain function and structure (Oscar-Berman & Marinkovic, 2007; Sullivan, Harris, & Pfefferbaum, 2010). As a consequence, there is an increased risk to develop alcohol-related dementia and alcohol-induced persisting amnesic syndrome (coined as Wernicke-Korsakoff syndrome; Ridley, Draper, & Withall, 2013). Neurobehavioral impairments are most pronounced in the subgroup of severe alcoholics suffering from the so-called Korsakoff syndrome, which may be a manifestation of Wernicke's encephalopathy (characterized by a

dense and persistent amnesic disorder due to severe thiamine deficiency; for overviews, see Butters, 1985; Oscar-Berman & Marinkovic, 2007). Nonetheless, also in non-Korsakoff alcoholics, neurocognitive deficiencies have been reported to be influenced by age, duration of alcoholism (and abstinence), amount and type of consumed alcohol.

Neurocognitive Sequelae of Alcoholism

Cognitive impairments associated with excessive alcohol intake mainly affect higher-order cognitive functions such as executive and (episodic) memory functions (Bernardin, Maheut-Bosser, & Paille, 2014; Sinforiani et al., 2011). The term "executive functions" refers to a set of quite heterogeneous functions such as inhibitory control, working memory, cognitive flexibility, fluency and problem solving (Friedman & Miyake, 2016).

Correspondence and reprint requests to: Liane Kaufmann, General Hospital Hall, Department of Psychiatry and Psychotherapy A, A-6060 Hall in Tyrol, Austria. E-mail: liane.kaufmann@tirol-kliniken.at

Impairments in these functions have been attributed to dysfunctions of frontocerebellar circuits (e.g., Oscar-Berman et al., 2014). Moreover, alcohol abuse has also been found to harm episodic memory (e.g., Beydoun et al., 2014) and related fronto-temporal brain structures.

However, brain damage caused by alcoholism may be rather widespread, affecting also the limbic system (including hippocampal structures) and the cerebellum (associated with motor/behavioral functioning, respectively; cf. Dupuy & Chanaud, 2016; Oscar-Berman & Marinkovic, 2007). In sum, chronic excessive alcohol intake has been frequently associated with poor performance on tasks tapping visuospatial perception (e.g., embedded figures), construction (copying), and (psycho)motor functions (e.g., eye–hand coordination, stability in gait and balance, speeded performance; Sullivan et al., 2010).

Alcohol, Age, and Dementia

Several recent systematic reviews and meta-analyses have been targeted at examining whether alcohol consumption has any influence on incident cognitive decline or dementia (Anstey, Mack & Cherbuin, 2009; Peters, Peters, Warner, Beckett, & Bulpitt, 2008). However, findings are somewhat controversial as far as the effects of alcohol consumption on vascular dementia and cognitive decline in general are concerned (Panza et al., 2012). Notably, the findings of a recent systematic review disclose that cognitive impairments in alcohol-related dementia mainly concern executive, memory, and visuospatial functions (Ridley et al., 2013). However, from previous studies, it remains unclear whether cognitive profiles of elderly with and without a history of alcohol abuse can be differentiated upon using standardized neuropsychological tests such as the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) (Morris et al., 1989).

Study Rationale

Although lasting alcohol abstinence may yield cognitive recovery, subtle cognitive impairments may persist and interfere with every-day-life activities and treatment compliance (thus enhancing the risk of relapse; Panza et al., 2012; Sullivan et al., 2010). Hence, there is a clear need for sensitive neuropsychological assessments suited to detect even subtle cognitive impairments in affected individuals. The main objectives of the present retrospective study were (i) to compare cognitive performance in elderly patients with (suspected) minor neurocognitive disorder (according to clinical psychiatric diagnosis based on DSM-V criteria; APA, 2013), with or without a history of alcohol abuse; and (ii) to evaluate whether a standard neuropsychological test battery administered to the elderly (i.e., CERAD) is sensitive to (possibly subtle) cognitive deficiencies induced by chronic alcohol abuse. Importantly, different from previous studies we used propensity score matching to match our study groups on possible confounds such as age, gender, education, and overall cognitive performance.

METHODS

Participants

Patient data were drawn from geriatric out- and inpatient services of a psychiatric general hospital. Please note that in Austria (and Germany alike), each kind of substance abuse, including alcohol dependency, is treated in general psychiatric hospitals because no special hospitals for the treatment of alcohol dependency exist in the public health system. Of a patient pool of $n = 738$ (thereof 78 elderly with and 660 without alcohol abuse), 144 elderly were included in the final data set (each group $n = 72$). All patients were referred for clinical neuropsychological examinations to assess whether cognitive performance warranted a diagnosis of minor neurocognitive disorder due to Alzheimer's disease or other forms of incipient dementias (APA, 2013; formerly called minimal cerebral dysfunction; APA, 1994).

Neuropsychological assessments were conducted between 2009 and 2015. Inclusion criteria for participation in the present retrospective study were (i) age above 55 years; (ii) German as first language; (iii) at least 7 years of formal education; (iv) Mini-Mental State Examination (MMSE) (Folstein, Folstein, & McHugh, 1975) score of ≥ 24 ; (v) absence of neurological disorders such as severe brain injury and neoplastic brain processes; (vi) absence of Wernicke encephalopathy as defined by clinical operational criteria (Caine, Halliday, Kril, & Harper, 1997); and (vii) absence of severe psychiatric disorders including major depressive disorders (see participant characteristics, Table 1).

Diagnoses of (suspected) minor neurocognitive disorder were based on clinical psychiatric examination in accordance with the rules and regulations of the APA (2013) and substantiated by the results of a standard rating instrument (i.e., Clinical Dementia Rating/CDR, Hughes, Berg, Danziger, Coben, & Martin, 1982; CDR box scores of 0.5 or 1 obtained by certified clinicians). Study participants with a recent history of alcohol abuse underwent neuropsychological examinations after alcohol abstinence of at least 10 days (but not more than 3 months) and reported a drinking history of >10 years with an alcohol consumption exceeding 100 g per day. Considering recent suggestions proposing that moderate drinking equals to no more than one or two drinks per day (for women and men, respectively; U.S. Department of Health and Human Services, 2010), corresponding to approximately 30 g of alcohol (Kerr et al., 2005; cited in Panza et al., 2012), our elderly patients can be considered heavy drinkers.

Tests Used

Neuropsychological functioning was measured using the CERAD (Morris et al., 1989). Here, we used the validated German version of the CERAD (Thalman et al., 2000). The CERAD is a standard clinical neuropsychological tool used to diagnose dementia which was repeatedly reported to be highly predictive for the early identification of and differential diagnosis of Alzheimer's disease, among others

Table 1. Patient characteristics

Matching variables	Elderly with alcohol abuse (<i>n</i> = 72)	Elderly without alcohol abuse (<i>n</i> = 72)
Age (in years)	66.71 (6.20)	66.75 (5.95)
Education (in years)	10.76 (2.09)	10.35 (1.86)
Sex (female)	40.28%	40.28%
MMSE (mean raw score, max. 30)	28.06 (1.59)	28.04 (1.83)
Disease-related variables ^a (no. of patients)		
Cerebrovascular incidents	37	35
Cardiovascular incidents	39	15
Coronary heart diseases	6	4
Cardiac insufficiency	1	2
Atrial fibrillation	6	6
Myocardial infarction	2	0
Stent or pacemaker implantation	3	8
Stroke/apoplexy	2	6
Transient ischemic attack	3	2
Subcortical arteriosclerotic encephalopathy	1	4
Hypertension	31	33
Diabetes mellitus type 2	11	10
COPD	10	8
Compensated cirrhoses of the liver	10	0
Benzodiazepine use and other dependencies (except alcohol)	20	8
Affective disorders ^o	40	39

^aAs derived from hospital patient charts; ^oexcluding major depressive disorders (see the Methods section).

(Mayer et al., 2015; Schmid, Ehrensperger, Berres, Beck, & Monsch, 2014; Wolfsgruber et al., 2014). Subtests of the CERAD are well-established neuropsychological tests adapted to the elderly. The CERAD is comprised of subtests that were shown to be sensitive to common cognitive impairments associated with Alzheimer's disease.

Subtests of the CERAD core battery are (in the order of administration): (1) Verbal Fluency Animals (tapping verbal semantic association skills; sum score of named animals), (2) Modified Boston Naming Test (BNT, measuring confrontation naming and thus, word finding difficulties; max. score of 15), (3) MMSE (cognitive screening; max. score of 30), (4) Word List Learning (sum of three learning trials; max. score of 30), (5) Constructional Praxis Copy (tapping visuo-constructional skills; max. score of 11), (6) Word List Recall (measuring verbal recall memory; max. score of 10), (7) Word List Discriminability (measuring verbal recognition memory; max. score of 10), (8) Constructional Praxis Recall (measuring figural memory, max. score of 11). Moreover, three additional subtests include the so-called CERAD-Plus tests (again in the order of test administration): (9) and (10) Trail Making Tests A and B (tapping visuo-motor speed and cognitive flexibility, respectively; scored is the processing time in seconds) as well as (11) Verbal Fluency S-Words (measuring verbal phonological association skills).

All subtests were administered according to standard procedures. For all subtests, raw scores were converted to T-scores ($M = 50$ and $SD = 10$) upon using the automated analysis program provided by the memory clinic of the

University Hospital Basel, Switzerland (<https://www.memoryclinic.ch/de/main-navigation/neuropsychologen/cerad-plus/auswertungprogramme/cerad-plus-10-excel/>). Moreover, we computed a total score according to the method proposed by Chandler et al. (2005). The Chandler total score (CTS) is derived by summing up the raw scores of all but two subtests of the CERAD core battery (i.e., excluding MMSE and Constructional Praxis Recall), with the Verbal Fluency raw score being limited to a maximum score of 24 and the Word List Discriminability raw score being calculated by subtracting the number of false positives from the number of true positives (each $n = 10$). Notably, subtests of the CERAD-Plus are not considered for calculating the CTS. The CTS has a maximum of 100 raw points. For each study participant, the CTS raw score was converted to a T-score.

Furthermore, to evaluate whether potential group differences regarding depressive mood might confound neuropsychological performance in our patients (e.g., DeFrancesco, Marksteiner, Deisenhammer, Hinterhuber, & Weiss, 2009), we also report patients' data on the 15-item short version of the Geriatric Depression Scale (GDS, Yesavage et al., 1983; for the German standardization, see Gauggel & Birkner, 1999). The GDS is based on self-reports and examines depressive symptomatology frequently observed in the elderly. Because GDS data were not available for all participants (elderly with alcohol abuse $n = 62$ vs. those without alcohol abuse $n = 38$), GDS could not be used as a matching variable.

Analyses

Propensity score matching was applied to match the groups of elderly with and without alcohol abuse based on age, years of education, gender, and overall cognitive ability (indexed by the MMSE; Folstein et al., 1975). We drew random subsamples of the experimental group (i.e., elderly with alcohol abuse) consisting of 70 up to 77 participants to increase the likelihood of finding a matched subsample of the control group (i.e., elderly without alcohol abuse). This process was repeated until the difference between mean ages as well as mean years of education was less than 0.5 years, the proportion of female and male participants differed by less than 1% and MMSE score differed by less than 0.1 between groups. To find matched groups, we used the programming language R (R Core Team, 2016) and the R package MatchIt (Ho, Imai, King, & Stuart, 2011), which implements various methods for matching (for a step-by-step guide for MatchIt, see Randolph, Falbe, Manuel, & Balloun, 2014). Applying this procedure, we found almost perfectly matched groups with 72 participants per group (Table 1).

To investigate potential sensitivity of standardized scores of the CERAD variables and the standardized CTS for the group differentiation between elderly with *versus* without alcohol abuse we used the following three-step procedure.

First, we compared the standardized scores by doing a Welch's *t* test, separately for each CERAD variable. *p*-Values were corrected for multiple testing by controlling the false discovery rate (Benjamini & Hochberg, 1995).

Second, we investigated how good the standardized scores of the CERAD variables were at differentiating between the two groups by using decision tree learning, more specifically classification trees that fall within the family of tree-based models. Classification trees are tree-like decision models commonly used in predictive modeling or data mining. They consist of a root node and interior nodes corresponding to splitting attributes and terminal nodes corresponding to a particular decision. Edges denote the outcome of an attribute test of a splitting attribute and are labeled with disjoint ranges.

Decision trees are built by recursively splitting attributes based on a splitting criterion. In the present study, we used the GINI index as splitting criterion (Gini, 1921). Classification trees were generated separately for each CERAD variable using the R package rpart (Therneau, Atkinson, & Ripley, 2015). To avoid overfitting of the data, we restricted the maximal depth of the trees to three, and the minimum number of participants required in a node in order for a split to be attempted to 10 participants. Moreover, we pruned the trees based on the complexity parameter for which the cross-validated error was smallest, available *via* the R function printcp from the R package rpart. Classification accuracy was evaluated by applying a cross-validation procedure: We randomly split the data 1000 times into a training set containing 128 participants and a test set containing 16 participants with 50% of the participants from each group.

Then, classification accuracy was averaged over 1000 runs. Thereafter, we evaluated whether the obtained average

classification accuracy was above chance level by running a binomial test using R (R Core Team, 2016) for each CERAD variable separately. Again, *p*-values were corrected for multiple testing using the procedure suggested by Benjamini and Hochberg (1995). Classification trees were drawn using the R package rpart.plot (Milborrow, 2016). For other figures, we used the R package ggplot2 (Wickham, 2009).

Third, we sought to investigate how well we can classify patients when using all CERAD variables as features. Therefore, we generated gradient boosted instead of decision trees by using the R package xgboost (Chen & Guestrin, 2016; Chen, He, & Benesty, 2016). Thus, in contrast to usual decision tree models, decisions are not based on a single tree, but on several decision trees (i.e., random forest). The decision trees select a random subset of the features (one of the CERAD variables) at each candidate split in the learning process. In case a feature is a very strong predictor for the output, it will be in many trees of the random forest. Thereby, boosted decision tree models enable us to measure *feature importance*.

In our case, the outputs were the two groups (elderly with and without alcohol abuse) and features were the CERAD variables. Furthermore, we used grid search to tune hyperparameters. Therefore, we fixed the parameters $\eta = 0.01$ (the learning rate), $\gamma = 0.01$ (the minimum loss reduction) and number of rounds = 75, but tried to find the best values for subsample ratio, subsample ratio of the columns, and the maximum depth of a tree. For subsample ratio, we considered the values 0.5, 0.75, and 0.9, for subsample ratio of the columns 0.5, 0.75, and 1 and for the maximum depth the values 3, 4, and 5. Ultimately, training error was minimal upon using 10-fold cross validation as well as a subsample ratio of 0.75, a subsample ratio of the columns of 1 and a max depth of 4.

Taken together, this three-step approach allowed us to investigate whether elderly with and without alcohol abuse can be differentiated regarding their cognitive performance. In particular, we aimed to investigate (i) whether a *single* CERAD variable can be used to classify the two groups above chance (decision trees) and (ii) which CERAD variables are most important when classifying the two groups based on *all* features (boosted trees).

Effect sizes are reported where appropriate to facilitate interpretation of results, in particular considering (i) methodological differences between studies making it difficult if not impossible to compare results otherwise and (ii) the potential tendency to yield underpowered results because despite strict matching criteria some heterogeneity may persist in clinical studies regarding disease-related variables. Importantly, we adhered to Ferguson's (2009) recommendations for interpreting effect sizes that are more conservative than those reported by Cohen (1992). Thus, non-significant results will be discussed only when effect sizes indicated the recommended minimum effect size representing a "practically" significant effect according to Ferguson (2009): effects may be considered practically relevant when Cohen's $d > .4$ (corresponding to a r of $> .2$ and to a $\eta^2/\text{adjusted } R^2$ of $> .04$).

The present study was conducted in compliance with the declaration of Helsinki and was approved by the local ethical committee of Innsbruck Medical University.

RESULTS

Descriptive statistics of CERAD variables and the CTS are given in Table 2. Results revealed significant group differences for the “Trail Making Test A” ($t(139.18) = -2.39$; $p = .018$; $d = .40$) and “Trail Making Test B” ($t(139.44) = -2.34$; $p = .021$; $d = .39$), indicating better performance of the control group (i.e., elderly without alcohol abuse) in the Trail Making Tests. However, group differences were no longer significant after correcting for multiple testing (both $p = .113$).

Notably, study groups did not differ significantly regarding the severity of cognitive impairment as indexed by the number of CERAD variables with a Z-score < -2 (elderly with alcohol abuse $M = 1.65$; $SD = 1.79$ vs. those without alcohol abuse $M = 1.33$; $SD = 1.62$; $t(140.66) = 1.12$; $p = .263$; $d = .19$). Likewise, after correcting for multiple testing, sex was not found to be a significant moderator of our results (Appendix A; adjusted $p > .174$). Nonetheless, though not reaching significance, effect size estimations indicate that sex was a practically significant moderator of performance on the CERAD subtest “Word List Discriminability” ($\eta^2_G = .042$).

Finally, group differences were not significant regarding self-reported depressive symptomatology (GDS mean raw scores in elderly with vs. without alcohol abuse: $M = 5.4$; $SD = 3.9$ vs. $M = 5.1$; $SD = 4.4$, respectively); $t(70.73) = 0.24$; $p = .812$; $d = .05$). However, across all participants (and after correcting for multiple testing) the GDS score correlated significantly with the CTS ($r(98) = -.22$; $p = .029$), indicating that higher GDS scores (reflecting more depressive symptoms) were associated with worse cognitive performance. However, the GDS score did not correlate significantly with any of the CERAD variables after correcting for multiple testing (neither

Table 3. Mean accuracies of cross-validation runs of classification tree models with standardized Z scores of CERAD subtests as attributes and p -values as well as adjusted p -values of binomial tests (against chance level)

Test variable	Accuracy (%)	p -Value	P_{adj}
Verbal Fluency	45	.211	.613
Boston Naming Test	46	.359	.659
Word List Learning	49	.934	1.000
Word List Recall	53	.560	.880
Word List Intrusions	50	1.000	1.000
Word List Discriminability	64	.001*	.012*
Constructional Praxis Copy	48	.803	1.000
Constructional Praxis Recall	55	.279	.613
Trail Making Test A	59	.037*	.203
Trail Making Test B	51	.934	1.000
Trail Making Test B/A	45	.279	.613

Note: p -Values were calculated by running binomial tests and corrected for multiple testing by controlling the false discovery rate (Benjamini & Hochberg, 1995).

*Significant at $p < .05$.

across all participants nor when calculated separately for the two study groups, all correlations $|r| < .3$).

Results of the classification trees are provided in Table 3. We were able to correctly classify a significant part of our patients upon using Z-scores for the CERAD variable “Word List Discriminability.” To inspect the classification tree for “Word List Discriminability,” we additionally computed it based on all data (Figure 1A). When using this classification tree, participants could be discriminated with 69% classification accuracy based on two splits: Patients were classified as belonging to the group with alcohol abuse, when their Z-score was between -0.85 and 0.86 , whereas patients were classified into the control group, when their Z-score lay outside this interval. Moreover, this classification tree yielded correct classifications for 61% of the elderly with alcohol abuse and 78% of the elderly without alcohol abuse.

Table 2. Means and SDs of standardized z-scores of the CERAD variables and the Chandler Total Score for both study groups

Test variable	Elderly with alcohol abuse ($n = 72$) Mean (SD)	Elderly without alcohol abuse ($n = 72$) Mean (SD)	p -Value (uncorr.)	p -Value (corr.)	Cohen's d
Verbal Fluency	-0.78 (0.88)	-0.65 (0.87)	.351	.644	.16
Boston Naming Test	0.35 (0.89)	0.33 (0.94)	.911	.930	.02
Word List Learning	-1.03 (1.27)	-1.26 (1.46)	.312	.644	.17
Word List Recall	-0.72 (1.38)	-0.70 (1.17)	.930	.930	.01
Word List Intrusions	0.30 (0.85)	0.25 (0.99)	.713	.872	.06
Word List Discriminability	-0.46 (1.37)	-0.73 (1.34)	.233	.640	.20
Constructional Praxis Copy	0.11 (1.15)	0.23 (1.00)	.499	.785	.11
Constructional Praxis Recall	-0.73 (1.59)	-0.38 (1.59)	.197	.640	.22
Trail Making Test A	-1.29 (1.11)	-0.81 (1.28)	.018*	.113	.40
Trail Making Test B	-1.13 (1.26)	-0.60 (1.45)	.021*	.113	.39
Trail Making Test B/A	-0.02 (1.14)	0.08 (1.10)	.584	.803	.09
Chandler Total Score	-1.55 (1.14)	-1.56 (1.22)	.966	.644	.01

Notes: uncorr. = not corrected for multiple testing; corr. = corrected for multiple testing.

*Significant at $p < .05$.

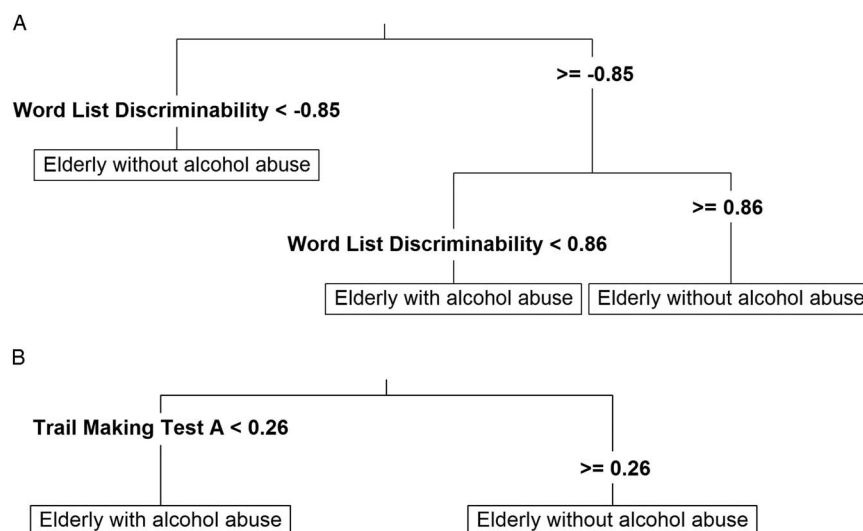


Fig. 1. Classification trees based on the CERAD variables “Word List Discriminability” (A) and “Trail Making Test A” (B).

Likewise, classification accuracy using the Z-scores of the “Trail Making Test A” was significantly above chance level before correcting for multiple testing. The classification tree indicated that study participants might be correctly classified into the two groups based on a binary split at a Z-score of 0.26 (Figure 1B). Using this split, 60% of all participants (94% of elderly with alcohol abuse and 26% of those without alcohol abuse) could be correctly classified.

Notably, cross-validated mean classification accuracy did not improve when including a subset of or all CERAD variables using classification trees. For instance, when including “Trail Making Test A” and “Word List Discriminability” as features, classification accuracy of decision trees with a maximum depth of 3, 4, 5, or 6 splits was 64% and hence, not better than the classification tree for “Word List Discriminability” as a single feature. Moreover, when including all CERAD variables, mean classification accuracy of decision trees with maximum depth of 3, 4, 5, or 6 splits was 58% and thus, even worse than relying only on “Word List Discriminability” exclusively.

However, we could improve classification accuracy to 65% when running gradient boosted trees and 10-fold cross validation considering all CERAD variables. In addition to “Word List Discriminability” and “Trail Making Test A,” also “Word List Recall” was among the three most important features (Figure 2). Hence, the latter three CERAD variables can be regarded as important features for classifying the two study groups. Moreover, when running xgboost using only these three features, classification accuracy increased to 68%.

Moreover, Figure 3 and Table 4 depict sensitivity, specificity, and receiver operating characteristic (ROC) curves for classification trees based on the CERAD variables “Word List Discriminability” (Fig. 3A) and “Trail Making Test A” (Fig. 3B) as well as boosted trees based on all CERAD variables (Fig. 3C). Areas under the curve (AUC) of classification trees based on the CERAD variables “Word List Discriminability” and “Trail Making Test A” as well as

boosted trees based on all CERAD variables were 0.69, 0.60, and 0.75, respectively.

DISCUSSION

The main purpose of the present retrospective matched case-control study was to examine whether the CERAD neuropsychological test battery may be sensitive to alcohol-related cognitive deficiencies in elderly patients with (suspected) minor neurocognitive disorder. As our patient pool comprised far more elderly without a history of alcohol abuse, propensity score matching was applied to match groups regarding several potential confounds such as age, education, gender, and MMSE (Table 1). Moreover, elderly with and without a history of alcohol abuse were comparable regarding CERAD (i.e., Chandler/CTS) total score and severity of cognitive impairment (indexed by the number of CERAD variables with a Z-score < -2 , indicating performance deficiencies of two *SD* below the normative sample). Results indicated that specific subtests of the CERAD seem to provide first indications for identifying patients with minor neurocognitive disorder due to a history of alcohol abuse.

In particular, our findings disclosed that, after correcting for multiple comparisons, group differences as examined by *t* tests did not become significant (Table 2). However, before controlling for multiple testing, two CERAD variables (i.e., “Trail Making Test A” and “Trail Making Test B”, tapping visuo-motor speed and cognitive flexibility, respectively) were found to differentiate between groups (favoring elderly without alcohol abuse, mean Z-score differences reaching .48 on Trails A and .53 on Trails B; see Table 2), thus suggesting considerable and according to Ferguson (2009) practically significant alcohol-related performance differences.

Of interest, upon using classification trees, our results revealed that patient groups could be classified based on two CERAD variables: “Word List Discriminability”

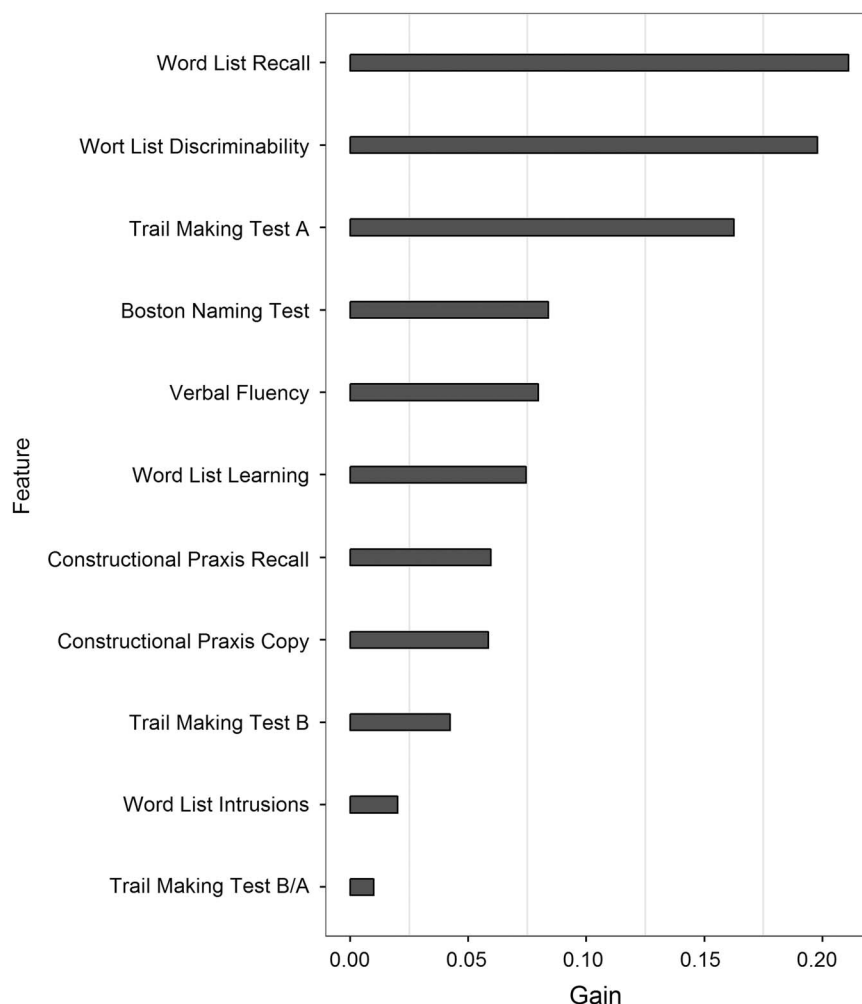


Fig. 2. Feature importance (indicating information gain) of the CERAD variables based on boosted trees.

(reaching a classification accuracy of 64%) and to a lesser degree “Trail Making Test A” (classification accuracy of 59%, significant only before controlling for multiple testing, Figure 1 and Table 3). These results were substantiated by the boosted tree analysis (Figure 3). Of interest, beyond the two aforementioned CERAD variables, “Word List Recall” (tapping verbal free recall memory) was found to contribute significantly to group classification when considering all CERAD variables as predictors in the boosted trees analysis (Figure 2). The latter results are largely consistent with previous findings suggesting that memory and psychomotor functions (including motor tasks with a speed factor) are detrimentally affected by excessive alcohol intake (Oscar-Berman et al., 2014; Sullivan et al., 2010).

The finding that decision tree models revealed only verbal recognition memory (indexed by the CERAD variable “Word List Discriminability”) to allow for significant group classification was unexpected. In particular so as another memory measure tapping active verbal recall (i.e., “Word List Recall”) did not contribute to group discrimination. A possible explanation for the latter discrepancy may be that task demands of the two CERAD variables call upon different levels of executive control mechanisms. Unlike in the

active memory recall task (“Word List Recall”) asking participants to recall as many words as possible from a previously learnt word list, verbal recognition memory requires participants to decide whether a presented word belonged to the previously learnt list of words or not. Notably, this places heavy demands on inhibitory control mechanisms known to be frequently deficient in alcoholics (Oscar-Berman et al., 2014; Panza et al., 2012; Sullivan et al., 2010). However, these differential aspects need to be interpreted carefully as verbal free recall was found to be a meaningful predictor for group classification in the boosted tree analysis (Figure 2).

Moreover, at first glance, it was rather unexpected that visuo-motor speed (“Trail Making Test A”) but not cognitive flexibility (a specific aspect of executive functions, and indexed by “Trail Making Test B” or the quotient “Trail Making Test A/Trail Making Test B”) seemed to be sensitive to alcohol-related cognitive impairments. However, it is important to note that the “Trail Making Test A” is a rather complex task requiring visual scanning and visuospatial processing (i.e., scanning randomly distributed letters in an attempt to allocate letters in the correct alphabetical order), planning and monitoring (i.e., which letter comes next in the alphabet), as well as visuo-motor coordination under speed

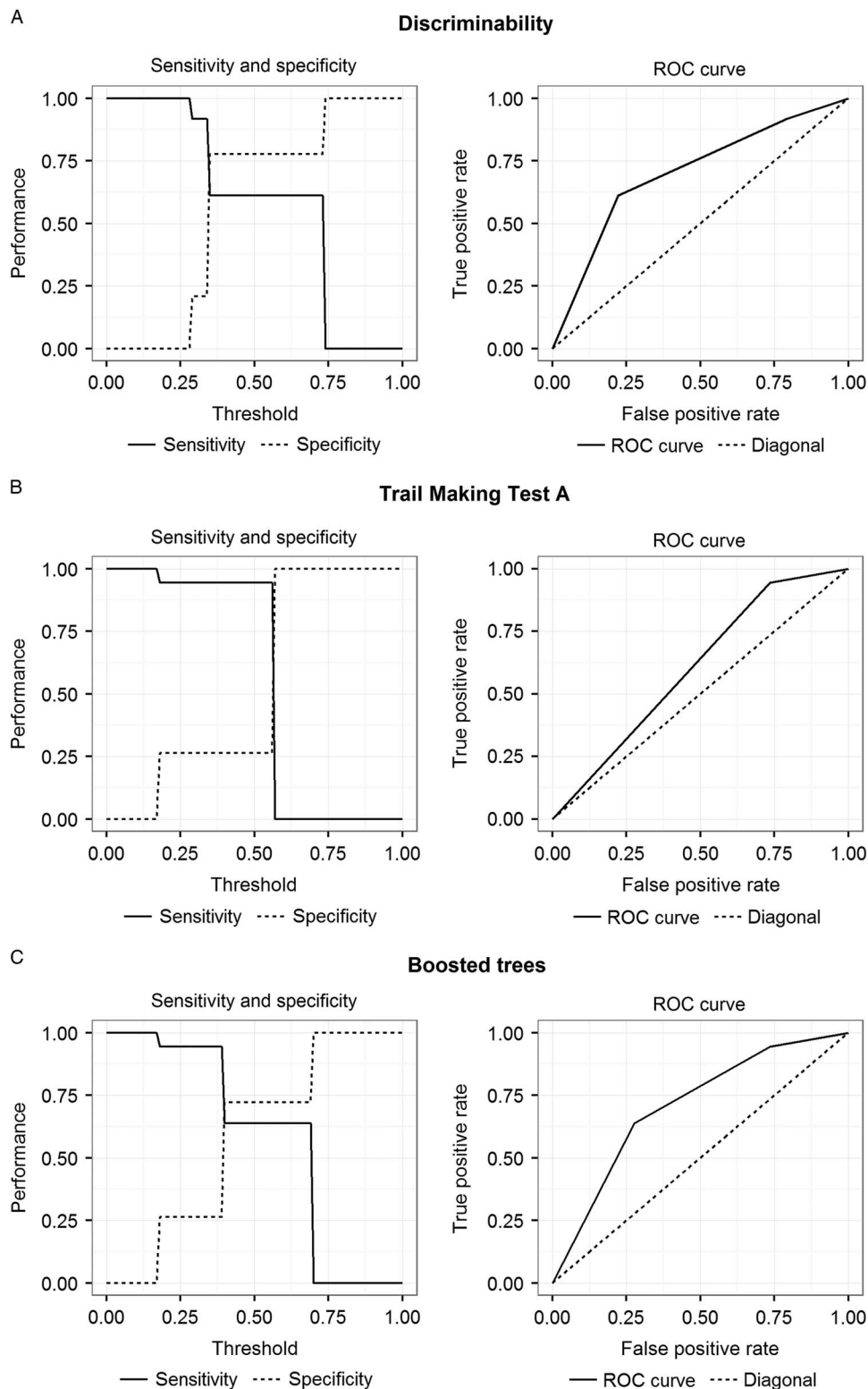


Fig. 3. Sensitivity, specificity, and ROC curves for classification trees based on the CERAD variables “Word List Discriminability” (A) and “Trail Making Test A” (B) and boosted trees based on all CERAD variables (C).

requirements (i.e., time taken to complete the test). Thus, poor performance on the “Trail Making Test A” nicely fits previous findings reporting alcohol-related impairments on visuospatial cognition, executive functions and (speeded) psychomotor abilities thought to originate from dysfunctions

in (parietal and) fronto-cerebellar brain regions (Butters, 1985; Oscar-Berman et al., 2014).

Altogether, our classification results based on classification tree and boosted tree models suggest that a combination of verbal free recall, verbal recognition memory and visuo-motor

Table 4. Sensitivity, specificity and accuracies of classification trees based on the CERAD variables “Word List Discriminability” and “Trail Making Test A” and boosted trees based on all CERAD variables for different thresholds

Model	Threshold	Sensitivity	Specificity	Accuracy
Decision tree:				
Word List Discriminability				
	0.00–0.28	1.00	0.00	0.50
	0.29–0.34	0.92	0.21	0.56
	<i>0.35–0.73</i>	<i>0.61</i>	<i>0.78</i>	<i>0.69</i>
	0.74–1.00	0.00	1.00	0.50
Decision tree:				
Trail Making Test A				
	0.00–0.17	1.00	0.00	0.50
	<i>0.18–0.56</i>	<i>0.94</i>	<i>0.26</i>	<i>0.60</i>
	0.57–1.00	0.00	1.00	0.50
Boosted trees:				
All CERAD subtests				
	0.00–0.17	1.00	0.00	0.50
	0.18–0.30	0.94	0.26	0.60
	<i>0.31–0.74</i>	<i>0.72</i>	<i>0.78</i>	<i>0.75</i>
	0.75–0.76	0.60	0.82	0.71
	0.77–1.00	0.00	1.00	0.50

Note. Best fit results are quoted in *italics*.

speed seem to be sensitive to alcohol-related performance specificities in our study group of elderly with (suspected) minor neurocognitive disorder. Notably, as our study groups were closely matched to demographic data, overall cognitive functioning and comorbid diseases, we are confident that the observed group differences reflect alcohol-related cognitive impairments. Our results are largely consistent with those reported by Bertoux and colleagues (2016) in patients suffering from the behavioral variant of frontotemporal dementia (bvFTD). Memory impairments of these patients were not solely mediated by fronto-executive dysfunctions (as indexed by encoding and strategic retrieval difficulties). Instead, they were also found to involve storage and consolidation processes being indicative of genuine (hippocampal type) memory impairments. Likewise, alcohol-related impairments in our study group were observed with regard to both subtests drawing on verbal free recall and recognition condition.

Furthermore, we found that the CERAD total score (CTS) was not sensitive to alcohol-related cognitive impairments in our study group. Rather, our patients exhibited comparable performance on the CTS irrespective of the absence or presence of a history of alcohol abuse. Previously, the CTS was reported to reliably detect (and even predict) patients suffering from Alzheimer’s disease (Wolfsgruber et al., 2014). However, patients examined in the latter study were considerably older than our study participants (mean age 84 vs. 67 years, respectively) and had a clinically confirmed diagnosis of Alzheimer’s disease while our patients suffered from (suspected) minor neurocognitive disorder.

Finally, it is important to acknowledge previously reported associations between (late-life) depression and cognitive dysfunction (DeFrancesco et al., 2009; Hesel et al., 2016;

Panza et al., 2010). Also in our study group (calculated across all participants and corrected for multiple testing), a higher GDS score (reflecting more self-reported depressive symptomatology) was significantly correlated with worse cognitive performance as indexed by a low CERAD total score. Notably, the strength of the latter correlation can be considered practically relevant according to Ferguson (2009).

On the contrary, in our study, group differences between elderly with *versus* without alcohol abuse regarding the GDS score were not significant. Additionally, GDS scores did not correlate significantly with any of the CERAD variables (neither across all participants nor when calculated separately for the two study groups). Because we were not able to retrieve GDS scores from all study participants, the latter findings should be interpreted carefully. Nevertheless, these results indicate that our main findings should not be driven by depressive mood but rather reflect alcohol-induced cognitive impairments in our elderly patients with (suspected) minor neurocognitive disorder.

A potential limitation of the present study may be missing information on which type of alcohol has been consumed. Previously, it has been claimed that, compared with spirits, wine and beer might have less detrimental effects on brain function and structure (Panza et al., 2012; Peters et al., 2008). A further limitation is that detailed information on drinking history was not available for each study participant. Importantly, however, all study participants reported to consume more than 100 g of alcohol per day and reported to have been drinking heavily for more than 10 years. Hence, our patients can clearly be considered heavy drinkers.

Furthermore, although rigorous inclusion/exclusion criteria and the applied propensity score matching procedure ensured that our study groups were as homogeneous as possible with regard to clinical comorbid disorders (including the presence of Korsakoff syndrome and/or Wernicke encephalopathy), groups differed regarding cardiovascular incidents, stroke history, and benzodiazepine use. In particular, stroke history was more frequently reported in elderly without alcohol abuse, while those with alcohol abuse had a higher incidence of cardiovascular episodes and benzodiazepine use. Although we were able to match our study groups on several critical variables, it has to be noted that it is almost impossible to avoid some heterogeneity on disease-related variables in clinical samples. Nonetheless, the general state of physical health may be considered comparable across our study groups (Table 1).

A clear asset of the present study is the application of a sophisticated matching method yielding closely matched groups. Thus, our results are not confounded by the matching variables. In particular, classification of elderly with alcohol abuse on verbal memory and visuo-motor speed is not attributable to age- or education-related group differences (Panza et al., 2012; Sullivan et al., 2010) and may neither be explained by poorer overall cognitive abilities (indexed by MMSE and CERAD/CTS; Ridley et al., 2013) nor by the severity of depressive symptomatology (GDS; DeFrancesco et al., 2009; Hesel et al., 2016).

Moreover, the present study is novel as to the best of our knowledge previous studies investigating cognitive performance in elderly with neurocognitive impairments mostly compared patient groups to normal controls (instead of to another patient group matched for overall cognitive performance; e.g., Alzheimer's disease: Ehrensperger, Berres, Taylor, & Monsch, 2010; mild cognitive impairment: Pajanen et al., 2010). Moreover, as none of these studies report Z-scores for the CERAD subtests, direct comparisons with our findings are difficult. As such, group differences reported in the literature may be more pronounced than group differences reported in our study, which compares closely matched and rather high-functioning groups of elderly with minor neurocognitive disorders (with and without alcohol abuse).

In summary, as far as we know, the present study is the first to systematically examine the effects of alcohol abuse in a large and well-matched sample of elderly patients with (suspected) minor neurocognitive disorder. Importantly, our findings suggest that also in rather high functioning elderly (average MMSE of 28) alcohol abuse may cause cognitive impairments that, albeit subtle and circumscribed, are detectable by means of a standard neuropsychological test battery frequently used to diagnose neurodegenerative diseases like dementia (i.e., CERAD). In particular, CERAD variables measuring verbal free recall and recognition memory as well as visuo-motor speed allowed discriminating elderly with and without alcohol abuse. This indicates that the respective CERAD variables seemed useful in identifying alcohol-related cognitive deficiencies in elderly patients with (suspected) minor neurocognitive disorder. Of course, future research endeavors are required to confirm our findings. Nonetheless, as the CERAD is widely used for standard neuropsychological assessment, the additional diagnostic value seems beneficial.

ACKNOWLEDGMENTS

The authors declare no conflicts of interest. The authors received no specific funding for this work.

REFERENCES

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.), Washington, DC: American Psychiatric Association Press.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.), Washington, DC: American Psychiatric Association Press.
- Anstey, K.J., Mack, H.A., & Cherbuin, N. (2009). Alcohol consumption as a risk factor for dementia and cognitive decline: Meta-analysis of prospective studies. *American Journal of Geriatric Psychiatry*, 17(7), 542–555.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300. Retrieved from <http://www.jstor.org/stable/2346101>
- Bernardin, F., Maheut-Bosser, A., & Paille, F. (2014). Cognitive impairments in alcohol-dependent subjects. *Frontiers in Psychiatry*, 5, 78.
- Bertoux, M., Ramanan, S., Slachevsky, A., Wong, S., Henriquez, F., Musa, G., ... Dubois, B. (2016). So close but yet so far: Executive contribution to memory processes in behavioral variant of frontotemporal dementia. *Journal of Alzheimers Disease*, 54(3), 1005–1014.
- Beydoun, M.A., Beydoun, H.A., Gamaldo, A.A., Teel, A., Zonderman, A.B., & Wang, Y. (2014). Epidemiological studies on modifiable factors associated with cognition and dementia: Systematic review and meta-analysis. *BMC Public Health*, 14, 643.
- Butters, N. (1985). Alcoholic Korsakoff's syndrome: Some unresolved issues concerning etiology, neuropathology, and cognitive deficits. *Journal of Clinical and Experimental Neuropsychology*, 7(2), 181–210.
- Caine, D., Halliday, G.M., Kril, J.J., & Harper, C.G. (1997). Operational criteria for the classification of chronic alcoholics: Identification of Wernicke's encephalopathy. *Journal of Neurology, Neurosurgery, and Psychiatry*, 62(1), 51–60.
- Chandler, M.J., Lacritz, L.H., Hynan, L.S., Barnard, H.D., Allen, G., Deschner, M., ... Cullum, C.M. (2005). A total score for the CERAD neuropsychological battery. *Neurology*, 65(1), 102–106.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). ACM, New York, NY. (pp. 785–794). doi:<http://dx.doi.org/10.1145/2939672.2939785>.
- Chen, T., He, T., & Benesty, M. (2016). xgboost: Extreme gradient boosting. R package version 0.4-4. Retrieved from <https://CRAN.R-project.org/package=xgboost>.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159.
- DeFrancesco, M., Marksteiner, J., Deisenhammer, E.A., Hinterhuber, H., & Weiss, E. M. (2009). Association of mild cognitive impairment (MCI) and depression. *Neuropsychiatrie*, 23(3), 144–150. [Article in German].
- Dupuy, M., & Chanaud, S. (2016). Imaging the addicted brain: Alcohol. *International Review of Neurobiology*, 129, 1–31.
- Ehrensperger, M.M., Berres, M., Taylor, K.I., & Monsch, A.U. (2010). Early detection of Alzheimer's disease with a total score of the German CERAD. *Journal of the International Neuropsychological Society*, 16, 910–920.
- Ferguson, C.J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, 40(5), 532–538.
- Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). 'Mini Mental State' - a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12, 189–198.
- Friedman, N.P., & Miyake, A. (2016). Unity and diversity of executive functions: Individual differences as a window to cognitive structure. *Cortex*, 86, 186–204.
- Gauggel, S., & Birkner, B. (1999). Validität und Reliabilität einer deutschen Version der Geriatrischen Depressionsskala (GDS). *Zeitschrift für Klinische Psychologie*, 28, 18–27. [Article in German].
- Gini, C. (1921). Measurement of inequality of incomes. *The Economic Journal*, 31(121), 124–126. <http://doi.org/10.2307/2223319>.
- Heser, K., Bleckwenn, M., Wiese, B., Mamone, S., Riedel-Heller, S.G., Stein, J., ... Wagner, M., for the AgeCoDe study group. (2016).

- Late-life depressive symptoms and lifetime history of major depression: Cognitive deficits are largely due to incipient dementia rather than depression. *Journal of Alzheimers Disease*, 54(1), 185–199.
- Ho, D.E., Imai, K., King, G., & Stuart, E. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 42(8), 1–28.
- Hughes, C.P., Berg, L., Danziger, W.L., Coben, L.A., & Martin, R.L. (1982). A new clinical scale for the staging of dementia. *British Journal of Psychiatry*, 140, 566–572.
- Mayer, D., Diwo, A., Imarhiagbe, D., Erler, S., Marksteiner, J., & Kaufmann, L. (2015). About the differential diagnostic utility of the CERAD neuropsychological assessment battery in patients with mild cognitive impairment with and without depressive symptomatology. *Zeitschrift für Neuropsychologie*, 26(2), 131–142. [Article in German].
- Milborrow, S. (2016). Rpart.plot: Plot 'rpart' models: An enhanced version of 'plot.rpart'. R package version 2.0.1. Retrieved from <https://CRAN.R-project.org/package=rpart.plot>.
- Morris, J.C., Heyman, A., Mohs, R.C., Hughes, J.P., van Belle, G., Fillenbaum, G., ... Clark, C. (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*, 39, 1159–1165.
- Oscar-Berman, M., & Marinkovic, K. (2007). Alcohol: Effects on neurobehavioral functions and the brain. *Neuropsychological Reviews*, 17(3), 239–257.
- Oscar-Berman, M., Valmas, M.M., Sawyer, K.S., Ruiz, S.M., Luhar, R.B., & Gravitz, Z.R. (2014). Profiles of impaired, spared and recovered neuropsychological processes in alcoholism. *Handbook of Clinical Neurology*, 125, 183–210.
- Paajanen, T., Hänninen, T., Tunnard, C., Mecocci, P., Sobow, T., Tsolaki, M., ... Soininen, H. (2010). CERAD neuropsychological battery total score in multinational mild cognitive impairment and control populations: The AddNeuroMed study. *Journal of Alzheimer's Disease*, 22, 1089–1097.
- Panza, F., Frisardi, V., Seripa, D., Logroscino, C., Santamato, A., Imbimbo, B.P., ... Solfrizzi, V. (2012). Alcohol consumption in mild cognitive impairment and dementia: Harmful or neuroprotective? *International Journal of Geriatric Psychiatry*, 27(12), 1218–1238.
- Panza, F., Frisardi, V., Capurso, D., D'Introno, A., Colacicco, A.M., Imbimbo, B.P., ... Solfrizzi, V. (2010). Late-life depression, mild cognitive impairment, and dementia: Possible continuum? *American Journal of Geriatric Psychiatry*, 18(2), 98–116.
- Peters, R., Peters, J., Warner, J., Beckett, N., & Bulpitt, C. (2008). Alcohol, dementia and cognitive decline in the elderly: A systematic review. *Age and Ageing*, 37, 505–512.
- R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R foundation for statistical computing. Retrieved from <http://www.r-project.org/>.
- Randolph, J.J., Falbe, K., Manuel, A.K., & Balloun, J.L. (2014). A step-by-step guide to propensity score matching in R information on the dataset used. *Practical Assessment, Research & Evaluation*, 19(18), 1–6.
- Ridley, N.J., Draper, B., & Withall, A. (2013). Alcohol-related dementia: An update of the evidence. *Alzheimer's Research & Therapy*, 5, 3.
- Schmid, N.S., Ehrensperger, M.M., Berres, M., Beck, I.R., & Monsch, A.U. (2014). The extension of the German CERAD neuropsychological assessment battery with tests assessing subcortical, executive and frontal functions improves accuracy in dementia diagnosis. *Dementia and Geriatric Cognitive Disorders Extra*, 4, 322–334.
- Sinforiani, E., Zucchella, C., Pasotti, C., Casoni, F., Bini, P., & Costa, A. (2011). The effects of alcohol on cognition in the elderly: From protection to neurodegeneration. *Functional Neurology*, 26(2), 103–106.
- Smith, A. (1995). Medical manifestations of alcoholism in the elderly. *The International Journal of the Addictions*, 30(13&14), 1749–1798.
- Sullivan, E.V., Harris, R.A., & Pfefferbaum, A. (2010). Alcohol's effects on brain and behavior. *Alcohol Research & Health*, 33(1), 127–143.
- Thalman, B., Monsch, A.U., Schneitter, M., Bernasconi, F., Aebi, C., Camachova-Davet, Z., ... Staehelin, H.B. (2000). The CERAD neuropsychological assessment battery (CERAD-NAB) - A minimal dataset as a common tool for German-speaking Europe. *Neurobiology of Aging*, 21, 30.
- Therneau, T., Atkinson, B., & Ripley, B. (2015). Rpart: Recursive partitioning and regression trees. Retrieved from <https://cran.r-project.org/package=rpart>.
- Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. New York, NY: Springer-Verlag.
- Wolfsgruber, S., Jessen, F., Wiese, B., Stein, J., Bicke, H., Mösch, E., ... Wagner, M., for the AgeCoDe study group. (2014). The CERAD neuropsychological assessment battery total score detects and predicts Alzheimer Disease dementia with high diagnostic accuracy. *American Journal of Geriatric Psychiatry*, 22(10), 1017–1028.
- Yesavage, J.A., Brink, T.L., Rose, T.L., Lum, O., Huang, V., Adey, M., ... von Leirer, O. (1983). Development and validation of a Geriatric Depression Scale: A preliminary report. *Journal of Psychiatric Research*, 17, 37–49.

APPENDIX A

Means and *SDs* of standardized *Z*-scores of the CERAD variables and the Chandler Total Score for male and female elderly with and without alcohol abuse as well as *p*-values and effect sizes of the interaction effects between study group and sex

Test variable	Elderly with alcohol abuse Mean (<i>SD</i>)		Elderly without alcohol abuse Mean (<i>SD</i>)		<i>p</i> -Value (uncorr.)	<i>p</i> -Value (corr.)	η^2_G
	Men (<i>n</i> = 43)	Women (<i>n</i> = 29)	Men (<i>n</i> = 43)	Women (<i>n</i> = 29)			
Verbal Fluency	-0.84 (0.79)	-0.70 (1.01)	-0.76 (0.89)	-0.48 (0.84)	.642	.705	.002
Boston Naming Test	0.19 (0.89)	0.59 (0.84)	0.22 (0.87)	0.51 (1.02)	.705	.705	.001
Word List Learning	-0.98 (1.25)	-1.11 (1.31)	-1.13 (1.56)	-1.45 (1.30)	.683	.705	.001
Word List Recall	-0.77 (1.35)	-0.64 (1.45)	-0.54 (1.24)	-0.94 (1.04)	.229	.705	.010
Word List Intrusions	0.26 (0.94)	0.38 (0.70)	0.28 (0.99)	0.21 (1.02)	.547	.705	.003
Word List Discriminability	-0.73 (1.52)	-0.05 (1.01)	-0.55 (1.46)	-0.99 (1.11)	.015*	.174	.042
Constructional Praxis Copy	0.12 (1.04)	0.09 (1.31)	0.40 (0.84)	-0.02 (1.16)	.287	.705	.008
Constructional Praxis Recall	-0.84 (1.71)	-0.55 (1.42)	-0.41 (1.62)	-0.35 (1.58)	.672	.705	.001
Trail Making Test A	-1.27 (1.03)	-1.31 (1.25)	-0.87 (1.20)	-0.71 (1.41)	.630	.705	.002
Trail Making Test B	-1.00 (1.20)	-1.32 (1.34)	-0.66 (1.50)	-0.52 (1.38)	.317	.705	.007
Trail Making Test B/A	0.07 (1.02)	-0.16 (1.31)	0.04 (1.13)	0.15 (1.08)	.381	.705	.005
Chandler Total Score	-1.64 (1.13)	-1.42 (1.17)	-1.54 (1.29)	-1.59 (1.14)	.497	.705	.003

Note. uncorr. = not corrected for multiple testing; η^2_G = generalized eta squared.

*Significant at $p < .05$.