universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Impairments in activities of daily living (ADL) and the prediction of conversion to dementia – a retrospective study“

verfasst von / submitted by

Ursula Hanl, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2021 / Vienna 2021

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 840
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Psychologie UG2002
Betreut von / Supervisor:	Univ.- Prof. Mag. Dr. Claus Lamm

Table of Contents

Impairments in Activities of Daily Living and the Prediction of Conversion to Dementia – a Retrospective Study

Dementia is a global public health issue, which impacts society on multiple levels. With age being the most common risk factor for the onset of dementia, the increment of life expectancy is a major cause for rising prevalence rates. While in 2015 it was estimated that there were 47 million cases of dementia worldwide, it is predicted that in 2050 around 132 million people will be affected (World Health Organization [WHO], 2015 & 2017). According to the Austrian Alzheimer's Association, there are currently about 100,000 people with dementia in Austria. Predictions state that this number is going to increase to 230,000 cases of dementia in Austria until 2050 (Österreichische Alzheimergesellschaft [ÖAG], n.d.). Such estimates of future trends are necessary, to prepare the healthcare system and plan medical and social capacities (Wancata et al., 2015).

The course of Alzheimer's disease (AD), which is the leading cause of dementia, is described in several stages. Early stages can be extremely difficult to detect, as they lack distinct clinical symptoms. However, there is empirical evidence that those presymptomatic stages are associated with neurological changes that are consistent with AD-pathology. Those findings suggest that when AD is diagnosed, there might already be advanced brain damage. Therefore, research on early stages of AD is crucial for the development of preventive strategies (Jessen et al., 2014; Stogmann et al., 2016). Several studies have attempted to determine, when exactly cognitive functioning begins to decline before the conversion to dementia. There is empirical evidence suggesting that impairments in activities of daily living (ADL) occur in prodromal stages of dementia and therefore can be instrumentalized to identify individuals who are at risk of developing dementia (Amieva et al., 2008; Pérès et al., 2008; Reppermund et al., 2011).

Theoretical Background

The Three-Staged Model of AD

In the past, various models have attempted to explain the progression of AD. While there is consent that the course of AD must be described in multiple stages, models differ in their conceptualization of early stages. The following study is based on the three-staged model of AD proposed by Jessen et al. (2014), who declare that subjective reports of cognitive decline as indicators for early manifestations of AD are often undervalued. The individual stages of the model are described below.

Stage 1: Subjective cognitive decline. The first stage of subjective cognitive decline (SCD) is characterized by two criteria. Firstly, individuals with SCD report a constant decline of their own cognitive capacities. The cognitive impairment is therefore based on self-perception and does not need to be observed by others. Secondly, the individual's performance on standardized cognitive tests remains intact, which means that from an objective perspective, cognitive capacities are unimpaired (Jessen et al., 2020). Previously, research on the role of SCD in preclinical AD has been complicated through a lack of standardized research procedures and concepts. For example, different terms have been used to describe self-reported cognitive decline, including the terms subjective memory decline (SMD), subjective cognitive impairment (SCI), subjective memory complaint (SMC) and subjective memory impairment (SMI). With the aim to create a collective understanding of SCD and facilitate research progress on the concept, Jessen et al. (2014) started the Subjective Cognitive Decline Initiative (SCD-I).

Multiple studies have stated that individuals with SCD, compared to individuals without SCD, are at greater risk for developing dementia in the future (Jessen, 2010; Mitchell, Beaumont, Ferguson, Yadegarfar, & Stubbs, 2014). Mitchell et al. (2014) investigated the association between SCD and future cognitive decline in a meta-analysis and discovered that 14% of the individuals with SCD converted to dementia. They further claimed that for individuals with SCD the risk of developing dementia is twice as high as for individuals without SCD. Although a standardised frame for SCD-research has been set (Jessen et al., 2014), investigations on SCD are still highly heterogenous concerning their terminology and methodology. As a result, it is problematic to make estimations about prevalence rates and conversion rates based on SCD only (Parnetti, Chipi, Salvadori, D'Andrea, & Eusebi, 2019).

Jessen et al. (2020) added that SCD can be caused by several medical conditions and concluded that it cannot be used as an indicator for future cognitive impairment in most individuals.

Stage 2: Mild cognitive impairment. In the meta-analysis of Mitchel et al. (2014), 27% of the individuals with SCD converted to mild cognitive impairment (MCI) later. Petersen (2004), who was the first one to define MCI as a separate diagnostic entity, positioned it on a continuum between cognitive changes caused by ageing processes and cognitive impairment as a sign of dementia. Individuals in the second stage of MCI demonstrate a decrease in cognitive function, which is not severe enough to be diagnosed with dementia. Unlike SCD, MCI is defined by impaired objective performance in neuropsychological tests. Depending on memory functioning, MCI is further divided into nonamnestic (naMCI) and amnestic (aMCI) subtypes. Both naMCI and aMCI can impact single or multiple cognitive domains. Petersen (2004) set following criteria for the diagnosis of aMCI:

1. Subjective memory complaint
2. Objective memory impairment
3. Largely preserved general cognitive function
4. Essentially preserved ADL
5. No presence of dementia

If possible, the first criterion should be confirmed by an informant, which can for example be a caregiver. Next, an actual impairment in cognitive performance must be detected, normally by conducting neuropsychological tests. The term general cognitive function refers to cognitive capacities not related to memory like executive functions, language abilities or visuospatial skills. Performance in memory and non-memory domains should be judged relative to age norms.

While aMCI is characterized by memory impairment as the main symptom, patients with naMCI show cognitive decline only in nonmemory domains, for example in the use of language. Research on MCI as precursor of AD has focused on the amnestic subtype, probably because the nonamnestic subtype is reportedly less common (Petersen, 2011). Moreover, individuals with nonamnestic MCI seem to be more likely to develop dementia with Lewy bodies (DLB), rather than AD (Dugger et al., 2015; Ferman et al., 2013).

Stage 3: Dementia. The third and last stage of the model is reached, when a patient develops dementia due to AD. It is much harder to distinguish between normal ageing and MCI than between MCI and dementia (Petersen, 2004). While daily functions are mostly preserved in patients with MCI, patients with AD show impaired functional activities to the degree that they cannot live independently within society any longer (Petersen, 2011). The areas most severely affected by AD are memory, language, temporal and spatial orientation, attention, visual perception, problem solving, social skills and activities of daily living (McKhann et al., 1948). Currently, there are three different approaches that are most commonly used for diagnosing AD (Oedekoven & Dodel, 2019):

- NINCDS-ADRDA criteria (McKhann et al., 1948),
- DSM-5 criteria (American Psychiatric Association [APA], 2013)
- and ICD-10 criteria (WHO, 1993).

The NINCDS-ADRDA criteria were proposed by the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) and the Alzheimer's Disease and Related Disorders Association (ADRDA). Together they founded a work group that aimed to develop criteria for the clinical diagnosis of AD. The work group published criteria for possible, probable and definite AD, as well as circumstances under which the diagnosis of probable AD is unlikely or uncertain. According to NINCDS-ADRDA criteria, the diagnosis of definite AD demands histopathologic confirmation, which can only be gained from biopsy or autopsy. Probable AD is diagnosed when the following criteria are met:

- determination of dementia by clinical assessment and neuropsychological testing
- deficits in two or more cognitive domains
- progressive nature of cognitive decline
- onset of cognitive deficits between ages of 40 and 90 (typically after age 65)
- no systemic or brain diseases that could cause the cognitive dysfunction (McKhann et al., 1948).

In the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), the term major neurocognitive disorder (NCD) was introduced as an alternative to the term dementia. Based on different causes, there are several subtypes of NCD. Those include for example NCD due to Alzheimer's disease, NCD due to Lewy body disease or NCD due to traumatic brain injury. The authors claim that the term dementia is often associated with older

patients and is sometimes mistakenly used as a synonym for AD. Therefore, they decided that a change in terminology would be a helpful first attempt to eliminate confusion and stigma related to dementia. The term dementia is however still used in the clinical context, as it has established itself there. As a result, major NCD and dementia are commonly used interchangeably (APA, 2013; Sachdev et al., 2014).

Finally, there is the International Statistical Classification of Diseases and Related Health Problems (ICD), which is published by the WHO. In clinical practice, the tenth revision of the ICD (ICD-10) is used to diagnose AD, as the eleventh revision (ICD-11) is still under development. However, a preliminary beta version of ICD-11 is already available online (WHO, 2019).

In the present study, diagnosis of AD is based on NINCDS-ADRDA criteria for probable AD, which have been widely used in clinical practice and research since their publication. Blacker et al. (1994) evaluated the criteria and concluded that they have good validity and reliability. Furthermore, the diagnosis of probable AD achieved good sensitivity and specificity across 13 studies, with an average sensitivity of 81% and an average specificity of 70% (Knopman et al., 2001). Conversely, the diagnostic precision of the NINCDS-ADRDA criteria has also been questioned by some researchers (Dubois et al., 2007; Varma et al., 1999). Above all, Dubois et al. (2007) criticized that knowledge about AD biomarkers is not included in the diagnostic process. McKhann et al. (2011) reviewed the NINCDS-ADRDA criteria and stated that biomarker evidence can indeed be helpful to increase certainty that a clinical case of dementia is caused by AD. First, however, it must be ensured that the core clinical criteria for probable AD are met. Further research is necessary to determine the value of biomarker evidence for AD diagnosis (McKhann et al., 2011).

Activities of Daily Living

Researchers are constantly trying to identify new approaches, which enable them to detect AD as early as possible. In the past years, conversion rates of patients with SCD have been analysed to determine whether they can be used in the early detection of AD. However, results on conversion rates are inconsistent across different research projects and scientists disagree on whether subjective memory complaints are useful at predicting dementia (Lehrner et al., 2016). Currently, it is not possible to predict conversion to dementia by using biomarkers

only. Still, research on the role of biomarkers in preclinical AD has made a lot of progress and biomarker-based approaches may be useful in combination with other detection methods (Jessen et al., 2020; Stogmann et al., 2016).

Recently, the assessment of ADL as an additional tool for the prediction of conversion to dementia has been discussed. ADL can be divided into instrumental (IADL) and basic ADL. IADL include more complex tasks of daily life, such as household chores, managing finances or taking medication that demand interaction with the environment. Those IADL enable individuals to live independently in a community and are important for their quality of life. Basic ADL are essential to self-care and refer to fundamental everyday tasks like dressing, washing and eating. Typically, impairments in IADL appear first, in early stages of AD. Afterwards, as the disease progresses, individuals begin to struggle to carry out basic ADL unattended (Hindmarch, Lehfeld, de Jongh, & Erzigkeit, 1998; Lively, 2015).

For the diagnosis of dementia, contrary to MCI, an impairment in ADL is a necessary criterion (Petersen, 2011). Although current diagnostic standards do consider ADL to be preserved in early stages of AD, several authors have claimed that declined ADL-performance is apparent much earlier than previously thought and may even be a first hint of later progression to dementia. When it comes to early detection of AD, impaired IADL play a major role as they typically appear first. In research, however, the term ADL is mostly used to refer to IADL. This also applies to the studies that are introduced below (Amieva et al., 2008; Pérès et al., 2008; Reppermund et al., 2011; Stogmann et al., 2016).

Amieva et al. (2008) compared control subjects with subjects who developed AD over a 14-year follow-up period. Five to six years before their diagnosis, the predemented subjects in the sample already had significantly higher ADL scores. At the same time, they started becoming dependent of assistance in their daily living activities, even though they did not meet diagnostic criteria for AD. The impairment in ADL increased sharply during the last three years before the individuals were diagnosed with AD.

In the study of Pérès et al. (2008), patients with two or more impaired ADL at baseline were at a higher risk of developing AD ten years later. However, it should be noted that only four IADL relating to the domains telephoning, transportation, medication and handling finances, were included in the analysis. Five years prior to diagnosing AD, an impairment in each of the four complex ADL significantly predicted the diagnosis. Only the ability to handle

finances, which appeared to be affected earliest, was impaired ten years before the AD diagnosis (Pérès et al., 2008).

Further support for a long preclinical phase of AD comes from Reppermund et al. (2011) who also found a significant association between MCI and ADL-performance. According to informants' evaluations, individuals with MCI had more difficulty with ADL than cognitively unimpaired individuals. Additionally, compared to non-impaired participants, subjects with MCI showed a declined performance in neuropsychological testing. The authors suspected that their tool for assessing functional impairment, the Bayer-ADL (B-ADL) scale, was multidimensional. Therefore, they conducted a factor analysis that revealed two factors, which they categorized as high cognitive demand (HCD) and low cognitive demand (LCD) factors. Consistent with their expectations, subjects with MCI showed major difficulties in B-ADL items that had their highest loading on the HCD factor. Reppermund et al. (2011) furthermore reported possible gender differences regarding functional abilities in the stage of MCI, as they found a significantly higher B-ADL score and a higher score for HCD items in male participants with MCI. Additional support for gender differences comes from Pérès et al. (2011) who found an association between ADL-impairment and risk for dementia in male, but not in female participants.

By investigating possible restrictions of ADL in all stages of AD, Stogmann et al. (2016) were the first ones to explicitly examine ADL-performance in the stage of SCD. In their cross-sectional comparison of SCD, MCI and AD patients with healthy control subjects, they found a significantly decreased performance of ADL in SCD patients. Those outcomes suggest that the use of declined ADL as predictor of conversion to dementia might already be possible in the early stage of SCD. They further studied the relationship between ADL impairment and depressive symptoms. Across all subgroups (SCD, naMCI, aMCI, AD and control group), individuals with depressive symptoms reported more problems with ADL than individuals without depressive symptomatic. When studying associations between ADL and cognitive function, correlations with domains of the Neuropsychological Test Battery Vienna (NTBV) were found. Although depressive symptoms were a stronger predictor of functional impairment, cognitive performances in the NTBV-domains verbal memory, attention and executive functioning were significant predictors as well (Stogmann et al., 2016).

Reliable Change Analysis

AD is a progressive neurodegenerative disease and decline in cognitive and functional abilities are typical over the course of AD. An important part of AD prevention and intervention are neuropsychological follow-up evaluations. In repeated testing situations, change in an individual's test score can be caused by multiple factors. Reliable change index (RCI) is used to differentiate effects of disease progression from change caused by systematic or measurement errors, like practice effects or test unreliability. In this study, RCI was used to determine changes in functional impairment due to components not related to functional abilities (Foki et al., 2018; Hinton-Bayre, 2010).

Jacobson and Truax (1991) introduced the RCI as a method for assessing whether an individual's test score has significantly changed over time. They published the following formula for calculating RCI:

$$RCI = \frac{X_2 - X_1}{SE_{diff}} \quad (1)$$

While X_2 is the retest score, X_1 describes the initial test score. SE_{diff} is the standard error of difference between the two scores and represents a distribution that would be anticipated if no actual change had occurred in individual test scores. It can be calculated via the standard error of measurement SE_m, which in turn can be derived from the standard deviation of the initial test SD and the test-retest reliability r_{tt} .

$$SE_{diff} = \sqrt{2 \cdot (SE_m)^2} \quad (2)$$

$$SE_m = SD \cdot \sqrt{1 - r_{tt}} \quad (3)$$

The concept of RCI was initially proposed by Jacobson and Truax (1991) for evaluating the outcomes of psychotherapy, therefore they did not consider practice effects. As soon as subjects are exposed to the same testing procedures repeatedly, which is very common in the neuropsychological context, a systematic bias in form of practice effects is created. To assess actual change independently from expected practice effects, Chelune, Naugle, Lüders, Sedlak and Awad (1993) added a correction factor, the mean practice effect of a control group, to the formula:

$$RCI = ((X_2 - X_1) - (M_2 - M_1)) \cdot SE_{diff} \quad (4)$$

The difference between the retest score X_2 and the initial test score X_1 stands for an individual's change in performance. M_2 is the mean retest score and M_1 the mean initial test score of the control group, their difference represents the practice effect (Chelune et al., 1993; Ringendahl, 2013).

However, the approach by Chelune et al. (1993) does not include any other parameters than the mean practice effect, which it has been criticised for. To address this problem, McSweeny, Naugle, Chelune and Lüders (1993) introduced a standardized regression based (SRB) model. The method of McSweeny et al. (1993) corrects for both, practice effects and regression to the mean on retesting.

$$RCI_{SRB} = \frac{Y - Y'}{SE_E} \quad (5)$$

$$SE_E = SD_Y \cdot \sqrt{1 - r^2_{XY}} \quad (6)$$

Y is the observed retest score, Y' is the predicted retest score from the regression equation and SE_E is the standard error of the estimate from the regression equation. SE_E can be computed via SD_Y, which describes the standard deviation of the retest score, and the test-retest reliability r_{XY} (Hinton-Byre, 2016). By using a regression-based model, different factors that might influence the retest score, including age, education level and test-retest interval, can be identified. The predicted retest score Y' can be calculated as follows:

$$Y' = b \cdot X + c \quad (7)$$

$$b = r \cdot (SD_Y / SD_X) \quad (8)$$

$$c = M_Y - b \cdot M_X \quad (9)$$

As a linear regression model is used, the initial test score X is multiplied by the regression coefficient b (the slope) and the constant c (the intercept) is added. For calculating b, standard deviations of initial test scores (SD_X) and retest scores (SD_Y) for controls are needed. The

constant *c* is derived from the mean of retest scores for controls M_Y and the mean of baseline scores for controls M_X (Hinton-Bayre, 2016; McSweeny et al., 1993).

SRB models, like the one proposed by McSweeny et al. (1993), have been criticised for using a biased estimate of the retest score. Therefore Maassen, Bossema and Brand (2006) offered a revised regression approach, in which an adjusted slope (b_{adj}) is used for estimating the predicted test score and can then be used to obtain an adjusted constant (c_{adj}; Hinton-Bayre, 2010; Maassen et al., 2006).

$$Y' = b_{adj} \cdot X + c_{adj} \quad (10)$$

$$b_{adj} = SD_Y / SD_X \quad (11)$$

$$c_{adj} = M_Y - b_{adj} \cdot M_X \quad (12)$$

Hinton-Byre (2010) compared several models for assessing reliable change (RC) and pointed out that they are based on the same fundamental structure. He stated that all versions of formulas for calculating RCI can be replaced by one simple formula:

$$RCI = \frac{Y - Y'}{SE} \quad (13)$$

In this expression, the difference between the real retest score *Y* and the predicted retest score *Y'* is divided by the standard error *SE*. The resulting standardized RC score can be interpreted in terms of a standard Z distribution or a *t* distribution if the control sample contains less than 50 participants (Hinton-Byre, 2010). A significant change is usually presumed from a z-value of ±1.645, which is consistent with a 90% level of confidence. While a z-value of -1.645 or lower is interpreted as significant decline, a z-value of +1.645 and higher is interpreted as a significant improvement (with $p = .05$, one-tailed and $p = .10$, two-tailed; Ringendahl, 2013). In his revision of RC methodology, Hinton-Byre (2010) concluded that it remains unclear whether there is one model best suited for general use. Therefore, it was decided to use three different models for calculating RCI and compare the results.

Study Aim

Based on the results of Stogmann et al. (2016), the described longitudinal study aimed to assess and compare ADL in individuals with SCD, naMCI, aMCI and AD as well as in a group of healthy control subjects. The main goal was to determine the prognostic value of ADL in predicting conversion to dementia. For that purpose, baseline ADL-scores of control subjects, patients with SCD, naMCI and aMCI were compared to the scores of patients who had developed AD after the follow-up period. While the B-ADL score represents the independent variable, conversion to dementia is the dependent variable. The dependent variable, conversion to dementia, is a dichotomous variable with the possible outcomes 0 (no conversion) and 1 (conversion to dementia). In addition, possible associations between impairment in ADL with demographic variables and with test scores of Mini Mental State Examination (MMSE), Neuropsychological Test Battery Vienna (NTBV), Beck Depression Inventory-II (BDI-II), Wortschatztest (WST) and Vienna Visuo-Constructional Test 3.0 Screening (VVT 3.0 Screening) were measured. Finally, different models for calculating reliable change index (RCI) for B-ADL were used to assess reliable change in the context of neuropsychological testing. In order to do so, the three methods proposed by Chelune et al. (1993), Maassen et al. (2006) and McSweeny et al. (1993) were selected.

Research Questions

Research question 1. The first and main research question of the present study is: *Can activities of daily living predict later conversion to dementia?* In accordance with previous findings (Amieva et al., 2008; Pérès et al., 2008; Reppermund et al., 2011; Stogmann et al., 2016) it was expected that activities of daily living could be used as a predictor for conversion to dementia, which lead us to the following hypotheses:

- H_0 (1): Activities of daily living (B-ADL) cannot predict conversion to dementia.
- H_1 (1): Activities of daily living (B-ADL) can predict conversion to dementia.

Research question 2. In the second research question, possible associations between B-ADL score and other variables of interest are addressed: *Is there an association between activities of daily living and demographic variables (age, gender, education) or the results of MMSE, NTBV, BDI-II, WST-IQ and VVT 3.0 Screening?* In order to study correlations between variables, an exploratory analysis was conducted. The following non-directional hypotheses were formulated:

- H_0 (2.1): There is no association between age and activities of daily living (B-ADL).
- H_1 (2.1): There is an association between age and activities of daily living (B-ADL).
- H_0 (2.2): There is no association between gender and activities of daily living (B-ADL).
- H_1 (2.2): There is an association between gender and activities of daily living (B-ADL).
- H_0 (2.3): There is no association between education level and activities of daily living (B-ADL).
- H_1 (2.3): There is an association between education level and activities of daily living (B-ADL).
- H_0 (2.4): There is no association between cognitive status (MMSE) and activities of daily living (B-ADL).
- H_1 (2.4): There is an association between cognitive status (MMSE) and activities of daily living (B-ADL).
- H_0 (2.5): There is no association between cognitive function (NTBV) and activities of daily living (B-ADL)
- H_1 (2.5): There is an association between cognitive function (NTBV) and activities of daily living (B-ADL).
- H_1 (2.6): There is no association between depressive symptomatic (BDI-II) and activities of daily living (B-ADL).
- H_0 (2.6): There is an association between depressive symptomatic (BDI-II) and activities of daily living (B-ADL).
- H_0 (2.7): There is no association between verbal intelligence (WST-IQ) and activities of daily living (B-ADL).
- H_1 (2.7): There is an association between verbal intelligence (WST-IQ) and activities of daily living (B-ADL).
- H_0 (2.8): There is no association between visuo-constructional abilities (VVT 3.0 Screening) and activities of daily living (B-ADL).
- H_1 (2.8): There is an association between visuo-constructional abilities (VVT 3.0 Screening) and activities of daily living (B-ADL).

Research Question 3. The third research question aims to descriptively compare reliable change methods: *Is there a difference between methods for assessing reliable change for activities of daily living in converters versus non-converters?* For that purpose, three different methods for calculating reliable change index (RCI) were conducted and results were compared (Chelune et al., 1993; Maassen et al., 2006; McSweeny et al., 1993).

Methods

Sample

Patients either came to the memory outpatient clinic at their own request or were referred by their physician, due to memory problems. First, patients took part in a detailed anamnestic interview, to make sure that they fulfil inclusion criteria. Only patients with a minimum age of 50 years and a MMSE score ≥ 23 were taken into consideration. Furthermore, patients were excluded from participation if any of the following criteria were met:

- history of stroke or severe head injury
- current diagnosis of a mental or behavioural disorder according to ICD-10 (Dilling, Mombour, & Schmidt, 2008), except for depressive symptoms
- dementia diagnosis at baseline
- medical conditions that could impair the performance in the neuropsychological assessment (including cardiac, renal, respiratory or hepatic disease as well as language, auditory, motor or visual deficits)

In total, the sample consists of 315 participants. All patients who fulfilled the inclusion criteria were, depending on their performance in neuropsychological testing, assigned to either the SCD, naMCI, aMCI or the control group. At baseline, 45 patients were diagnosed with SCD, 126 patients were diagnosed with naMCI, 98 patients were diagnosed with aMCI and 46 participants were assigned to the control group. Because (sub-) depressive symptoms are common in the elderly population, patients with depressive symptoms were not excluded from participation (Stogmann et al., 2016). Control subjects were required to have a MMSE-score equal to or higher than 27.

Study Design

This study is a retrospective, longitudinal study and part of the research project Vienna Conversion to Dementia Study (VCD). The VCD is a prospective cohort study, as it includes community-dwelling patients who come to the memory clinic for assessment of a possible cognitive impairment. Data was collected between June 2001 and October 2020 at the Department of Neurology at the Medical University of Vienna, with follow-up testing being conducted 12 to 48 months after the initial testing procedure.

After completing the neuropsychological evaluation, patients were allocated to one of four clinical subgroups. Patients who experienced subjective cognitive decline and had a mean z-score above -1.5 standard deviations (*SD*) in each domain of NTBV were assigned to the SCD group (Stogmann et al., 2016). MCI diagnosis was given according to Petersen criteria (Petersen, 2011), with a cut-off score of -1.5 *SD* below the age, gender and education matched mean of a normative sample. The diagnosis of naMCI was given, when the z-score in at least one non-memory domain was lower than -1.5 *SD*. Patients were diagnosed with aMCI if their z-score in the memory domain was below -1.5 *SD*. AD was diagnosed in accordance with NINCDS-ADRDA criteria (McKhann et al., 1984).

The study plan was submitted to the Ethics Committee of the Medical University of Vienna and was discussed in their conference on December 15th in 2020. After requested information had been added, the Ethics Committee gave its approval on January 4th, 2021. Provided that the study will be conducted in accordance with the ICH-Guidline for Good Clinical Practice, this decision is valid for one year. The Ethics Committee was continuously informed about additions to the study plan. In general, there have been no changes regarding the study procedure.

Assessment Procedure

Mini Mental State Examination. In the first part of the assessment, patients were screened using the German version of the Mini Mental State Examination (MMSE), which is a diagnostic instrument for the assessment of cognitive impairments in elderly patients. As it is quick and easy to implement, it is a useful tool for getting a first impression of a patient's cognitive status and for detecting changes in cognitive function. It has a maximum score of 30,

with lower scores indicating weaker cognitive performance. The first part of the MMSE requires vocal responses only and tests the dimensions of attention, memory and orientation. In the second part, patients' abilities to name, follow verbal instructions, follow written instructions, write a sentence and copy a complex figure are measured. The original MMSE has a test-retest reliability of r = .89 and an inter-rater reliability of r = .83 (Folstein, Folstein, & McHugh, 1975). An evaluation of the German version resulted in a Cronbach's alpha of α = .79 (Beyermann, Trippe, Bähr, & Püllen, 2013). As soon as patients reached a MMSE-score equal to or higher than 23, further neuropsychological tests were conducted.

Vienna Visuo-Constructional Test 3.0 Screening. The screening process also included the Vienna Visuo-Constructional Test 3.0 Screening (VVT 3.0 Screening), which is a measure for assessing visuo-constructional problems in patients with cognitive dysfunction. It is an abbreviated version of the original Vienna Visuo-Constructional Test (VVT) and therefore allows a more economical implementation compared to the full version. In the first part of the VVT 3.0 Screening, patients are instructed to copy three figures as precisely as possible. The evaluation is then based on the accuracy of the drawings. In the second part, patients are asked to copy the figures again as accurately as possible, this time from memory (Valencia & Lehrner, 2018). Certain criteria, such as the size of the figures or length and parallelism of the lines, are then used to assess accuracy. The higher the test score, the higher the patients visuo-constructional abilities (Lehrner et al., 2015). An evaluation of psychometric criteria resulted in an internal consistency (Cronbach's alpha) of α = .84 (Numrich, 2017 in Valencia & Lehrner, 2018).

Neuropsychological Test Battery Vienna. The Neuropsychological Test Battery Vienna (NTBV) was constructed to detect dementia in a clinical setting. It measures cognitive performance in the domains of psychomotor speed, attention, language, memory, and executive functioning, which are domains commonly affected by AD. In total, the NTBV consists of 13 subtests, which can be assigned to the five domains. For the measurement of psychomotor speed, the NTBV uses the Trail Making Test A (TMT-A; Reitan, 1979) and the symbol counting task from the cerebral insufficiency test (c.I.; Lehrl & Fischer, 1997). To assess the language domain, semantic verbal fluency (SVF) and phonematic verbal fluency (PVF) tasks (Goodglass & Kaplan, 1983) as well as the modified Boston Naming Test (BNT; Morris et al., 1989) are included. The Verbal Selective Reminding Test (VSRT; Lehrner, Gleiß, Maly, Auff, & Dal-Bianco, 2006) is used to test the memory domain. In order to measure attention, the

Alters-Konzentrations-Test (AKT; Gatterer, 1990), the Trail Making Test B (TMT-B; Reitan, 1979) and the digit symbol test from the German version of the revised Wechsler Adult Intelligence Scale (WAIS-R; Tewes, 1994) were selected. To study executive function, the NTBV contains the Five-Point Test (Regard, Strauss, & Knapp, 1982), the interference test from the C.I. (Lehrl & Fischer, 1997) and two tests from the Nuremberg Aging Inventory (NAI), the Maze Test and the Stroop-Test (Oswald & Fleischmann, 1997). Additionally, the difference between Trail Making Test B and Trail Making Test A (TMT-B − TMT-A) can be used as an indicator for executive functioning (Lehrner, Maly, Gleiß, Auff, & Dal-Bianco, 2007).

Reportedly, among patients complaining about cognitive decline, the NTBV can differentiate between those with and without AD (Lehrner et al., 2016). Overall internal consistency is high across different study populations, with a Cronbach's alpha of α = .83-.93. Within a sample of patients with dementia diagnosis it ranged from α =.87 to α =.89. Test-retest reliability across different populations reached values of r = .69-.94 (Hitzl, 2015 as cited in Lehrner et al., 2016; Lehrner et al., 2007; Macher, 2013 as cited in Lehrner et al., 2016).

Wortschatztest. With the Wortschatztest (WST) verbal intelligence and speech comprehension can be measured. It consists of 40 items, with each item containing one target word and five made up distractor words. The more target words the patient recognises, the higher the WST-score. Resulting WST-scores can further be transformed into IQ-scores. The WST is a standardized vocabulary test with an internal consistency (Cronbach's alpha) of α = .94 and a split-half reliability of r = .95 (Schmidt & Metzler, 1992).

Beck Depression Inventory-II. The German revision of the Beck Depression Inventory-II (BDI-II) was used to assess self-rated severity of depression. It consists of 21 items, which patients must rate on a four-point Likert scale. A score higher than 13 indicates that a patient has clinically relevant depressive symptoms. Evaluations including clinical and non-clinical samples propose that the BDI-II has an internal consistency (Cronbach's alpha) of $\alpha \geq .84$ and a test-retest reliability of $r \geq .75$ (Hautzinger, Keller, & Kühner, 2006; Kühner, Bürger, Keller, & Hautzinger, 2007).

Bayer Activities of Daily Living Scale. In this study, activities of daily living were assessed using the Bayer activities of daily living (B-ADL) scale, as this is an instrument particularly suitable for patients in early stages of dementia. The B-ADL consists of 25 items, which were designed for community residing patients with MCI or mild to moderate dementia and must be rated on a 10-point Likert scale. Additionally, items can be rated "not applicable" or "unknown". It can either be completed by patients themselves or by a person sufficiently familiar with the patient, like a family member or a caregiver. A higher B-ADL score stands for an increased functional impairment in daily life. To calculate a global B-ADL score, the scores of the individual items are summed up. The sum score is then divided by the number of items rated from 1 to 10, which means that items rated "not applicable" or "unknown" are not included in the computation of the global scale score. The resulting B-ADL score ranges between the values 1.00 and 10.00 (Hindmarch et al., 1998). According to several validation studies, the B-ADL scale has an internal consistency (Cronbach's alpha) of $\alpha > .97$ and a test-retest reliability of $r = .757$ (Erzigkeit et al., 2001; Erzigkeit & Lehfeld, 2010).

Statistical Analysis

Data was analysed using IBM SPSS (version 26) with the significance level set to $\alpha = .05$. As a first step, data was checked for plausibility. In this procedure, missing diagnoses were added and implausible values were corrected. Next, the variable conversion to dementia was calculated, which divides the sample into converters (1) and non-converters (0), based on their diagnosis at follow-up. Descriptive statistics were used to visualize the distribution of the main parameters (test scores of B-ADL, MMSE, BDI-II, WST-IQ and VVT 3.0 Screening) as well as the demographic variables (age, gender and education level).

Research question 1. The main research question of this study aims to answer whether ADL can be used to predict conversion to dementia. Since conversion to dementia is a binary outcome variable, logistic regression was used to investigate this question. As regression models are most efficient when all major predictors are considered, it was decided to include B-ADL score, demographic variables and NTBV variables into hierarchical logistic regression. In order to avoid multicollinearity and achieve a dimensional reduction without losing informational content of the NTBV, a principal components analysis (PCA) was conducted. Within the PCA, uncorrelated factor scores were calculated using the Anderson-Rubin method.

The standardized factor scores with $\mu = 0$ and $\sigma = 1$ were then added as predictors in block three of the logistic regression. To investigate the isolated contribution of the main variable of interest, the B-ADL score, it was entered in the first block. Afterwards, demographic variables were added in block two and NTBV factors in block three of the model. In this way it was possible to examine whether the influence of B-ADL on conversion changes by adding demographic variables and NTBV factor scores to the regression model.

Next, a receiver operator characteristic (ROC) curve for B-ADL was calculated. The cut-off score for determining sensitivity and specificity of B-ADL self-rating was set according to Youden's index (J = Sensitivity + Specificity – 1; Youden, 1950). For estimating overall accuracy of B-ADL, the area under the curve (AUC) was interpreted. Sensitivity and specificity from the chosen cut-off point were further used to determine positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (LR+) and negative likelihood ratio (LR-). The LR+ describes the ratio of the probability for receiving an AD diagnosis among individuals with AD, to the probability for receiving an AD diagnosis among subjects without AD. LR- on the other hand, defines the ratio of the probability for getting a negative test result among participants with AD, to the probability for receiving an AD diagnosis among participants without AD. Resulting likelihood ratios were interpreted as proposed by Jaeschke et al. (1994):

- LR+ > 10 and LR- < 0.1 indicate high and often convincing diagnostic accuracy
- LR+ 5 – 10 and LR- 0.1 – 0.2 indicate moderate diagnostic accuracy
- LR+ 2 – 5 and LR- 0.2 – 0.5 indicate weak diagnostic accuracy
- LR+ 1 – 2 and LR- 0.5 – 1 indicate scarce diagnostic accuracy

Research question 2. In order to study possible associations between B-ADL baseline and follow-up scores with demographic variables and different test scores, correlation coefficients were calculated. When visualizing the relationships between the variables of interest with scatterplots, several non-linear relationships were discovered. Therefore, it was decided to use Spearman´s correlation coefficient (r_S), which is more robust to violations of assumptions than Pearson´s correlation coefficient (*r;* Field, 2018). Correlational effect sizes were interpreted according to criteria by Cohen (1988), with $r \geq .10$ indicating a small, $r \geq .30$ a medium and $r \geq .50$ a large effect.

Research question 3. To address the third and final research question, RCI was calculated as proposed by Chelune et al (1993), Maassen et al. (2006) and McSweeny et al. (1993). Calculations were conducted using the Reliable Change Calculator 1.0 by Hinton-Bayre (2010). To detect upper and lower limits for significant change, 90% confidence intervals were computed. Under the assumption that a RCI more extreme than ± 1.64 implies reliable change, the following formula was used:

$$CI = \pm 1.64 \cdot SED + (M_2 - M_1) \quad (14)$$

Confidence intervals were determined for the diagnostic subgroups, with *SED* representing the standard error of difference, M_2 the mean at follow-up and M_1 the baseline-mean of the respective group. Depending on whether their RCI was within or outside the confidence interval, individuals were then divided into subjects who deteriorated, improved or showed no change in RCI. Afterwards, the three RCI models were compared regarding their responsiveness to positive and negative change.

Results

First, demographic variables and variables of interest were analysed regarding their distribution within the diagnostic subgroups and the total sample. For that purpose, the Shapiro-Wilk test was used, as it has higher statistical power than the Kolmogorov-Smirnov test (Field, 2018). Results then were interpreted in combination with histograms, Q-Q plots and values of skewness and kurtosis. It was decided to describe all variables uniformly via median and interquartile range (IQR) because of the distributional characteristics.

Descriptive Statistics

The sample consists of 315 subjects, including 160 female (50.8%) and 155 male (49.2%) participants. It has a median age of 67 years (IQR = 61−74) and a median education level of 12 years (IQR = 8−16). At the time of the baseline evaluation, 46 subjects (14.6%) were assigned to the control group and 45 (14.3%) to the SCD condition. In total, 224 participants (71.1%) received the MCI diagnosis, with 126 (40%) individuals being diagnosed with naMCI and 98 (31.1%) with aMCI. In Table 1, demographic characteristics are compared between diagnostic groups, based on the diagnosis given at baseline. It also shows MMSE scores according to diagnostic subgroup, as MMSE scores are available for all 315 participants.

Some neuropsychological tests were not completed by all participants, including the VVT 3.0 Screening, the BDI-II and the WST. At baseline, the VVT 3.0 Screening was performed by 105 subjects, the BDI-II by 309 subjects and the WST by 310 subjects. The median VVT 3.0 Screening, BDI-II and WST-IQ scores of the total sample and the diagnostic subgroups are shown in Table 2.

Table 1

Baseline Characteristics for the Diagnostic Subgroups

	n	Age	f/m	Education	MMSE
CG	46	61 (58.75−62)	16/30	12 (8−18)	29 (28−30)
SCD	45	66 (61−73.5)	21/24	12 (9−15)	29 (28−30)
naMCI	126	70 (63.75−75)	74/52	10.5 (8−14.25)	28 (27−29)
aMCI	98	69 (62−75)	49/49	12 (8.75−16)	28 (26−29)
Total (N)	315	67 (61−74)	160/155	12 (8−16)	28 (27−29)

Note. variables are described via median and interquartile range; f/m = proportion of female and male participants; CG = control group; SCD = subjective cognitive decline; naMCI = nonamnestic mild cognitive impairment; aMCI = amnestic mild cognitive impairment; MMSE = Mini Mental State Examination.

Follow-up testing was conducted between 12 to 48 months after the first assessment. At the second testing point, 62 participants (19.7%) were diagnosed with SCD and 172 (54.6%) with a form of MCI. Of the patients with MCI, 90 (28.6%) received the naMCI and 82 (26%) the aMCI diagnosis. At follow-up, the control group consisted of 50 healthy subjects (15.9%). Since 31 individuals were diagnosed with AD at follow-up, the conversion rate of this study is 9.8%. For both testing points, the number of participants within each diagnostic subgroup is shown in Table 3.

Table 2

Baseline VVT 3.0 Screening, BDI-II and WST-IQ Scores for the Diagnostic Subgroups

	VVT 3.0 Screening		BDI-II		WST-IQ	
	n	*Mdn*	*n*	*Mdn*	*n*	*Mdn*
CG	17	10 (10−10)	46	2 (1−5.25)	46	116 (107−122)
SCD	20	10 (10−10)	44	8 (5.25−13)	44	116 (107−122)
naMCI	42	9.5 (9−10)	121	10 (6−14.5)	124	110 (99−118)
aMCI	26	9 (9−10)	98	8.5 (4−14)	96	110 (99−122)
Total	105	10 (9−10)	309	8 (4−13)	310	110 (101−122)

Note. Variables are described via median *(Mdn)* and interquartile range; VVT 3.0 Screening = Vienna Visuo-Constructional Test 3.0 Screening; BDI-II = Beck Depression Inventory-II; WST = Wortschatztest; CG = control group; SCD = subjective cognitive decline; naMCI = nonamnestic mild cognitive impairment; aMCI = amnestic mild cognitive impairment.

Table 3

Diagnostic Subgroups at Baseline and Follow-up

		Follow-up					
		CG	SCD	naMCI	aMCI	AD	Total
Baseline	CG	46 (14.6)	−	−	−	−	46 (14.6)
	SCD	2 (0.6)	25 (7.9)	10 (3.2)	8 (2.5)	−	45 (14.3)
	naMCI	2 (0.6)	28 (8.9)	61 (19.4)	31 (9.8)	4 (1.3)	126 (40)
	aMCI	−	9 (2.9)	19 (6)	43 (13.7)	27 (8.6)	98 (31.1)
Total		50 (15.9)	62 (19.7)	90 (28.6)	82 (26)	31(9.8)	315 (100)

Note. CG = control group; SCD = subjective cognitive decline; naMCI = nonamnestic mild cognitive impairment; aMCI = amnestic mild cognitive impairment; AD = Alzheimer's disease; % in brackets.

Depending on whether patients have developed AD at follow-up or not, they can be divided into converters and non-converters. The sample consists of 31 converters (9.8%) and 284 non-converters (90.2%). In the total sample, the second testing took place after a median test-retest interval of 24 months (16−36). The median test-retest interval for converters was 18 months (14− 32) and for non-converters 25 months (16− 37). A comparison between demographic characteristics and MMSE scores of converters and non-converters is shown in Table 4. Since not all individuals have completed VVT 3.0 Screening, BDI-II and WST, the numbers of participating converters and non-converters with their median test scores are given in Table 5. At baseline, the B-ADL scale was completed by all 315 participants. After the follow-up interval 273 subjects, including 14 converters and 259 non-converters, filled out the B-ADL scale. Median B-ADL scores of converters versus non-converters at baseline and follow-up are presented in Table 6.

Table 4

Baseline Characteristics for Converters and Non-converters

	Converters	Non-converters
n	31	284
Age	73 (67−77)	67 (60−73)
f/m	17/14	143/141
Education	12 (8−16)	12 (8−16)
MMSE	26 (25−27)	29 (27−29)

Note. Variables are described via median and interquartile range; f/m = proportion of female and male participants; MMSE = Mini Mental State Examination.

Table 5

Baseline VVT 3.0 Screening, BDI-II and WST-IQ Scores for Converters and Non-converters

	Converters		Non-Converters	
	n	*Mdn*	*n*	*Mdn*
VVT 3.0 Screening	8	8.50 (8−10)	97	10 (9−10)
BDI-II	31	7 (4−10)	278	8 (3−13)
WST-IQ	30	108.5 (97−122)	280	110 (101.75−120.25)

Note. Variables are described via median *(Mdn)* and interquartile range; VVT 3.0 Screening = Vienna Visuo-Constructional Test 3.0 Screening; BDI-II = Beck Depression Inventory-II; WST = Wortschatztest.

Table 6

B-ADL Scores for Converters and Non-converters at Baseline and Follow-up

	B-ADL Baseline		B-ADL Follow-up	
	n	*Mdn*	*n*	*Mdn*
Converters	31	1.6 (1.32−3)	14	2.02 (1.19−3.57)
Non-converters	284	1.6 (1.21−2.34)	259	1.56 (1.28−2.2)
Total	315	1.6 (1.24−2.36)	273	1.6 (1.28−2.27)

Note. Variables are described via median *(Mdn)* and interquartile range; B-ADL = Bayer Activities of Daily Living.

The comparison of B-ADL in the individual subgroups shows that the median B-ADL score is continuously increasing across the diagnostic groups. At baseline, the lowest B-ADL score can be found in the control group and the highest value was reached by individuals with aMCI. The median B-ADL score at baseline was 1.30 for control group, 1.50 for SCD group, 1.66 for naMCI group and 1.78 for individuals with aMCI. Figure 1 shows the distribution of self-rated B-ADL across the diagnostic subgroups at the first testing point. The same pattern applies at follow-up, with a median B-ADL score of 1.23 for control group, 1.60 for SCD group, 1.65 for naMCI group, 1.92 for aMCI group and 2.02 for subjects with AD diagnosis. The distribution of B-ADL score across the diagnostic subgroups at the second testing point is visualized in Figure 2. Outliers are displayed as circles and asterisks show extreme values. From the length of the whiskers, it can be concluded that the distribution of B-ADL score within the subgroups at both testing points is skewed.

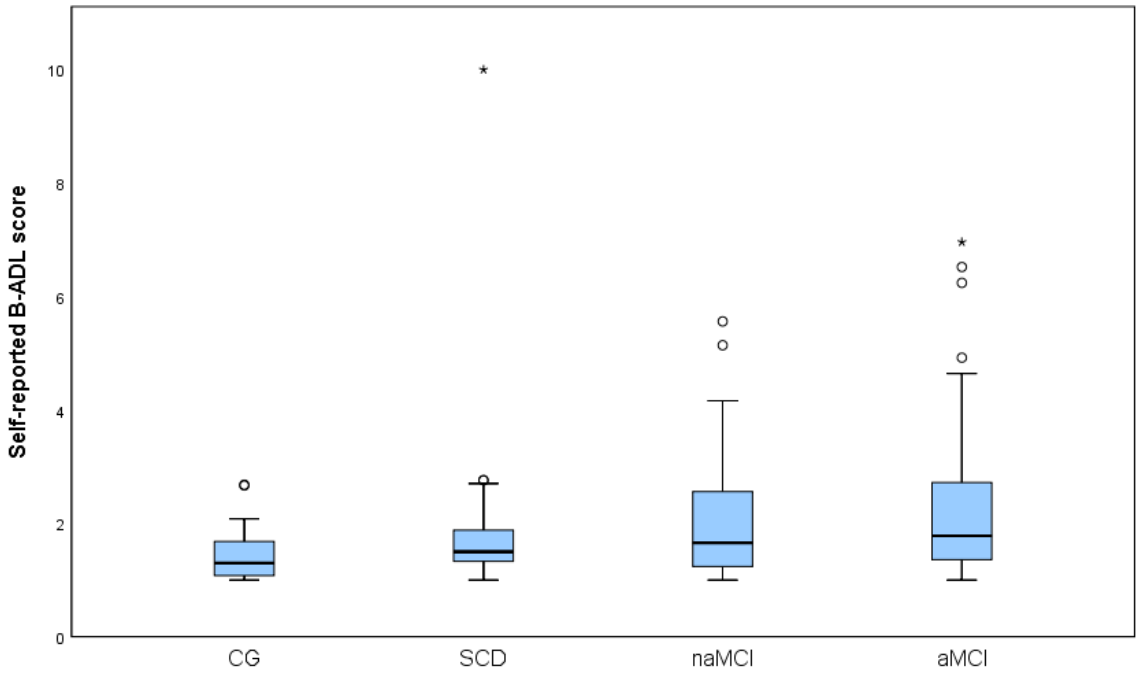

Figure 1. Boxplots of Bayer activities of daily living (B-ADL) score across diagnostic subgroups at baseline.

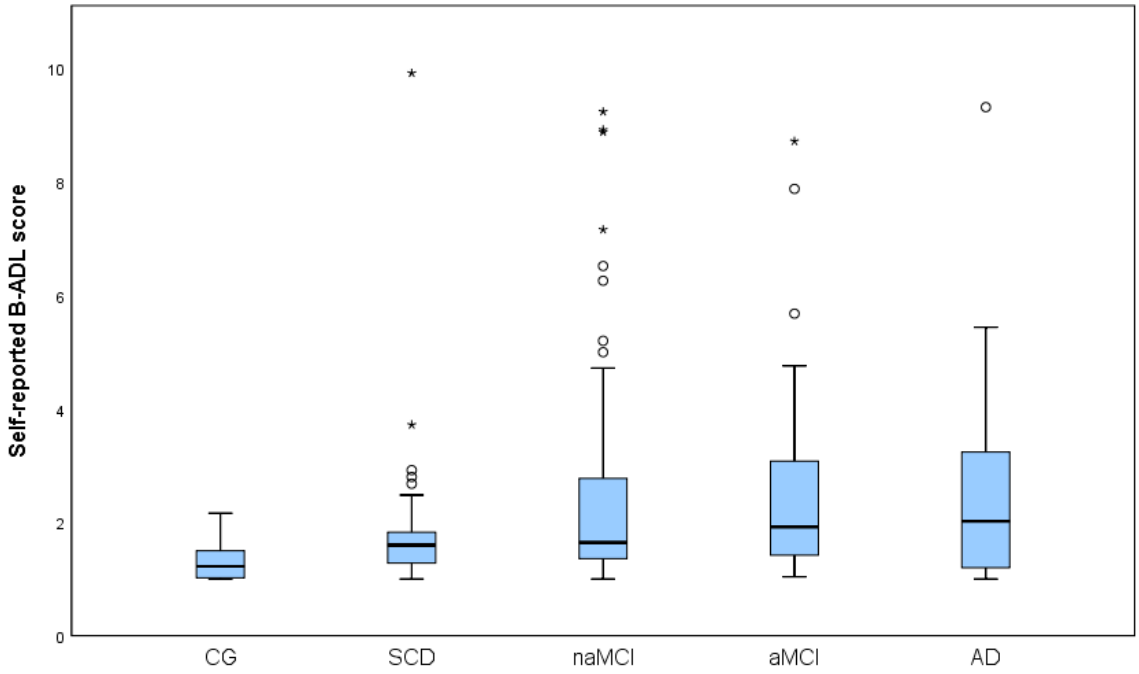

Figure 2. Boxplots of Bayer activities of daily living (B-ADL) score across diagnostic subgroups at follow-up.

Logistic Regression

The use of logistic regression is tied to certain assumptions that mostly agree with the assumptions from the general linear model. However, the linearity assumption in logistic regression does not refer to the relationship between outcome and predictor, but to linearity between the predictor and the logit of the outcome variable. Linearity is automatically given for dichotomous variables, which only applies to the predictor variable gender. The other predictors are continuous variables, therefore it needs to be ensured that each of them is linearly related to the logit of the outcome variable. In order to do so, the interaction of each continuous predictor with its own natural logarithm (ln) must be included into the logistic regression model. If the interaction is not significant, it can be assumed that linearity is given. All interaction terms had p-values $\geq .999$, which indicates that linearity of the logit is met for all continuous predictors.

Some of the variables from the NTBV-subtests correlate highly with each other, which signals that there is multicollinearity within the used data. As parameters of logistic regression

can be affected by collinearity, the variance inflation factor (VIF) was computed first. The VIF shows whether there is a strong linear relationship between predictors and should not include values greater than 10. The tolerance statistic, which is the reciprocal of the VIF and should not reach a value below .10, was also considered (Field, 2018). VIF and tolerance statistic were calculated for all variables of interest, including variables from NTBV subtests, demographic variables and B-ADL self-rating. The following variables from NTBV subtests had VIF values > 10 and tolerance statistics < .10: AKT Time, Stroop-Test Time Colours, Stroop-Test Time Words, Stroop-Test Total / Time, Stroop-Test Interference, Interference Time (c.I.), Interference Total / Time (c.I.).

In order to combine collinear variables and by that reduce the total number of variables, it was decided to perform PCA. To assure that the variables are suitable for PCA, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and the Bartlett´s test of sphericity were calculated first. With a value of .83, the KMO statistic was above the cut-off score of .50 and the Bartlett´s test was highly significant ($p < .001$). The production of KMO values for individual variables showed that the variable Five-Point Test Perseverations had a KMO value below .50, therefore it was excluded from the analysis.

In PCA, there are several approaches for determining the number of factors to extract, which often lead to different results. Therefore, three different criteria were applied and the outcomes were compared. First, the eigenvalues were presented graphically by using a scree plot. The scree plot showed a point of inflexion at the fourth data point, hence it suggests a three-component solution that explains 61.35% of total variance. Secondly, it was examined how many factors had eigenvalues > 1, since the Kaiser's criterion proposes only extracting those factors. Five factors had eigenvalues greater than one, recommending a five-component solution that explains 70.5% of total variance. Finally, to use a method that minimizes over- and under-extraction of factors, parallel analysis was conducted. In parallel analysis, a random data set, which has the same characteristics as the original data, is generated. Then the eigenvalues of the real data are compared to the eigenvalues of the artificial data. Only those factors are retained whose observed eigenvalue is greater than the eigenvalue of the factors from the random data set (O'connor, 2000). In the parallel analysis, the eigenvalue of the random data was greater than the observed data eigenvalue from the third component onwards, which would suggest a two-component solution that explains 55.36% of total variance.

Table 7

Factor Loadings for Principal Components Analysis with Orthogonal Varimax Rotation of NTBV Variables

	Factor				
	1	2	3	4	5
AKT Total / Time	**.545**	.173	-.297	.389	-.261
AKT Time	**-.503**	-.156	.365	-.356	.298
Digit-Symbol Test (WAIS-R)	**.621**	.265	-.240	.309	-.262
Five-Point Test Total Correct	**.535**	.145	-.126	.092	-.230
Interference Time (c.I.)	**-.743**	-.139	.340	-.212	.186
Interference Total / Time (c.I.)	**.752**	.187	-.320	.195	-.157
PVF Total Words	**.535**	.145	-.126	.092	-.230
Stroop-Test Time Colours	**-.655**	-.223	.365	-.036	.138
SVF Total Words	**.513**	.380	-.177	.229	-.315
Symbols Counting (c.I.)	**-.709**	-.025	-.001	-.187	.068
VSRT Word Span	.225	**.775**	-.027	.044	-.223
VSRT Learning Performance	.232	**.866**	-.153	.105	-.209
VSRT Delayed Recall	.149	**.848**	-.210	.133	-.184
VSRT Recognition	.052	**.705**	-.134	.113	.129
Stroop-Test Time Words	-.374	-.229	**.829**	-.197	.195
Stroop-Test Total / Time	.398	.157	**-.810**	.138	-.099
Stroop-Test Interference	-.160	-.178	**.879**	-.227	.168
Maze Time (NAI)	-.192	-.096	.197	**-.854**	.221
Maze Total / Time (NAI)	.256	.124	-.121	**.844**	-.155
TMT-A	-.404	-.186	.217	**-.494**	.169
BNT	.126	.153	-.021	.129	**-.513**
TMT-B	-.349	-.110	.251	-.294	**.791**
TMT-B − TMT-A	-.251	-.056	.210	-.151	**.864**
Eigenvalues	10.75	1.99	1.38	1.06	1.04
% of variance	46.72	8.64	6.00	4.63	4.52

Note. Highest factor loadings appear in bold. NTBV = Neuropsychological Test Battery Vienna; TMT-B = Trail Making Test Version B; AKT = Alters-Konzentrations-Test; WAIS-R = Wechsler Adult Intelligence Scale-Revised; NAI = Nürnberger Alters Inventar; c.I. = Cerebral Insufficiency Test; SVF = Semantic Verbal Fluency; BNT = Boston Naming Test; PVF = Phonematic Verbal Fluency; VSRT = Verbal Selective Reminding Test.

Since the NTBV was constructed to measure five cognitive domains, it was decided to extract five factors, as suggested by Kaiser's criterion. Although Kaiser's criterion has been criticized for causing overextraction, it is suitable for samples with N > 250 and an average communality ≥ .60. As the sample consists of 315 participants and has an average communality of .68, it is appropriate to use Kaiser's criterion. Additionally, when using PCA to avoid multicollinearity in regression, it is recommended to rather accept over- than underextraction of factors (Field, 2018).

PCA was conducted on 23 variables from NTBV subtests. To make interpretation easier and gain uncorrelated factor scores, which could then be used as predictors in the logistic regression model, orthogonal varimax rotation was chosen. In Table 7, the rotated factor loadings of the NTBV variables on the five extracted factors are presented. For each variable, the highest factor loading is written in bold. Moreover, the eigenvalues and the percent of explained variance for each factor are displayed. To identify the constructs behind the retained factors, the variables with the highest loadings per factor were examined for common themes. The five NTBV factors were then named as follows:

- Factor 1: Attention and Verbal Fluency
- Factor 2: Memory
- Factor 3: Cognitive Control
- Factor 4: Processing Speed
- Factor 5: Executive Functioning and Word Finding

After the factor extraction, VIF values and tolerance statistics of all variables that were considered as predictors were checked again. VIF values were between 1.09 and 1.66, tolerance statistics were ≥ .60, therefore it can be assumed that there is no multicollinearity between the predictors of the logistic regression model.

Before logistic regression can be conducted, it needs to be checked whether there are cases in the data that fit the model poorly and therefore could influence the logistic regression model. For that purpose, residuals and influence statistics were examined. As it is recommended to combine different influence statistic methods, Cook´s distance, leverage statistics and DFBeta were used (Field, 2018). The results of the residual analysis were acceptable, as no cases had values above 3 and six cases had values above 2, out of which two cases had values greater than 2.5. Regarding influential statistics, two cases with DFBeta > 1 and two cases with

Cook's distance > 1 were found. In addition, 30 cases had leverage values three times the average leverage value, indicating those data points might have an influence on the regression model. Finally, it was examined whether there are cases in the data that were identified as influential cases by several methods. There was only one case with Cook's distance > 1 and a leverage value three times greater than the average leverage value. With increasing sample size, individual data points have less influence on the model, which is why it was decided not to eliminate any cases.

In the first block of logistic regression, the B-ADL score was entered as a single predictor. Compared to a model without predictors, no significant improvement in model fit was achieved by adding B-ADL as predictor for conversion to dementia ($\chi^2(1) = 1.17$, $p = .279$). In block two, demographic variables age, gender and education level were added. The overall model summary showed that the model in block two fits the data significantly better than a model with no predictors ($\chi^2(4) = 13.54$, $p = .009$) and it also fits the data significantly better than the model from block one ($\chi^2(3) = 12.36$, $p = .009$). From the predictor variables in block two, age was identified as significant predictor ($b = 0.08$, Wald $\chi^2(1) = 10.85$, $p = .001$). However, with a Nagelkerke's R^2 of .09, the model of block two seems to explain only a very small amount of outcome variance. In the third block of logistic regression, the NTBV factor scores, which had been extracted in PCA, were included as predictors. Compared to the previous model, adding the NTBV factors in block three significantly improved the model fit ($\chi^2(5) = 84.21$, $p \leq .001$).

Table 8 includes the Wald statistics for all predictor variables and their *p*-values. It further contains the regression coefficient *b* and the *OR*, which is the exponential of *b*. A significant relationship between predictor and outcome can only be assumed, if the confidence interval of the *OR* does not include the value 1. As significant *p*-values were found for all NTBV factors and none of the *OR* included the value 1, it can be concluded that the five NTBV factor scores function as significant predictors for conversion to dementia.

Table 8

Block 3 of Hierarchical Logistic Regression

						95% CI for *OR*	
	b	*SE*	Wald	Sig.	*OR*	*LL*	*UL*
B-ADL	-0.454	0.267	2.893	.089	0.635	0.376	1.072
Age	-0.085	0.040	4.535	.033*	0.919	0.850	0.993
Gender	-1.276	0.637	4.009	.045*	0.279	0.080	0.973
Education Level	0.065	0.073	0.789	.374	1.067	0.925	1.230
F1 Attention & Verbal Fluency	-1.236	0.329	14.073	.000**	0.291	0.152	0.554
F2 Memory	-2.385	0.419	32.417	.000**	0.092	0.041	0.209
F3 Cognitive Control	1.049	0.252	17.349	.000**	2.856	1.743	4.680
F4 Processing Speed	-0.764	0.288	7.032	.008*	0.466	0.265	0.819
F5 Executive Function & Word Fluency	0.982	0.252	15.234	.000**	2.671	1.631	4.374
Constant	2.284	2.929	0.608	.435	9.818		

Note. R^2 = .271 (Cox-Snell), .575 (Nagelkerke); Model χ^2 (9) = 97.744, $p \leq .001$; * $p < .05$; ** $p \leq .001$; *b* = regression coefficient; *SE* = standard error; *OR* = odds ratio; CI = confidence interval; *LL* = lower limit; *UL* = upper limit; B-ADL = Bayer Activities of Daily Living.

For the main variable of interest, the B-ADL score, no significant result was found in logistic regression analysis. To further assess the prognostic validity of self-rated ADL and determine whether there is a difference in self-rated ADL between subjects who do and who do not convert to dementia, a ROC curve for B-ADL score was calculated. In Figure 3, the ROC curve for the B-ADL self-rating is shown and compared to a reference line. On the blue ROC curve for B-ADL score, all possible cut-off points are included, with each being assigned specific values for sensitivity and specificity. It can be compared to the red reference line, which represents a test with both sensitivity and specificity of zero, that allows no discrimination between individuals with and without dementia. Based on the maximum value of Youden´s index (J), a cut-off score of 3.28 was selected. With that cut-off point, B-ADL had a sensitivity of .23 and a specificity of .93. These values for sensitivity and specificity were then used to calculate a PPV of .27 and a NPV of .92. The resulting LR- was 0.83, suggesting scarce and rarely important diagnostic evidence. The LR+ was 3.37, suggesting weak diagnostic evidence, which however can be important under some circumstances (Jaeschke et al., 1994).

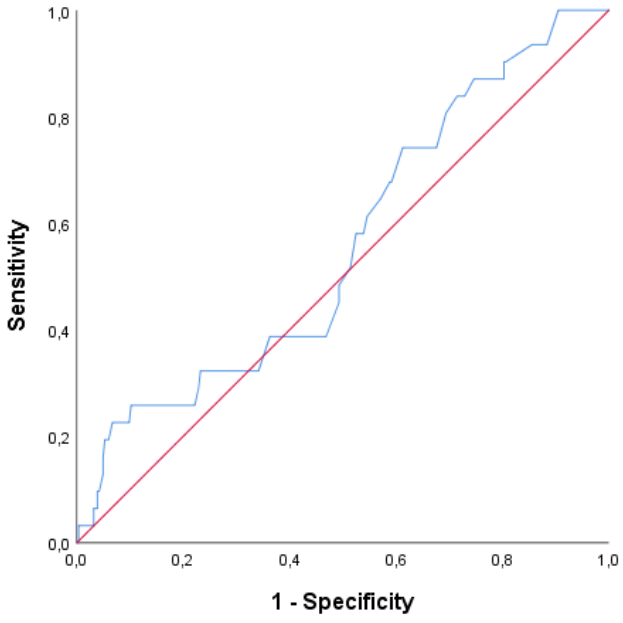

Figure 3. ROC curve for B-ADL self-rating.

As sensitivity, specificity, PPV, NPV, LR+ and LR- are specific for the chosen cut-off point, AUC was interpreted to determine the overall accuracy of B-ADL. It can range between 0 and 1, with 0 indicating that the test hast no discriminative abilities and 1 indicating that a test is perfectly accurate (Mandrekar, 2010). The AUC for B-ADL was .56, 95% CI [.46, .66], which means that the chance that B-ADL self-rating correctly distinguishes between patients with and without dementia is 56%.

Correlation Analysis

Since the visualization of the data revealed several non-linear relationships between variables, possible associations between baseline and follow-up B-ADL scores with other variables were investigated with Spearman´s correlation coefficient (r_S). First, it was tested if B-ADL scores at baseline and follow-up correlate with each other. The result was highly significant with a large effect size ($r_S = .66$, $p < .001$, 95% CI for r_S [.58, .74]). Next, associations between B-ADL baseline and follow-up with demographic characteristics and scores of MMSE, VVT 3.0 Screening, BDI-II and WST-IQ were investigated. The resulting correlation coefficients are presented in Table 9.

B-ADL baseline and B-ADL follow-up both significantly correlated with the variables MMSE and BDI-II. While correlations between B-ADL scores and MMSE correspond to a small effect, correlations between BDI-II and B-ADL scores at both testing points can be interpreted as large. Additionally, a small correlation with age was found but only for the follow-up B-ADL score. A significant relationship between two variables can only be assumed if the confidence interval for the correlation coefficient does not include the value 0 (Field, 2018). None of the confidence intervals for significant correlation coefficients contain 0, therefore the assumption that the associations are significant can be retained.

Like B-ADL, NTBV was conducted at baseline and at follow up. In Table 10, correlations for each NTBV subtest with B-ADL baseline score are shown. Except for two variables, Maze Time and Maze Total / Time, all NTBV subtests correlated with B-ADL score at baseline. With effect sizes ranging from $|r_S| = .114$ to $|r_S| = .286$, all significant correlations can be interpreted as small.

Table 9

Correlation Coefficients between Demographic Variables, MMSE, VVT 3.0 Screening, BDI-II and WST-IQ with B-ADL Scores

	B-ADL Baseline	B-ADL Follow-up
	r_S (95% CI[a])	r_S (95% CI[a])
Age	.091	.284** (.083, .475)
Gender	-.149	-.164
Education	-.059	-.039
MMSE	-.198* (-.381, -.017)	-.291** (-.468, -.093)
VVT 3.0 Screening	-.119	-.195
BDI-II	.551** (.386, .691)	.550** (.366, .702)
WST-IQ	-.014	.119

Note. Confidence intervals are based on 2000 bootstrap samples; * $p < .05$; ** $p \leq .001$; B-ADL = Bayer Activities of Daily Living; r_S = Spearman's correlation coefficient; CI = confidence interval; MMSE = Mini Mental State Examination; VVT 3.0 S. = Vienna Visuo-Constructional Test 3.0 Screening; WST = Wortschatztest; BDI-II = Beck Depression Inventory-II; [a] given for significant values only.

Table 11 displays correlations between NTBV variables and B-ADL at follow-up. All NTBV subtests, except Five-Point Test Perseverations, correlated significantly with B-ADL follow-up score. The effect sizes reached values between $|r_S| = .154$ and $|r_S| = .372$. With correlation coefficients above .30, seven correlations had medium effect sizes. The other significant correlation coefficients were at least above .10 and thus indicate a small effect. For both, Table 10 and 11, confidence intervals for significant values were calculated. None of them contained the value 0, suggesting significant correlations indeed represent significant associations between variables.

Table 10

Spearman Correlation Coefficients (r_S) for NTBV Subtests and B-ADL at Baseline

NTBV Baseline	B-ADL Baseline
TMT-B	.207**
AKT Time	.217**
AKT Total / Time	-.208**
Digit-Symbol Test (WAIS-R)	-.286**
Maze Time (NAI)	.106
Maze Total / Time (NAI)	-.084
Interference Time (c.I.)	.211**
Interference Total / Time (c.I.)	-.210**
SVF Total Words	-.256**
TMT-B − TMT-A	.163**
TMT-A	.182**
Stroop-Test Time Colours	.219**
Symbols Counting (c.I.)	.123*
SVF Tools	-.203**
BNT	-.139*
PVF Total Words	-.215**
PVF L-Words	-.196**
PVF F-Words	-.120*
PVF B-Words	-.206**
SVF Animals	-.212**
VSRT Learning Performance	-.229**
VSRT Delayed Recall	-.222**
VSRT Word Span	-.179**
VSRT Recognition	-.139*
SVF Supermarket Items	-.248**
Stroop-Test Total / Time	-.176**
Stroop-Test Time Words	.202**
Stroop-Test Interference	.134*
Five-Point Test Total Correct	-.217**
Five-Point Test Perseverations	-.114*

Note. * $p < .05$; ** $p \leq .001$; B-ADL = Bayer Activities of Daily Living; r_S = Spearman's correlation coefficient; CI = confidence interval; NTBV = Neuropsychological Test Battery Vienna; TMT-B = Trail Making Test Version B; AKT = Alters-Konzentrations-Test; WAIS-R = Wechsler Adult Intelligence Scale-Revised; NAI = Nürnberger Alters Inventar; c.I. = Cerebral Insufficiency Test; SVF = Semantic Verbal Fluency; BNT = Boston Naming Test; PVF = Phonematic Verbal Fluency; VSRT = Verbal Selective Reminding Test; [a] given for significant values only.

Table 11

Spearman Correlation Coefficients (r_S) for NTBV Subtests and B-ADL at Follow-up

NTBV Follow-up	B-ADL Follow-up
TMT-B	.274**
AKT Time	.346**
AKT Total / Time	-.354**
Digit-Symbol Test (WAIS-R)	-.367**
Maze Time (NAI)	.204**
Maze Total / Time (NAI)	-.191**
Interference Time (c.I.)	.184**
Interference Total / Time (c.I.)	-.174**
SVF Total Words	-.365**
TMT-B − TMT-A	.185**
TMT-A	.306**
Stroop-Test Time Colours	.265**
Symbols Counting (c.I.)	.273**
SVF Tools	-.292**
BNT	-.193**
PVF Total Words	-.223**
PVF L-Words	-.190**
PVF F-Words	-.154*
PVF B-Words	-.245**
SVF Animals	-.293**
VSRT Learning Performance	-.372**
VSRT Delayed Recall	-.298**
VSRT Word Span	-.292**
VSRT Recognition	-.179**
SVF Supermarket Items	-.323**
Stroop-Test Total / Time	-.242**
Stroop-Test Time Words	.251**
Stroop-Test Interference	.182**
Five-Point Test Total Correct	-.262**
Five-Point Test Perseverations	-.070

Note. * $p < .05$; ** $p \leq .001$; B-ADL = Bayer Activities of Daily Living; r_S = Spearman's correlation coefficient; CI = confidence interval; NTBV = Neuropsychological Test Battery Vienna; TMT-B = Trail Making Test Version B; AKT = Alters-Konzentrations-Test; WAIS-R = Wechsler Adult Intelligence Scale-Revised; NAI = Nürnberger Alters Inventar; c.I. = Cerebral Insufficiency Test; SVF = Semantic Verbal Fluency; BNT = Boston Naming Test; PVF = Phonematic Verbal Fluency; VSRT = Verbal Selective Reminding Test; [a] given for significant values only.

Reliable Change Index

To assess differences in B-ADL score between first and second testing that are not caused by changes in functional abilities, RCI was calculated. While all 315 participants completed the B-ADL self-rating at baseline, only 273 individuals took part at the follow-up. Table 12 contains means, standard deviations and test-retest reliabilities of the diagnostic subgroups at both testing points, which were then used to generate RCI´s for the individual participants. The control group was used as reference group. As B-ADL scores are not normally distributed among the groups, it was decided to use Spearman´s correlation coefficient for calculating test-retest reliability (r_{tt}).

Table 12

B-ADL Characteristics at Baseline and Follow-up for CG, SCD, naMCI, aMCI, Converters and Non-converters

	B-ADL Baseline			B-ADL Follow-up			
	n	*M*	*SD*	*n*	*M*	*SD*	r_{tt}
CG	46	1.42	0.43	46	1.30	0.32	.65**
SCD	45	1.84	1.35	40	1.67	0.55	.65**
naMCI	126	1.94	0.88	114	2.30	1.62	.64**
aMCI	98	2.21	1.20	73	2.61	1.95	.57**

Note. B-ADL = Bayer Activities of Daily Living; *M* = mean; *SD* = standard deviation; r_{tt} = test-retest reliability; CG = control group; SCD = subjective cognitive decline; naMCI = nonamnestic mild cognitive impairment; aMCI = amnestic mild cognitive impairment; * $p < .05$; ** $p \leq .001$.

Next, confidence intervals were calculated to determine limiting values for the identification of statistically significant change in B-ADL score. For computing the confidence interval, the standard error of difference (*SED*) and the standard error of measurement (*SEM*) were used. Table 13 contains *SEM*, *SED*, upper and lower limits of the confidence intervals for SCD, naMCI and aMCI group.

Table 13

SEM, SED and Limiting Values for SCD, naMCI, aMCI, Converters and Non-converters

	SEM	*SED*	Limiting Value Deterioration	Limiting Value Improvement
SCD	0.80	1.13	-2.02	1.68
naMCI	0.53	0.75	-0.87	1.59
aMCI	0.79	1.12	-1.44	2.24
Converters	0.72	1.02	-1.01	2.33
Non-converters	0.80	1.13	-1.66	2.04

Note. SEM = standard error of measurement; SED = standard error of difference; SCD = subjective cognitive decline; naMCI = nonamnestic mild cognitive impairment; aMCI = amnestic mild cognitive impairment.

Afterwards it was examined how many individual RCI values are below, within and above the confidence interval. This procedure was carried out using the methods for RCI calculation by Chelune et al. (1993), Maassen et al. (2006) and McSweeny et al. (1993), creating three different RCI values for each subject. All RCI values were calculated with the Reliable Change Calculator 1.0 by Hinton-Bayre (2010). The results of the comparison of the different RCI methods are shown in Table 14.

While 2.5% to 5% of the participants from the SCD group deteriorated, 17.5% to 32.5% showed an improvement. The B-ADL score declined in 7.9% to 14.9% and increased in 25.4% to 37.7% of individuals from the naMCI group. In the aMCI group between 4.1% to 15% showed a deterioration and 16.4% to 42.5% an increase in functional impairment. When comparing RCI models across diagnostic subgroups, the method most responsive to negative change was the model by Chelune et al. (1993). The model by McSweeny et al. (1993) was most responsive to positive change.

Table 14

Comparison of Methods by Chelune, Maassen and McSweeny for Assessing Reliable Change in B-ADL across Diagnostic Subgroups

	SCD (n = 40)			naMCI (n = 114)			aMCI (n = 73)		
	⇓	⇔	⇑	⇓	⇔	⇑	⇓	⇔	⇑
Chelune	5	75	20	14.9	59.7	25.4	15	68.5	16.4
	(2)	(30)	(8)	(17)	(68)	(29)	(11)	(50)	(12)
Maassen	2.5	80	17.5	9.7	61.4	29	9.6	60.3	30.1
	(1)	(32)	(7)	(11)	(70)	(33)	(7)	(44)	(22)
McSweeny	2.5	65	32.5	7.9	54.4	37.7	4.1	53.4	42.5
	(1)	(26)	(13)	(9)	(62)	(43)	(3)	(39)	(31)

Note. SCD = subjective cognitive decline; aMCI = amnestic mild cognitive impairment; naMCI = nonamnestic mild cognitive impairment; ⇓ = deterioration in %, ⇔ = no change in %, ⇑ = improvement in %; n in brackets.

The same procedure was repeated to assess reliable change in converters versus non-converters with the control group being used as reference group. The mean B-ADL score of converters was 2.17 at baseline and 2.83 at follow-up. The group of non-converters had a mean score of 1.91 at baseline and a mean score of 2.10 at follow-up testing. *SEM* and *SED*, which were used to calculate confidence intervals, as well as upper and lower limits of the confidence intervals for converters and non-converters are included in Table 13. The results of the comparison of individual RCI values with the three different RCI-methods are shown in Table 15. Between 14.29% and 21.43% of converters deteriorated, 28.57% to 50% showed an increase in functional impairment. In non-converters 2.32% to 9.27% declined, between 14.67% to 24.71% showed a rise in B-ADL score. Again, the model by Chelune et al. (1993) was most responsive to negative change and the model by McSweeny et al. (1993) was most responsive to positive change.

Table 15

Comparison of Methods by Chelune, Maassen and McSweeny for Assessing Reliable Change in B-ADL in Converters versus Non-converters

	Converters (n = 14)			Non-converters (n = 259)		
	⇓	⇔	⇑	⇓	⇔	⇑
Chelune	21.43 (3)	50 (7)	28.57 (4)	9.27 (24)	76.10 (197)	14.67 (38)
Maassen	14.29 (2)	42.86 (6)	42.86 (6)	6.18 (16)	72.97 (189)	20.85 (54)
McSweeny	21.43 (3)	28.57 (4)	50 (7)	2.32 (6)	72.97 (189)	24.71 (64)

Note. ⇓ = deterioration in %, ⇔ = no change in %, ⇑ = improvement in %; n in brackets.

Discussion

Out of the 315 study participants, 31 were diagnosed with AD at follow-up, which corresponds to an overall conversion rate of 9.8%. All 31 converters were diagnosed with MCI at baseline, therefore the conversion rate within the MCI group is 13.8%. As four participants with naMCI and 27 with aMCI converted at the second testing point, the conversion rate for naMCI is 3.2% and the conversion rate for aMCI group is 28.6% in this sample. Those findings are in line with previous study results, which also discovered that conversion rates are higher among participants with aMCI than in subjects with naMCI (Lehrner et al., 2005; Lehrner et al., 2016; Silva et al., 2016). The fact that only subjects who were diagnosed with MCI at baseline converted at follow-up, supports the theory of Jessen et al. (2014) that AD develops across three stages.

Contrary to the primary hypothesis of this study, no evidence was found that ADL can be used to predict conversion to dementia. Therefore, the alternative hypothesis could not be confirmed. These results differ from previous evidence, on which the hypothesis was based. When comparing the study design to former research projects, it becomes apparent that methodology is heterogenous across studies. Amieva et al. (2008) and Pérès et al. (2008) used a French version of the Lawton IADL scale (Lawton & Brody, 1969) to assess self-rated impairment in ADL. The results of Reppermund et al. (2011) on the other hand are based on informants' ratings of patients on the B-ADL scale. Only Stogmann et al. (2016) used the same assessment procedure that was also used in this study, a self-rating of study participants on the B-ADL scale. Like in the study of Stogmann et al. (2016), an increase in B-ADL score across diagnostic subgroups was found, with the lowest B-ADL score in the control group and the highest value in the AD group. It must be mentioned, however, that the differences in median B-ADL score were relatively low between adjacent diagnostic groups and are probably not statistically significant. In contrast to this study, the investigation by Stogmann et al. (2016) was not a longitudinal study. Additionally, while in the present study B-ADL raw scores were used, the calculations by Stogmann et al. (2016) were based on z-scores.

Another aspect of this investigation that distinguishes it from most previous studies is the age of the participants. Individuals from a minimum age of 50 years were recruited and a mean age of 67 years was reached in the total sample. In the studies of Amieva et al. (2008) and Pérès et al. (2008) subjects had to be at least 65 years old to participate, Reppermund et al.

(2011) only included individuals between 70 and 90 years of age. Compared to Amieva et al. (2008), Pérès et al. (2008) and Reppermund et al. (2011), whose participants were on average between 75 and 86 years old, the sample of this study was younger. Regarding participant age, only the sample of Stogmann et al. (2016), with a mean age of 68 years, was comparable to the present studies' sample.

The calculated ROC curve was located close to the diagonal reference line, indicating that the ability of B-ADL to distinguish between subjects with and without AD is comparable to a randomly drawn distinction. Using the cut-off score based on Youden's Index, a high specificity (.93) and a low sensitivity (.23) were achieved. This suggests that B-ADL can correctly identify healthy subjects, who do not have AD, with a high probability. However, the risk of B-ADL scale not being able to identify participants, who do have AD, is also high. In a next step, sensitivity and specificity were used to calculate PPV and NPV. The resulting PPV of .27 and NPV of .92 indicate that the probability of a patient with AD diagnosis for having the disease, is relatively low. On the other hand, if a patient receives a negative test result, the probability that this individual does not have AD is comparably high. The LR+ of 3.37 implies that, compared to participants without AD, individuals with AD were 3.37 times more likely to receive a diagnosis. The reciprocal of the LR- can be calculated to assess the probability of a negative result in participants without AD, compared to subjects with AD. In this sample, a subject without AD is 1.20 times more likely to receive a negative test result, than an individual with AD. With an AUC of .56, the overall diagnostic accuracy of B-ADL in discriminating between subjects with and without AD is not much higher than the probability of a random prediction.

While B-ADL self-rating could not be identified as a predictor for conversion to dementia, significant results were found for the variables age, gender and for the five NTBV factors in block three of the logistic regression. Although there are only approximate measures for explained variance in logistic regression, it can be observed that pseudo R^2 values have sharply risen from block two to block three. It is therefore assumed that by adding the NTBV factors, the amount of explained variance has greatly increased. Hence, like in the study of Stogmann et al. (2016), it is concluded that NTBV domains are useful in predicting conversion to dementia.

It has already been recognised that some of the variables from the different NTBV subtests correlate highly with each other. Considering that the 30 NTBV variables can be assigned to the same five cognitive domains, it seems plausible that similar results were obtained for the individual variables. As in the investigation of Stogmann et al. (2016), except for two NTBV variables, small correlations between B-ADL and NTBV subtests at baseline were found. At follow-up, 29 of the 30 variables from NTBV correlated significantly with B-ADL self-rating. Medium correlation coefficients were revealed for three variables from the attention domain (AKT Time, AKT Total / Time, Digit-Symbol Test), two from the language domain (SVF Total Words, SVF Supermarket Items), one from the psychomotor speed domain (TMT-A) and one variable from the memory domain (VSRT Learning Performance).

Concerning the variables age and gender, ambiguous evidence was found. As a positive correlation between the variable age and the B-ADL score at follow-up was found, an increase in age seems to be related to an increase in functional impairment. Even though age was identified as significant predictor in the logistic regression, indicators for model fit suggest that it does not contribute much to the prediction of conversion to dementia. Unlike Reppermund et al. (2011) and Pérès et al. (2011), who reported possible associations between ADL-performance and gender, no connection between B-ADL score and the variable gender could be found in the correlation analysis. It is questionable, whether gender is relevant as a predictor of conversion to dementia, as it was only significant in block three of the logistic regression. Future studies need to investigate, how exactly the variables age and gender are related to ADL and what role they play in predicting conversion to dementia.

Regarding the relationship between functional impairment and cognitive abilities, a small negative association between MMSE score and B-ADL score at both testing points was discovered. Thereby it can be concluded that a decrease in cognitive function is accompanied by an increase in functional impairment and vice versa. These results coincide with previous evidence from Reppermund et al. (2011) and Stogmann et al. (2016). Moreover, a strong positive relationship between B-ADL at baseline and follow-up with BDI-II was found in the correlation analysis, indicating that an increase in functional impairment is associated with an increase in depressive symptoms. In the study of Stogmann et al. (2016) the same trend was found, as patients with depressive symptoms reported a higher B-ADL score.

For the interpretation of RCI it is important to be aware that an improvement in RCI describes an increase in B-ADL score and thus a decrease in ADL-performance. RCI was calculated to determine whether individuals have significantly deteriorated or improved in B-ADL between the first and the second testing point. As it is still unclear whether there is one RCI model that is best suited for general use, three commonly used methods were compared. Although there is no consensus, some authors prefer regression-based models (Hinton-Bayre, 2016) and it has been reported that the SRB model by McSweeny et al. (1993) is most responsive to change (Duff, 2012). A similar pattern was found in this study, with the model by McSweeny et al. (1993) being most responsive, but only to positive change. Regarding negative change, the mean practice model by Chelune et al. (1993) was most responsive.

There are several factors that might affect the responsiveness of different RCI methods. The Chelune model shows the highest response to negative change if the standard deviation of the baseline scores (SD_X) exceeds the standard deviation of the retest scores (SD_Y), however this does not apply to the data ($SD_X < SD_Y$). Another factor that can influence model responsiveness is the position of individual baseline scores in relation to the mean baseline score from the control group (M_X). According to Hinton-Bayre (2016), Chelune's model is most responsive to negative change if the retest variance is low and individual baseline scores exceed the normative mean. Our sample includes 188 subjects with baseline scores > M_X, which could explain why the RCI by Chelune et al. (1993) was most responsive.

Using the methods of McSweeny et al. (1993) and Maassen et al. (2006), the highest proportion of individuals with significant increases in B-ADL were in the aMCI group, followed by naMCI group. The SCD subgroup included the lowest proportion of subjects with significant improvement. The order was different using the Chelune model, with naMCI containing the highest proportion of significant improvements, followed by SCD, followed by aMCI. With all three methods, the percentage of individuals who showed an improvement and consequently a worsening in ADL was higher than the percentage of subjects who deteriorated. When comparing RCI across diagnostic subgroups, regardless of the RCI method used, the majority of participants showed no significant change in any direction. In the comparison of RCI from converters and non-converters, there were more converters with significantly increased B-ADL score than converters who showed no change or deterioration. However, RCI could only be calculated for the 14 converters, who completed the B-ADL scale at both testing points.

Limitations

Unfortunately, not all neuropsychological tests were completed by all participants at both testing points. Regarding the B-ADL scale, from the 315 subjects that completed B-ADL at baseline, only 273 took part at follow-up. This corresponds to a completion rate of 86.7% for B-ADL. This is problematic, as the B-ADL score represents the independent variable in the main research question. When interpreting the study outcomes, it must be deliberated that those results might be biased by systematic dropouts.

Moreover, the results cannot be generalized, as the participants were not randomly selected from the general population. Instead, only preselected subjects, who had to meet certain criteria, were considered. Since the study participants were patients in the memory clinic of the Vienna General Hospital, the results can only be related to a clinical context.

Although a self-evaluation is highly economical in its implementation, it is questionable whether it is a reliable source of information on ADL-performance. The individual response behaviour could be influenced by various factors. Firstly, participants could respond in a way they think is socially desirable. Secondly, subjects could be motivated to intentionally over- or underestimate their ADL-performance.

In this study, the B-ADL scale was part of a test battery that contained several neuropsychological tests. Usually, the test battery was presented to each participant in the same order, with B-ADL being implemented after MMSE, VVT 3.0 Screening, NTBV, BDI-II and WST. Therefore, the B-ADL self-rating could have been influenced by the participants' subjective perception of their own performance in the preceding tests. In addition, the completion of the whole test battery took approximately one to two hours. It is possible that participants were already exhausted or unmotivated by the time they filled out the B-ADL scale, which could have affected their B-ADL score. Consequently, it cannot be excluded that the order or duration of the test battery had an influence on B-ADL self-evaluation.

Future Research

The focus of this study was on ADL in an elderly population in the clinical context, which is why a relatively high dropout rate was to be expected. In order to reduce the negative effects of missing data and participant dropouts on the interpretability of the results, a possible solution could be to recruit a larger sample at the first testing point. Alternatively, patients could be encouraged to complete the B-ADL scale at home and submit it at their next appointment. This could prevent participants at follow-up from dropping out, before they fill out the B-ADL scale.

So far, there has only been little research concerning the relationship between ADL and dementia. Additionally, the existing investigations are inconsistent in their methodology, which makes the comparability of study outcomes difficult. In a first step, the different methods for assessing ADL could be compared regarding their psychometric criteria. This would enable the identification of the most reliable tool for measuring ADL and set common standards for future research.

While in this study the assessment of ADL was based on participants' self-evaluations, the B-ADL scale can also be completed by an informant who is sufficiently familiar with the patient. Although the self-rated version is more economical in its implementation, it is necessary to investigate whether there are significant discrepancies between B-ADL scores resulting from self and external assessment. Future research projects could compare the results from B-ADL self-evaluation and informant ratings, to check if one B-ADL version is superior to the other.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed.). Washington, DC: American Psychiatric Publishing. doi:10.1176/appi.books.9780890425596

Amieva, H., Le Goff, M., Millet, X., Orgogozo, J. M., Pérès, K., Barberger-Gateau, P. ... & Dartigues, J. F. (2008). Prodromal Alzheimer's disease: Successive emergence of the clinical symptoms. *Annals of Neurology, 64*(5), 492−498. doi:10.1002/ana.21509

Beyermann, S., Trippe, R. H., Bähr, A. A., & Püllen, R. (2013). Mini-Mental State Examination in geriatrics: An evaluation of diagnostic quality. *Zeitschrift für Gerontologie und Geriatrie, 46*(8), 740–747. doi:10.1007/s00391-013-0488-6

Blacker, D., Albert, M. S., Bassett, S. S., Go, R. C., Harrell, L. E., & Folstein, M. F. (1994). Reliability and validity of NINCDS-ADRDA criteria for Alzheimer's disease: The National Institute of Mental Health Genetics Initiative. *Archives of Neurology, 51*(12), 1198−1204. doi:10.1001/archneur.1994.00540240042014

Chelune, G. J., Naugle, R. I., Lüders, H., Sedlak, J., & Awad, I. A. (1993). Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology, 7*(1), 41–52. doi:10.1037/0894-4105.7.1.41

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). New York: Routledge. doi:10.4324/9780203771587

Dilling, H., Mombour, W., & Schmidt, M. H. (2008). *Internationale Klassifikation psychischer Erkrankungen–ICD 10*. Bern: WHO.

Dubois, B., Feldman, H. H., Jacova, C., DeKosky, S. T., Barberger-Gateau, P., Cummings, J. ... & Scheltens, P. (2007). Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS–ADRDA criteria. *The Lancet Neurology, 6*(8), 734−746. doi:10.1016/S1474-4422(07)70178-3

Duff, K. (2012). Evidence-based indicators of neuropsychological change in the individual patient: Relevant concepts and methods. *Archives of Clinical Neuropsychology, 27*(3), 248−261. doi:10.1093/arclin/acr120

Dugger, B. N., Davis, K., Malek-Ahmadi, M., Hentz, J. G., Sandhu, S., Beach, T. G. ... & Sabbagh, M. N. (2015). Neuropathological comparisons of amnestic and nonamnestic mild cognitive impairment. *BMC Neurology*, *15*(1), 1−8. doi:10.1186/s12883-015-0403-4

Erzigkeit, H., Lehfeld, H., Peña-Casanova, J., Bieber, F., Yekrangi-Hartmann, C., Rupp, M. ... & Hindmarch, I. (2001). The Bayer-Activities of Daily Living Scale (B-ADL): Results from a validation study in three European countries. *Dementia and Geriatric Cognitive Disorders*, *12*(5), 348−358. doi:10.1159/000051280

Erzigkeit, H., & Lehfeld, H. (2010). *Bayer ADL-Skala (B-ADL): Eine Skala zur Erfassung von Beeinträchtigungen der Alltagskompetenz bei älteren Patienten mit Einbußen der kognitiven Leistungsfähigkeit*. Spardorf: Pearson.

Ferman, T. J., Smith, G. E., Kantarci, K., Boeve, B. F., Pankratz, V. S., Dickson, D. W. ... & Petersen, R. C. (2013). Nonamnestic mild cognitive impairment progresses to dementia with lewy bodies. *Neurology*, *81*(23), 2032−2038. doi:10.1212/01.wnl.0000436942.55281.47

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). London: Sage publications.

Foki, T., Hitzl, D., Pirker, W., Novak, K., Pusswald, G., & Lehrner, J. (2018). Individual cognitive change after DBS-surgery in parkinson's disease patients using reliable change index methodology. *Neuropsychiatrie*, *32*(3), 149−158. doi:10.1007/s40211-018-0271-4

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*(3), 189−198. doi:10.1016/0022-3956(75)90026-6

Gatterer, G. (1990). *Alters-Konzentrations-Test (AKT*). Göttingen: Hogrefe.

Goodglass, H. & Kaplan, P. (1983). *The assessment of aphasia and related disorders* (2nd ed.). Philadelphia: Lea & Febiger.

Hautzinger, M., Keller, F., & Kühner, C. (2006). *Beck Depressions-Inventar (BDI-II)*. Frankfurt: Harcourt Test Services.

Hindmarch, I., Lehfeld, H., de Jongh, P., & Erzigkeit, H. (1998). The Bayer activities of daily living scale (B-ADL). *Dementia and Geriatric Cognitive Disorders*, *9*(2), 20−26. doi:10.1159/000051195

Hinton-Bayre, A. D. (2010). Deriving reliable change statistics from test–retest normative data: Comparison of models and mathematical expressions. *Archives of Clinical Neuropsychology*, *25*(3), 244−256. doi:10.1093/arclin/acq008

Hinton-Bayre, A. D. (2016). Clarifying discrepancies in responsiveness between reliable change indices. *Archives of Clinical Neuropsychology*, *31*(7), 754−768. doi:10.1093/arclin/acw064

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*(1), 12–19. doi:10.1037/10109-042

Jaeschke, R., Guyatt, G. H., Sackett, D. L., Guyatt, G., Bass, E., Brill-Edwards, P. ... & Wilson, M. (1994). Users' guides to the medical literature: III. How to use an article about a diagnostic test B. What are the results and will they help me in caring for my patients? *Jama*, *271*(9), 703−707. doi:10.1001/jama.1994.03510330081039

Jessen, F., Amariglio, R. E., Van Boxtel, M., Breteler, M., Ceccaldi, M., Chételat, G. ... & Glodzik, L. (2014). A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. *Alzheimer's & Dementia*, *10*(6), 844−852. doi:10.1016/j.jalz.2014.01.001

Jessen, F., Amariglio, R. E., Buckley, R. F., Van der Flier, W. M., Han, Y., Molinuevo, J. L. ... & Wagner, M. (2020). The characterisation of subjective cognitive decline. *Lancet Neurology*, *19*(3), 271−278. doi:10.1016/S1474-4422(19)30368-0

Jessen, F., Wiese, B., Bachmann, C., Eifflaender-Gorfer, S., Haller, F., Kölsch, H. ... & Bickel, H. (2010). Prediction of dementia by subjective memory impairment: Effects of severity and temporal association with cognitive impairment. *Archives of General Psychiatry*, *67*(4), 414−422. doi:10.1001/archgenpsychiatry.2010.30

Knopman, D. S., DeKosky, S. T., Cummings, J. L., Chui, H., Corey–Bloom, J., Relkin, N. ... & Stevens, J. C. (2001). Practice parameter: Diagnosis of dementia (an evidence-based review). *Neurology*, *56*(9), 1143−1153. doi:10.1212/WNL.56.9.1143

Kühner, C., Bürger, C., Keller, F., & Hautzinger, M. (2007). Reliabilität und Validität des revidierten Beck-Depressionsinventars (BDI-II). Befunde aus deutschsprachigen Stichproben. *Nervenarzt*, *78*(6), 651−6. doi:10.1007/s00115-006-2098-7

Lawton, M. P., & Brody, E. M. (1969). Assessment of older people: Self-maintaining and instrumental activities of daily living. *The Gerontologist*, *9*(3), 179−186.

Lehrl, S., & Fischer, B. (1997). *Kurztest für cerebrale Insuffizienz (cI-Test)*. Ebersberg*:* Vless.

Lehrner, J., Bodendorfer, B., Lamm, C., Moser, D., Dal-Bianco, P., Auff, E., & Pusswald, G. (2016). Subjective memory complaints and conversion to dementia in patients with subjective cognitive decline and patients with mild cognitive impairment. *Zeitschrift für Neuropsychologie, 27*, 85−93. doi:10.1024/1016-264X/a000175.

Lehrner, J., Gleiß, A., Maly, J., Auff, E., & Dal-Bianco, P. (2006). Der Verbale Selektive Reminding Test (VSRT): Ein Verfahren zur Überprüfung verbaler Gedächtnisfunktionen. *Neuropsychiatrie, 20*(3), 204–214.

Lehrner, J., Gufler, R., Guttmann, G., Maly, J., Gleiß, A., Auff, E., & Dal-Bianco, P. (2005). Annual conversion to Alzheimer disease among patients with memory complaints attending an outpatient memory clinic: The influence of amnestic mild cognitive impairment and the predictive value of neuropsychological testing. *Wiener Klinische Wochenschrift*, *117*(18), 629−635. doi:10.1007/s00508-005-0428-6

Lehrner, J., Krakhofer, H., Lamm, C., Macher, S., Moser, D., Klug, S. ... & Pusswald, G. (2015). Visuo-constructional functions in patients with mild cognitive impairment, Alzheimer's disease, and Parkinson's disease. *Neuropsychiatrie, 29*(3), 112−119. doi:10.1007/s40211-015-0141-2

Lehrner, J., Maly, J., Gleiß, A., Auff, E., & Dal-Bianco, P. (2007). The Vienna Neuropsychological Test Battery (VNTB) for detecting dementia–standardization, norms, and validation. *Psychologie in Österreich*, *4*(5), 358−365.

Lively, S. T. (2015). *Activities of daily living (ADL): Cultural differences, impacts of disease and long-term health effects*. New York: Nova Science Publishers, Incorporated.

Maassen, G. H., Bossema, E. R., & Brand, N. (2006). Reliable change assessment with practice effects in sport concussion research: A comment on Hinton-Bayre. *British Journal of Sports Medicine, 40*(10), 829–833. doi:10.1136/bjsm.2005.023713

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*, *5*(9), 1315–1316. doi:10.1097/JTO.0b013e3181ec173d

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). 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, 34*(7), 939−939. doi: 10.1212/WNL.34.7.939

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack Jr, C. R., Kawas, C. H. ... & Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia, 7*(3), 263−269. doi:10.1016/j.jalz.2011.03.005

McSweeny, A. J., Naugle, R. I., Chelune, G. J., & Lüders, H. (1993). "T scores for change": An illustration of a regression approach to depicting change in clinical neuropsychology. *The Clinical Neuropsychologist, 7*(3), 300–312. doi:10.1080/13854049308401901

Mitchell, A. J., Beaumont, H., Ferguson, D., Yadegarfar, M., & Stubbs, B. (2014). Risk of dementia and mild cognitive impairment in older people with subjective memory complaints: Meta-analysis. *Acta Psychiatrica Scandinavica, 130*(6), 439−451. doi:10.1111/acps.12336

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): I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology, 39*(9), 1159–1165. doi:10.1212/WNL.39.9.1159

O'connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers, 32*(3), 396−402. doi:10.3758/BF03200807

Oedekoven, C., & Dodel, R. (2019). Diagnostische Kriterien und Diagnose der Demenz vom Alzheimer-Typ. *Neurologie Up2date, 2*(1), 91−105. doi:10.1055/a-0803-4606

Österreichische Alzheimergesellschaft (n.d.). Zahlen & Statistik. Retrieved July 17, 2021, from http://www.alzheimer-gesellschaft.at/informationen/zahlen-statistik/

Oswald, W. D., & Fleischmann, U. M. (1997). *Nürnberger-Alters-Inventar (NAI*). Göttingen: Hogrefe.

Parnetti, L., Chipi, E., Salvadori, N., D'Andrea, K., & Eusebi, P. (2019). Prevalence and risk of progression of preclinical Alzheimer's disease stages: A systematic review and

meta-analysis. *Alzheimer's Research & Therapy, 11*(1), 1−13. doi:10.1186/s13195-018-0459-7

Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine, 256*(3), 183−194. doi:10.1111/j.1365-2796.2004.01388.x

Petersen, R. C. (2011). Mild cognitive impairment. *New England Journal of Medicine, 364*(23), 2227−2234. doi:10.1056/NEJMcp0910237

Pérès, K., Helmer, C., Amieva, H., Orgogozo, J. M., Rouch, I., Dartigues, J. F., & Barberger-Gateau, P. (2008). Natural history of decline in instrumental activities of daily living performance over the 10 years preceding the clinical diagnosis of dementia: A prospective population-based study. *Journal of the American Geriatrics Society, 56*(1), 37−44. doi:10.1111/j.1532-5415.2007.01499.x

Pérès, K., Helmer, C., Amieva, H., Matharan, F., Carcaillon, L., Jacqmin-Gadda, H. ... & Dartigues, J. F. (2011). Gender differences in the prodromal signs of dementia: Memory complaint and IADL-restriction. A prospective population-based cohort. *Journal of Alzheimer's Disease, 27*(1), 39−47. doi:10.3233/JAD-2011-110428

Regard, M., Strauss, E., & Knapp, P. (1982). Children's production on verbal and non-verbal fluency tasks. *Perceptual and Motor Skills*, *55*(3), 839–844. doi:10.2466/pms.1982.55.3.839

Reitan, R. M. (1979). *Trail Making Test (TMT)*. Tucson: Reitan Neuropsychology Laboratory.

Reppermund, S., Sachdev, P. S., Crawford, J., Kochan, N. A., Slavin, M. J., Kang, K. ... & Brodaty, H. (2011). The relationship of neuropsychological function to instrumental activities of daily living in mild cognitive impairment. *International Journal of Geriatric Psychiatry, 26*(8), 843−852. doi:10.1002/gps.2612

Ringendahl, H. (2013). Neuropsychologische Verlaufsdiagnostik bei Parkinson-Patienten – Darstellung kritischer Test-Retestdifferenzen für die Einzelfalldiagnostik. *Zeitschrift für Neuropsychologie, 24*(4), 253−266. doi:10.1024/1016-264X/a000103

Sachdev, P. S., Blacker, D., Blazer, D. G., Ganguli, M., Jeste, D. V., Paulsen, J. S., & Petersen, R. C. (2014). Classifying neurocognitive disorders: The DSM-5 approach. *Nature Reviews Neurology, 10*(11), 634−642. doi:10.1038/nrneurol.2014.181

Schmidt, K. H., & Metzler, P. (1992). *Wortschatztest (WST)*. Bern: Huber.

Silva, M. R., Moser, D., Pflüger, M., Pusswald, G., Stögmann, E., Dal-Bianco, P. ... & Lehrner, J. (2016). Self-reported and informant-reported memory functioning and awareness in patients with mild cognitive impairment and Alzheimer´ s disease. *Neuropsychiatrie, 30*(2), 103−112. doi:10.1007/s40211-016-0185-y

Stogmann, E., Moser, D., Klug, S., Gleiss, A., Auff, E., Dal-Bianco, P. ... & Lehrner, J. (2016). Activities of daily living and depressive symptoms in patients with subjective cognitive decline, mild cognitive impairment, and Alzheimer's disease. *Journal of Alzheimer's Disease*, *49*(4), 1043−1050. doi:10.3233/JAD-150785

Tewes, U. (1994). *Hamburg-Wechsler-Intelligenztest für Erwachsene – Revision* (*HAWIE-R*). Bern: Verlag Hans Huber.

Valencia, N., & Lehrner, J. (2018). Screening for dementia with the Vienna Visuo-Constructional Test 3.0 screening (VVT 3.0 screening). *Neuropsychiatrie, 32*(4), 196−203. doi:10.1007/s40211-018-0279-9

Varma, A. R., Snowden, J. S., Lloyd, J. J., Talbot, P. R., Mann, D. M. A., & Neary, D. (1999). Evaluation of the NINCDS-ADRDA criteria in the differentiation of Alzheimer's disease and frontotemporal dementia. *Journal of Neurology, Neurosurgery & Psychiatry*, *66*(2), 184−188. doi:10.1136/jnnp.66.2.184

Wancata, J., Freidl, M., Unger, A., Jahn, R., Soulier, N., Fellinger, M., & Daniel, R. (2015). Projections about the future number of dementia sufferers: Increasing life expectancy not sufficiently considered?. *Psychiatria Danubina*, *27*(4), 452−457.

World Health Organization. (1993). *The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research*. World Health Organization. Geneva: World Health Organization.

World Health Organization (2015). *The epidemiology and impact of dementia: Current state and future trends*. Retrieved July 17, 2021, from https://www.who.int/mental_health/neurology/dementia/dementia_thematicbrief_epidemiology.pdf

World Health Organization. (2017). *Global action plan on the public health response to dementia 2017–2025*. Retrieved July 17, 2021, from https://apps.who.int/iris/bitstream/handle/10665/259615/9789241513487-eng.pdf

World Health Organization (2019). *International statistical classification of diseases and related health problems* (11th ed.). Retrieved July 17, 2021, from https://icd.who.int/en

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*(1), 32–35. doi:10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3

List of Figures

List of Tables

Appendix

Abstract English

There is empirical evidence suggesting that impairments in activities of daily living (ADL) occur in prodromal stages of dementia and can be instrumentalized to identify individuals who are at risk for dementia. In the present retrospective longitudinal study, self-evaluated ADL were assessed using the Bayer-ADL (B-ADL) scale at two testing points. To determine the prognostic value of ADL in predicting conversion to dementia, baseline ADL-scores of patients with SCD, naMCI and aMCI were compared to the scores of patients who had developed AD at follow-up. Additionally, possible associations between B-ADL with the results from other neuropsychological tests and demographic parameters were investigated. Finally, three different methods for calculating reliable change index (RCI) for B-ADL were used and the results were compared. At follow-up, 31 of the 315 participants had converted to AD, which corresponds to a conversion rate of 9.8%. While no evidence was found that self-rated B-ADL can be used for predicting conversion to dementia, domains of the Neuropsychological Test Battery Vienna (NTBV) were identified as significant predictors. ADL were strongly associated with depressive symptoms and weakly with overall cognitive functioning. Moreover, small to medium correlations between the B-ADL score and the results of NTBV subtests were found. While the regression-based method by McSweeny was most responsive to positive change, Chelune's method showed the highest responsiveness to negative change. As it is possible that there are discrepancies between external- and self-assessed B-ADL, future research should investigate whether informant-based evaluations of ADL can be used in predicting conversion to AD.

Keywords: Activities of daily living, Alzheimer's disease, Mild cognitive impairment, Subjective cognitive decline, depressive symptoms, Reliable change index

Abstract Deutsch

Es gibt empirische Belege dafür, dass Beeinträchtigungen der Aktivitäten des täglichen Lebens (ATL) in prodromalen Stadien der Demenz auftreten und instrumentalisiert werden können, um Personen zu identifizieren, bei denen ein Risiko für Demenz besteht. In der vorliegenden retrospektiven Längsschnittstudie wurde die Bayer-ADL (B-ADL) Skala verwendet, um selbstbewertete ATL zu zwei Testzeitpunkten zu erfassen. Zur Bestimmung des prognostischen Wertes von ATL in der Vorhersage der Konversion zur Demenz, wurden ATL Baseline-Werte von Patient*innen mit SCD, naMCI und aMCI mit den Scores von Patient*innen verglichen, die nach dem Follow-up Zeitraum Alzheimer Demenz (AD) entwickelt hatten. Zusätzlich wurden mögliche Assoziationen zwischen B-ADL mit den Ergebnissen anderer neuropsychologischer Testverfahren und demografischen Parametern untersucht. Zuletzt wurden drei verschiedene Methoden zur Berechnung des Reliable Change Index (RCI) für B-ADL angewandt und die Ergebnisse verglichen. Bei der Follow-up Untersuchung waren 31 der 315 Teilnehmer*innen zur Demenz konvertiert, was einer Konversionsrate von 9.8% entspricht. Während nicht nachgewiesen werden konnte, dass selbstbewertete B-ADL zur Vorhersage der Konversion zur Demenz verwendet werden können, wurden die Domänen der Neuropsychologischen Testbatterie Vienna (NTBV) als signifikante Prädiktoren identifiziert. Zu Studienbeginn und bei der Nachuntersuchung waren ATL stark mit depressiven Symptomen und schwach mit der kognitiven Gesamtfunktion assoziiert. Darüber hinaus wurden kleine bis mittlere Korrelationen zwischen dem B-ADL Score und den Ergebnissen von NTBV-Subtests gefunden. Während die regressionsbasierte Methode von McSweeny am stärksten auf positive Veränderungen reagierte, zeigte die Methode von Chelune die höchste Responsivität auf negative Veränderungen. Da es möglicherweise Diskrepanzen zwischen extern und selbstbewerteten B-ADL gibt, sollte in zukünftigen Forschungsprojekten untersucht werden, ob auf Informanten basierte Bewertungen von ATL zur Vorhersage der Konversion zu Alzheimer Demenz verwendet werden können.

Schlagwörter: Aktivitäten des täglichen Lebens, Alzheimer Demenz, Mild Cognitive Impairment, Subjective Cognitive Decline, depressive Symptome, Reliable Change Index