Early recognition of neurocognitive disorders:
Dementia screening in primary care
and the detection of mild cognitive impairment via verbal fluency tests

Réka Balogh

Ph.D. Thesis

Szeged

2022

Early recognition of neurocognitive disorders: Dementia screening in primary care and the detection of mild cognitive impairment via verbal fluency tests

Ph.D. Thesis

Réka Balogh, M.A.

Doctoral School of Clinical Medicine

Department of Psychiatry, Albert Szent-Györgyi Health Centre

Faculty of Medicine, University of Szeged

Supervisors:

János Kálmán, M.D., Ph.D., D.Sc.
Professor

Department of Psychiatry, Albert Szent-Györgyi Health Centre
Faculty of Medicine, University of Szeged

Gábor Gosztolya, Ph.D.
Senior Research Fellow

ELKH-SZTE Research Group on Artificial Intelligence

Szeged

2022

Papers the thesis is based on:

I. **Balogh, R.**, Imre, N., Papp, E., Kovács, I., Heim, S., Karádi, K., Hajnal, F., Pákáski, M., & Kálmán, J. (2019). Dementia in Hungary: General practitioners' routines and perspectives regarding early recognition. *European Journal of General Practice*, *26*(1),1–7.
SJR Indicator: D1
IF: 1.904

II. **Balogh, R.**, Imre, N., Gosztolya, G., Hoffmann, I., Pákáski, M., & Kálmán, J., (2022). The role of silence in verbal fluency tasks - a new approach for the detection of mild cognitive impairment. *Journal of the International Neuropsychological Society*, 24;1-13.
SJR Indicator: Q2
IF: 3.114

Cumulative impact factor of publications related to this thesis: **5.018**

TABLE OF CONTENTS

I. ABBREVIATIONS

AD	Alzheimer’s disease
ADAS-Cog	Alzheimer’s Disease Assessment Scale – Cognitive Subscale
aMCI	amnestic mild cognitive impairment
ANOVA	analysis of variance
AUC	area under the curve
CDT	Clock Drawing Test
CI	confidence interval
CT	computed tomography
DSM-5	Diagnostic and Statistical Manual of Mental Disorders (5^{th} Edition)
GDS	Global Deterioration Scale
GDS-15	15-item Geriatric Depression Scale
GP	general practitioner
ICD-11	International Classification of Diseases (11^{th} Revision)
HC	healthy control individuals
M	mean
MCI	mild cognitive impairment
MMSE	Mini-Mental State Examination
MoCA	Montreal Cognitive Assessment
MRI	magnetic resonance imaging
n	number of individuals
naMCI	non-amnestic mild cognitive impairment
NCD	neurocognitive disorders
NIA-AA	National Institute on Aging and Alzheimer’s Association
PET	positron emission tomography
PVF	phonemic verbal fluency
SD	standard deviation
SVF	semantic verbal fluency

II. SCOPE OF THE THESIS

Cognitive disorders represent a worldwide problem: at present, more than 55 million people are living with dementia, and the number of cases is estimated to reach around 150 million by the year 2050 (Nichols et al., 2022). Early diagnosis of the disease is pivotal as it provides patients with a higher chance of benefiting from their treatment, while also aiding them and their relatives to access relevant information regarding the condition, to cope, and to plan for the future.

Mild cognitive impairment (MCI) is a heterogenous syndrome, characterized by a subtle deficit of memory, language, and executive skills, and is often considered the prodromal stage of dementia. Thus, it plays a major role in the identification of individuals at risk of dementia.

Primary care is another significant factor in the recognition process: as in many health systems, GPs are the first contact point for elderly patients seeking health care, they are in a unique position and as gateways, they play a major role in the recognition of cognitive decline. Despite this, high rates of undetected dementia in primary care are a widespread problem. The evaluation of GPs' routines could help us to enhance the currently low recognition rates. Since barriers toward effective dementia detection include the low use of cognitive assessment methods, it is crucial to offer healthcare providers cognitive screening tools that are quick, simple, and can yield objective and reliable results.

This thesis comprises the results and conclusions drawn from two original research articles. In the first paper, our main goal was to assess Hungarian GPs' routines and views regarding the screening of dementia in primary care practices. The aim of the second study was to introduce a new approach to the analysis of verbal fluency tests. In this method, we created temporal parameters based on the silent segments, the hesitations, and the irrelevant utterances found in the fluency voice recordings.

III. INTRODUCTION

1. Aging and cognitive health

The advances in medical sciences, the availability and the development of better public health care allow people worldwide to live longer, resulting in a higher average life expectancy than in previous decades and centuries. According to estimations, by 2030, 1.4 billion individuals – one in every six people – will be 60 years or older (WHO, 2022). In relation to this, the number of the oldest old is also rising – people over the age of 85 are currently the fastest-growing age group in many countries (National Institute on Aging, 2007). Besides the general benefits of the globally increasing life expectancy, this trend also presents challenges, as the health care system now has to face more cases of chronic diseases, and age-associated neurocognitive disorders that mainly affect the elderly (National Institute on Aging, 2007).

Dementia is one of these conditions, and it is currently affecting around 55 million people worldwide. Based on predictions, by 2050, the number of dementia cases will reach 150 million globally (Nichols et al., 2022), while in Europe, the number of dementia cases is estimated to increase to 15.9 million by 2040 (compared to the number of 7.7 million in 2001) (Meijer et al., 2022). In Hungary, the approximated number of patients with dementia lies between 150,000 and 300,000 registered cases (Ersek et al., 2010; Takacs et al., 2015). In 2018, 1.49% of the population was suffering from dementia in the country. Additionally, despite the decrease in the population of Hungary, the cases of dementia are estimated to rise in the following years (Alzheimer Europe, 2019).

Being a public health priority in more and more countries, dementia represents a substantial economic burden worldwide. With the increasing number of individuals affected by the disease, the economic impact is also predicted to rise (Meijer et al., 2022). When describing the economic costs of dementia regarding the patients and their families, two types of costs have to be mentioned. One is the medical and long-term care cost, while the other is the value of unpaid (or informal) care which is most commonly provided by close family or friends. This estimated cost per year is about four times higher than the cost required by similarly aged persons without the condition (National Academies of Sciences, 2021). While dementia-related costs vary significantly from country to country, a study conducted in Europe showed that the cost per patient per year can range from 162.9 to 32,606.9 EUR (Meijer et al., 2022).

2. Neurocognitive disorders

Cognitive abilities tend to peak at around the age of 30. From early adulthood, even in the absence of cognitive disorders, one's cognitive skills start to decline gradually, which phenomenon we can refer to as normal cognitive aging. It is associated with the decline of several cognitive abilities, such as certain aspects of memory (e.g., delayed free recall, source memory, prospective memory), language, processing speed, visuospatial abilities, and executive functions (Harada et al., 2013). The decline affecting memory and reasoning shows a modest speed until around the age of 65, from which the deterioration accelerates (Salthouse, 2019). Understanding the magnitude of these cognitive changes is pivotal to be able to distinguish the age-associated neurocognitive decline from the symptoms of neurocognitive disorders (Harada et al., 2013).

In 2014, the concept and the term *subjective cognitive decline* was described. It has been defined as a self-perceived cognitive decline in any cognitive domain, which does not need to be confirmed by objective tests or by an informant, and is not associated with a specific disease (Jessen et al., 2014).

The term *neurocognitive disorders* (NCD) was introduced in the *5th Edition of the Diagnostic and Statistical Manual of Mental Disorders* (DSM-5). NCD refers to a group of disorders (including delirium, major NCD and mild NCD), in which cognitive deficits are the most prominent and defining feature, and the impairment in cognition is acquired, which means that there is a decline in cognition compared to a previous level of cognitive functioning (American Psychiatric Association, 2013; Sachdev et al., 2014).

2.1. Dementia (major neurocognitive disorder) – definition and description

The word dementia derives from the Latin word *demens*, literally meaning "out of someone's mind" (Assal, 2019). *Dementia* is increasingly referred to as *major NCD*, due to the stigmatizing effect of the former; however, the term has not yet gained much currency and is still commonly used in the scientific literature as dementia, therefore, it will be referred to as such in this thesis as well. Based on the DSM-5, its diagnostic criteria include a "significant cognitive decline from a previous level of performance in one or more cognitive domains" which deficits "interfere with independence in everyday activities" (American Psychiatric Association, 2013).

Dementia is not a specific disease; it is rather a group of symptoms. In patients with dementia, multiple higher cortical functions are disturbed, including memory, thinking, orientation, comprehension, calculation, learning, language, and judgment. These cognitive symptoms are often accompanied or sometimes preceded by the decline of emotional control, social behavior, or motivation (World Health Organization, 2019).

Dementias are often progressive and irreversible, however, depending on the etiology, some may be reversible, in which case the underlying condition can be successfully treated. There are several modifiable factors (including cardiovascular, metabolic, endocrine, and lifestyle factors) that can modulate susceptibility to dementia: up to 50% of dementia cases can be attributable to these changeable factors. Knowing this rate, from a preventive point of view, it is pivotal to be aware of the health and lifestyle factors and habits that can be avoided or managed (Barnes & Yaffe, 2011).

Among reversible dementias, drug and alcohol toxicity and depression (i.e., pseudodementia) are the most significant causes, while nutritional deficiencies (e.g., Vitamin B12 deficiency), metabolic diseases, or even infections can also be mentioned (Rone-Adams et al., 2013; Tripathi & Vibha, 2009). Regarding progressive dementias, the most common types are Alzheimer's disease (AD) (accounting for 60-70% of the cases), frontotemporal lobar degeneration, Lewy body disease, and vascular disease (Rone-Adams et al., 2013; World Health Organization, 2017). Nevertheless, the underlying pathologies often overlap: for example, around 80% of patients with AD show vascular pathology (e.g., vascular lesions, such as large or microinfarcts or atherosclerosis) as well (Toledo et al., 2013).

Although dementia is often considered a disease of the elderly, and indeed, old age is one of its greatest risk factors, not all elderly people fall victim to it. It is pertinent to note, that the disease may occur at a younger age, in which case (i.e., before the age of 65) it is referred to as *early-onset dementia* (Alzheimer's Association, 2013; Vieira et al., 2013), which may account for 9% of dementia cases (World Health Organization, 2022).

2.2. Mild cognitive impairment – definition and description

The earliest concept of minor cognitive decline, a "*grey zone*" between healthy cognitive aging and major cognitive decline was first reported in patients in the late 1980s and 1990s (Geda & Nedelska, 2012). In the past decades, there has been a growing interest in the research of the condition.

Using the Global Deterioration Scale (GDS), Reisberg and colleagues introduced the expression *mild cognitive impairment* (MCI) in 1988. GDS is a test administered by the clinician, and it is based on subjective complaints and objective observation of the memory deficit, as well as a clinical interview and a functional ability assessment of the patient. In their article, mild cognitive impairment corresponds to the severity level 3 on the scale, which describes “with only minimal functional impairment” (Reisberg et al., 1988).

Ten years later, Petersen and colleagues further developed the term MCI (Petersen et al., 1999), and they proposed criteria for the condition, as follows: (1) subjective complaint – preferably corroborated by an informant, (2) objective memory impairment for age, (3) relatively preserved general cognition for age, (4) intact activities of daily living, and (5) not demented (Petersen, 2004). In 2011, the National Institute on Aging and Alzheimer’s Association (NIA-AA) also created a recommendation for the diagnosis of the preclinical stages of dementia (Albert et al., 2011).

MCI is now widely defined and viewed as an intermediate or transitional stage between healthy cognitive aging and dementia. *Mild neurocognitive disorder*, a diagnostic entity introduced in the DSM-5 shows strong similarities with MCI, although the diagnostic approaches related to them are not identical (Bermejo-Pareja et al., 2021; Stokin et al., 2015).

MCI is a heterogenous condition (Winblad et al., 2004), and its characteristics vary in terms of subtypes. Based on the observed symptoms, MCI can be classified into amnestic (aMCI) or non-amnestic (naMCI) forms, while, considering the observed cognitive deficits, it can be divided into single- or multiple-domain subtypes. In aMCI impairments are observed predominantly in memory, while in the case of naMCI, negative changes occur in executive functions, attention, visuospatial ability, or language (Senanayake et al., 2016). In single-domain MCI, only one major cognitive domain is impaired (e.g., memory or executive functions), while in multiple-domain MCI more than one area is affected.

The incidence of MCI is fairly high among the elderly: approximately 15-20% of people at age 65 or older have MCI (Roberts & Knopman, 2013). Even though its outcome is not certain, one of the significant aspects of this condition is that it is associated with an increased risk of developing dementia later on (Alzheimer’s Association, 2018; Roberts et al., 2014). Before reaching the diagnostic threshold of probable Alzheimer’s disease, most patients experience a subtle cognitive decline, the characterization of MCI (Petersen et al., 2001), which presymptomatic phase can last for several years (Jack et al., 2013; Liss et al., 2021). Compared to cognitively healthy subjects, the conversion rate of Alzheimer’s dementia can be 3.1 times higher in persons with MCI (Bennett et al., 2002). In a recent study, researchers found an 18.4% 1-year conversion rate from MCI to dementia (Thaipisuttikul et al., 2022). However, the progression to dementia is not inevitable: some MCI patients remain in a state of mild memory impairment or even recover (Winblad et al., 2004). Furthermore, in a longitudinal study with 12 years of follow-up, the authors found the reversion rate to be as high as 58%, suggesting that patients with MCI could have a high chance of a positive prognosis (Overton et al., 2019).

2.3. The importance of early recognition

Although the two terms are often used interchangeably, it is worth mentioning briefly the difference between the early and the timely diagnosis of dementia. The term early diagnosis is usually used for a diagnosis that is made in the earliest stages, i.e., at the very first signs, or even before the manifestation of cognitive symptoms – often by relying on biomarkers. In contrast, timely detection is defined as "disclosure of the diagnosis at the right time for the individual with consideration of their preferences and unique circumstances" (Watson et al., 2018, p. 2). Compared to early diagnosis, which emphasizes the benefits that can stem from early interventions, timely diagnosis is considered a more person-centered approach (Ausó et al., 2020; Dhedhi et al., 2014; Watson et al., 2018).

In July 2021, the U.S. Food and Drug Administration approved *Aduhelm* (aducanumab) for the treatment of AD. Despite the insecurities regarding its significant clinical effects, it shows promising results based on the clinical trials conducted so far (Golde, 2022; Haddad et al., 2022). However, at present, there are still very limited disease-modifying treatment options for dementia: most of them only offer symptomatic treatment (Perneczky, 2019). However, early recognition is crucial in MCI and dementia, because it can provide an opportunity to reduce the rate of cognitive decline (Hahn & Andel, 2011), as interventions applied at the early stage of the disease are more likely to be effective (Sindi et al., 2015). In parallel, early detection allows better patient follow-up and helps to observe the disease mechanism as well (Brodaty et al., 2017). It also benefits the patients and their family significantly: it supports maintaining the independence of the patient, (e.g., helps to find strategies and tools to maximize independence), offers the opportunity to treat or control any comorbid conditions or factors that influence the cognitive decline (e.g., major depressive disorder, metabolic disorders, or certain lifestyle factors) and offers a chance for the family and relatives to start planning for the future (Knopman & Petersen, 2014; Liss et al., 2021). The diagnosis can also provide an explanation to the patients and their families regarding the recent cognitive or affective changes they may experience in their daily life (Ismail et al., 2010).

Despite its value, the difficulty of early dementia recognition is a global problem: research suggests that the rate of undetected dementia can reach 60% in the community setting or residential or primary care; moreover, many cases of AD remain undiagnosed even after years of the symptom manifestation (Boustani et al., 2003; Lang et al., 2017).

2.4. The road to diagnosis – the role of primary care

General practitioners (GPs) are greatly involved in the early stages of the dementia recognition process, as most patients visit them first to have their initial cognitive examination (Wilkinson et al., 2004).

The first step of the diagnostic process is usually a subjective complaint – voiced by the patients themselves, or by a close family member (often referred to as an informant). These complaints often regard difficulties in remembering things (e.g. forgetting names of acquaintances), in language use (e.g. word-finding difficulties), in orienting oneself in not familiar environments, misplacing personal items, losing track in conversations or losing track of the train of thoughts, or forgetting the aim of an ongoing activity (e.g., going into a room to fetch something) (Nelson & O'Connor, 2008). Even though subjective memory complaints play an important role in the detection of cognitive decline, a study found no relationship between the subjective feeling of deterioration and the actual level of cognitive functioning. Rather, including the complaints in the diagnostic process may lead to misclassification. Another important observation is that while cognitively intact subjects tend to overestimate their cognitive problems, MCI patients underestimate their cognitive difficulties (Edmonds et al., 2014).

It is also worth noting that when patients with early complaints go through a cognitive evaluation, a substantial proportion of them perform normally on global cognitive tests (for example on the Mini-Mental State Examination (MMSE)), which can delay the recognition of the condition. Thus, applying detailed cognitive assessments that are not only sensitive but cover multiple cognitive domains (such as memory, language, attention/executive functions, visuoperceptual/visuoconstructional performance) is pivotal (Lopez, 2013).

One of the main obstacles to effective dementia case-finding in primary care however is the low use of standardized cognitive tests. Although there are several available tools to guide the diagnostic process, the clinical diagnosis of MCI is still very often determined by a doctor's professional judgment regarding the causes of the patient's symptoms (Dementia Australia, 2020). Not only is dementia a taboo topic for many GPs (Kaduszkiewicz et al., 2008), but some of them also experience ambivalence regarding the advantages of early diagnosis (Hansen et al., 2008). The recognition rates are further influenced by cultural background and even gender: the awareness and concerns for cognitive deficits vary greatly between different ethnic and occupational groups and sexes (American Psychiatric Association, 2013).

Besides the complaints reported by the individual or their family members, GPs' concerns about signs of dementia during patient consultation, targeted case-finding, and population screening can also be potential pathways to the identification of cognitive disorders. Based on the literature and the related guidelines, the views on the value of cognitive impairment screening are controversial. In their recent review, the US Preventive Services Task Force concluded that there is a lack of evidence to determine the balance of advantages and disadvantages of screening (Owens et al., 2020). Even though numerous studies concluded that there is no evidence of the negative impact of screening, recent guidelines on the diagnosis of dementia do not support the routine screening of asymptomatic individuals (Ismail et al., 2020; Ranson et al., 2018).

In Hungary, after the first consultation, GPs can decide, if needed, to carry out basic neuropsychological tests (of which the MMSE and the Clock Drawing Test (CDT) are financially reimbursed) and/or refer potential dementia patients to secondary care (e.g., memory clinics, psychiatric care, neurology) for further investigation. It is important to note, that the above-mentioned brief cognitive tests are not designed for the diagnosis of dementia or MCI: rather, their role is to highlight the need for further, targeted examination or referral of the individual with positive results (Owens et al., 2020). The establishment of the diagnosis, the identification of the etiology based on the *International Classification of Diseases – 10th revision* (ICD-10), and the prescription of the necessary medications are the tasks of psychiatrists or neurologists.

3. Identification of neurocognitive disorders

3.1. Clinical characterization

There are several approaches for identifying the presence of MCI or dementia. Clinical characterizations can guide physicians, and, for better diagnostic accuracy, may be combined with neuropsychological or laboratory tests (Chun et al., 2021). Among the most widely-known sets of criteria for MCI, there are Petersen's criteria (Petersen et al., 1999), the NIA-AA Criteria (Albert et al., 2011), and the Jak-Bondi criteria (Jak et al., 2009). These sets of criteria are also commonly used as comparison standards when evaluating the diagnostic accuracy of other cognitive tests (Chun et al., 2021).

3.2. Biological approaches

Since the cognitive deterioration in MCI is so subtle that the onset of the condition can be hardly identified only by cognitive evaluation, the assay of biological markers has been an increasingly common practice in the last decades (Takeda et al., 2006; Tucker-Drob, 2019). Neuropathological measurements not only aid the identification of the specific subtypes of dementia but are also useful in the examination of disease progression (Tucker-Drob, 2019).

The deposition of amyloid can be observed 10-20 years before the clinical manifestation of AD, and thus can act as a useful marker for the presence of MCI (Takeda et al., 2006). Currently, the routinely used biomarkers for MCI and AD are cerebrospinal fluid (CSF) biomarkers tau and Aβ1–42 (Giau et al., 2019; Ma et al., 2022; Shaw et al., 2009).

Regarding neuroimaging, several modalities have been used targeting the identification of MCI or AD: diffusion sensor imaging (DTI), structural magnetic resonance imaging (MRI), functional MRI (fMRI), or positron emission tomography (PET) (Wee et al., 2012). Using PET, glucose hypometabolism in the parietal and temporal regions can indicate neurodegeneration (Anchisi et al., 2005). In patients with AD as well as MCI, MRI studies have shown atrophy in the hippocampal and entorhinal cortices and grey matter loss in the thalamic regions (van de Mortel et al., 2021; Wolz et al., 2011). In a longitudinal neuroimaging study, even though tissue loss was present in non-demented individuals as well, the observed change – in whole brain volume, ventricular CSF, temporal grey matter, orbitofrontal and temporal association cortices – was significantly accelerated in the case of MCI patients. Thus, these regions help to differentiate MCI from age-related changes (Driscoll et al., 2009).

Besides the decline of memory and executive functions, language impairment is another significant characteristic of dementia and is present even in the preclinical phase of the disease (Cuetos et al., 2007). As such, language deficits are also considered as promising candidates as biomarkers for the diagnosis of MCI. In speech analysis studies, the goal is to focus on and identify speech features that later can be feasible to differentiate between the healthy and the cognitively impaired population. The investigated features include temporal parameters (such as speech and phonation time, number and proportion of pauses, or prosodic rate), phonological variables (such as spectrum features or syllabic variability), voice quality measures, and amplitude parameters (Martínez-Nicolás et al., 2021).

3.3. Cognitive and neuropsychological tests

Despite the growing availability regarding the assessments of the above-mentioned biological measures, evaluating the cognitive changes remains the driving factor in the diagnosis of dementia and MCI (Tucker-Drob, 2019). According to a systematic review of clinical guidelines, the application of neuropsychological tests is the most often recommended approach besides biomarker assessments when it comes to MCI detection. Several neurocognitive tests have been proposed for the screening of MCI, such as the MMSE, the Montreal Cognitive Assessment (MoCA), the California Verbal Learning Test, or the Boston Naming Test (Y.-X. Chen et al., 2021). The most recent guidelines in Hungary recommend the following screening tests: MMSE, Addenbrooke's Cognitive Examination, Addenbrooke's Cognitive Examination-Revised, Addenbrooke's Cognitive Examination-III, MoCA, Alzheimer's Disease Assessment Scale – Cognitive Subscale (ADAS-Cog), and the CDT (Egészségügyi Szakmai Kollégium, 2020).

Since there are no specific and widely-accepted practical guidelines – in terms of tests to use and cut-off scores to implement –, there is a high heterogeneity and inconsistency both in research and clinical practice when it comes to the diagnosis of MCI. Even though regarding neuropsychological tests, there is no definitive operationalization for the MCI diagnosis, an impairment of 1.0–2.0 standard deviation (SD) below adjusted norm scores (with regards to age, education, and cultural background) on at least one test (assessing memory, executive functioning, attention, language or visuospatial skills) has been put forward (American Psychiatric Association, 2013). The decision about the applicable screening tool is influenced by several factors: the referral question, the functional status of the patient, or the theoretical background of the specialist, among others (Nelson & O'Connor, 2008).

Because the implementation of a complex and extensive diagnostic protocol can take up a considerable amount of time, the routine method for the diagnostic process is utilizing brief cognitive screening tests (Chun et al., 2021). The optimal screening test is short and easy to administer, and at the same time effective, free from biases associated with demographic factors, and is also acceptable for elderly patients (Lorentz et al., 2002). Besides showing high sensitivity, a screening tool at the primary care level is preferably not too specific, to ensure high yield (Abd Razak et al., 2019). Moreover, because of the high prevalence of MCI (ranging from 16% to 20%), it is favorable to have a screening method that allows targeting a large number of potentially affected individuals at frequent intervals (L. Chen et al., 2020; Roberts & Knopman, 2013). In their systematic review, Abd

Razak and colleagues separated three main approaches to the screening measures used at the primary healthcare level: the instrument can be (1) administered by healthcare providers, (2) by the patient, or (3) the caretaker can fill out a self-administered questionnaire (Abd Razak et al., 2019).

Many studies had cataloged, evaluated, and ranked dementia screening tools based on different aspects. The most frequently used tools for the screening for MCI are the MoCA, the CDT, and the MMSE (Abd Razak et al., 2019; Chun et al., 2021), while the Mini-Cog is also widely used (Fernandes et al., 2021). Even though MMSE is the most commonly used tool in research and clinical settings, MoCA seems to have a better ability to detect the subtle cognitive decline of MCI (Abd Razak et al., 2019; Ciesielska et al., 2016). Based on a review of the diagnostic accuracy of MMSE regarding MCI, MMSE seems to have a sensitivity ranging between 45% and 77%, and a specificity ranging between 53% and 92% for the detection of MCI (Lin et al., 2013). Compared to this, MoCA seems to have a higher (83-97%) sensitivity to the presence of MCI (Abd Razak et al., 2019).

4. Verbal fluency tests

4.1. Types and characterization

Verbal fluency tests are among the most common neuropsychological tests, administered both in research and clinical settings. While their asset requirement is minimal, their significant advantage is that they can be administered to individuals of various ages and levels of education (Oberg & Ramírez, 2006). Verbal fluency tests can be divided into two subtypes: *phonemic* (PVF) and *semantic* (SVF) verbal fluency, also known as letter fluency and category fluency, respectively. In the standard versions of the tests, subjects are given 60 seconds to recall as many words as they can, which begin with a given letter (PVF) or belong to a given semantic category (SVF). Their administration can vary based on the phonological or semantic restriction set by the administrator.

PVF tests usually include 3 trials (with 3 different starting letters) (Lehtinen et al., 2021). The starting letters in PVF have a significant effect, as they determine the number and frequency of the eligible words (Strauss et al., 2006). According to the results of a cross-linguistic meta-analysis of PVF, letters with high frequency in the target language result in a higher number of words (Oberg & Ramírez, 2006). Mainly in English but in the case of several other languages as well, the most commonly used letter combination for

PFV is f, a, and s (Oberg & Ramírez, 2006; Olabarrieta-Landa et al., 2017). In the SVF tasks, frequently used semantic categories are animals, fruits, vegetables, and clothes, while vehicles, objects, food items, and items found at home or in a supermarket are also applied, although less frequently (Olabarrieta-Landa et al., 2017).

Besides these above-mentioned verbal fluency tasks, another, later developed fluency task type, action fluency (or verb fluency) is also worth mentioning. When performing action fluency, the participants have to produce as many verbs ("things that people do") as they can (Piatt et al., 1999).

Despite their simplicity, performing verbal fluency tasks requires the simultaneous activation of multiple cognitive processes (Troyer et al., 1997). Besides evaluating knowledge and memory, both PVF and SVF tests rely on other cognitive processes as well: they assess the executive functions (divergent reasoning for generating category example, flexibility while searching subcategories) and engage the working memory (the subjects need to keep the exact instruction and prior responses in mind) (Mueller et al., 2015). Cognitive control processes also play a major role in the execution of verbal fluency tests, as during the test one must repress the repetitions and any potentially incorrect or irrelevant responses (Shao et al., 2014). While both PVF and SVF require rapid associative exploration, two different cognitive areas are involved in the process of performing them. Since SVF relies more on semantic associations it reflects more on the integrity of semantic memory, while PVF is more dependent on search strategies based on lexical representation (Henry & Crawford, 2004; Teng et al., 2013). Furthermore, while both SVF and verb fluency tasks are content-oriented ("guided by meaning") speech tasks (Östberg et al., 2005; Vita et al., 2014), verb fluency may be more sensitive to the functions of frontal-subcortical circuits (Cappa et al., 2002; Davis et al., 2010).

The validity of fluency tests for the assessment of verbal and executive skills has been confirmed by multiple studies (Shao et al., 2014). Since these abilities, among others, are proven to be altered in dementia and other forms of cognitive impairments, fluency tests have great potential to become effective screening tools.

According to the results of a meta-analysis on verbal fluency performance, the deficit of both PVF and SVF is related to the severity of dementia measured by the MMSE (Henry & Crawford, 2004). It is worth mentioning however, that, fluency tests, and especially SVF tests have significant advantages over MMSE. In contrast with the SVF tests, MMSE is insensitive to some important cognitive domains, which are impaired in dementia: for example, it does not include any tasks measuring executive functions.

According to the research of Kim and colleagues, MMSE supplemented with verbal fluency task (task type not specified) results in a significantly better screening ability for MCI than using MMSE alone (Kim et al., 2014).

4.2. Verbal fluency analysis methods

The most traditional and widely-used method used to assess fluency performance requires the researcher or clinician to score the number of unique and correct words that are produced by the participant, while also counting the repetitions, perseverations, and intrusive (incorrect) words. Especially in the case of PVF tasks, both error and repetition scores offer cues for the detection of MCI and AD (Wajman et al., 2019).

If one wants to examine the task performance solely based on word count, the traditional method can be refined by comparing the scores of the different time-intervals (e.g., 0-20, 21-40, 41-60 secs) (Demetriou & Holtzer, 2017; Jacobs et al., 2020) or by only considering the number of produced words regarding one interval (Venegas & Mansur, 2011). These methods have the advantage of enabling us to gain information about the temporal dynamics of the word production. Based on observations, the number of recalled words falls progressively during the time-span of the task (Cho et al., 2021; Demetriou & Holtzer, 2017; Venegas & Mansur, 2011). This can be due to the different cognitive processes that are dominant in the different stages of the task (mostly automatic word-retrieval at the beginning and more controlled and effortful word-retrieval towards the end) (Crowe, 1998; Fernaeus & Almkvist, 1998).

Although it is a less frequent approach, some studies focus on lexical-semantic variables regarding the words, such as word frequency, familiarity, or typicality. A research conducted among cognitively intact, aMCI, and AD patients showed that the cognitively impaired groups produced words with higher typicality than control subjects, while high typicality was also related to conversion to AD (Vita et al., 2014). Word frequency can also be examined regarding the time span of the tasks. Dor example, according to the results of a study, the frequency of words in fluency tests decreases over time, i.e., while participants list more common words at the beginning of the task, the words produced towards the end of the task are less common (both in the healthy control and in the cognitively impaired groups) (Linz et al., 2019).

A more elaborate fluency analysis method, the so-called *cluster-analysis* or *clustering* is based on grouping the consecutive words that are similar in some respects (e.g., rhyming words, homonyms, words beginning with the same two first letters in case of any phonemic fluency; pets or farm animals in case of animal fluency tasks) (Troyer et al., 1997). As the executive functions involved in the test deteriorate with age, cluster size (the number of words belonging to one subcategory) and the number of switching (calculated as the number of clusters -1) also show a decreasing pattern with age (Zhao et al., 2013). In SVF tasks, the number of switches seems to be able to differentiate between subjects with healthy cognition and MCI (Oh et al., 2019). Impaired switching performance in animal SVF test could be an effective precursor sign for the later conversion to AD. Based on the results of a 17-year longitudinal study, lower switching index in the case of future AD patients could be observed 5 years before the clinical diagnosis (Raoux et al., 2008). Although the method of clustering can provide more in-depth information about the underlying mental processes involved during the task, the procedure is rather lengthy and burdensome. Most of the time it requires the manual coding and grouping of words, which, besides being rather time-consuming can raise reliability issues, as it depends on how raters determine certain subcategories (Cho et al., 2021; Taler et al., 2020). Furthermore, in the case of SVF, the priori-determined subcategorization schemes cannot include all the possible subcategories an individual may create (Woods et al., 2016).

4.3. Computational approaches

Owing to the fast development of mobile technology, increasingly more researchers examine the way mobile platforms could aid cognitive assessment among the elderly. Based on the level of innovation, Koo and colleagues suggested three main categories of mobile assessments: (1) mobile or computerized versions of existing neuropsychological tests, (2) novel cognitive tools developed specifically for using them via mobile platforms, or (3) the use of new data types (e.g. game performance metrics or physical movement changes) for cognitive assessment (Koo & Vizer, 2019).

In the past years, aiming to address the limitations of the manual methods and to achieve large-scale analysis with objective and quick results, there have been multiple attempts to automatize the application and analysis of verbal fluency tests. Most of these attempts are focused on the automatization of scoring or the automatization of cluster analysis. For the latter, the main goal is to automatize the identification of clusters to make the process faster and less prone to inter-rater variability and subjectivity.

Cho and colleagues used automated analyses of letter fluency data: their algorithm counted the number of correct responses from manual fluency transcripts (the number of errors, e.g., proper nouns or numbers were subtracted using automated part-of-speech category tagging). By aligning audio signals using the transcript of verbal fluency, they also extracted temporal measures, such as word start, word duration, and inter-word reaction time (Cho et al., 2021).

In their pilot study, Ryan and colleagues presented a system for automated phonetic clustering analysis. Their system used two methods for determining phonetic clusters (or phonetic similarity): the common-biphone check (based on binary similarity values) and the edit-distance method (based on phonetic-similarity score). According to the results, their automated approach (using a common-biphone check) proved to be more sensitive to brain damage or degeneration than the manual cluster-analysis (Ryan et al., 2013).

To be able to automatize the clustering process in the case of SVF, the strength of semantic relatedness has to be measured automatically. For this purpose, numerous researchers apply a technique called latent semantic analysis (LSA), which is based on the co-occurrence of words in large corpora of natural text (including articles, books, and speeches), representing the semantic context of a word. Based on these contexts, a numeric value (between 0 and 1) might be derived to indicate the strength of semantic relatedness between words (Ledoux et al., 2014; Pakhomov & Hemmy, 2014).

Woods and colleagues used another computational method called explicit semantic analysis (ESA), which defines the strength of the semantic association between words on a continuously varying scale utilizing word concept vectors derived from the analysis of Wikipedia entries. According to the authors, the advantage of this method is that it quantifies semantic relationships based on multiple conceptual similarities (e.g. taxonomic, cultural, economic), and it can be applied to any semantic category (Woods et al., 2016).

In the past few years, researchers also introduced the analysis of temporal dynamics of verbal fluency performances. Temporal information of the tasks is mainly combined with semantic information, which is based on the idea that there is an association between the meaning of the words (i.e., their relatedness) and the tempo at which they are generated (Cho et al., 2021; Holmlund et al., 2019; Tröger et al., 2019). In their research, Holmlund and colleagues, following the manual transcription of the voice samples, used a forced temporal alignment method to timestamp response-words, and evaluated the semantic associations between individual words utilizing GloVe word vectors. Their results showed that there was a correlation between the speed of speech and the semantic coherence between successive words, indicating longer pauses between semantically less related words (Holmlund et al., 2019). In their article, they highlight the fact that, by utilizing a calibrated model, automatic transcription of digitally collected verbal fluency data is achievable with a relatively low error rate. Furthermore, by using automatic speech recognition (ASR) system with high resolution (+/- 10 ms) and applying forced alignment tools, one can gain valuable temporal information on verbal fluency tasks by time-stamping each utterance in the recording (Holmlund et al., 2019).

Despite the multiple experiments on computational approaches of verbal fluency analysis, there is no standardized tool for application yet. A major difficulty regarding the fully automatic end-to-end analysis of audio fluency recordings stems from the characteristics of the general 1-minute response. Voice recordings of fluency test performances often comprise more than solely a sequence of task-relevant words: they usually also contain extraneous speech, like filler words or hesitations ("er", or "uhm"), irrelevant comments ("oh it's not as easy as I thought…"), questions directed at the experimenter ("is there still time left?"), utterances that express loud thinking ("I'm not sure, maybe I said this one before…", "then there's lion, and… lion, lion…"), or other parts of speech, like conjunctions. To be able to automatically analyze the relevant parts, fluency recordings need to go through an often time-consuming preparation process prior to analysis: the words irrelevant to the tasks need to be removed from the recording/transcription and in some cases, words need to be lemmatized (i.e., to be converted to their stem) (L. Chen et al., 2020; Holmlund et al., 2019).

Given the substantial amount of task-irrelevant content in most fluency recordings, the question arises whether the analysis of these segments could provide valuable information regarding the overall performance of the patient.

In summary: Low rates of dementia and MCI detection in primary care is a global problem. Since primary care practices act as the first step in the identification process, examining GPs' views and approaches towards the topic of cognitive screening could help us to enhance the current ineffectual routines and thus, the low detection rates.

According to numerous studies conducted in various countries, both at primary and at clinical health care levels, the most widely used, conventional evaluation process for the detection of cognitive decline are traditional pen and paper testing methods. Even though some of these brief cognitive tools show sufficient sensitivity, their administration and scoring can be time-consuming for everyday use in clinical settings, and they can also pose difficulty when their re-assessment is needed to monitor disease progression. Because of their short and rather simple administration, verbal fluency tasks could be optimal screening tools, however, their evaluation often requires a substantial amount of time, and some of the methods of analysis also raise inter-rater reliability issues. Thus, there is a great need for low-cost and at the same time rapid methods that would allow the effective and objective recognition and follow-up of the early stages of cognitive decline.

IV. AIMS OF THE STUDIES

The *first study* focuses on dementia screening and detection in Hungarian primary care. Therefore, in this study, we aimed to:

(I) examine Hungarian GPs' views regarding the early recognition and the current recognition rates of dementia

(II) identify the methods GPs use for dementia screening

(III) evaluate GPs' satisfaction with the available and most widespread neurocognitive and dementia screening tests

(IV) explore GPs' ideas about an ideal test for early recognition and those optimal circumstances that could contribute to more effective dementia identification in Hungarian primary care.

The focus of the *second exploratory study* of the thesis was to examine PVF and SVF audio recordings by moving beyond the words listed by the participants and thus, by exploring the additional, previously unharvested data present in the fluency recordings. Our aims were to:

(I) examine whether the derived temporal parameters differ between participants classified as healthy control (HC) and as MCI

(II) compare the traditional, word count-based method and the temporal parameters regarding their ability to detect differences in the performance of the HC and MCI groups

(III) compare the different (phonemic and semantic) types of fluency tasks investigating their sensitivity to the presence of MCI.

V. MATERIALS AND METHODS

1. Study 1

1.1. Study questionnaire

As part of a national research project in collaboration between the University of Szeged and the University of Pécs, a self-administered questionnaire was designed specifically to explore a broad range of aspects regarding GPs' views on dementia and their role in its detection and management in Hungary. In the survey, several significant topics were investigated, including GPs' routines and perspectives regarding dementia screening and detection, which topic is covered by the present study. Further items of the questionnaire targeted GPs' factual knowledge of dementia (see: Imre et al., 2019) as well as their attitudes regarding dementia patients and their management (see: Heim, 2022; Heim et al., 2019). The development and validation of the questionnaire was a multi-stage process, taking up to one year (*Figure 1*). The questions analyzed in the present paper were fixed-response (single or multiple choice) and Likert-type questions; open-ended questions were not applied.

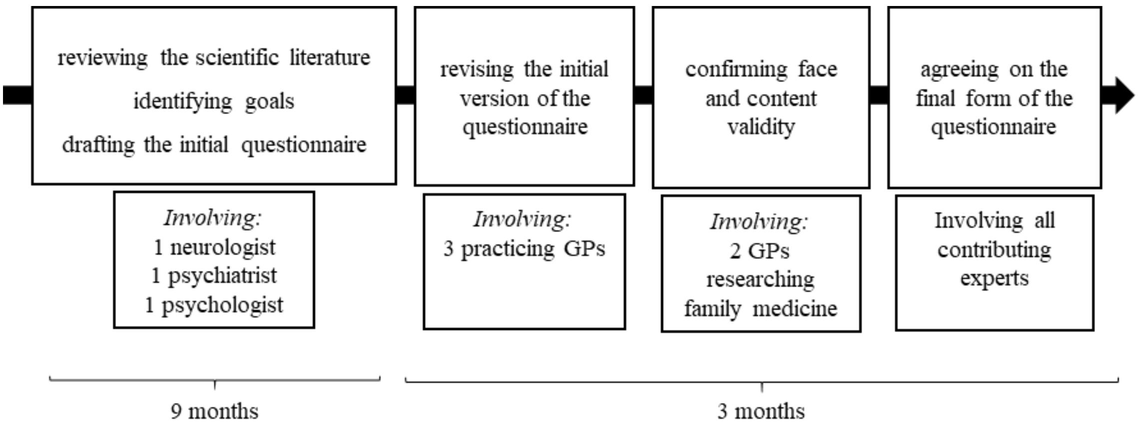

Figure 1. The multi-stage process of the questionnaire development.

1.2. Participants

In Hungary, all GPs are obligated to participate in a continuous postgraduate education program, which means attending one professional training course every 5 years. Since the aim was to reach as many GPs as possible from every region of the country, the questionnaires were distributed at six major mandatory training courses and at three national conferences (ensuring that GPs from all 19 counties of Hungary could be represented among the attendees). The events were held within a 10-month time frame, between February and November 2014. To avoid the courses' influence on the results of the study, we selected events that did not provide any specific education about dementia during our recruitment period. The questionnaires were distributed along with a written informative. Participation was entirely voluntary and anonymous. Ethical approval was obtained from the Regional and Institutional Research Ethics Committee of the University of Pécs (reference number: 5244).

1.3. Data analysis

Data were analyzed using the SPSS v.24 statistical analysis software package (*IBM SPSS Statistics for Windows*, 2016). Descriptive statistics (mean, percentage, standard deviation) were applied for all items on the questionnaire. Comparative analysis was executed for one question, using the Wilcoxon signed ranks test. The significance level was set at 0.05.

2. Study 2

2.1. Participants

Participants (patients and their relatives, scheduled for consultations) were recruited at the Memory Clinic of the Department of Psychiatry, University of Szeged. Data collection was carried out between February 2018 and March 2020. Participation in the study was voluntary. All participants were informed about the aims of the study and gave their written consent. The experiment was conducted according to the ethical principles of the Declaration of Helsinki, and it was approved by the Regional Human Biomedical Research Ethics Committee of the University of Szeged (Reference No. 231/2017-SZTE).

The required sample size for the study was assessed a priori using G * Power v.3.2.9.7. (Faul et al., 2007) with the settings of effect size $d = 0.8$, alpha error probability: 0.05, power (1-beta error probability): 0.8. Based on this, the optimal sample size was calculated as 52, which later (due to COVID-19 regulations halting data collection in clinical research) was limited to 50. Initially, a total of 79 individuals were recruited to take part in the study.

Inclusion criteria were listed as follows: at least 50 years of age, a minimum of 8 years of formal education, and Hungarian as the native language. The two main exclusion criteria were the presence of dementia or major cognitive deficits and depression. To rule out possible cases of dementia, the MMSE (Folstein et al., 1975) was applied as a screening tool: participants with a score of 24 or below were excluded from the study. The possibility of depression was assessed using the 15-item version of the Geriatric Depression Scale

(GDS-15) (Yesavage & Sheikh, 1986): participants scoring 7 or above on the test were excluded. In addition, individuals were excluded from the study if they had any past or present neuropsychological, psychotic, or affective disorders, head injuries, stroke, substance abuse disorders, major (uncorrected) hearing loss, or language problems (e.g., stutter), based on patient history and medical records. Participants with MRI or CT records showing evidence of micro- or macrohemorrhages, lacunar or other infarctions, cerebral contusion, encephalomalacia, aneurysm, vascular malformation, or space-occupying lesions were also excluded. After reviewing and evaluating the criteria, 50 subjects were considered eligible for inclusion in the study (*Figure 2*).

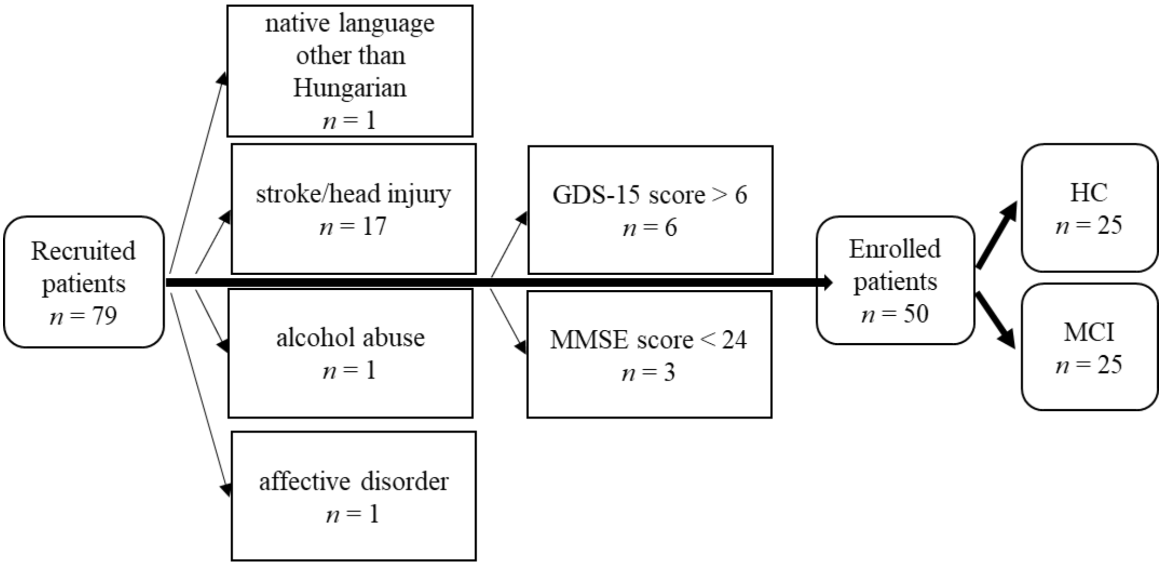

Figure 2. Flowchart of the participant exclusion process.
GDS-15: 15-item Geriatric Depression Scale; MMSE: Mini-Mental State Examination; HC: healthy control; MCI: mild cognitive impairment

2.2. Study protocol

Each participant performed a series of neuropsychological tests: six fluency tasks, the Digit Span Test – Forward and Backward (Wechsler, 1981), the Non-Word Repetition Test (Gathercole et al., 1994), the Listening Span Test (Daneman & Carpenter, 1980), the CDT (Shulman et al., 1986), and the ADAS-Cog (Rosen et al., 1984). The fluency tasks were implemented in a fixed order, separated by the five shorter cognitive tests, while ADAS-Cog was administered at the very end of the study protocol to prevent fatigue. We also ensured that tasks assessing the same cognitive domain did not follow each other directly. In the three PVF tasks, the participants were asked to list as many words as they can, starting with the letters ‘k’, ‘t’, and ‘a’, while avoiding proper nouns. The starting letters in this study were chosen based on previous studies conducted with Hungarian-speaking population (e.g., Mészáros et al., 2011; Tánczos et al., 2014).

For the SVF tasks, participants had to produce as many items belonging to the category of animals, food items, and verbs (i.e., actions – "things that people do") as they could. In the current study, for the sake of simplicity, action fluency will be regarded as a SVF task, because both semantic fluencies and action fluency are content-oriented speech tasks. Regarding the SVF tasks, the participants were instructed to avoid saying variations of the same word stem (e.g., horse, horses; go, goes). For all six verbal fluency tasks, participants had one minute to perform the task. The one minute-interval began with the investigator saying: 'Start.' Every verbal fluency task was recorded using an Olympus Digital Voice Recorder (16 kHz sampling rate, 16-bit resolution). The recordings were also transcribed manually for the calculation of the traditional scores.

2.2.1. Analysis based on temporal parameters

Voice recordings of all fluency tasks were manually transcribed in Praat, a free language software enabling speech analysis (Boersma & Weenink, 2020). The transcription process was supervised by a linguist specialized in language pathologies, while quality control was ensured by an expert in the field of computational speech processing. Due to the quality of their recordings, an HC participant's animal category fluency task and an MCI participant's 'k' letter fluency task were unsuitable for transcription; therefore, these recordings were not considered in the analysis of temporal parameters, but they were included in the traditional analysis.

The transcriptions of the fluency recordings contain not only the task-relevant answers of the participants (the recalled words – including correct, incorrect, and repeated words), but also silent pauses, and paralinguistic phenomena: hesitation sounds (filled pauses, like "hmm" and "er"), and irrelevant utterances, such as comments or loud thinking said by the subjects. False starts ("te- … tiger"), as well as laughing and coughing sounds were also annotated. The laughter, coughs, and false starts were considered unintentional, and, as the number of their occurrences was negligible, were discarded from further analysis.

For each recording, task-relevant words, silent segments, hesitation sounds, and irrelevant utterances were annotated based on their boundaries (i.e., their exact start and end times), providing their duration measures. Based on this, the total number, the average length, and the total length of *silent pauses*, the total number, the average length, and the total length of *hesitations*, and the total number, the average length, and the total length of *irrelevant utterances* were calculated. Besides these parameters, the mean time between two consecutive task-relevant words (*average word transition time*) was also calculated based on the transcript. Not only correct words, but also the errors and repetitions, were considered as task-relevant words. The average word transition time (irrelevant of its content, such as silent pause, hesitation, or irrelevant utterance) provided information about the average time the participant needed to produce a new task-relevant word. The parameters used in the study are listed and defined in *Table 1*; two waveform extracts from a fluency task performed by an HC and an MCI subject are shown in *Figure 3*.

Temporal fluency parameters	Description
Silent pause parameters	
Total number of silent pauses (count)	Number of silent segments
Average length of silent pauses (s)	Average length of silent segments
Total length of silent pauses (s)	Total length of silent segments
Hesitation parameters	
Total number of hesitations (count)	Total number of filled pauses (e.g., 'hmm', 'umm')
Average length of hesitations (s)	Average length of filled pauses (e.g., 'hmm', 'umm')
Total length of hesitations (s)	Total length of filled pauses (e.g., 'hmm', 'umm')
Irrelevant utterances parameters	
Total number of irrelevant utterances (count)	Total number of filler words and comment blocks (including articles and conjunctions)
Average length of irrelevant utterances (s)	Average length of filler words and comment blocks (including articles and conjunctions)
Total length of irrelevant utterances (s)	Total length of filler words and comment blocks (including articles and conjunctions)
Average word transition time (s)	Mean period of time between two consecutive 'task-oriented' words

Table 1. List and definitions of the temporal parameters.

2.2.2. Traditional fluency analysis based on word count

In the traditional scoring method (Lezak, 2012), we calculated the number of correct words, the number of errors, and the number of repetitions or perseverations; the last two were considered as one variable. In the case of animal fluency, when a participant recalled synonymous words (e.g., cat and kitten), variations in gender (e.g., hen and rooster), or an animal and its offspring (e.g., horse and foal), words were only scored as one. The participants did not receive points for naming a subcategory if they also gave specific examples of it (e.g., in the case of food items: fruit (0 points), apple (1 point), pear (1 point)).

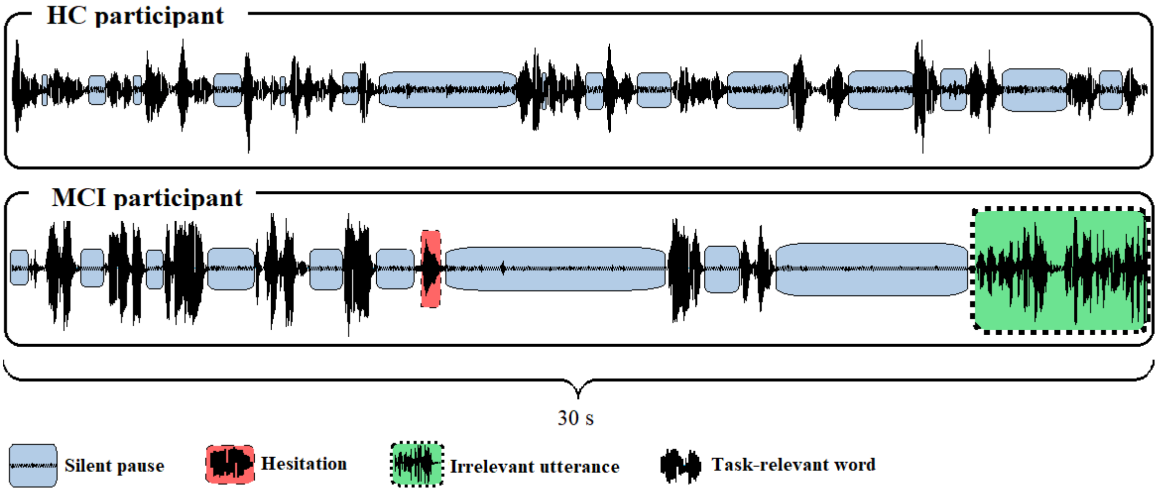

Figure 3. Waveforms extracted from the food item fluency recordings of two participants.
Extracted from Praat; HC: healthy control; MCI: mild cognitive impairment

2.3. Data analysis

Descriptive statistical analysis was used to examine the demographic features, the neuropsychological test scores, and the fluency measures of the participants. The assumption of normality was not met according to the results of the Shapiro-Wilk test in more than two-thirds of the cases, therefore, to obtain comparable statistical measures, comparisons between the HC and the MCI groups were executed using the Mann-Whitney U test. Categorical variables were compared using the Chi-square test. Effect sizes were calculated using the Pearson correlation coefficient (Rosenthal, 1991). Receiver operating characteristic (ROC) analysis was applied to assess the classification abilities of the temporal parameters and the traditional scores. Sensitivity and specificity were calculated using threshold values that yielded the highest possible sensitivity (while keeping specificity at a minimum of 50%). For the comparison of classification abilities, the differences between the area under the curve variables (AUCs) were compared based on the method of DeLong, DeLong, and Clarke-Pearson (1998). For all statistical comparisons, the level of significance was set at 0.05. All analyses were performed using SPSS v.24 (*IBM SPSS Statistics for Windows*, 2016), except for the comparison of AUCs, for which the MedCalc Statistical Software v.19.6. (*MedCalc Software*, 2020) was utilized. For the a priori assessment of the required sample size, G * Power v.3.2.9.7. was used (Faul et al., 2007).

VI. RESULTS

1. Study 1

1.1. Demographics and practice characteristics

Altogether 402 GPs handed back their completed questionnaire, which is more than 8% of all 4,850 GPs practicing in Hungary in 2014 (Hungarian Central Statistical Office, n.d.). The completion rate varied for each question, therefore, in the *Results* section, the numbers of responses are indicated in brackets for each question. Demographic information and characteristics of practices are presented in *Table 2.*

Gender		**Age**		**Estimated number of patients/day**		**Estimated number of dementia patients**	
(*n* = 387)	%	(*n* = 393)	%	(*n* = 393)	%	(*n* = 383)	%
male	46.3	25-35	5.9	0-30	2.0	0-50	49.9
female	53.7	36-45	12.5	31-40	16.9	51-100	38.1
		46-55	24.9	41-50	27.9	101-150	8.4
		56-65	40.2	51-60	25.1	151-200	2.6
		65+	16.5	60+	25.9	200+	1.0

Table 2. GPs' demographics and practice characteristics.

1.2. Ways of dementia evaluation and views on cognitive tests

The vast majority of GPs reported that they ask the patient general questions (91%; *n* = 355) or they gather information from relatives (64%; *n* = 253). Only a quarter of them (24%; *n* = 95) indicated that they utilize cognitive tests for this purpose and some did not perform any examinations at all to test for the possible occurrence of dementia (5%; *n* = 22).

Two of the most widely used tests for dementia evaluation, the MMSE and the CDT, are fairly well-known among respondents: most GPs reported that they knew CDT (89%; *n* = 307) and slightly fewer people stated familiarity with MMSE (76%; *n* = 265). One-fifth (18%; *n* = 63) of the respondents said that they knew the Early Mental Test (Kálmán et al., 2013), and only a few GPs stated that they were familiar with Mini-Cog (4%; *n* = 17) or GPCOG (1%; *n* = 4). Of them, more than two-thirds indicated that they were completely or mostly satisfied with the CDT (69%; *n* = 152) while a slightly lower percentage of them expressed satisfaction with the MMSE (65%; *n* = 98).

1.3. Views regarding dementia identification and management

Supporting the importance of dementia recognition in its early stages, the vast majority (90%; *n* = 352) believed that early therapy could slow down symptom progression. GPs also held the view (97%; *n* = 374) that early detection enhanced both the patients' and their relatives' well-being.

Regarding their views on dementia testing and management, participants were required to mark their answers on a 5-point Likert-scale (strongly agree/mostly agree and strongly disagree/mostly disagree responses are presented together). Three-fourths (75%; *n* = 290) of the GPs believed that managing dementia patients and their caregivers took more time than they could afford in their practice. Provided that conditions were suitable, the majority (79%; *n* = 298) would implement standardized cognitive tests for early detection; however, half of the respondents (56%; *n* = 210) felt that currently available anti-dementia therapies were ineffective *(Table 3)*.

	Level of agreement						
Statement	Strongly agree (%)	Mostly agree (%)	Can not decide (%)	Mostly disagree (%)	Strongly disagree (%)	*n*	*M* *(SD)*
If conditions were suitable, I would implement screening tests for early detection of dementia.	25.7	53.3	6.4	13.0	1.6	377	2.11 (0.987)
Managing dementia patients and their caregivers takes more time than I can afford at my practice.	31.9	43.4	5.2	15.6	3.9	385	2.16 (1.150)
Currently available anti-dementia therapies are effective.	1.9	14.8	26.7	37.2	19.4	371	3.57 (1.022)

Table 3. GPs' views of the detection and management of dementia.
Points of the Likert-scale: 1: Strongly agree; 2: Mostly agree; 3: Can not decide; 4: Mostly disagree; 5: Strongly disagree. M: mean, SD: standard deviation

1.4. Suggestions for the improvement of dementia detection

From a list of five contributing factors to a more effective dementia examination routine, GPs marked the items as necessary with the following percentages: more time for patients (81%; *n* = 311), up-to-date tests (with a maximum of 5 minutes needed for administration and evaluation) (77%; *n* = 297), help from assistants (50%; *n* = 192), more staff (44%; *n* = 170), and, lastly, more examination rooms (26%; *n* = 103).

Regarding an optimal, up-to-date instrument, GPs preferred a pen-and-paper test that could be administered by an assistant or the patients themselves and would include information from the patients' caregivers (*Table 4*).

Aspects	Options	*n*	%
Test administrator (*n* = 230)	assistant	87	37.8
	patient	86	37.4
	GP	54	23.5
	caregiver	3	1.3
Caregiver information (*n* = 317)	should contain	278	87.7
	should not contain	39	12.3
Test format (*n* = 321)	pen-and-paper test	265	82.6
	computer-based test program	48	15.0
	online test	8	2.5
Administration time (*n* = 330)	up to 5 minutes	189	57.3
	up to 10 minutes	110	33.3
	up to 15 minutes	31	9.4

Table 4. GPs' ideas about an optimal cognitive screening tool.

1.5. Estimated recognition of dementia

GPs were asked to estimate the recognition rate of dementia in Hungarian primary care and in their practice. Regarding primary care, almost two-thirds of them (62%; *n* = 226) thought that case recognition is under 30% and only very few (7%; *n* = 27) estimated that dementia is recognized in more than 60% of the cases. However, when asked about their recognition rate, half of them (49%; *n* = 180) said that they recognize a maximum of 30%, meanwhile, one-sixth (16%; *n* = 61) reported that they detect more than 60%. Wilcoxon signed ranks test was performed and results suggested that GPs' estimation of dementia recognition rate in their practice was significantly higher than their estimations of recognition rate in primary care ($Z = -7.806$; $p < 0.001$).

2. Study 2

2.1. Demographics and neuropsychological test scores

Participants were split into two groups based on their MMSE scores. MMSE cut-off scores were determined based on the results of previous research conducted by our research group: in these studies, the mean scores of MMSE emerged as 29.17 ± 0.71 and 29.24 ± 0.523 for the HC group and 26.97 ± 0.96 and 27.16 ± 0.898 for the MCI group (Gosztolya et al., 2019; Toth et al., 2018). Hence, in the present study, participants achieving a score of 29 to 30 points were considered healthy control (HC) subjects, while participants achieving a score of 25 to 28 points formed the MCI group. The subtypes of MCI were not considered.

The two participant groups showed no significant difference in gender and years of education. However, the mean age of participants was significantly higher in the MCI group in comparison with those in the HC group. Regarding the GDS-15 score, no significant difference was found between the two groups (*Table 5*).

	HC (n = 25)	**MCI (n = 25)**		
	M (SD)		**Comparative test statistics**	***p***
Demographics				
Gender (male/female)	8/17	7/18	$\chi^2(1) = 0.095$	0.758
Age (years)	67.32 (8.300)	71.72 (5.435)	U = 187.000; Z = -2.440	**0.015**
Education (years)	13.48 (2.632)	12.36 (2.827)	U = 255.500; Z = -1.136	0.256
Neuropsychological test scores				
MMSE	29.44 (0.507)	26.96 (1.060)	U = 0.000; Z = -6.202	**< 0.001**
GDS-15	1.84 (1.724)	2.40 (1.225)	U = 232.500; Z = -1.587	0.112

Table 5. Descriptive and comparative statistics for the demographic characteristics and neuropsychological test scores of the study participants.
Significant p-values (p < 0.05) are in ***bold****; HC: healthy control; MCI: mild cognitive impairment; MMSE: Mini-Mental State Examination; GDS-15: 15-item Geriatric Depression Scale*

2.2. Temporal parameters of verbal fluency performance

Considering the PVF tasks, in the ‘a’ fluency, the average length and the total length of irrelevant utterances were significantly higher in the MCI group, while none of the temporal parameters differed between the two groups in the case of the ‘k’ and ‘t’ PVF (*Table 6*).

Regarding the three SVFs, the total number of silent pauses was significantly higher in the HC group in the animal and action fluency tasks, whereas the average length of silent pauses and the average word transition time were significantly higher in the MCI group throughout all the three tasks (*Table 7*).

Phonemic fluency tasks	**HC**	**MCI**				
Temporal parameters	***M*** **(*SD*)**		**Mann-Whitey *U* Test**			**Effect size**[r]
Letter 'k'	$n = 25$	$n = 24^{*}$	***U***	***Z***	***p***	r
Total number of silent pauses (count)	19.040 (4.485)	17.291 (4.591)	230.500	-1.394	0.163	0.19
Average length of silent pauses (s)	2.438 (0.941)	2.767 (1.031)	244.000	-1.120	0.263	0.16
Total length of silent pauses (s)	42.569 (6.571)	43.532 (5.688)	278.000	-0.440	0.660	0.06
Total number of hesitations (count)	2.000 (2.645)	1.708 (2.095)	281.500	-0.382	0.702	0.05
Average length of hesitations (s)	0.482 (0.448)	0.398 (0.382)	279.500	-0.421	0.674	0.06
Total length of hesitations (s)	1.350 (1.737)	1.137 (1.334)	283.000	-0.349	0.727	0.05
Total number of irrelevant utterances (count)	3.280 (4.559)	4.333 (3.818)	225.000	-1.517	0.129	0.22
Average length of irrelevant utterances (s)	1.021 (0.666)	1.242 (0.851)	274.000	-0.520	0.603	0.07
Total length of irrelevant utterances (s)	4.283 (7.149)	4.889 (3.609)	213.000	-1.742	0.082	0.25
Average word transition time (s)	4.505 (2.687)	5.159 (3.979)	230.000	-1.400	0.162	0.20
Letter 't'	$n = 25$	$n = 25$				
Total number of silent pauses (count)	18.320 (4.269)	16.920 (5.259)	257.000	-1.081	0.280	0.15
Average length of silent pauses (s)	2.521 (0.879)	2.993 (1.542)	261.000	-0.999	0.318	0.14
Total length of silent pauses (s)	42.847 (5.770)	43.834 (5.666)	278.000	-0.669	0.503	0.07
Total number of hesitations (count)	1.480 (2.023)	1.720 (2.051)	290.000	-0.455	0.649	0.06
Average length of hesitations (s)	0.520 (0.509)	0.443 (0.348)	293.500	-0.375	0.708	0.05
Total length of hesitations (s)	1.128 (1.504)	1.069 (1.326)	312.500	0.000	1.000	0.00
Total number of irrelevant utterances (count)	3.240 (3.562)	3.720 (2.806)	256.500	-1.097	0.273	0.16
Average length of irrelevant utterances (s)	0.967 (0.580)	1.228 (0.616)	231.500	-1.573	0.116	0.21
Total length of irrelevant utterances (s)	4.154 (5.656)	4.825 (3.379)	234.500	-1.515	0.130	0.21
Average word transition time (s)	3.816 (1.739)	4.944 (3.045)	250.000	-1.213	0.225	0.17
Letter 'a'	$n = 25$	$n = 25$				
Total number of silent pauses (count)	13.920 (3.639)	14.120 (4.876)	298.000	-0.283	0.778	0.04
Average length of silent pauses (s)	3.636 (1.446)	3.853 (2.834)	268.000	-0.863	0.388	0.12
Total length of silent pauses (s)	45.881 (5.219)	43.042 (7.551)	235.00	-1.504	0.133	0.01
Total number of hesitations (count)	1.040 (1.059)	1.200 (1.354)	311.000	-0.031	0.976	0.00
Average length of hesitations (s)	0.640 (0.565)	0.462 (0.485)	263.500	-0.974	0.330	0.14
Total length of hesitations (s)	0.973 (1.316)	0.985 (1.254)	301.500	-0.219	0.827	0.03
Total number of irrelevant utterances (count)	3.480 (4.154)	4.560 (3.292)	214.500	-1.918	0.055	0.27
Average length of irrelevant utterances (s)	1.065 (0.701)	1.630 (0.725)	180.000	-2.572	**0.010**	0.36
Total length of irrelevant utterances (s)	4.637 (5.286)	7.160 (5.322)	204.000	-2.106	**0.035**	0.30
Average word transition time (s)	5.115 (2.651)	5.224 (2.839)	286.000	-0.514	0.607	0.07

Table 6. Descriptive measures and statistical comparison of the temporal parameters in the phonemic fluency tasks.
Significant p-values ($p < 0.05$) are in ***bold****.* **: one fluency voice recording was unsuitable for transcription.* ^{r}*: effect size is calculated as Pearson's r, expressed in absolute value. Strength of association: 0.1 to 0.3: small, 0.3 to 0.5: medium, 0.5 to 1.0: large* (Cohen, 1988). *M: mean; SD: standard deviation; HC: healthy control; MCI: mild cognitive impairment, s: second*

Semantic fluency tasks	HC	MCI				
Temporal parameters	***M* (*SD*)**		**Mann-Whitney *U* Test**			**Effect size^r**
Animals	*n* = 24*	*n* = 25	***U***	**Z**	***p***	**r**
Total number of silent pauses (count)	25.666 (4.603)	21.760 (4.968)	156.000	-2.890	**0.004**	0.41
Average length of silent pauses (s)	1.437 (0.445)	1.883 (0.718)	179.000	-2.420	**0.016**	0.34
Total length of silent pauses (s)	35.489 (6.485)	37.982 (8.193)	229.000	-1.420	0.156	0.20
Total number of hesitations (count)	3.166 (2.371)	3.240 (3.620)	271.000	-0.586	0.558	0.08
Average length of hesitations (s)	0.564 (0.290)	0.460 (0.358)	237.500	-1.255	0.209	0.18
Total length of hesitations (s)	2.195 (1.982)	2.139 (2.820)	264.500	-0.713	0.476	0.10
Total number of irrelevant utterances (count)	3.333 (3.595)	5.120 (4.850)	231.500	-1.380	0.167	0.20
Average length of irrelevant utterances (s)	1.019 (0.641)	1.146 (0.727)	277.000	-0.461	0.645	0.07
Total length of irrelevant utterances (s)	4.379 (6.116)	6.562 (5.647)	220.000	-1.603	0.109	0.23
Average word transition time (s)	2.021 (0.756)	2.852 (0.841)	128.000	-3.440	**0.001**	0.49
Food items	*n* = 25	*n* = 25				
Total number of silent pauses (count)	25.400 (6.062)	21.720 (5.926)	216.000	-1.877	0.061	0.26
Average length of silent pauses (s)	1.395 (0.504)	1.888 (0.937)	201.000	-2.163	**0.031**	0.30
Total length of silent pauses (s)	33.192 (6.464)	36.368 (7.200)	242.000	-1.368	0.171	0.19
Total number of hesitations (count)	2.600 (2.432)	2.600 (2.661)	307.000	-0.109	0.913	0.02
Average length of hesitations (s)	0.444 (0.348)	0.494 (0.435)	306.000	-0.128	0.898	0.02
Total length of hesitations (s)	1.636 (1.544)	1.855 (2.015)	302.000	-0.207	0.836	0.03
Total number of irrelevant utterances (count)	3.600 (3.905)	4.360 (4.733)	294.000	-0.362	0.717	0.05
Average length of irrelevant utterances (s)	0.772 (0.581)	1.051 (1.028)	273.000	-0.770	0.441	0.11
Total length of irrelevant utterances (s)	3.716 (4.898)	5.210 (5.353)	259.000	-1.044	0.297	0.15
Average word transition time (s)	1.755 (0.770)	2.630 (1.356)	171.000	-2.746	**0.006**	0.40
Actions	*n* = 25	*n* = 25				
Total number of silent pauses (count)	24.240 (6.332)	19.080 (5.597)	184.000	-2.502	**0.012**	0.35
Average length of silent pauses (s)	1.600 (0.565)	2.373 (1.439)	192.000	-2.338	**0.019**	0.33
Total length of silent pauses (s)	35.898 (5.605)	38.524 (7.485)	230.000	-1.601	0.109	0.22
Total number of hesitations (count)	2.720 (2.282)	2.840 (2.511)	309.000	-0.069	0.945	0.01
Average length of hesitations (s)	0.547 (0.362)	0.554 (0.477)	292.000	-0.401	0.689	0.06
Total length of hesitations (s)	1.963 (1.741)	2.096 (2.290)	302.000	-0.205	0.837	0.03
Total number of irrelevant utterances (count)	4.040 (3.920)	4.160 (3.681)	307.500	-0.098	0.922	0.01
Average length of irrelevant utterances (s)	1.069 (0.626)	1.153 (0.760)	290.500	-0.427	0.669	0.06
Total length of irrelevant utterances (s)	4.302 (4.600)	5.188 (4.351)	273.500	-0.757	0.449	0.11
Average word transition time (s)	2.258 (0.996)	2.989 (1.199)	196.000	-2.260	**0.024**	0.32

Table 7. Descriptive measures and statistical comparison of the temporal parameters in the semantic fluency tasks.

Significant p-values ($p < 0.05$) are in ***bold****. *: one fluency voice recording was unsuitable for transcription. ^r: effect size is calculated as Pearson's r, expressed in absolute value. Strength of association: 0.1 to 0.3: small, 0.3 to 0.5: medium, 0.5 to 1.0: large* (Cohen, 1988)*. M: mean; SD: standard deviation; HC: healthy control; MCI: mild cognitive impairment, s: second*

2.3. Traditional word count measures of verbal fluency performance

In the three PVF tasks, no statistically significant difference was found between the groups regarding the number of correct words and the number of repetitions or perseverations.

However, in the ‘a’ PVF task, participants from the MCI group produced more errors than participants from the HC group (*Table 8*).

	HC	MCI				
Traditional fluency scores of the phonemic fluency tasks	***M* (*SD*)**		**Mann-Whitney *U* Test**			**Effect size[r]**
	n = 25	*n* = 25	***U***	***Z***	***p***	***r***
Letter ’k’						
Correct words	13.68 (4.571)	11.52 (4.700)	227.000	-1.667	0.096	0.24
Errors	0.04 (0.200)	0.16 (0.374)	275.000	-1.400	0.162	0.20
Repetitions/perseverations	0.16 (0.374)	0.32 (0.690)	294.000	-0.537	0.591	0.08
Letter ’t’						
Correct words	12.88 (4.314)	10.76 (4.371)	233.000	-1.547	0.122	0.22
Errors	0.20 (0.408)	0.28 (0.614)	307.500	-0.139	0.889	0.02
Repetitions/perseverations	0.48 (0.653)	0.28 (0.678)	248.500	-1.577	0.115	0.22
Letter ’a’						
Correct words	8.68 (3.424)	7.32 (3.987)	240.000	-1.416	0.157	0.20
Errors	0.12 (0.332)	0.72 (1.208)	231.500	-2.106	**0.035**	0.30
Repetitions/perseverations	0.20 (0.577)	0.20 (0.408)	292.500	-0.609	0.542	0.09

Table 8. Descriptive measures and statistical comparison of the traditional fluency scores in the phonemic fluency tests.
Significant p-values (p < 0.05) are in ***bold****. *: one fluency voice recording was unsuitable for transcription. [r]: effect size calculated as Pearson’s r, expressed in absolute value. Strength of association: 0.1 to 0.3: small, 0.3 to 0.5: medium, 0.5 to 1.0: large (Cohen, 1988). M: mean; SD: standard deviation; HC: healthy control; MCI: mild cognitive impairment*

As for the SVF tests, participants from the HC group had a significantly higher number of correct words in the case of all three (animals, food items, and actions) tasks. In the number of repetitions or perseverations, there was no statistically significant difference between the two study groups (*Table 9*).

	HC	MCI				
Traditional fluency scores of the semantic fluency tasks	***M* (*SD*)**		**Mann-Whitney *U* Test**			**Effect size[r]**
	n = 25	*n* = 25	***U***	***Z***	***p***	***r***
Animals						
Correct words	20.54 (4.412)	14.76 (3.358)	99.000	-4.154	**<0.001**	0.59
Errors	0.00 (0.000)	0.04 (0.200)	300.000	-1.000	0.317	0.14
Repetitions/perseverations	0.42 (0.584)	0.48 (0.963)	298.000	-0.343	0.731	0.05
Food items						
Correct words	22.72 (6.073)	17.16 (5.249)	156.500	-3.034	**0.002**	0.43
Errors	0.04 (0.200)	0.04 (0.200)	312.500	0.000	1.000	0.00
Repetitions/perseverations	0.28 (0.458)	0.40 (0.764)	311.000	-0.038	0.970	0.01
Actions						
Correct words	18.72 (6.175)	14.40 (4.916)	194.500	-2.293	**0.022**	0.32
Errors	0.04 (0.200)	0.04 (0.200)	312.500	0.000	1.000	0.00
Repetitions/perseverations	0.40 (0.764)	0.48 (0.918)	308.500	-0.098	0.922	0.01

Table 9. Descriptive measures and statistical comparison of the traditional fluency scores in the semantic fluency tests.

*Significant p-values (p < 0.05) are in **bold**. [r]: effect size calculated as Pearson's r, expressed in absolute value. Strength of association: 0.1 to 0.3: small, 0.3 to 0.5: medium, 0.5 to 1.0: large (Cohen, 1988). M: mean; SD: standard deviation; HC: healthy control; MCI: mild cognitive impairment*

2.4. ROC analysis of the significant temporal parameters

ROC analysis of the temporal parameters was carried out in the case of the five parameters that showed significant differences between the HC and MCI groups based on the previously conducted comparative tests (*Table 10*). For every ROC analysis, sensitivity and specificity were determined using threshold values optimal for early screening, i.e., maximizing the sensitivity, while keeping specificity greater than or equal to 50%.

The analysis revealed that the average length and the total length of irrelevant utterances had a significant classification ability in the case of the 'a' PVF, with the same sensitivity (80%) and specificity (52%) for both parameters. In the SVF tests, the number of silent pauses had significant classification ability both in the animal and in the action fluency tests, while the average length of silent pauses and the average word transition time were shown to be able to discriminate between the groups in the case of all three SVF tests. Sensitivity was the highest in the case of the average word transition time in the animal fluency test (sensitivity: 96.0%; specificity: 62.5%).

Fluency tasks	Temporal parameters	Accuracy measures					
		p	**AUC**	**95% CI-**	**95% CI+**	**Sensitivity (%)**	**Specificity (%)**
Letter 'a'	Average length of irrelevant utterances (s)	**0.010**	0.712	0.569	0.855	80.0	52.0
	Total length of irrelevant utterances (s)	**0.035**	0.674	0.523	0.824	80.0	52.0
Animals	Total number of silent pauses (count)	**0.004**	0.740	0.598	0.882	76.0	50.0
	Average length of silent pauses (s)	**0.016**	0.702	0.549	0.855	72.0	50.0
	Average word transition time (s)	**0.001**	0.787	0.651	0.922	96.0	62.5
Food items	Average length of silent pauses (s)	**0.031**	0.678	0.528	0.828	68.0	52.0
	Average word transition time (s)	**0.006**	0.726	0.587	0.866	76.0	52.0
Actions	Total number of silent pauses (count)	**0.013**	0.706	0.562	0.849	72.0	52.0
	Average length of silent pauses (s)	**0.019**	0.693	0.544	0.841	72.0	52.0
	Average word transition time (s)	**0.024**	0.686	0.536	0.837	80.0	52.0

Table 10. Accuracy measures of those temporal parameters that significantly differed between the two groups based on the previous comparative statistic tests.
Significant p-values ($p < 0.05$) indicate that the measure is significantly better than chance at discriminating between individuals of the two groups. AUC: area under the curve; CI: confidence interval

2.5. ROC analysis of the significant traditional measures

ROC analysis was also executed on the traditional measures that showed significant differences between the HC and MCI groups, to determine the classification ability of these measures. The analysis revealed that the number of errors in the 'a' PVF test had no significant classification ability. Concerning the SVF tests, the number of correct words showed significant classification abilities in the case of the animal, the food item, and the action fluencies. The animal naming fluency showed the highest sensitivity of 100% (specificity: 56%). Accuracy measures of the traditional fluency scores that showed significant differences between the groups are given in *Table 11*.

		Accuracy measures					
Fluency tasks	**Traditional measures**	***p***	**AUC**	**95% CI-**	**95% CI+**	**Sensitivity (%)**	**Specificity (%)**
Letter 'a'	Number of errors	0.116	0.630	0.474	0.785	36.0	88.0
Animals	Number of correct words	**< 0.001**	0.842	0.734	0.949	100.0	56.0
Food items	Number of correct words	**0.002**	0.750	0.616	0.884	76.0	64.0
Actions	Number of correct words	**0.022**	0.689	0.543	0.834	68.0	52.0

Table 11. Accuracy measures of those traditional fluency measures that significantly differed between the two groups based on the previous comparative statistic tests. *Significant p-values ($p < 0.05$) indicate that the measure is significantly better than chance at discriminating between individuals of the two groups. AUC: area under the curve; CI: confidence interval*

2.6. Comparison classification abilities

Pairwise comparisons of AUCs were executed to compare the classification ability of the three significant temporal parameters (total number of silent pauses, average length of silent pauses, average word transition time) and the significant traditional measure (number of correct words) regarding the SVF tasks. In the animal category fluency, the results indicated no significant differences regarding AUCs between the number of correct words and the total number of silent pauses ($z = 1.433$, $p = 0.151$) or the average word transition time ($z = 1.579$, $p = 0.114$), however, the classification ability of the average length of silent pauses was smaller ($z = 2.043$, $p = 0.041$) compared to the correct word count. In the case of the food item fluency, no difference was found between the AUCs of the number of correct words and the average length of silent pauses ($z = 0.978$, $p = 0.328$), and the average word transition time ($z = 0.662$, $p = 0.508$). Furthermore, in action fluency, the classification ability of correct word-count did not differ from either the total number of silent pauses ($z = 0.267$, $p = 0.789$), the average length of silent pauses ($z = 0.056$, $p = 0.954$) or the average word transition time ($z = 0.046$, $p = 0.962$).

VII. DISCUSSION

1. Study 1

1.1. Main findings

Based on the results of the first study in this thesis, Hungarian GPs are aware of the low dementia recognition rates and they accept the idea of cognitive evaluation for signs of possible dementia in primary care practices. Furthermore, they know the most commonly used cognitive screening tests, and more than two-thirds of them are satisfied with them. However, only a quarter of them use standardized cognitive tests in their practices, while the vast majority apply general questions to the patient or the informant. What may be driving only a smaller proportion of them to apply effective dementia testing (i.e., standardized testing) is probably their doubts about effective anti-dementia therapies. In addition to this, the most important barriers to effective dementia case-finding appear to be insufficient conditions: mainly lack of time and assistance, as well as cost-effective instruments.

1.2. Interpretation of the results and clinical implications

Our results revealed a discrepancy between GPs' overall views regarding testing for dementia versus their actual habits. Even though GPs seem to agree with the benefits of early recognition and know the available cognitive tools, only a quarter of them apply these tests for the purpose of dementia detection, while a few do not perform any examinations at all. A similar conflict was found regarding Dutch GPs' views and habits, who reported taking action at a more progressed stage of dementia, despite knowing the importance of early intervention (van Hout et al., 2000). The rare application of formal tests has been also observed in other European countries: many GPs (85% of French, 79% of Swiss, 53% of Italian, and 33% of Scottish) reported that they did not regularly perform standard procedures in their diagnostic evaluation (Giezendanner et al., 2019; McIntosh et al., 1999; Somme et al., 2013; Veneziani et al., 2016), with some preferring the use of non-standardized, general questions (Somme et al., 2013). Even when cognitive concerns are present, only 50% of American GPs (or their staff) administer cognitive screening (Bernstein et al., 2019). There are, however, some exceptions: 92% of Irish GPs self-reported in a survey that they used an appropriate tool to evaluate their patients' cognition (Dyer et al., 2017) and only 10% of German GPs did not use any screening instrument (Thyrian & Hoffmann, 2012). It is important to note, nevertheless, that not only the willingness to test and the utilization of standardized cognitive tools but also the chosen tool can have a major impact on successful case-finding. Presuming that we disregard the low rates of test usage, part of the reason for the under-diagnosis of dementia may be the widespread use of tests that are not sensitive enough to detect the early stages of the condition (Breton et al., 2019).

As only a fraction of Hungarian GPs seem to trust the effectiveness of the anti-dementia medication, the often insufficient habits regarding cognitive screening may reflect their therapeutic nihilism. This supposition is supported by previous studies: for example, in a research conducted in France, around 50% of the participating GPs felt that it was not worth making a dementia diagnosis because of the ineffective pharmacological treatment (Harmand et al., 2018). In another study, 33% of the surveyed Swiss GPs believed that a dementia diagnosis is not "clinically actionable", as only 20% of them had the opinion that anti-dementia drugs positively influence the course of the disease (Giezendanner et al., 2019). Probably related to this, helplessness is a common feeling GPs experience when working with dementia patients (Heim et al., 2019).

The findings of our study indicated that the main obstacle to testing for dementia might be short consultation time with patients (which is on average 6 minutes in Hungary) (Irving et al., 2017). Besides the shortage of time, GPs mentioned the need for quickly administrable cognitive tools and more help from healthcare staff. These concerns are reflected by previous studies from several different countries as well, such as Germany, the Netherlands, Spain, Sweden, Slovenia, and the UK (Chithiramohan et al., 2019; Koch & Iliffe, 2010; Prins et al., 2016; Sannemann et al., 2021). Despite the time restrictions, current views of scientific literature advocate for the integration of a targeted case-finding approach into primary care, prompting early dementia identification (Ranson et al., 2018).

In addition to the circumstances described above, GPs' knowledge and awareness of dementia and MCI also have a significant impact on their willingness to conduct cognitive testing, as well as to comply with practical guidelines (de Levante Raphael, 2022; Lu et al., 2022).

The underutilization of validated cognitive tests might be partly due to the lack of agreements on the most effective ways of dementia recognition, leaving the GPs without an unambiguous source of reference. A crucial way to improve recognition rates would be the regular update of international and national dementia guidelines (e.g., the latest Hungarian version was released in June 2022, and will be in effect until May 2025), which usually give suggestions on the most adequate testing methods for dementia recognition.

The underuse of standardized instruments and the underdiagnosis of dementia in primary care may also be attributed to the prioritization of somatic diseases over cognitive problems among the elderly. Since more than 65% of people over the age of 65 have multiple chronic conditions (Lehnert et al., 2011), the examination of memory functions might end up at the bottom of the priority list (Boise et al., 1999). Furthermore, the progression of dementia is a slow process and thus is less obvious than the sudden onset of a somatic, sometimes painful symptom requiring urgent examination. In relation to this, a substantial part of GPs may even not consider MCI as a disease (Lu et al., 2022). Healthcare providers might also be reluctant to screen for dementia because they feel that a person with such a diagnosis will place an additional burden on an already overburdened health system (de Levante Raphael, 2022).

Cultural differences in attitudes towards age-related memory problems may also affect the success of dementia detection. In Hungary, dementia symptoms (especially in the earlier stages) are often overlooked and thus do not prompt taking steps toward recognition. The tolerance for cognitive decline associated with older age may be higher in Hungary compared to other countries, where elderly people live far from their families and lead a more independent life (e.g., the USA) which would be greatly endangered by a mental illness.

Although the present research explored the issue from the GPs' perspective, it is important to note that the attitudes of the patient and the family can also influence by hindering or delaying the diagnostic process. Many of those influential factors are similar to those found in the case of GPs, such as skepticism about the effectiveness of the therapy, or the belief that cognitive decline is a natural part of aging. From the patients' standpoint, fear of the dementia diagnosis or lack of communication (for example, not mentioning their cognitive complaints during an examination) is also a significant factor. For caregivers and family, factors also include the fear of stigmatization or concerns about the diagnosis' impact on the patient's autonomy (Bradford et al., 2009).

1.3. Strengths and limitations

Apart from an international study with limited sample sizes (Petrazzuoli et al., 2017), no extensive research has been conducted on GPs' routines and views regarding dementia, neither in Hungary nor in most Central and Eastern European countries. As part of a national research project, this study sheds light on the dementia screening practices of Hungarian GPs. The results provide a valuable benchmark for future dementia-related research conducted in Hungary: in particular, they can be used as a basis for comparison in studies measuring the impact of the recently published dementia guidelines. The findings may also facilitate the undertaking of comparative studies among different countries, supporting the development of effective national dementia strategies. An additional value of the study is its sample since it describes results drawn from a relatively large proportion of the population of interest (8% of all practicing GPs in Hungary in 2014), with participants from all regions of the country.

However, when interpreting the results, some limitations should be considered. Firstly, as our findings were based on the answers of self-recruited, voluntary participants, the results might represent the views and routines of a more motivated and competent sample of Hungarian GPs. Secondly, given the somewhat sensitive topic of dementia detection practices, the probable effect of social desirability bias should also be taken into account. Thirdly, since a pen-and-paper questionnaire was applied, it could not be ensured that each question was answered. It resulted in different numbers of missing responses throughout the survey, which consequently limited the validity of questions with fewer responses. Regarding future works, it would be useful to recruit a representative sample of Hungarian GPs and to apply qualitative methods to further deepen our understanding of the topic.

2. Study 2

2.1. Main findings

The second study of this thesis presented a new practical framework for verbal fluency analysis. To the best of our knowledge, we were the first to report on verbal fluency performance beyond the recalled words, focusing on the pauses and task-irrelevant content of speech in the fluency recordings. In the study, we quantitatively analyzed several temporal parameters that were calculated based on silent pauses, hesitations, and irrelevant speech segments annotated in the recordings. The main finding was that in the case of the three SVF tests, some of the temporal parameters (total number of silent pauses, average length of silent pauses, and average word transition time) could discriminate between individuals with cognitive impairment and individuals with healthy cognition. However, no other temporal measure differed systematically between the two groups.

These results suggest that when examining SVF tests, the analysis of the temporal parameters based on silent pauses may complement or even substitute the widely applied, but more time-consuming and labor-intensive traditional word-scoring method, while still providing comparable classification ability.

2.2. Interpretation of the results

2.2.1. Diagnostic value of the temporal parameters

The inconsistency in the direction of differences regarding the silence-based parameters between the two groups is worth highlighting: the average lengths of the silent pauses and the average word transition times were longer in the MCI group, whereas HC participants had a higher number of silent pauses in the case of the SVF tasks. Since silent pauses were defined as the absence of speech/sound regardless of length, every detectable silent segment found in the recordings was annotated as a silent pause, including even the brief transitions between words. Therefore, the number of silent pauses was increased by the number of words uttered by the participant. Since the HC group produced significantly more correct words in the SVF tasks, the number of silent pauses was also significantly increased in this group. The average word transition time parameter also had a direct link with the number of correct words. Since this parameter contains every speech segment except task-relevant words, the increase in the average word transition time led, by definition, to a decrease in the number of recalled words. Therefore, it could be considered that these two parameters were somewhat inversely proportional.

In their study, applying automated analysis of PVF tests, Cho and colleagues measured inter-word response time, which is a similar concept to our average word transition time variable; however, besides partial words and non-verbal vocalizations they also included the filled pauses in this measure. Similar to our outcomes, their results showed that there was a strong negative correlation between mean inter-word response time and the number of correct words recalled by the participants, indicating that better-performing subjects were able to retrieve words more quickly from their mental lexicon (Cho et al., 2021).

The importance of silent pauses has also been highlighted in the area of connected (spontaneous) speech analysis: studies have shown that compared to HC subjects, participants with MCI produce more and longer silent pauses in their speech (Sluis et al., 2020; Toth et al., 2018). Even though spontaneous speech samples provide ecologically valid data, utilizing verbal fluency tests may be even more advantageous, as they can be combined with already standardized qualitative and quantitative approaches. The regulated verbal fluency tasks also have the benefit of shorter administration time compared to spontaneous speech tasks, as in the latter the task duration time can greatly vary (e.g., Imre et al., 2022). To be able to compare the results of the two, it is important to note the difference between spontaneous speech and verbal fluency performances. Compared to the connected speech samples where pauses appear more unevenly and randomly, silent pauses (with varying lengths) appear between every word in the verbal fluency recordings, therefore producing a "word-pause-word-pause"-like sequence. Because of these distinct characteristics, the number of silent pauses needs to be interpreted based on the methodology of the specific study.

Most of the recent approaches to verbal fluency analysis focus on the semantic content when evaluating fluency performance (Tröger et al., 2019; Woods et al., 2016). In contrast, our work focused on the examination of more easily quantifiable, objective variables. Nevertheless, we were able to achieve classification abilities comparable to those reported in previous studies (e.g., AUC: 0.758 (König et al., 2018), AUC: 0.77 (L. Chen et al., 2020)). The significant classification ability of the silent pause parameters in our study suggests that differentiation between HC and MCI patients' SVF performance may be possible by examining only the silent pauses in their speech. This can be achieved simply by dividing the voice recordings into voiced and unvoiced segments (Lopez-de-Ipina et al., 2015). This method would not require additional time-consuming steps, such as the manual transcription and preparation of the answers, nor their identification as correct words, errors, or repetitions, as opposed to the majority of fluency analysis techniques. This could make the analysis procedure considerably faster and easier. However, since this does not provide any semantic information, it can be viewed for instance as an alternative, inverse approach to the traditional analyses based on word count: instead of considering the number of recalled correct words, it focuses on the silent pauses between the words. Focusing on silent segments and paralinguistic phenomena has the advantage of making the processing of verbal fluency tasks swifter and more undemanding, as variables based on them do not need to be adapted to different languages and cultures.

2.2.2. The role of semantic networks in the detection of MCI

Our results confirmed the advantage of SVF over PVF in the detection of minor cognitive impairment. In all three SVF tests (animal, food item, and action), the same three temporal parameters (number of silent pauses, average length of silent pauses, average word transition time), and one of the traditional measures (correct word count) showed a difference between the two groups. In contrast, regarding the PVF tests, differences were only observed in the case of the 'a' PVF, where two temporal parameters (the average and total length of irrelevant utterances) and one of the traditional measures (incorrect words) showed significant differences. These results are consistent with those of earlier studies, confirming that SVF tasks may be more appropriate for detecting the cognitive changes that occur in MCI (McDonnell et al., 2020; Nikolai et al., 2018). Furthermore, when compared to other SVF test categories (plants, clothes, vehicles), the animal fluency test has shown the highest sensitivity (98.8%) in discriminating between HC and MCI participants (García-Herranz et al., 2020). In agreement with the results of García-Herranz and colleagues, animal fluency achieved the highest accuracy scores in the present study as well, not only with the traditional scoring method but also when examining the temporal parameters.

The difference between the SVF and PVF tasks regarding their sensitivity to the presence of cognitive decline may be due to the different cognitive demands that the two task types require. It is known, that PVF tasks are in general more difficult to perform since besides the more restrictive rules of the tasks, they require the search of lexical representation. In comparison, the performance of SVF relies to a greater extent on semantic associations, and since the necessary words are already organized in semantic clusters, the recalling of items is facilitated (Chasles et al., 2020; Murphy et al., 2006). According to most psycholinguistic models, human word representations are based on semantic associations, thus, performing SVF resembles ordinary language use (Luo et al., 2010). The utility of SVF tasks is therefore further underpinned by numerous studies concluding that in MCI, the earliest symptoms regarding speech impairment often manifest in everyday language use (Mueller et al., 2018).

The fact that MCI participants were not able to perform at the level of HC participants (as measured by the number of correct words) in either of the semantic fluency tasks probably reflects the semantic deficits that have been observed in MCI patients (Chang et al., 2022; Taler et al., 2020), and which is caused by the degradation of semantic networks. This impairment might indicate that the neuropathology of MCI extends beyond the hippocampal region affecting the cortical areas related to semantic memory. Based on this performance pattern, authors of a cross-sectional study concluded that executive functions may be a key in differentiating MCI patients from cognitively intact individuals (Pakzad et al., 2018).

2.3. Implementation for future research

Based on the preliminary outcomes of this novel verbal fluency analysis, further projects should be focused on the collection of more and higher quality data in order to define precise reference values for the quantity of silent pauses associated with MCI. In the future, this could allow for the development of an automated tool for MCI screening, based on the analysis of temporal speech parameters. In addition, it remains to be determined whether combining this method of temporal parameter-analysis with automated cluster analysis or the analysis of acoustic features could provide additional value for classification.

Due to the recent pandemic, the frequent application of telemedicine, including telepsychiatry, rose sharply. Using remote methods for examination and outpatient care has become more and more widespread, and even though conflicting results exist regarding the topic, telemedicine practice generally shows a good acceptance among the patients. Telemedicine has the advantage of being able to produce diagnostic and treatment results comparable to face-to-face consultations, which is why it may be worthwhile to further develop and extend the use of remote application-based, computerized tests (Hubley et al., 2016; Munro Cullum et al., 2014). A significant advantage of fluency tests in this regard is that they can be self-administered, and, if the appropriate tools are available, recorded and evaluated remotely. In a recently published study, Kwon et colleagues (2021) introduced a smartphone application for a semi-automatic, self-administrated version of the SVF test. Even though the diagnostic accuracy of the software was lower than the one of MMSE, the method still delivered acceptable accuracy, sensitivity, and specificity regarding the early screening of AD (Kwon et al., 2021). Even if not applied remotely, given their short and easy administration and the quick evaluation time, computerized fluency tests would also meet the needs expressed by GPs in our study.

2.4. Strengths and limitations

The main strength of this study is that it presents a novel practical framework: to the best of our knowledge, this is the first research reporting on verbal fluency performance with the focus on previously unexploited information present in the fluency recordings. The theoretical and methodological foundation established in this study can provide a strong basis for future research in this field.

Since a total of six different fluency tasks were used in the study protocol, we were able to compare not only the semantic and phonemic task types, but also subtypes of these tasks. Furthermore, the application of ROC-analysis enabled us to directly compare our potential diagnostic variables to those of other studies, despite the different methodologies. The results of this study encourage further investigation on the utilization of different silence-based parameters in verbal fluency tasks.

Nevertheless, this study has limitations, of which the significant age difference between the two study groups is the most relevant as it has been shown that age has a significant influence on verbal fluency abilities (Kempler et al., 1998; Rodriguez-Aranda & Martinussen, 2006). The results are, however, inconsistent regarding the matter: some studies conducted among the elderly suggest that the level of education has a greater impact on their verbal fluency performance (Esteves et al., 2015; Lubrini et al., 2022; Mathuranath et al., 2003). Either way, we cannot rule out the possibility that the age of the participants might have affected their verbal fluency performance regardless of their cognitive state. Nevertheless, this sample would closely represent the affected population in case of a potential real-life application, especially since elderly age itself is a primary risk factor for MCI.

In the study, the MCI group was not subdivided along the subtypes of MCI (i.e., aMCI, naMCI). Since verbal fluency performance varies across these subtypes (Teng et al., 2013; Weakley et al., 2013), it would be worthwhile to design a study investigating the effect of the MCI subtype on the temporal parameters. In relation to this, the etiology of MCI should also be considered in future research.

To follow the clinic's protocol, and to ensure that our outcomes can be comparable with the previous results of our research team, we relied on the MMSE to assess the cognitive status of the participants. However, according to some reviews, MMSE is not optimal for the detection of MCI, as it has a relatively low sensitivity for the impairment of frontal executive functions (Kim et al., 2014). Even though not designed to accurately detect MCI, MMSE can be used to distinguish cognitively intact and impaired subjects, and as this, it was sufficient to assist our aims.

When interpreting the results of the study, it is also important to take into consideration that, due to its exploratory nature, corrections for multiple comparisons were not applied during the statistical analysis. As one of the main goals of the study was to investigate and identify all temporal fluency parameters that can differentiate between the groups, confirmatory studies are required to further confirm to the discriminatory ability and clinical utility of the significant temporal parameters.

VIII. CONCLUSIONS

The present thesis showed that Hungarian GPs are aware of the benefits of early dementia detection and of the concurrently low recognition rates as well. However, the majority of them do not use formal cognitive tests for dementia case-finding. In accordance with the goals of international dementia plans, improving the rates of detected dementia in Hungarian primary care should be a major objective. The most effective way to achieve this is to lend sufficient support to GPs: besides providing more favorable conditions and emphasizing the benefits of early treatment, it would be also important to construct remote or automated screening tools applicable in GPs' practices.

Based on the assessed needs of GPs, we introduced an alternative method of verbal fluency analysis. This study was the first in the literature to examine verbal fluency performance beyond the retrieved words. The results revealed the discriminatory ability of the silent pause parameters in the case of SVF tests. Since silence-related parameters could be relatively easily extracted from fluency voice recordings, this approach shows promising potential; however, the calibration and validation of the method remain to be solved. Building on the results, the next goal would be to construct an automated instrument capable of identifying cognitively impaired patients based on their speech/silence ratio.

The novel findings presented in this thesis are the following:

I. Although Hungarian GPs are aware of the benefits of dementia detection and the concurrently low recognition rate in primary care, and most of them reported satisfaction with the available cognitive tools, only a minority uses standardized cognitive tests for dementia screening.

II. Time pressure and the lack of cost- and time-effective instruments seem to be the most significant barriers to effective dementia screening in primary care.

III. When analyzing SVF tasks, silence-related temporal parameters can differentiate between the HC and MCI groups; however, hesitations and irrelevant utterances showed no systematic differences between the groups.

IV. The discriminatory ability of some silence-related temporal parameters is comparable to the most traditionally applied count-based measures.

V. In line with previous studies, our results confirmed the superiority of SVF over PVF tests regarding the differentiation between the HC and MCI groups, highlighting the important role of the semantic networks in the detection of MCI.

The development of remote and computerized tools is especially important, seeing the growing necessity of telemedicine-based health consultations. Considering the globally heavy burden on the healthcare systems, an automated and cost-effective telemedical tool would be a valuable addition to any primary care or clinical practices, and would likely improve the detection rates of MCI.

IX. ACKNOWLEDGEMENTS

Throughout the years of my doctoral studies and during the writing of this dissertation I have received a lot of support from several people who, in one way or another, contributed and extended their valuable assistance. I hereby would like to take the opportunity to thank them all.

First of all, I wish to express my gratitude to my supervisor, Prof. Dr. János Kálmán, for allowing me to conduct my research at the Department of Psychiatry, University of Szeged, as well as for his guidance and support throughout the process. I am especially grateful to my supervisor, Dr. Gábor Gosztolya, for the many phone consultations, and his valuable contribution and constructive suggestions on my manuscripts and the dissertation. I would like to thank Dr. Ildikó Hoffmann and Dr. Magdolna Pákáski for their help and professional advice during my work. I would like to extend my sincere thanks to all my co-authors and I would also like to thank all those who participated in the research.

I am very grateful for the friendship of Nóra Imre, who has stood (sat) beside me for the past ten years, and with whom I have been able to share all the joys and sorrows of the life of a Ph.D. student. I would like to express my deepest gratitude to all my friends who have supported me from near and far, offline and online, who, depending on the stage of this journey, inspired me, consoled me, or celebrated with me.

Last but not least, I would like to thank my family for the endless love, support, encouragement, and patience they have given me, not only in the past years but throughout my entire life.

X. REFERENCES

Abd Razak, M. A., Ahmad, N. A., Chan, Y. Y., Mohamad Kasim, N., Yusof, M., Abdul Ghani, M. K. A., Omar, M., Abd Aziz, F. A., & Jamaluddin, R. (2019). Validity of screening tools for dementia and mild cognitive impairment among the elderly in primary health care: a systematic review. *Public Health*, *169*, 84–92. https://doi.org/10.1016/j.puhe.2019.01.001

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., Gamst, A., Holtzman, D. M., Jagust, W. J., Petersen, R. C., Snyder, P. J., Carrillo, M. C., Thies, B., & Phelps, C. H. (2011). The diagnosis of mild cognitive impairment 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), 270–279. https://doi.org/10.1016/j.jalz.2011.03.008

Alzheimer Europe. (2019). *Dementia in Europe Yearbook 2019: Estimating the prevalence of dementia in Europe*.

Alzheimer's Association. (2013). *Early Onset Dementia: A National Challenge, a Future Crisis.*

Alzheimer's Association. (2018). 2018 Alzheimer's Disease Facts and Figures. *Alzheimer's & Dementia*, 367–429.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. American Psychiatric Association.

Anchisi, D., Borroni, B., Franceschi, M., Kerrouche, N., Kalbe, E., Beuthien-Beumann, B., Cappa, S., Lenz, O., Ludecke, S., Marcone, A., Mielke, R., Ortelli, P., Padovani, A., Pelati, O., Pupi, A., Scarpini, E., Weisenbach, S., Herholz, K., Salmon, E., … Perani, D. (2005). Heterogeneity of Brain Glucose Metabolism in Mild Cognitive Impairment and Clinical Progression to Alzheimer Disease. *Archives of Neurology*, *62*(11), 1728. https://doi.org/10.1001/archneur.62.11.1728

Assal, F. (2019). *History of Dementia* (pp. 118–126). https://doi.org/10.1159/000494959

Ausó, E., Gómez-Vicente, V., & Esquiva, G. (2020). Biomarkers for Alzheimer's Disease Early Diagnosis. *Journal of Personalized Medicine*, *10*(3), 114. https://doi.org/10.3390/jpm10030114

Barnes, D. E., & Yaffe, K. (2011). The projected effect of risk factor reduction on Alzheimer's disease prevalence. *The Lancet Neurology*, *10*(9), 819–828. https://doi.org/10.1016/S1474-4422(11)70072-2

Bennett, D. A., Wilson, R. S., Schneider, J. A., Evans, D. A., Beckett, L. A., Aggarwal, N. T., Barnes, L. L., Fox, J. H., & Bach, J. (2002). Natural history of mild cognitive impairment in older persons. *Neurology*, *59*(2), 198–205. https://doi.org/10.1212/WNL.59.2.198

Bermejo-Pareja, F., Contador, I., del Ser, T., Olazarán, J., Llamas-Velasco, S., Vega, S., & Benito-León, J. (2021). Predementia constructs: Mild cognitive impairment or mild neurocognitive disorder? A narrative review. *International Journal of Geriatric Psychiatry*, *36*(5), 743–755. https://doi.org/10.1002/gps.5474

Bernstein, A., Rogers, K. M., Possin, K. L., Steele, N. Z. R., Ritchie, C. S., Kramer, J. H., Geschwind, M., Higgins, J. J., Wohlgemuth, J., Pesano, R., Miller, B. L., & Rankin, K. P. (2019). Dementia assessment and management in primary care settings: a survey of current provider practices in the United States. *BMC Health Services Research*, *19*(1), 919. https://doi.org/10.1186/s12913-019-4603-2

Boersma, P., & Weenink, D. (2020). *Praat: doing phonetics by computer [Computer program]*.

Boise, L., Camicioli, R., Morgan, D. L., Rose, J. H., & Congleton, L. (1999). Diagnosing dementia: perspectives of primary care physicians. *The Gerontologist*, *39*(4), 457–464. https://doi.org/10.1093/GERONT/39.4.457

Boustani, M., Peterson, B., Hanson, L., Harris, R., & Lohr, K. N. (2003). Screening for dementia in primary care: a summary of the evidence for the U.S. Preventive Services Task Force. *Annals of Internal Medicine*, *138*(11). https://doi.org/10.7326/0003-4819-138-11-200306030-00015

Bradford, A., Kunik, M. E., Schulz, P., Williams, S. P., & Singh, H. (2009). Missed and Delayed Diagnosis of Dementia in Primary Care. *Alzheimer Disease & Associated Disorders*, *23*(4), 306–314. https://doi.org/10.1097/WAD.0b013e3181a6bebc

Breton, A., Casey, D., & Arnaoutoglou, N. A. (2019). Cognitive tests for the detection of mild cognitive impairment (MCI), the prodromal stage of dementia: Meta-analysis of diagnostic accuracy studies. *International Journal of Geriatric Psychiatry*, *34*(2), 233–242. https://doi.org/10.1002/gps.5016

Brodaty, H., Aerts, L., Crawford, J. D., Heffernan, M., Kochan, N. A., Reppermund, S., Kang, K., Maston, K., Draper, B., Trollor, J. N., & Sachdev, P. S. (2017).

Operationalizing the Diagnostic Criteria for Mild Cognitive Impairment: The Salience of Objective Measures in Predicting Incident Dementia. *The American Journal of Geriatric Psychiatry*, *25*(5), 485–497. https://doi.org/10.1016/j.jagp.2016.12.012

Cappa, S. F., Sandrini, M., Rossini, P. M., Sosta, K., & Miniussi, C. (2002). The role of the left frontal lobe in action naming: rTMS evidence. *Neurology*, *59*(5), 720–723. https://doi.org/10.1212/WNL.59.5.720

Chang, H.-T., Chiu, M.-J., Chen, T.-F., Liu, M.-Y., Fan, W.-C., Cheng, T.-W., Lai, Y.-M., & Hua, M.-S. (2022). Deterioration and predictive values of semantic networks in mild cognitive impairment. *Journal of Neurolinguistics*, *61*, 101025. https://doi.org/10.1016/j.jneuroling.2021.101025

Chasles, M.-J., Tremblay, A., Escudier, F., Lajeunesse, A., Benoit, S., Langlois, R., Joubert, S., & Rouleau, I. (2020). An Examination of Semantic Impairment in Amnestic MCI and AD: What Can We Learn From Verbal Fluency? *Archives of Clinical Neuropsychology*, *35*(1), 22–30. https://doi.org/10.1093/arclin/acz018

Chen, L., Asgari, M., Gale, R., Wild, K., Dodge, H., & Kaye, J. (2020). Improving the assessment of mild cognitive impairment in advanced age with a novel multi-feature automated speech and language analysis of verbal fluency. *Frontiers in Psychology, 11*, 535. https://doi.org/10.3389/fpsyg.2020.00535

Chen, Y.-X., Liang, N., Li, X.-L., Yang, S.-H., Wang, Y.-P., & Shi, N.-N. (2021). Diagnosis and Treatment for Mild Cognitive Impairment: A Systematic Review of Clinical Practice Guidelines and Consensus Statements. *Frontiers in Neurology*, *12*. https://doi.org/10.3389/fneur.2021.719849

Chithiramohan, A., Iliffe, S., & Khattak, I. (2019). Identifying barriers to diagnosing dementia following incentivisation and policy pressures: General practitioners' perspectives. *Dementia (London, England), 18*(2), 514–529. https://doi.org/10.1177/1471301216682625

Cho, S., Nevler, N., Parjane, N., Cieri, C., Liberman, M., Grossman, M., & Cousins, K. A. Q. (2021). Automated Analysis of Digitized Letter Fluency Data. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.654214

Chun, C. T., Seward, K., Patterson, A., Melton, A., & MacDonald-Wicks, L. (2021). Evaluation of Available Cognitive Tools Used to Measure Mild Cognitive Decline: A Scoping Review. *Nutrients*, *13*(11), 3974. https://doi.org/10.3390/nu13113974

Ciesielska, N., Sokołowski, R., Mazur, E., Podhorecka, M., Polak-Szabela, A., & Kędziora-Kornatowska, K. (2016). Is the Montreal Cognitive Assessment (MoCA) test better suited than the Mini-Mental State Examination (MMSE) in mild cognitive impairment (MCI) detection among people aged over 60? Meta-analysis. *Psychiatria Polska*, *50*(5), 1039–1052. https://doi.org/10.12740/PP/45368

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). L. Erlbaum Associates.

Crowe, S. F. (1998). Decrease in Performance on the Verbal Fluency Test as a Function of Time: Evaluation in a Young Healthy Sample. *Journal of Clinical and Experimental Neuropsychology*, *20*(3), 391–401. https://doi.org/10.1076/jcen.20.3.391.810

Cuetos, F., Arango-Lasprilla, J. C., Uribe, C., Valencia, C., & Lopera, F. (2007). Linguistic changes in verbal expression: A preclinical marker of Alzheimer's disease. *Journal of the International Neuropsychological Society*, *13*(03). https://doi.org/10.1017/S1355617707070609

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

Davis, C., Heidler-Gary, J., Gottesman, R. F., Crinion, J., Newhart, M., Moghekar, A., Soloman, D., Rigamonti, D., Cloutman, L., & Hillis, A. E. (2010). Action versus animal naming fluency in subcortical dementia, frontal dementias, and Alzheimer's disease. *Neurocase*, *16*(3), 259–266. https://doi.org/10.1080/13554790903456183

de Levante Raphael, D. (2022). The Knowledge and Attitudes of Primary Care and the Barriers to Early Detection and Diagnosis of Alzheimer's Disease. *Medicina*, *58*(7), 906. https://doi.org/10.3390/medicina58070906

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1998). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, *44*(3), 837–845.

Dementia Australia. (2020). *Mild cognitive impairment [Web page] Retrieved from: https://www.dementia.org.au/sites/default/files/helpsheets/Helpsheet-DementiaQandA13-MildCognitiveImpairment_english.pdf.*

Demetriou, E., & Holtzer, R. (2017). Mild Cognitive Impairments Moderate the Effect of Time on Verbal Fluency Performance. *Journal of the International Neuropsychological Society*, *23*(1), 44–55. https://doi.org/10.1017/S1355617716000825

Dhedhi, S. A., Swinglehurst, D., & Russell, J. (2014). 'Timely' diagnosis of dementia: what does it mean? A narrative analysis of GPs' accounts. *BMJ Open*, *4*(3), e004439. https://doi.org/10.1136/bmjopen-2013-004439

Driscoll, I., Davatzikos, C., An, Y., Wu, X., Shen, D., Kraut, M., & Resnick, S. M. (2009). Longitudinal pattern of regional brain volume change differentiates normal aging from MCI. *Neurology*, *72*(22), 1906–1913. https://doi.org/10.1212/WNL.0b013e3181a82634

Dyer, A., Foley, T., O'Shea, B., & Kennelly, S. (2017). Dementia Diagnosis and Referral in General Practice: A Representative Survey of Irish GPs. *Age and Ageing*, *46*(Suppl_3), iii13–iii59. https://doi.org/10.1093/AGEING/AFX144.57

Edmonds, E. C., Delano-Wood, L., Galasko, D. R., Salmon, D. P., & Bondi, M. W. (2014). Subjective Cognitive Complaints Contribute to Misdiagnosis of Mild Cognitive Impairment. *Journal of the International Neuropsychological Society*, *20*(8), 836–847. https://doi.org/10.1017/S135561771400068X

Egészségügyi Szakmai Kollégium. (2020). Egészségügyi szakmai irányelv: A demencia kórismézése, kezelése és gondozása. *Egészségügyi Közlöny*.

Ersek, K., Kárpáti, K., Kovács, T., Csillik, G., Gulácsi, A. L., & Gulácsi, L. (2010). [Epidemiology of dementia in Hungary]. *Ideggyogyaszati Szemle*, *63*(5–6), 175–182.

Esteves, C. S., Oliveira, C. R., Moret-Tatay, C., Navarro-Pardo, E., Carli, G. A. de, Silva, I. G., Irigaray, T. Q., & Argimon, I. I. de L. (2015). Phonemic and semantic verbal fluency tasks: normative data for elderly Brazilians. *Psicologia: Reflexão e Crítica*, *28*(2), 350–355. https://doi.org/10.1590/1678-7153.201528215

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fernaeus, S.-E., & Almkvist, O. (1998). Word Production: Dissociation of Two Retrieval Modes of Semantic Memory Across Time. *Journal of Clinical and Experimental Neuropsychology*, *20*(2), 137–143. https://doi.org/10.1076/jcen.20.2.137.1170

Fernandes, B., Goodarzi, Z., & Holroyd-Leduc, J. (2021). Optimizing the diagnosis and management of dementia within primary care: a systematic review of systematic reviews. *BMC Family Practice*, *22*(1), 166. https://doi.org/10.1186/s12875-021-01461-5

Folstein, M. F., Folstein, S. E., & Mchugh, 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. https://doi.org/10.1016/0022-3956(75)90026-6

García-Herranz, S., Díaz-Mardomingo, M. C., Venero, C., & Peraita, H. (2020). Accuracy of verbal fluency tests in the discrimination of mild cognitive impairment and probable Alzheimer's disease in older Spanish monolingual individuals. *Aging, Neuropsychology, and Cognition*, *27*(6), 826–840. https://doi.org/10.1080/13825585.2019.1698710

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory*, *2*(2), 103–127. https://doi.org/10.1080/09658219408258940

Geda, Y. E., & Nedelska, Z. (2012). Mild Cognitive Impairment: A Subset of Minor Neurocognitive Disorder? *The American Journal of Geriatric Psychiatry*, *20*(10), 821–826. https://doi.org/10.1097/JGP.0b013e31826abc00

Giau, V. van, Bagyinszky, E., & An, S. S. A. (2019). Potential Fluid Biomarkers for the Diagnosis of Mild Cognitive Impairment. *International Journal of Molecular Sciences*, *20*(17), 4149. https://doi.org/10.3390/ijms20174149

Giezendanner, S., Monsch, A. U., Kressig, R. W., Mueller, Y., Streit, S., Essig, S., Zeller, A., & Bally, K. (2019). General practitioners' attitudes towards early diagnosis of dementia: A cross-sectional survey. *BMC Family Practice*, *20*(1). https://doi.org/10.1186/S12875-019-0956-1

Golde, T. E. (2022). Disease-Modifying Therapies for Alzheimer's Disease: More Questions than Answers. *Neurotherapeutics*, *19*(1), 209–227. https://doi.org/10.1007/s13311-022-01201-2

Gosztolya, G., Vincze, V., Tóth, L., Pákáski, M., Kálmán, J., & Hoffmann, I. (2019). Identifying Mild Cognitive Impairment and mild Alzheimer's disease based on spontaneous speech using ASR and linguistic features. *Computer Speech & Language*, *53*, 181–197. https://doi.org/10.1016/j.csl.2018.07.007

Haddad, H. W., Malone, G. W., Comardelle, N. J., Degueure, A. E., Poliwoda, S., Kaye, R. J., Murnane, K. S., Kaye, A. M., & Kaye, A. D. (2022). Aduhelm, a novel anti-amyloid monoclonal antibody, for the treatment of Alzheimer's Disease: A comprehensive review. *Health Psychology Research*, *10*(2). https://doi.org/10.52965/001c.37023

Hahn, E. A., & Andel, R. (2011). Nonpharmacological therapies for behavioral and cognitive symptoms of mild cognitive impairment. *Journal of Aging and Health*, *23*(8), 1223–1245. https://doi.org/10.1177/0898264311422745

Hansen, E. C., Hughes, C., Routley, G., & Robinson, A. L. (2008). General practitioners' experiences and understandings of diagnosing dementia: Factors impacting on early diagnosis. *Social Science & Medicine*, *67*(11), 1776–1783. https://doi.org/10.1016/J.SOCSCIMED.2008.09.020

Harada, C. N., Natelson Love, M. C., & Triebel, K. L. (2013). Normal Cognitive Aging. *Clinics in Geriatric Medicine*, *29*(4), 737–752. https://doi.org/10.1016/j.cger.2013.07.002

Harmand, M. G. C., Meillon, C., Rullier, L., Taddé, O. B., Pimouguet, C., Dartigues, J. F., Bergua, V., & Amieva, H. (2018). Description of general practitioners' practices when suspecting cognitive impairment. Recourse to care in dementia (Recaredem) study. *Aging & Mental Health*, *22*(8), 1040–1049. https://doi.org/10.1080/13607863.2017.1330871

Heim, S. (2022). *Az időskori demencia kórképek háziorvosi vonatkozásai: A háziorvosok attitűdjének, szerepvállalásának vizsgálata a demenciák felismerésében és gondozásában, valamint véleményük az ellátást segítő és gátló tényezők tekintetében*.

Heim, S., Busa, C., Pozsgai, É., Csikós, Á., Papp, E., Pákáski, M., Kálmán, J., Hajnal, F., & Karádi, K. (2019). Hungarian general practitioners' attitude and the role of education in dementia care. *Primary Health Care Research & Development*, *20*, e92. https://doi.org/10.1017/S1463423619000203

Henry, J. D., & Crawford, J. R. (2004). A Meta-Analytic Review of Verbal Fluency Performance Following Focal Cortical Lesions. *Neuropsychology*, *18*(2), 284–295. https://doi.org/10.1037/0894-4105.18.2.284

Holmlund, T. B., Cheng, J., Foltz, P. W., Cohen, A. S., & Elvevåg, B. (2019). Updating verbal fluency analysis for the 21st century: Applications for psychiatry. *Psychiatry Research*, *273*, 767–769. https://doi.org/10.1016/j.psychres.2019.02.014

Hubley, S., Lynch, S. B., Schneck, C., Thomas, M., & Shore, J. (2016). Review of key telepsychiatry outcomes. *World Journal of Psychiatry*, *6*(2), 269. https://doi.org/10.5498/wjp.v6.i2.269

Hungarian Central Statistical Office. (n.d.). *Number of practicing general practitioners in Hungary (2007-2017). [Data base]. Retrieved from* *http://statinfo.ksh.hu/Statinfo/index.jsp*. Retrieved October 16, 2022, from http://statinfo.ksh.hu/Statinfo/index.jsp.

IBM SPSS Statistics for Windows. (2016). IMB Corporation.

Imre, N., Balogh, R., Gosztolya, G., Tóth, L., Hoffmann, I., Várkonyi, T., Lengyel, C., Pákáski, M., & Kálmán, J. (2022). Temporal Speech Parameters Indicate Early Cognitive Decline in Elderly Patients With Type 2 Diabetes Mellitus. *Alzheimer Disease & Associated Disorders*, *36*(2), 148–155. https://doi.org/10.1097/WAD.0000000000000492

Imre, N., Balogh, R., Papp, E., Kovács, I., Heim, S., Karádi, K., Hajnal, F., Kálmán, J., & Pákáski, M. (2019). Knowledge of general practitioners on dementia and mild cognitive impairment: a cross-sectional, questionnaire study from Hungary. *Educational Gerontology*, *45*(8), 495–505. https://doi.org/10.1080/03601277.2019.1660137

Irving, G., Neves, A. L., Dambha-Miller, H., Oishi, A., Tagashira, H., Verho, A., & Holden, J. (2017). International variations in primary care physician consultation time: a systematic review of 67 countries. *BMJ Open*, *7*(10), e017902. https://doi.org/10.1136/BMJOPEN-2017-017902

Ismail, Z., Black, S. E., Camicioli, R., Chertkow, H., Herrmann, N., Laforce, R., Montero-Odasso, M., Rockwood, K., Rosa-Neto, P., Seitz, D., Sivananthan, S., Smith, E. E., Soucy, J., Vedel, I., & Gauthier, S. (2020). Recommendations of the 5th Canadian Consensus Conference on the diagnosis and treatment of dementia. *Alzheimer's & Dementia*, *16*(8), 1182–1195. https://doi.org/10.1002/alz.12105

Ismail, Z., Rajji, T. K., & Shulman, K. I. (2010). Brief cognitive screening instruments: an update. *International Journal of Geriatric Psychiatry*, *25*(2), 111–120. https://doi.org/10.1002/gps.2306

Jack, C. R., Knopman, D. S., Jagust, W. J., Petersen, R. C., Weiner, M. W., Aisen, P. S., Shaw, L. M., Vemuri, P., Wiste, H. J., Weigand, S. D., Lesnick, T. G., Pankratz, V. S., Donohue, M. C., & Trojanowski, J. Q. (2013). Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *The Lancet Neurology*, *12*(2), 207–216. https://doi.org/10.1016/S1474-4422(12)70291-0

Jacobs, S., Mercuri, G., & Holtzer, R. (2020). Assessing within-task verbal fluency performance: the utility of individual time intervals in predicting incident mild cognitive impairment. *Aging, Neuropsychology, and Cognition*, 1–15. https://doi.org/10.1080/13825585.2020.1817306

Jak, A. J., Bondi, M. W., Delano-Wood, L., Wierenga, C., Corey-Bloom, J., Salmon, D. P., & Delis, D. C. (2009). Quantification of Five Neuropsychological Approaches to Defining Mild Cognitive Impairment. *The American Journal of Geriatric Psychiatry*, *17*(5), 368–375. https://doi.org/10.1097/JGP.0b013e31819431d5

Jessen, F., Amariglio, R. E., Boxtel, M., Breteler, M., Ceccaldi, M., Chételat, G., Dubois, B., Dufouil, C., Ellis, K. A., Flier, W. M., Glodzik, L., Harten, A. C., Leon, M. J., McHugh, P., Mielke, M. M., Molinuevo, J. L., Mosconi, L., Osorio, R. S., Perrotin, A., … Wagner, M. (2014). A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. *Alzheimer's & Dementia*, *10*(6), 844–852. https://doi.org/10.1016/j.jalz.2014.01.001

Kaduszkiewicz, H., Wiese, B., & van den Bussche, H. (2008). Self-reported competence, attitude and approach of physicians towards patients with dementia in ambulatory care: Results of a postal survey. *BMC Health Services Research*, *8*(1), 1–10. https://doi.org/10.1186/1472-6963-8-54/TABLES/3

Kálmán, J., Pákáski, M., Hoffmann, I., Drótos, G., Darvas, G., Boda, K., Bencsik, T., Gyimesi, A., Gulyás, Z., Bálint, M., Szatlóczki, G., & Papp, E. (2013). Korai Mentális Teszt: az enyhe kognitív zavar szűrõtesztjének fejlesztése. *Ideggyógyászati Szemle*, *66*(1–2), 43–52.

Kempler, D., Teng, E. L., Dick, M., Taussig, I. M., & Davis, D. S. (1998). The effects of age, education, and ethnicity on verbal fluency. *Journal of the International Neuropsychological Society*, *4*(6), 531–538. https://doi.org/10.1017/S1355617798466013

Kim, J. W., Lee, D. Y., Seo, E. H., Sohn, B. K., Choe, Y. M., Kim, S. G., Park, S. Y., Choo, I. H., Youn, J. C., Jhoo, J. H., Kim, K. W., & Woo, J. I. (2014). Improvement of Screening Accuracy of Mini-Mental State Examination for Mild Cognitive Impairment and Non-Alzheimer's Disease Dementia by Supplementation of Verbal Fluency Performance. *Psychiatry Investigation*, *11*(1), 44–51. https://doi.org/10.4306/pi.2014.11.1.44

Knopman, D. S., & Petersen, R. C. (2014). Mild cognitive impairment and mild dementia: a clinical perspective. *Mayo Clinic Proceedings*, *89*(10), 1452–1459. https://doi.org/10.1016/j.mayocp.2014.06.019

Koch, T., & Iliffe, S. (2010). Rapid appraisal of barriers to the diagnosis and management of patients with dementia in primary care: a systematic review. *BMC Family Practice*, *11*(1), 52. https://doi.org/10.1186/1471-2296-11-52

König, A., Linz, N., Tröger, J., Wolters, M., Alexandersson, J., & Robert, P. (2018). Fully automatic speech-based analysis of the semantic verbal fluency task. *Dementia and Geriatric Cognitive Disorders*, *45*(3–4), 198–209. https://doi.org/10.1159/000487852

Koo, B. M., & Vizer, L. M. (2019). Mobile Technology for Cognitive Assessment of Older Adults: A Scoping Review. *Innovation in Aging*, *3*(1). https://doi.org/10.1093/geroni/igy038

Kwon, S. J., Kim, H. S., Han, J. H., Bae, J. bin, Han, J. W., & Kim, K. W. (2021). Reliability and Validity of Alzheimer's Disease Screening With a Semi-automated Smartphone Application Using Verbal Fluency. *Frontiers in Neurology*, *12*. https://doi.org/10.3389/fneur.2021.684902

Lang, L., Clifford, A., Wei, L., Zhang, D., Leung, D., Augustine, G., Danat, I. M., Zhou, W., Copeland, J. R., Anstey, K. J., & Chen, R. (2017). Prevalence and determinants of undetected dementia in the community: a systematic literature review and a meta-analysis. *BMJ Open*, *7*(2), e011146. https://doi.org/10.1136/bmjopen-2016-011146

Ledoux, K., Vannorsdall, T. D., Pickett, E. J., Bosley, L. v, Gordon, B., & Schretlen, D. J. (2014). Capturing additional information about the organization of entries in the lexicon from verbal fluency productions. *Journal of Clinical and Experimental Neuropsychology*, *36*(2), 205–220. https://doi.org/10.1080/13803395.2013.878689

Lehnert, T., Heider, D., Leicht, H., Heinrich, S., Corrieri, S., Luppa, M., Riedel-Heller, S., & König, H. H. (2011). Review: health care utilization and costs of elderly persons with multiple chronic conditions. *Medical Care Research and Review : MCRR*, *68*(4), 387–420. https://doi.org/10.1177/1077558711399580

Lehtinen, N., Luotonen, I., & Kautto, A. (2021). Systematic administration and analysis of verbal fluency tasks: Preliminary evidence for reliable exploration of processes underlying task performance. *Applied Neuropsychology: Adult*, 1–13. https://doi.org/10.1080/23279095.2021.1973471

Lezak, M. D. (2012). *Neuropsychological Assessment (5th ed.)*. Oxford University Press.

Lin, J. S., O'Connor, E., Rossom, R. C., Perdue, L. A., Burda, B. U., Thompson, M., & Eckstrom, E. (2013). *Screening for Cognitive Impairment in Older Adults: An Evidence Update for the U.S. Preventive Services Task Force*. Agency for Healthcare Research and Quality (US).

Linz, N., Lundholm Fors, K., Lindsay, H., Eckerström, M., Alexandersson, J., & Kokkinakis, D. (2019). Temporal Analysis of the Semantic Verbal Fluency Task in Persons with Subjective and Mild Cognitive Impairment. *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, 103–113. https://doi.org/10.18653/v1/W19-3012

Liss, J. L., Seleri Assunção, S., Cummings, J., Atri, A., Geldmacher, D. S., Candela, S. F., Devanand, D. P., Fillit, H. M., Susman, J., Mintzer, J., Bittner, T., Brunton, S. A., Kerwin, D. R., Jackson, W. C., Small, G. W., Grossberg, G. T., Clevenger, C. K., Cotter, V., Stefanacci, R., … Sabbagh, M. N. (2021). Practical recommendations for timely, accurate diagnosis of symptomatic Alzheimer's disease (MCI and dementia) in primary care: a review and synthesis. *Journal of Internal Medicine*, *290*(2), 310–334. https://doi.org/10.1111/joim.13244

Lopez-de-Ipina, K., Martinez-de-Lizarduy, U., Barroso, N., Ecay-Torres, M., Martinez-Lage, P., Torres, F., & Faundez-Zanuy, M. (2015). Automatic analysis of Categorical Verbal Fluency for Mild Cognitive impartment detection: A non-linear language independent approach. *2015 4th International Work Conference on Bioinspired Intelligence (IWOBI)*, 101–104. https://doi.org/10.1109/IWOBI.2015.7160151

Lopez, O. L. (2013). Mild Cognitive Impairment. *CONTINUUM: Lifelong Learning in Neurology*, *19*(2), 411–424. https://doi.org/10.1212/01.CON.0000429175.29601.97

Lorentz, W. J., Scanlan, J. M., & Borson, S. (2002). Brief screening tests for dementia. *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie*, *47*(8), 723–733. https://doi.org/10.1177/070674370204700803

Lubrini, G., Periáñez, J. A., Laseca-Zaballa, G., Bernabéu-Brotons, E., & Ríos-Lago, M. (2022). Verbal Fluency Tasks: Influence of Age, Gender, and Education and Normative Data for the Spanish Native Adult Population. *Archives of Clinical Neuropsychology*, *37*(2), 365–375. https://doi.org/10.1093/arclin/acab056

Luo, L., Luk, G., & Bialystok, E. (2010). Effect of language proficiency and executive control on verbal fluency performance in bilinguals. *Cognition*, *114*(1), 29–41. https://doi.org/10.1016/j.cognition.2009.08.014

Lu, Y., Liu, C., Fawkes, S., Wang, Z., & Yu, D. (2022). Knowledge, attitudes, and practice of general practitioners toward community detection and management of mild cognitive impairment: a cross-sectional study in Shanghai, China. *BMC Primary Care*, *23*(1), 114. https://doi.org/10.1186/s12875-022-01716-9

Martínez-Nicolás, I., Llorente, T. E., Martínez-Sánchez, F., & Meilán, J. J. G. (2021). Ten Years of Research on Automatic Voice and Speech Analysis of People With Alzheimer's Disease and Mild Cognitive Impairment: A Systematic Review Article. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.620251

Mathuranath, P. S., George, A., Cherian, P. J., Alexander, A., Sarma, S. G., & Sarma, P. S. (2003). Effects of Age, Education and Gender on Verbal Fluency. *Journal of Clinical and Experimental Neuropsychology*, *25*(8), 1057–1064. https://doi.org/10.1076/jcen.25.8.1057.16736

McDonnell, M., Dill, L., Panos, S., Amano, S., Brown, W., Giurgius, S., Small, G., & Miller, K. (2020). Verbal fluency as a screening tool for mild cognitive impairment. *International Psychogeriatrics*, *32*(9), 1055–1062. https://doi.org/10.1017/S1041610219000644

McIntosh, I. B., Swanson, V., Power, K. G., & Rae, A. C. (1999). General practitioners' and nurses' perceived roles, attitudes and stressors in the management of people with dementia. *Health Bulletin*, *57*(1), 35–43. https://www.researchgate.net/publication/10703484_General_practitioners'_and_nurses'_perceived_roles_attitudes_and_stressors_in_the_management_of_people_with_dementia

MedCalc Software. (2020). MedCalc Software Ltd.

Meijer, E., Casanova, M., Kim, H., Llena-Nozal, A., & Lee, J. (2022). Economic costs of dementia in 11 countries in Europe: Estimates from nationally representative cohorts of a panel study. *The Lancet Regional Health - Europe*, *20*, 100445. https://doi.org/10.1016/j.lanepe.2022.100445

Mészáros, A., Kónya, A., & Kas, B. (2011). A verbális fluenciatesztek felvételének és értékelésének módszertana. *Alkalmazott Pszichológia*, *2*, 53–76.

Mueller, K. D., Hermann, B., Mecollari, J., & Turkstra, L. S. (2018). Connected speech and language in mild cognitive impairment and Alzheimer's disease: A review of picture description tasks. *Journal of Clinical and Experimental Neuropsychology*, *40*(9), 917–939. https://doi.org/10.1080/13803395.2018.1446513

Mueller, K. D., Koscik, R. L., LaRue, A., Clark, L. R., Hermann, B., Johnson, S. C., & Sager, M. A. (2015). Verbal Fluency and Early Memory Decline: Results from the Wisconsin Registry for Alzheimer's Prevention. *Archives of Clinical Neuropsychology*, *30*(5), 448–457. https://doi.org/10.1093/arclin/acv030

Munro Cullum, C., Hynan, L. S., Grosch, M., Parikh, M., & Weiner, M. F. (2014). Teleneuropsychology: Evidence for Video Teleconference-Based Neuropsychological Assessment. *Journal of the International Neuropsychological Society, 20*(10), 1028–1033. https://doi.org/10.1017/S1355617714000873

Murphy, K. J., Rich, J. B., & Troyer, A. K. (2006). Verbal fluency patterns in amnestic mild cognitive impairment are characteristic of Alzheimer's type dementia. *Journal of the International Neuropsychological Society*, *12*(04). https://doi.org/10.1017/S1355617706060590

National Academies of Sciences, E. and M. (2021). *Reducing the Impact of Dementia in America*. National Academies Press. https://doi.org/10.17226/26175

National Institute on Aging, N. I. of H. (2007). *Why Population Aging Matters: A Global Perspective*.

Nelson, A. P., & O'Connor, M. G. (2008). Mild cognitive impairment: a neuropsychological perspective. *CNS Spectrums*, *13*(1), 56–64. https://doi.org/10.1017/S1092852900016163

Nichols, E., Steinmetz, J. D., Vollset, S. E., Fukutaki, K., Chalek, J., Abd-Allah, F., Abdoli, A., Abualhasan, A., Abu-Gharbieh, E., Akram, T. T., al Hamad, H., Alahdab, F., Alanezi, F. M., Alipour, V., Almustanyir, S., Amu, H., Ansari, I., Arabloo, J., Ashraf, T., … Vos, T. (2022). Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. *The Lancet Public Health*, *7*(2), e105–e125. https://doi.org/10.1016/S2468-2667(21)00249-8

Nikolai, T., Bezdicek, O., Markova, H., Stepankova, H., Michalec, J., Kopecek, M., Dokoupilova, M., Hort, J., & Vyhnalek, M. (2018). Semantic verbal fluency impairment is detectable in patients with subjective cognitive decline. *Applied Neuropsychology: Adult*, *25*(5), 448–457. https://doi.org/10.1080/23279095.2017.1326047

Oberg, G., & Ramírez, M. (2006). Cross-linguistic meta-analysis of phonological fluency: Normal performance across cultures. *International Journal of Psychology*, *41*(5), 342–347. https://doi.org/10.1080/00207590500345872

Oh, S. J., Sung, J. E., Choi, S. J., & Jeong, J. H. (2019). Clustering and Switching Patterns in Semantic Fluency and Their Relationship to Working Memory in Mild Cognitive Impairment. *Dementia and Neurocognitive Disorders*, *18*(2), 47. https://doi.org/10.12779/dnd.2019.18.2.47

Olabarrieta-Landa, L., Torre, E. L., López-Mugartza, J. C., Bialystok, E., & Arango-Lasprilla, J. C. (2017). Verbal fluency tests: Developing a new model of administration and scoring for Spanish language. *NeuroRehabilitation*, *41*(2), 539–565. https://doi.org/10.3233/NRE-162102

Östberg, P., Fernaeus, S.-E., Hellström, Å., Bogdanović, N., & Wahlund, L.-O. (2005). Impaired verb fluency: A sign of mild cognitive impairment. *Brain and Language*, *95*(2), 273–279. https://doi.org/10.1016/j.bandl.2005.01.010

Overton, M., Pihlsgård, M., & Elmståhl, S. (2019). Diagnostic Stability of Mild Cognitive Impairment, and Predictors of Reversion to Normal Cognitive Functioning. *Dementia and Geriatric Cognitive Disorders*, *48*(5–6), 317–329. https://doi.org/10.1159/000506255

Owens, D. K., Davidson, K. W., Krist, A. H., Barry, M. J., Cabana, M., Caughey, A. B., Doubeni, C. A., Epling, J. W., Kubik, M., Landefeld, C. S., Mangione, C. M., Pbert, L., Silverstein, M., Simon, M. A., Tseng, C.-W., & Wong, J. B. (2020). Screening for Cognitive Impairment in Older Adults. *JAMA*, *323*(8), 757. https://doi.org/10.1001/jama.2020.0435

Pakhomov, S. V. S., & Hemmy, L. S. (2014). A computational linguistic measure of clustering behavior on semantic verbal fluency task predicts risk of future dementia in the Nun Study. *Cortex*, *55*, 97–106. https://doi.org/10.1016/j.cortex.2013.05.009

Pakzad, S., Ringuette, J., Bourque, P., & Sepehry, A. A. (2018). Executive Functions And Screening For Mild Cognitive Impairment And Alzheimers Disease: A Cross-Sectional Study. *Acta Psychopathologica*, *04*(03). https://doi.org/10.4172/2469-6676.100171

Perneczky, R. (2019). Dementia treatment versus prevention. *Dialogues in Clinical Neuroscience*, *21*(1), 43–51. https://doi.org/10.31887/DCNS.2019.21.1/rperneczky

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

Petersen, R. C., Doody, R., Kurz, A., Mohs, R. C., Morris, J. C., Rabins, P. v., Ritchie, K., Rossor, M., Thal, L., & Winblad, B. (2001). Current concepts in mild cognitive impairment. *Archives of Neurology*, *58*(12), 1985–1992. https://doi.org/10.1001/ARCHNEUR.58.12.1985

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild Cognitive Impairment. *Archives of Neurology*, *56*(3), 303. https://doi.org/10.1001/archneur.56.3.303

Piatt, A. L., Fields, J. A., Paolo, A. M., & Tröster, A. I. (1999). Action (verb naming) fluency as an executive function measure: convergent and divergent evidence of validity. *Neuropsychologia*, *37*(13), 1499–1503. https://doi.org/10.1016/S0028-3932(99)00066-4

Prins, A., Hemke, F., Pols, J., & van Charante, E. P. M. (2016). Diagnosing dementia in Dutch general practice: a qualitative study of GPs' practices and views. *The British Journal of General Practice : The Journal of the Royal College of General Practitioners*, *66*(647), e416–e422. https://doi.org/10.3399/BJGP16X685237

Ranson, J. M., Kuźma, E., Hamilton, W., Lang, I., & Llewellyn, D. J. (2018). Case-finding in clinical practice: An appropriate strategy for dementia identification? *Alzheimer's & Dementia : Translational Research & Clinical Interventions*, *4*, 288. https://doi.org/10.1016/J.TRCI.2018.04.011

Raoux, N., Amieva, H., le Goff, M., Auriacombe, S., Carcaillon, L., Letenneur, L., & Dartigues, J.-F. (2008). Clustering and switching processes in semantic verbal fluency in the course of Alzheimer's disease subjects: Results from the PAQUID longitudinal study. *Cortex*, *44*(9), 1188–1196. https://doi.org/10.1016/j.cortex.2007.08.019

Reisberg, B., Ferris, S. H., de Leon, M. J., Franssen, E. S. E., Kluger, A., Mir, P., Borenstein, J., George, A. E., Shulman, E., Steinberg, G., & Cohen, J. (1988). Stage-specific behavioral, cognitive, and in vivo changes in community residing subjects with age-associated memory impairment and primary degenerative dementia of the Alzheimer type. *Drug Development Research*, *15*(2–3), 101–114. https://doi.org/10.1002/DDR.430150203

Roberts, R. O., & Knopman, D. S. (2013). Classification and epidemiology of MCI. *Clinics in Geriatric Medicine*, *29*(4), 753–772. https://doi.org/10.1016/j.cger.2013.07.003

Roberts, R. O., Knopman, D. S., Mielke, M. M., Cha, R. H., Pankratz, V. S., Christianson, T. J. H., Geda, Y. E., Boeve, B. F., Ivnik, R. J., Tangalos, E. G., Rocca, W. A., & Petersen, R. C. (2014). Higher risk of progression to dementia in mild cognitive impairment cases who revert to normal. *Neurology*, *82*(4), 317–325. https://doi.org/10.1212/WNL.0000000000000055

Rodriguez-Aranda, C., & Martinussen, M. (2006). Age-Related Differences in Performance of Phonemic Verbal Fluency Measured by Controlled Oral Word

Association Task (COWAT): A Meta-Analytic Study. *Developmental Neuropsychology*, *30*(2), 697–717. https://doi.org/10.1207/s15326942dn3002_3

Rone-Adams, S., Stern, D. F., Olivier, T. W., & Blodgett, M. G. (2013). Understanding Dementia. *Strength & Conditioning Journal*, *35*(4), 88–98. https://doi.org/10.1519/SSC.0b013e31829ed301

Rosenthal, R. (1991). *Meta-analytic procedures for social research (Rev. ed.)*. Sage.

Rosen, W. G., Mohs, R. C., & Davis, K. L. (1984). A new rating scale for Alzheimer's disease. *American Journal of Psychiatry*, *141*(11), 1356–1364. https://doi.org/10.1176/ajp.141.11.1356

Ryan, J. O., Pakhomov, S., Marino, S., Bernick, C., & Banks, S. (2013). *Computerized Analysis of a Verbal Fluency Test* (pp. 884–889). https://aclanthology.org/P13-2153

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. https://doi.org/10.1038/nrneurol.2014.181

Salthouse, T. A. (2019). Trajectories of normal cognitive aging. *Psychology and Aging*, *34*(1), 17–24. https://doi.org/10.1037/pag0000288

Sannemann, L., Müller, T., Waterink, L., Zwan, M., Wimo, A., Stomrud, E., Pinó, S., Arrufat, J., Rodríguez-Gomez, O., Benaque, A., Bon, J., Ferreira, D., Johansson, G., Dron, A., Dumas, A., Georges, J., Kramberger, M. G., Visser, P. J., Winblad, B., … Jessen, F. (2021). General practitioners' attitude toward early and pre-dementia diagnosis of AD in five European countries—A MOPEAD project survey. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *13*(1). https://doi.org/10.1002/dad2.12130

Senanayake, U., Sowmya, A., Dawes, L., Kochan, N. A., Wen, W., & Sachdev, P. (2016). Classification of Mild Cognitive Impairment Subtypes using Neuropsychological Data. *Proceedings of the 5th International Conference on Pattern Recognition Applications and Methods*, 620–629. https://doi.org/10.5220/0005747806200629

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00772

Shaw, L. M., Vanderstichele, H., Knapik-Czajka, M., Clark, C. M., Aisen, P. S., Petersen, R. C., Blennow, K., Soares, H., Simon, A., Lewczuk, P., Dean, R., Siemers, E., Potter, W., Lee, V. M.-Y., & Trojanowski, J. Q. (2009). Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Annals of Neurology*, *65*(4), 403–413. https://doi.org/10.1002/ana.21610

Shulman, K. I., Shedletsky, R., & Silver, I. L. (1986). The challenge of time: Clock-drawing and cognitive function in the elderly. *International Journal of Geriatric Psychiatry*, *1*(2), 135–140. https://doi.org/10.1002/gps.930010209

Sindi, S., Mangialasche, F., & Kivipelto, M. (2015). Advances in the prevention of Alzheimer's Disease. *F1000Prime Reports*, *7*. https://doi.org/10.12703/P7-50

Sluis, R. A., Angus, D., Wiles, J., Back, A., Gibson, T. (Amy), Liddle, J., Worthy, P., Copland, D., & Angwin, A. J. (2020). An automated approach to examining pausing in the speech of people with dementia. *American Journal of Alzheimer's Disease & Other Dementiasr*, *35*, 153331752093977. https://doi.org/10.1177/1533317520939773

Somme, D., Gautier, A., Pin, S., & Corvol, A. (2013). General practitioner's clinical practices, difficulties and educational needs to manage Alzheimer's disease in France: analysis of national telephone-inquiry data. *BMC Family Practice*, *14*. https://doi.org/10.1186/1471-2296-14-81

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., Iwatsubo, T., Jack, C. R., Kaye, J., Montine, T. J., Park, D. C., Reiman, E. M., Rowe, C. C., Siemers, E., Stern, Y., Yaffe, K., Carrillo, M. C., Thies, B., Morrison-Bogorad, M., … Phelps, C. H. (2011). Toward defining the preclinical stages of 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), 280–292. https://doi.org/10.1016/j.jalz.2011.03.003

Stokin, G. B., Krell-Roesch, J., Petersen, R. C., & Geda, Y. E. (2015). Mild Neurocognitive Disorder. *Harvard Review of Psychiatry*, *23*(5), 368–376. https://doi.org/10.1097/HRP.0000000000000084

Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary* (3rd edition). Oxford University Press.

Takacs, R., Ungvari, G. S., & Gazdag, G. (2015). Reasons for acute psychiatric admission of patients with dementia. *Neuropsychopharmacologia Hungarica : A Magyar Pszichofarmakologiai Egyesulet Lapja = Official Journal of the Hungarian Association of Psychopharmacology*, *17*(3), 141–145.

Takeda, M., Kudo, T., Tanaka, T., Kamino, K., Okochi, M., & Tagami, S. (2006). Biological markers for diagnosis of MCI and neurodegenerative dementia. *International Congress Series*, *1290*, 101–107. https://doi.org/10.1016/j.ics.2005.12.076

Taler, V., Johns, B. T., & Jones, M. N. (2020). A Large-Scale Semantic Analysis of Verbal Fluency Across the Aging Spectrum: Data From the Canadian Longitudinal Study on Aging. *J Gerontol B Psychol Sci Soc Sci*, *75*(9), e221–e230. https://doi.org/10.1093/geronb/gbz003

Tánczos, T., Janacsek, K., & Németh, D. (2014). [Verbal fluency tasks I. Investigation of the Hungarian version of the phonemic fluency task between 5 and 89 years of age]. *Psychiatria Hungarica*, *29*(2), 158–180.

Teng, E., Leone-Friedman, J., Lee, G. J., Woo, S., Apostolova, L. G., Harrell, S., Ringman, J. M., & Lu, P. H. (2013). Similar Verbal Fluency Patterns in Amnestic Mild Cognitive Impairment and Alzheimer's Disease. *Archives of Clinical Neuropsychology*, *28*(5), 400–410. https://doi.org/10.1093/arclin/act039

Thaipisuttikul, P., Jaikla, K., Satthong, S., & Wisajun, P. (2022). Rate of conversion from mild cognitive impairment to dementia in a Thai hospital-based population: A retrospective cohort. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *8*(1). https://doi.org/10.1002/trc2.12272

Thyrian, J. R., & Hoffmann, W. (2012). Dementia care and general physicians--a survey on prevalence, means, attitudes and recommendations. *Central European Journal of Public Health*, *20*(4), 270–275. https://doi.org/10.21101/CEJPH.A3751

Toledo, J. B., Arnold, S. E., Raible, K., Brettschneider, J., Xie, S. X., Grossman, M., Monsell, S. E., Kukull, W. A., & Trojanowski, J. Q. (2013). Contribution of cerebrovascular disease in autopsy confirmed neurodegenerative disease cases in the National Alzheimer's Coordinating Centre. *Brain*, *136*(9), 2697–2706. https://doi.org/10.1093/brain/awt188

Toth, L., Hoffmann, I., Gosztolya, G., Vincze, V., Szatloczki, G., Banreti, Z., Pakaski, M., & Kalman, J. (2018). A Speech Recognition-based Solution for the Automatic Detection of Mild Cognitive Impairment from Spontaneous Speech. *Current Alzheimer Research*, *15*(2), 130–138. https://doi.org/10.2174/1567205014666171121114930

Tripathi, M., & Vibha, D. (2009). Reversible dementias. *Indian Journal of Psychiatry, 51 Suppl 1*(Suppl1), S52-5.

Tröger, J., Linz, N., König, A., Robert, P., Alexandersson, J., Peter, J., & Kray, J. (2019). Exploitation vs. exploration—computational temporal and semantic analysis explains semantic verbal fluency impairment in Alzheimer's disease. *Neuropsychologia*, *131*, 53–61. https://doi.org/10.1016/j.neuropsychologia.2019.05.007

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology*, *11*(1), 138–146. https://doi.org/10.1037/0894-4105.11.1.138

Tucker-Drob, E. M. (2019). Cognitive Aging and Dementia: A Life-Span Perspective. *Annual Review of Developmental Psychology*, *1*(1), 177–196. https://doi.org/10.1146/annurev-devpsych-121318-085204

van de Mortel, L. A., Thomas, R. M., & van Wingen, G. A. (2021). Grey Matter Loss at Different Stages of Cognitive Decline: A Role for the Thalamus in Developing Alzheimer's Disease. *Journal of Alzheimer's Disease*, *83*(2), 705–720. https://doi.org/10.3233/JAD-210173

van Hout, H., Vernooij-Dassen, M., Bakker, K., Blom, M., & Grol, R. (2000). General practitioners on dementia: tasks, practices and obstacles. *Patient Education and Counseling*, *39*(2–3), 219–225. https://doi.org/10.1016/S0738-3991(99)00034-8

Venegas, M. J., & Mansur, L. L. (2011). Verbal fluency: Effect of time on item generation. *Dementia & Neuropsychologia*, *5*(2), 104–107. https://doi.org/10.1590/S1980-57642011DN05020008

Veneziani, F., Panza, F., Solfrizzi, V., Capozzo, R., Barulli, M. R., Leo, A., Lozupone, M., Fontana, A., Arcuti, S., Copetti, M., Cardinali, V., Grasso, A., Tursi, M., Iurillo, A., Imbimbo, B. pietro, Seripa, D., & Logroscino, G. (2016). Examination of level of knowledge in Italian general practitioners attending an education session on diagnosis and management of the early stage of Alzheimer's disease: pass or fail? *International Psychogeriatrics*, *28*(7), 1111–1124. https://doi.org/10.1017/S1041610216000041

Vieira, R. T., Caixeta, L., Machado, S., Silva, A. C., Nardi, A. E., Arias-Carrión, O., & Carta, M. G. (2013). Epidemiology of early-onset dementia: a review of the literature. *Clinical Practice & Epidemiology in Mental Health*, *9*(1), 88–95. https://doi.org/10.2174/1745017901309010088

Vita, M. G., Marra, C., Spinelli, P., Caprara, A., Scaricamazza, E., Castelli, D., Canulli, S., Gainotti, G., & Quaranta, D. (2014). Typicality of Words Produced on a Semantic Fluency Task in Amnesic Mild Cognitive Impairment: Linguistic Analysis and Risk of Conversion to Dementia. *Journal of Alzheimer's Disease*, *42*(4), 1171–1178. https://doi.org/10.3233/JAD-140570

Wajman, J. R., Cecchini, M. A., Bertolucci, P. H. F., & Mansur, L. L. (2019). Quanti-qualitative components of the semantic verbal fluency test in cognitively healthy controls, mild cognitive impairment, and dementia subtypes. *Applied Neuropsychology: Adult*, *26*(6), 533–542. https://doi.org/10.1080/23279095.2018.1465426

Watson, R., Bryant, J., Sanson-Fisher, R., Mansfield, E., & Evans, T.-J. (2018). What is a 'timely' diagnosis? Exploring the preferences of Australian health service consumers regarding when a diagnosis of dementia should be disclosed. *BMC Health Services Research*, *18*(1), 612. https://doi.org/10.1186/s12913-018-3409-y

Weakley, A., Schmitter-Edgecombe, M., & Anderson, J. (2013). Analysis of Verbal Fluency Ability in Amnestic and Non-Amnestic Mild Cognitive Impairment. *Archives of Clinical Neuropsychology*, *28*(7), 721–731. https://doi.org/10.1093/arclin/act058

Wechsler, D. (1981). *The Wechsler Adult Intelligence Scale - Revised.* The Psychological Corporation.

Wee, C.-Y., Yap, P.-T., Zhang, D., Denny, K., Browndyke, J. N., Potter, G. G., Welsh-Bohmer, K. A., Wang, L., & Shen, D. (2012). Identification of MCI individuals using structural and functional connectivity networks. *NeuroImage*, *59*(3), 2045–2056. https://doi.org/10.1016/j.neuroimage.2011.10.015

WHO. (2022). *WHO Fact Sheet - Aging and Health*. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health

Wilkinson, D., Stave, C., Keohane, D., & Vincenzino, O. (2004). The role of general practitioners in the diagnosis and treatment of Alzheimer's disease: a multinational survey. *The Journal of International Medical Research*, *32*(2), 149–159. https://doi.org/10.1177/147323000403200207

Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L. O., Nordberg, A., Bäckman, L., Albert, M., Almkvist, O., Arai, H., Basun, H., Blennow, K., de Leon, M., Decarli, C., Erkinjuntti, T., Giacobini, E., Graff, C., Hardy, J., … Petersen, R. C. (2004). Mild cognitive impairment--beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *Journal of Internal Medicine*, *256*(3), 240–246. https://doi.org/10.1111/J.1365-2796.2004.01380.X

Wolz, R., Julkunen, V., Koikkalainen, J., Niskanen, E., Zhang, D. P., Rueckert, D., Soininen, H., & Lötjönen, J. (2011). Multi-Method Analysis of MRI Images in Early Diagnostics of Alzheimer's Disease. *PLoS ONE*, *6*(10), e25446. https://doi.org/10.1371/journal.pone.0025446

Woods, D. L., Wyma, J. M., Herron, T. J., & Yund, E. W. (2016). Computerized Analysis of Verbal Fluency: Normative Data and the Effects of Repeated Testing, Simulated Malingering, and Traumatic Brain Injury. *PLOS ONE*, *11*(12), e0166439. https://doi.org/10.1371/journal.pone.0166439

World Health Organization. (2017). *Global action plan on the public health response to dementia 2017-2025*.

World Health Organization. (2019). *International statistical classification of diseases and related health problems (11th ed.).*

World Health Organization. (2022). *Dementia [Fact sheet] Retrieved from: https://www.who.int/news-room/fact-sheets/detail/dementia*.

Yesavage, J. A., & Sheikh, J. I. (1986). 9/Geriatric Depression Scale (GDS): Recent Evidence and Development of a Shorter Version. *Clinical Gerontologist*, *5*(1–2), 165–173. https://doi.org/10.1300/J018v05n01_09

Zhao, Q., Guo, Q., & Hong, Z. (2013). Clustering and switching during a semantic verbal fluency test contribute to differential diagnosis of cognitive impairment. *Neuroscience Bulletin*, *29*(1), 75–82. https://doi.org/10.1007/s12264-013-1301-7

I.

EUROPEAN JOURNAL OF GENERAL PRACTICE
2020, VOL. 26, NO. 1, 7–13
https://doi.org/10.1080/13814788.2019.1673723

ORIGINAL ARTICLE

OPEN ACCESS Check for updates

Dementia in Hungary: General practitioners' routines and perspectives regarding early recognition

Réka Balogh[a], Nóra Imre[a], Edina Papp[a], Ildikó Kovács[a], Szilvia Heim[b], Kázmér Karádi[c], Ferenc Hajnal[d], Magdolna Pákáski[a] and János Kálmán[a]

[a]Department of Psychiatry, Faculty of Medicine, University of Szeged, Szeged, Hungary; [b]Department of Primary Health Care, Medical School, University of Pécs, Pécs, Hungary; [c]Institute of Behavioral Sciences, Medical School, University of Pécs, Pécs, Hungary; [d]Department of Family Medicine, Faculty of Medicine, University of Szeged, Szeged, Hungary

KEY MESSAGES

- Hungarian GPs were aware of the benefits of early dementia recognition.
- Most GPs do not use cognitive tests for case-finding.
- Besides providing longer consultation times, the primary way to improve the efficacy of recognition would be to construct a cost- and time-effective dementia identification strategy applicable in GPs' practices.

ABSTRACT

Background: Undetected dementia in primary care is a global problem. Since general practitioners (GPs) act as the first step in the identification process, examining their routines could help us to enhance the currently low recognition rates.

Objectives: The study aimed to explore, for the first time in Hungary, the dementia identification practices and views of GPs.

Methods: In the context of an extensive, national survey (February-November 2014) 8% of all practicing GPs in Hungary ($n = 402$) filled in a self-administered questionnaire. The questions (single, multiple-choice, Likert-type) analysed in the present study explored GPs' methods and views regarding dementia identification and their ideas about the optimal circumstances of case-finding.

Results: The vast majority of responding GPs (97%) agreed that the early recognition of dementia would enhance both the patients' and their relatives' well-being. When examining the possibility of dementia, most GPs (91%) relied on asking the patients general questions and only a quarter of them (24%) used formal tests, even though they were mostly satisfied with both the Clock Drawing Test (69%) and the Mini-Mental State Examination (65%). Longer consultation time was chosen as the most important facet of improvement needed for better identification of dementia in primary care (81%). Half of the GPs (49%) estimated dementia recognition rate to be lower than 30% in their practice.

Conclusions: Hungarian GPs were aware of the benefits of early recognition, but the shortage of consultation time in primary care was found to be a major constraint on efficient case-finding.

ARTICLE HISTORY
Received 8 February 2019
Revised 19 July 2019
Accepted 19 September 2019

KEYWORDS
General practitioners; primary care; dementia; case-finding; cognitive tests

Introduction

General practitioners (GPs) are greatly involved in the early stages of the dementia recognition process, as most patients visit them first to have their initial cognitive examination [1]. In Hungary, the estimated number of patients with dementia lies between 150,000 and 300,000 registered cases [2,3]. Due to the rapidly aging population, GPs in primary care are prone to see even more dementia patients in the future.

In Hungary, the dementia identification process depends on multiple professionals. Potential pathways to the identification of dementia could involve the

CONTACT Réka Balogh balogh.reka@med.u-szeged.hu Department of Psychiatry, Faculty of Medicine, University of Szeged, Szeged, Hungary
Supplemental data for this article can be accessed here.

patients' subjective complaints and/or their family members' reports on cognitive problems, GPs' concerns about signs of dementia during patient consultation, targeted case-finding and population screening [4]. If needed, GPs can decide to carry out basic neuropsychological tests (of which the Mini-Mental State Examination (MMSE) and the Clock Drawing Test (CDT) are financially reimbursed) and/or refer potential dementia patients to secondary care (memory clinic, psychiatry, neurology) for further investigation. Establishment of the diagnosis, identification of the etiology based on the International Classification of Diseases – 10th revision (ICD-10) and the prescription of the necessary anti-dementia medications are the tasks of psychiatrists or neurologists. After the diagnostic work-up, the specialists usually schedule patients for regular follow-up as well.

The difficulty of early dementia recognition is a global problem: research suggests that a substantial amount of dementia cases (up to 66%) is missed in primary care [5]. One of the main obstacles towards effective dementia case-finding in primary care is the low use of standardised cognitive tests. Not only is dementia a taboo topic for many GPs [6], some of them also experience ambivalence regarding the advantages of early diagnosis [7], thinking that treatment options are limited or non-existent, while some even believe that nothing could be done for patients with dementia [8].

Apart from an international study with limited sample sizes [9], no extensive research has been conducted on GPs' routines and views regarding dementia management neither in Hungary nor in many East-Central European countries. To address the lack of research, experts of two Hungarian universities collaborated on a large-scale project to examine several aspects of GPs in dementia care (see Methods section). As part of this project, the present study's main aim was (1) to identify the methods currently being used by Hungarian GPs for the recognition of dementia; (2) to observe GPs' satisfaction with the most widespread dementia screening tests; (3) to examine GPs' views regarding dementia and its management and (4) to explore their ideas about an ideal test for early recognition and those optimal circumstances that could contribute to the establishment of more efficient and effective ways of dementia identification in Hungarian primary care.

Methods

Instrument

To meet the aims of the project, a self-administered questionnaire was designed specifically to explore a broad range of aspects regarding GPs' role in dementia detection and management in Hungary. The project investigated several significant topics, including GPs' routines and perspectives regarding dementia detection in Hungary (which is covered by the present paper); GPs' factual knowledge of dementia [10] and also their attitudes regarding dementia patients and their management [11]. The development and validation of the questionnaire was a multistage process, taking up one year (Figure 1). The questions analysed in the present paper were fixed-response (single or

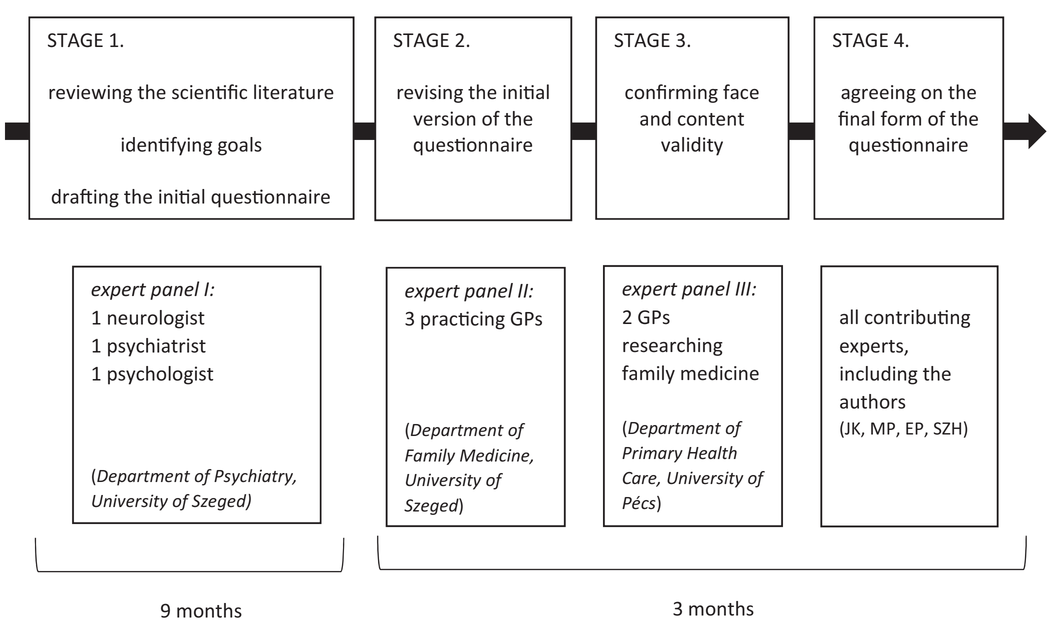

Figure 1. The multistage process of the questionnaire development.

multiple choice) and Likert-type questions; open-ended questions were not applied. (For the list of questions, refer Supplementary Material).

Participants and data collection

In Hungary, all practicing GPs must participate in a continuous education program, which means attending one professional training course in every 5 year period. Since our aim was to reach as many GPs as possible from every region of the country, the questionnaires were distributed at six major mandatory training courses and at three national conferences within a 10-month time frame, between February and November 2014. In order to avoid the courses' influence on the results of the study, we selected events that did not provide any specific education about dementia during our recruitment period. Ethical approval was obtained from the Regional and Institutional Research Ethics Committee of the University of Pécs (reference number: 5244).

The questionnaires were distributed on-site among the GP-attendees at the selected trainings and conferences, along with a written informative. Participation was entirely voluntary and anonymous. Completion rate varied for each question (the questions were completed on average by 86% of the respondents); therefore, in Results, the numbers of responses are indicated in brackets for each question.

Data was analysed using the SPSS v.24 statistical analysis software package (SPSS Inc., Chicago, IL, USA). Descriptive statistics (mean, percentage, standard deviation) were applied for all items on the questionnaire. Comparative analysis was executed for one question, using the Wilcoxon signed ranks test (statistical significance was set at the 5% level).

Results

Demographic properties and practice characteristics

Altogether 402 GPs handed back their completed questionnaire, which is more than 8% of all 4,850 GPs practicing in Hungary in 2014 [12]. Demographic information and characteristics of practices are presented in Table 1.

Table 1. GPs' demographics and characteristics of practices.

Gender ($n = 387$)	%	Place of practice ($n = 372$)	%	Dementia training (last 2 years) ($n = 366$)	%
Male	46.3	Urban	66.1	yes	19.4
Female	53.7	rural	33.9	no	80.6
Age ($n = 393$)	**%**	**Number of patients/day ($n = 393$)**	**%**	**Number of dementia patients ($n = 383$)**	**%**
25–35	5.9	0–30	2.0	0–50	49.9
36–45	12.5	31–40	16.9	51–100	38.1
46–55	24.9	41–50	27.9	101–150	8.4
56–65	40.2	51–60	25.1	151–200	2.6
65+	16.5	60+	25.9	200+	1.0

Ways of dementia evaluation and views on cognitive tests

The vast majority of GPs reported that they ask the patient general questions (91%; $n = 355$) or they gather information from relatives (64%; $n = 253$). Only a quarter of them (24%; $n = 95$) indicated that they utilise cognitive tests and some did not perform any examinations at all to test for the possible occurrence of dementia (5%; $n = 22$) (Table 2).

Two of the most widely used tests for dementia evaluation, the MMSE and the CDT, are fairly well-known among respondents: most GPs reported that they knew CDT (89%; $n = 307$) and fewer people stated familiarity with MMSE (76%; $n = 265$). One-fifth (18%; $n = 63$) of the respondents said that they knew Early Mental Test, however, only a few GPs stated they were familiar with Mini-Cog (4%; $n = 17$) or GPCOG (1%; $n = 4$). More than two-thirds of respondents indicated they were (completely or mostly) satisfied with the CDT (69%; $n = 152$) while a slightly lower percentage of them expressed satisfaction with the MMSE (65%; $n = 98$) (Table 3).

Table 2. GPs' ways of dementia evaluation at their practices.

Dementia evaluation method	n	%
Asking general questions	355	91.0
Gathering information from relatives	253	64.9
Taking cognitive tests	95	24.4
No examination	22	5.6

Table 3. GPs' satisfaction regarding the Mini-Mental State Examination (MMSE) and the Clock Drawing Test (CDT).

	Level of satisfaction							
Test	1	2	3	4	5	n	M	SD
CDT	1.8%	13.8%	14.7%	46.3%	23.4%	218	3.76	1.021
MMSE	1.3%	17.3%	16.0%	49.3%	16.0%	150	3.61	0.995

1: not satisfied at all; 5: completely satisfied; M and SD: mean and standard deviation of the Likert-type scale values.

Table 4. GPs' views of the detection and management of dementia.

Statement	Level of agreement: Strongly agree	Mostly agree	Cannot decide	Mostly disagree	Strongly disagree	n	M (SD)
Screening in primary care leads to more effective outcomes in therapy.	35.1%	33.2%	19.8%	8.3%	3.5%	373	2.12 (1.089)
If conditions were suitable, I would implement screening tests for early detection of dementia.	25.7%	53.3%	6.4%	13.0%	1.6%	377	2.11 (0.987)
Managing dementia patients and their caregivers takes more time than I can afford at my practice.	31.9%	43.4%	5.2%	15.6%	3.9%	385	2.16 (1.150)
Currently available anti-dementia therapies are effective.	1.9%	14.8%	26.7%	37.2%	19.4%	371	3.57 (1.022)

1: Strongly agree; 2: Mostly agree; 3: Cannot decide; 4: Mostly disagree; 5: Strongly disagree. *M* and SD: mean and standard deviation of the Likert-type scale values.

Views regarding dementia identification and management

Supporting the importance of dementia recognition in the early stages, the vast majority (90%; $n = 352$) believed that early therapy could slow down symptom progression. GPs also held the view (97%; $n = 374$) that early detection enhanced both the patients' and their relatives' well-being.

Regarding their views on dementia testing and managing, participants were required to mark their answers on a 5-point Likert-type scale (strongly agree/ mostly agree and strongly disagree/mostly disagree responses are presented together). Three-fourths (75%; $n = 290$) of the GPs believed that managing dementia patients and their caregivers took more time than they could afford in their practice. Provided that conditions were suitable, the majority (79%; $n = 298$) would implement standardised cognitive tests for the early detection. Despite that half of the respondents (56%; $n = 210$) felt that currently available anti-dementia therapies were ineffective, two-thirds of them (68%; $n = 255$) still believed that dementia already detected in primary care would lead to more effective outcomes in therapy (Table 4).

Suggestions for improvement of dementia detection: contributing factors and an optimal instrument

From a list of five contributing factors to a more effective dementia examination routine, GPs marked the items as necessary with the following percentages: more time for patients (81%; $n = 311$), up-to-date tests (with a maximum of 5 min needed for administration and evaluation) (77%; $n = 297$), help from assistants (50%; $n = 192$), more staff (44%; $n = 170$), and finally, more examination rooms (26%; $n = 103$). Regarding an optimal, up-to-date instrument, GPs preferred a pen-and-paper test that could be administered by an assistant or the patients themselves and would include information from the patients' caregivers (detailed results are provided in Table 5).

Table 5. GPs' ideas about an optimal cognitive screening tool.

Aspects	Options	n	%
Test administrator ($n = 230$)	Assistant	87	37.8
	Patient	86	37.4
	GP	54	23.5
	Caregiver	3	1.3
Caregiver information ($n = 317$)	Containing	278	87.7
	Not containing	39	12.3
Test format ($n = 321$)	Pen-and-paper test	265	82.6
	Computer-based test program	48	15.0
	Online test	8	2.5
Maximum administration time ($n = 330$)	Up to 5 min	189	57.3
	Up to 10 min	110	33.3
	Up to 15 min	31	9.4

Estimated recognition of dementia

Finally, GPs were asked to estimate the recognition rate of dementia in Hungarian primary care and their practice. Regarding primary care, almost two-thirds (62%; $n = 226$) thought case recognition is under 30% and only very few (7%; $n = 27$) estimated that dementia is recognised in more than 60% of the cases. However, when asked about their recognition rate, half of them (49%; $n = 180$) said that they recognise a maximum of 30%, meanwhile, one-sixth (16%; $n = 61$) reported that they detect more than 60%. Wilcoxon signed ranks test was performed and results suggested that GPs' estimation of their own dementia recognition rate was significantly higher than their estimations of recognition rate in primary care ($Z = -7.806$; $p < .000$).

Discussion

Main findings

Hungarian GPs are generally accepting of the idea of cognitive examinations for signs of possible dementia in primary care and more than two-thirds of them are satisfied with the most commonly used cognitive screening tests (MMSE and CDT). However, only a quarter of them uses standardised cognitive tests in their practices. GPs feel that early detection of dementia leads to more effective outcomes in therapy and serves the well-being of both patients and their families, however they remain ambivalent about the effectiveness of anti-dementia therapies. The most critical barriers towards effective dementia case-finding appear to be the insufficient conditions: mainly lack of time and quickly administrable instruments.

Interpretation of the study results

Our results revealed a discrepancy between GPs' overall attitudes regarding testing for dementia in primary care versus their actual habits. Even though GPs seem to be aware of the benefits of timely dementia detection and they know the most commonly used cognitive tools, only a quarter of them actually apply these tests for the purpose of dementia detection, while a few do not perform any examinations at all. A similar conflict was found regarding Dutch GPs' views and habits, who reported taking action at a more progressed stage of dementia, despite that they know the importance of early intervention [13]. The rare application of formal tests has been also observed in other European studies: many (85% of French, 79% of Swiss, 53% of Italian and 33% of Scottish) GPs reported that they did not regularly perform standard procedures in their diagnostic evaluation [14–17], with many preferring the use of non-standardised, general questions [18]. Although there are some exceptions: 92% of Irish GPs self-reported in a survey that they used an appropriate tool to evaluate their patients' cognition [19] and only 10% of German GPs did not use any screening instrument [20]. Although the trend of not performing formal tests seems to be widespread, missing data, especially from the East-Central European region only provides us with an incomplete image on the topic and raises difficulties with international comparisons.

Since Hungarian GPs seem to be ambivalent regarding the effectiveness of anti-dementia medication, their screening habits may reflect therapeutic nihilism. Some previous studies suggest so: e.g. half of the French GPs felt that it was not worth making a dementia diagnosis because of the ineffective pharmacological treatment [21].

Findings of the present study indicated that the main obstacle to testing for dementia might be short consultation time with patients (which is approximately 6 min in Hungary) [22]. Besides the shortage of time, GPs mentioned the need for quickly administrable cognitive tools and more help from health care staff. All of these concerns are reflected by previous studies (e.g. from the UK and the Netherlands) [23,24]. Despite the time restrictions, current views of scientific literature advocate for the integration of targeted case-finding approach into primary care, prompting early dementia identification [4].

Cultural differences in the attitudes towards age-related memory problems may also affect the success of dementia detection. In Hungary, dementia symptoms (especially in the earlier stages) are often overlooked and thus do not prompt taking steps towards recognition. The tolerance for cognitive decline associated with older age may be higher in Hungary compared to other countries, where elderly people live far from their families and lead a more independent life (e.g. the USA) that would be greatly endangered by a mental illness.

Implications for clinical practice

The underutilisation of validated cognitive tests might be partly due to the lack of agreements on the most effective ways of dementia recognition, leaving the GPs without an unambiguous source of reference. A crucial way to improve recognition rates would be the regular update of international and national dementia guidelines (e.g. the latest Hungarian version was in effect until 2008 and is about to be updated in 2019), which usually give suggestions on the most adequate testing methods for dementia recognition.

The underuse of standardised instruments and the underdiagnoses of dementia in primary care may also be attributed to the prioritisation of somatic diseases over cognitive problems among the elderly. Since more than 65% of people over the age of 65 have multiple chronic conditions [25], the examination of memory functions might end up at the bottom of the priority list [8]. Furthermore, the progression of dementia is a slow process and thus is less obvious than the sudden onset of a somatic, sometimes painful complaint requiring urgent examination.

Strengths and limitations

To the best of our knowledge, this was the first Hungarian study in which GPs were questioned about their routines and views regarding dementia recognition. Our study describes results drawn from a relatively large sample (8% of all practicing GPs in Hungary), with participants from all regions of Hungary.

When interpreting the results, some limitations should be considered. First, as our findings were based on the answers of self-recruited, voluntary participants, the results might represent the views and routines of a more motivated and competent sample of Hungarian GPs. Second, given the sensitive topic of dementia detection practices, the effect of social desirability bias should also be taken into account when interpreting the results. Third, since a pen-and-paper questionnaire was applied, it could not be ensured that all 402 participants filled out all the questions, thus resulting in different numbers of missing responses throughout the survey and limiting the validity of questions with less responses. Regarding future works, it would be useful to recruit a representative sample of Hungarian GPs and also to apply qualitative methods to deepen our understanding on the topic further.

Conclusion

Although GPs in our sample seem to be aware of the benefits of dementia detection in primary care and also the concurrently low recognition rate in the country, the majority does not use formal cognitive tests for case-finding. Besides providing more favourable conditions (e.g. time and professional help), proper education and emphasising the benefits of early identification and treatment, the main way to improve the efficacy of recognition in primary care would be to construct a cost- and time-effective dementia identification strategy applicable in GPs' practices. With sufficient help, GPs could significantly improve the rate of detected dementias in Hungary, which also corresponds with the goals of international dementia plans.

Acknowledgments

The authors wish to thank all participating general practitioners for their contribution.

Author contributions

RB and NI analysed and interpreted all data and also wrote the manuscript. EP and SZH were members of the expert panel that designed the questionnaire and were also responsible for data collection. IK assisted with drafting and critically reviewing the manuscript. KK participated in the expert panel that designed the questionnaire and supervised the development of the paper. FH organised the data collection and also supervised the development of the paper. JK and MP were experts who participated in designing the questionnaire; they also assisted with drafting and critically reviewing the manuscript. All authors read and approved the final manuscript.

Disclosure statement

The authors report no conflicts of interest.

Funding

The authors RB and NI were supported by the University of Szeged, Faculty of Medicine (grant number: EFOP-3.6.3-VEKOP-16-2017-00009).

ORCID

Réka Balogh http://orcid.org/0000-0002-0983-9561
Nóra Imre http://orcid.org/0000-0003-4212-8638

References

[1] Wilkinson D, Stave C, Keohane D, et al. The role of general practitioners in the diagnosis and treatment of Alzheimer's disease: a multinational survey. J Int Med Res. 2004;32(2):149–159.
[2] Érsek K, Kárpáti K, Kovács T, et al. A dementia epidemiológiája Magyarországon [Epidemiology of dementia in Hungary]. Ideggyogy Sz. 2010;63:175–182.
[3] Takacs R, Ungvari GS, Gazdag G. Demenciában szenvedő páciensek akut pszichiátriai osztályra történő felvételének okai [Reasons for acute psychiatric admission of patients with dementia]. Neuropsychopharmacol Hung. 2015;17:141–145.
[4] Ranson JM, Kuźma E, Hamilton W, et al. Case-finding in clinical practice: an appropriate strategy for dementia identification? Alzheimers Dement. 2018;4: 288–296.
[5] Boustani M, Peterson B, Hanson L, et al. Screening for dementia in primary care: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med. 2003;138(11):927–937.
[6] Kaduszkiewicz H, Wiese B, van den Bussche H. Self-reported competence, attitude and approach of physicians towards patients with dementia in ambulatory care: results of a postal survey. BMC Health Serv Res. 2008;8.
[7] Hansen EC, Hughes C, Routley G, et al. General practitioners' experiences and understandings of diagnosing dementia: factors impacting on early diagnosis. Soc Sci Med. 2008;67(11):1776–1783.
[8] Boise L, Camicioli R, Morgan DL, et al. Diagnosing dementia: perspectives of primary care physicians. Gerontologist. 1999;39(4):457–464.

[9] Petrazzuoli F, Vinker S, Koskela TH, et al. Exploring dementia management attitudes in primary care: a key informant survey to primary care physicians in 25 European countries. Int Psychogeriatr. 2017;29(9): 1413–1423.
[10] Imre N, Balogh R, Papp E, et al. Knowledge of general practitioners on dementia and mild cognitive impairment: a cross-sectional, questionnaire study from Hungary. Educ Gerontol. 2019;45(8):495–505.
[11] Heim S, Busa C, Pozsgai É, et al. Hungarian general practitioners' attitude and the role of education in dementia care. Prim Health Care Res Dev. 2019; 20(e92):1–6.
[12] Hungarian Central Statistical Office. Number of practicing general practitioners in Hungary (2007–2017). [Data base]. [cited 2019 Sept 25]. Available from: http://statinfo.ksh.hu/Statinfo/index.jsp.
[13] Van Hout H, Vernooij-Dassen M, Bakker K, et al. General practitioners on dementia: task, practices and obstacles. Patient Educ Couns. 2000;39(2-3):219–225.
[14] Giezendanner S, Monsch AU, Kressig RW, et al. Early diagnosis and management of dementia in general practice – how do Swiss GPs meet the challenge? Swiss Med Wkly. 2018;148:w14695.
[15] McIntosh IB, Swanson V, Power KG, et al. General practitioners' and nurses' perceived roles, attitudes and stressors in the management of people with dementia. Health Bull. 1999;57(1):35–43.
[16] Somme D, Gautier A, Pin S, et al. General practitioner's clinical practices, difficulties and educational needs to manage Alzheimer's disease in France: analysis of national telephone-inquiry data. BMC Fam Pract. 2013;14(1):81.
[17] Veneziani F, Panza F, Solfrizzi V, et al. Examination of level of knowledge in Italian general practitioners attending an education session on diagnosis and management of the early stage of Alzheimer's disease: pass or fail? Int Psychogeriatr. 2016;28(7): 1111–1124.
[18] Allen M, Ferrier S, Sargeant J, et al. Alzheimer's disease and other dementias: an organizational approach to identifying and addressing practices and learning needs of family physicians. Educ Gerontol. 2005;31(7):521–529.
[19] Dyer AH, Foley T, O'Shea B, et al. Dementia diagnosis and referral in general practice: a representative survey of Irish general practitioners. Irish Med J. 2018; 111(4):735.
[20] Thyrian JR, Hoffmann W. Dementia care and general physicians – a survey on prevalence, means, attitudes and recommendations. Cent Eur J Public Health. 2012;20(4):270–275.
[21] Harmand M-C, Meillon C, Rullier L, et al. Description of general practitioners' practices when suspecting cognitive impairment. Recourse to care in dementia (Recaredem) study. Aging Ment Health. 2018;22(8): 1046–1055.
[22] Irving G, Neves AL, Dambha-Miller H, et al. International variations in primary care physician consultation time: a systematic review of 67 countries. BMJ Open. 2017;7(10):e017902.
[23] Chithiramohan A, Iliffe S, Khattak I. Identifying barriers to diagnosing dementia following incentivisation and policy pressures: general practitioners' perspectives. Dementia (London). 2019;18(2):514–529.
[24] Prins A, Hemke F, Pols J, et al. Diagnosing dementia in Dutch general practice: a qualitative study of GPs' practices and views. Br J Gen Pract. 2016;66(647): e416–e422.
[25] Lehnert T, Heider D, Leicht H, et al. Review: health care utilization and costs of elderly persons with multiple chronic conditions. Med Care Res Rev. 2011; 68(4):387–420.

II.

Journal of the International Neuropsychological Society (2022), 1–13

doi:10.1017/S1355617721001454

The Role of Silence in Verbal Fluency Tasks – A New Approach for the Detection of Mild Cognitive Impairment

Réka Balogh[1,*], Nóra Imre[1], Gábor Gosztolya[2], Ildikó Hoffmann[3,4], Magdolna Pákáski[1] and János Kálmán[1]

[1]Department of Psychiatry, University of Szeged, Szeged, Hungary
[2]ELRN-SZTE Research Group on Artificial Intelligence, Szeged, Hungary
[3]Department of Hungarian Linguistics, University of Szeged, Szeged, Hungary
[4]Hungarian Research Centre for Linguistics, ELRN, Budapest, Hungary

(Received March 5, 2021; Final revision November 15, 2021; Accepted December 15, 2021)

Abstract

Objective: Most recordings of verbal fluency tasks include substantial amounts of task-irrelevant content that could provide clinically valuable information for the detection of mild cognitive impairment (MCI). We developed a method for the analysis of verbal fluency, focusing not on the task-relevant words but on the silent segments, the hesitations, and the irrelevant utterances found in the voice recordings. **Methods:** Phonemic (‘k’, ‘t’, ‘a’) and semantic (animals, food items, actions) verbal fluency data were collected from healthy control (HC; $n = 25$; $M_{age} = 67.32$) and MCI ($n = 25$; $M_{age} = 71.72$) participants. After manual annotation of the voice samples, 10 temporal parameters were computed based on the silent and the task-irrelevant segments. Traditional fluency measures, based on word count (correct words, errors, repetitions) were also employed in order to compare the outcome of the two methods. **Results:** Two silence-based parameters (the number of silent pauses and the average length of silent pauses) and the average word transition time differed significantly between the two groups in the case of all three semantic fluency tasks. Subsequent receiver operating characteristic (ROC) analysis showed that these three temporal parameters had classification abilities similar to the traditional measure of counting correct words. **Conclusion:** In our approach for verbal fluency analysis, silence-related parameters displayed classification ability similar to the most widely used traditional fluency measure. Based on these results, an automated tool using voiced-unvoiced segmentation may be developed enabling swift and cost-effective verbal fluency-based MCI screening.

Keywords: Cognitive aging, Mild cognitive impairment, Neuropsychology, Verbal fluency, Semantic memory, Speech parameters

INTRODUCTION

Mild cognitive impairment (MCI) is a heterogeneous clinical syndrome, often considered a transitional stage between healthy cognitive aging and dementia (Petersen, 2004), and it is also associated with an increased risk of developing dementia later on (Roberts et al., 2014). Early recognition and timely diagnosis are crucial in MCI, because they can provide an opportunity to reduce the rate of cognitive decline (Hahn & Andel, 2011), while also offering a chance for the patients and their relatives to start planning for the future (Knopman & Petersen, 2014). Considering the high prevalence of MCI (Roberts & Knopman, 2013) and especially the constantly overburdened clinical settings, it would be beneficial to replace the current labor-intensive and time-consuming assessments of cognitive functioning with swift, low-cost, and preferably automated tools.

Verbal fluency tests are neuropsychological tests, extensively used both in research and in the clinical practice. In the standard versions of the fluency tests, participants are given 60 s to list as many words as they can, beginning with a given letter (phonemic fluency) (Borkowski, Benton, & Spreen, 1967) or belonging to a given semantic category (semantic fluency) (Newcomb, 1969). There is an additional, third type of verbal fluency task: action fluency (or verb fluency), where the patients have to produce as many verbs (‘things that people do’) as they can (Piatt, Fields, Paolo, & Troster, 1999). However, in the current study, for the sake of simplicity, action fluency will be regarded as a semantic fluency task, because both semantic fluency and action fluency are content-oriented speech tasks (Östberg, Fernaeus, Hellstrom, Bogdanovic, & Wahlund, 2005).

*Correspondence and reprint requests to: Réka Balogh, Department of Psychiatry, University of Szeged, Korányi Fasor 8-10, Szeged, H-6720, Hungary. E-mail: balogh.reka@med.u-szeged.hu

Both phonemic and semantic fluency tasks require rapid associative exploration; however, semantic fluency relies more on semantic associations and reflects more on the integrity of semantic memory. On the other hand, phonemic fluency depends more on search strategies based on lexical representation (Henry, Crawford, & Phillips, 2004; Teng et al., 2013). Executive control processes also play a major role in the execution of verbal fluency tests, because during the task, subjects not only need to remember the exact instruction and keep the already used responses in mind, but they must also repress the repetitions and other potentially incorrect or irrelevant responses (Shao, Janse, Visser, & Meyer, 2014). Fluency tests have been validated in the assessment of verbal and executive skills (Shao et al., 2014), and both of these abilities have been reported to deteriorate in dementia and in other forms of cognitive impairments. Therefore, fluency tests have a great potential as effective screening tools for MCI (García-Herranz, Diaz-Mardomingo, Venero, & Peraita, 2020; McDonnell et al., 2020).

The traditional, most common approach for the assessment of verbal fluency performance requires the clinician to count the number of unique and correct words, along with the number of errors and the number of repetitions produced by the participant. This analysis can be refined by scoring the number of correct words based on time intervals (e.g., 0–20, 21–40, 41–60 s) (Demetriou & Holtzer, 2017; Jacobs, Mercuri, & Holtzer, 2021). Moving beyond simple word counts, a more sophisticated, qualitative method can be applied, which is called clustering. In this method, consecutive words are clustered based on linguistic similarity or a shared category (e.g., rhyming words in the case of phonemic fluency tasks, or pets in the case of the animal fluency task). Thus, the average sizes of the clusters and the number of switches between these clusters can be examined (Troyer, Moscovitch, & Winocur, 1997). Even though this approach may provide more information about the underlying mental processes, it is also relatively time-consuming. Furthermore, compared to the most widespread, word count-based assessment, this method requires the manual coding and grouping of words, which may even raise reliability issues (Taler, Johns, & Johns, 2020).

Recently, there have been multiple attempts with different approaches to overcome the disadvantages of the above-mentioned methods by introducing automated analyses. These approaches have the benefit of being objective, repeatable, and they also yield quick output (König et al., 2018). The majority of these methods focus on the computation and analysis of semantic clusters. Latent semantic analysis (LSA) can be applied to examine the strength of the semantic relationship of two consecutive words by constructing a co-occurrence matrix for all of the words found in a given corpus of text (Ledoux et al., 2014; Pakhomov & Hemmy, 2014). A more recent computational method, called explicit semantic analysis (ESA), examines Wikipedia entries for the quantification of relationships between words based on different types of similarities (e.g., taxonomic, geographic, or linguistic) (Woods, Wyma, Herron, & Yund, 2016). It is also possible to combine semantic measures with temporal information. In this approach, the recalled words are organized in clusters defined semantically and also in clusters based on the temporal proximity of the words (Tröger et al., 2019). Verbal fluency tasks can also be analyzed by exploring certain speech features that can be automatically extracted from fluency voice recordings (Lopez-de-Ipina et al., 2015).

However, there is a major obstacle in the application of the automatic analysis of fluency recordings that stems from the general characteristics of the responses produced by the participants: most voice recordings of fluency test performances contain more than just a sequence of task-relevant words. The recordings also contain speech segments irrelevant in terms of the task, including filler words or hesitations, irrelevant comments, questions directed at the examiner, or loud thinking. To be able to automatically analyze the task-relevant words, fluency recordings need to go through a time-consuming preparation process prior to the analysis: the words irrelevant to the task need to be removed from the recording or transcript, and some words need to be lemmatized (i.e., converted to their stem) (Chen et al., 2020; Holmlund, Cheng, Foltz, Cohen, & Elvevag, 2019).

Given the substantial amount of task-irrelevant content in most fluency recordings, the question arises whether the analysis of these segments could provide valuable information regarding the overall verbal fluency performance of the patient. After manually annotating the recordings, we derived temporal parameters that, instead of targeting the task-relevant words, contained the silent segments, the hesitations, and the utterances irrelevant to the task. Therefore, the focus of this exploratory study was to move beyond the words recalled by the participants and explore the additional, previously unharvested information present in the fluency recordings. It should be noted that this approach, similarly to the previously summarized methods, required substantial manual work. However, in the future (depending on the characteristics of the given parameter) it could allow the development of automatic analysis.

Our main goal was: (1) to examine whether these parameters can differentiate between participants classified as healthy control (HC) and as MCI (*temporal analysis method*). Besides the temporal parameters, traditional fluency scores (number of correct words, errors, and repetitions) were also calculated for the same fluency recordings (*traditional analysis method*). We sought; (2) to compare the two methods of analysis regarding their ability to detect differences in the performance of the HC and MCI groups. The inclusion of both phonemic and semantic fluency tasks in the research protocol also allowed us; and (3) to compare the different types of fluency tasks to investigate their sensitivity to the presence of MCI.

METHODS

Participants

Participants (patients and their relatives, scheduled for consultations) were recruited at the Memory Clinic of the Department of Psychiatry, University of Szeged (Szeged, Hungary). Data collection was carried out between February 2018 and March 2020.

The required sample size for the study was assessed a priori using G * Power v.3.2.9.7. (Faul, Erdfelder, Lang, & Buchner, 2007) with the settings of effect size $d = 0.8$; alpha error probability: 0.05, power (1-beta error probability): 0.8. Based on this, the optimal sample size was calculated as 52, which later (due to COVID-19 regulations halting data collection in clinical research) was limited to 50. Initially, a total of 79 individuals were recruited to take part in the study.

Inclusion criteria included at least 50 years of age, a minimum of 8 years of formal education, and Hungarian as a native language. Individuals were excluded if they had any past or present neuropsychological, psychotic or mood disorders, head injuries, stroke, substance abuse disorders, major (uncorrected) hearing loss, or language problems (e.g., stutter), based on patient history and medical records. Participants with MRI or CT records showing evidence of micro- or macrohemorrhages, lacunar or other infarctions, cerebral contusion, encephalomalacia, aneurysm, vascular malformation, or space-occupying lesions were also excluded.

In addition, the two main exclusion criteria were the presence of dementia or major cognitive deficits and depression. To rule out possible cases of dementia, the Mini-Mental State Examination (MMSE) (Folstein, Folstein, & Mchugh, 1975) was applied as a screening tool: participants with a score of 24 or below were excluded from the study. The possibility of depression was assessed using the 15-item version of the Geriatric Depression Scale (GDS-15) (Yesavage & Sheikh, 1986): participants scoring 7 or above on the test were excluded. After reviewing and evaluating the criteria, 50 subjects were considered eligible for inclusion in the study (Figure 1).

Participants were split into two groups based on their MMSE scores. MMSE cut-off scores were determined based on the results of previous studies conducted by our research group: in these works, the mean scores of MMSE emerged as $29.17 \pm 0.71/29.24 \pm 0.523$ for the HC and $26.97 \pm 0.96)/27.16 \pm 0.898$ for the MCI group (Gosztolya et al., 2019; Toth et al., 2018). Hence, participants achieving a score of 29 to 30 points were considered as healthy control (HC) subjects, while participants achieving a score of 25 to 28 points formed the MCI group. The subtypes of MCI (amnestic or non-amnestic) were not considered. The two groups showed no significant difference in gender and years of education. However, participants of the MCI group were significantly older than the participants enrolled in the HC group. No significant difference was found in the GDS-15 score between the two groups (Table 1).

Study Protocol

Each participant performed a series of neuropsychological tests: six fluency tasks, the Digit Span Test – Forward and Backward (Wechsler, 1981), the Non-Word Repetition Test (Gathercole, Willis, Baddeley, & Emslie, 1994), the Listening Span Test (Daneman & Carpenter, 1980), the Clock Drawing Test (Shulman, Shedletsky, & Silver, 1986) and the Alzheimer's Disease Assessment Scale – Cognitive Subscale (ADAS-Cog) (Rosen, Mohs, & Davis, 1984). The fluency tasks were implemented in a fixed order, separated by the five shorter cognitive tests, while ADAS-Cog was administered at the very end of the study protocol to prevent fatigue. We also ensured that tasks assessing the same cognitive domain did not follow each other directly.

In the three phonemic fluency tasks, the participants were asked to list as many words as they can, starting with the letters 'k', 't', and 'a', respectively, while avoiding proper nouns. For the semantic fluency tasks, participants had to name as many animals, food items, and actions (verbs – 'things that people do') as they could. The participants were instructed to avoid saying variations of the same word stem (e.g., horse, horses; go, goes). For all 6 verbal fluency tasks, participants had 1 min to perform the task. The 1-min interval began with the investigator saying: 'Start.' Every verbal fluency task was recorded using an Olympus Digital Voice Recorder (16 kHz sampling rate, 16-bit resolution). The recordings were also transcribed manually for the calculation of the traditional scores. Therefore, fluency performances were analyzed in two ways: by implementing the novel temporal parameters, and also by using the traditional method, based on word count.

Analysis Method Based on Temporal Parameters

Manual transcription process of the fluency recordings

Voice recordings of all fluency tasks were manually transcribed in Praat, a free language software enabling speech analysis (Boersma & Weenink, 2020). The transcription process was supervised by a linguist specialized in language pathologies (I. H.), while quality control was ensured by an expert in the field of computational speech processing (G. G.). Due to the quality of their recordings, an HC participant's animal category fluency task and an MCI participant's 'k' letter fluency task were unsuitable for transcription; therefore, these recordings were not considered in the analysis of temporal parameters, but they were included in the traditional analysis.

Annotation of speech features in the verbal fluency recordings

The transcriptions of the fluency recordings contained not only the task-relevant answers of the participants (the recalled words – including correct, incorrect, and repeated words), but also silent pauses, hesitation sounds (filled pauses, like 'hmm' and 'er'), and irrelevant utterances, such as comments or loud thinking said by the subjects (e.g., 'did I say this before?', 'uh, it's not an easy task, let me think . . . '). False starts ('te-. . . tiger'), as well as laughing and coughing sounds were also annotated. The laughing, coughing, and false starts parameters were considered unintentional and

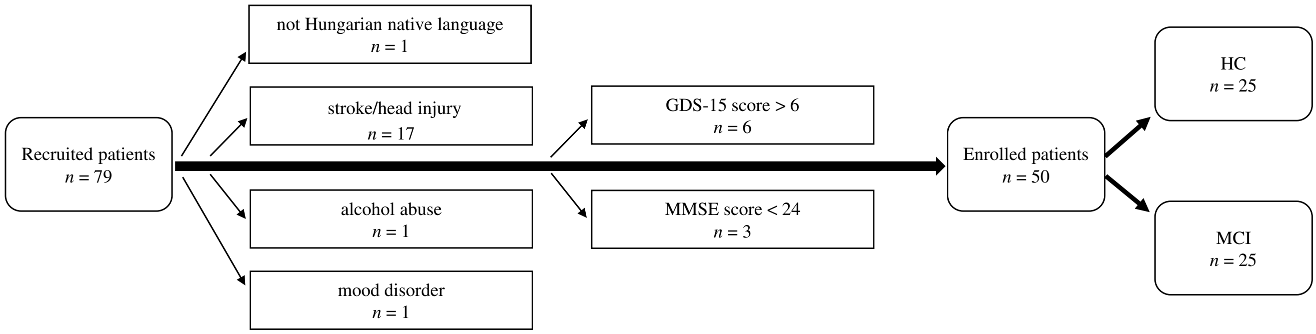

Fig. 1. Flowchart of the participant exclusion process. (GDS-15:15-item Geriatric Depression Scale; MMSE: Mini-Mental State Examination; HC: healthy control; MCI: mild cognitive impairment).

Table 1. Descriptive and comparative statistics for the demographic characteristics and neuropsychological test scores of the study participants

	HC ($n = 25$)	MCI ($n = 25$)		
	M (*SD*)		Comparative test statistics	*p*
Demographics				
Gender (male/female)	8/17	7/18	$\chi^2(1) = 0.095$	0.758
Age (years)	67.32 (8.300)	71.72 (5.435)	$U = 187.000$; $Z = -2.440$	**0.015**
Education (years)	13.48 (2.632)	12.36 (2.827)	$U = 255.500$; $Z = -1.136$	0.256
Neuropsychological test scores				
MMSE	29.44 (0.507)	26.96 (1.060)	$U = 0.000$; $Z = -6.202$	< **0.001**
GDS-15	1.84 (1.724)	2.40 (1.225)	$U = 232.500$; $Z = -1.587$	0.112

HC: healthy control; MCI: mild cognitive impairment; MMSE: Mini-Mental State Examination; GDS-15: 15-item Geriatric Depression Scale.
Significant *p*-values ($p < 0.05$) are in **bold.**

were discarded from further analysis because we found that the number of these occurrences was negligible.

Calculation of temporal parameters based on the speech features

For each recording, task-relevant words, silent segments, hesitation sounds, and irrelevant utterances were annotated based on their boundaries (their exact start and end times), providing their duration measures. Based on this, *the total number, the average length,* and *the total length of silent pauses*; *the total number, the average length,* and *the total length of hesitations*; and *the total number, the average length, and the total length of irrelevant utterances* were calculated. Besides these parameters, the mean time between two consecutive task-relevant words (*average word transition time*) was also calculated based on the transcript. Not only correct words but also the errors and repetitions were considered task-relevant words. The average word transition time, irrelevant of its content, such as silent pause, hesitation, or irrelevant utterance, provided information about the average time the participant needed to produce a new task-relevant word, and because of this, it had a positive association with the average and total length of silent pauses, hesitations, and irrelevant utterances.

It is worth noting that because of the distinctive regular rhythm that is inherent in verbal fluency performances, each of the task-relevant words listed by the participants was separated by a silent pause (irrelevant of its length). Consequently, the number of silent pauses increased in parallel with the number of task-relevant words said by the participant. Therefore, analyzing the number of silent pauses can be viewed as the converse of the traditional approach of counting only the task-relevant words.

The parameters used in the study are listed and defined in Table 2; two waveform extracts from a fluency task performed by an HC and an MCI subject are shown in Figure 2.

Traditional Fluency Analysis Based on Word Count

In the traditional scoring method (Lezak, 2012), we calculated the number of correct words, the number of errors, and the number of repetitions or perseverations; the last two were considered as one variable. In the case of animal fluency, when a participant recalled synonymous words (e.g., cat and kitten), variations in gender (e.g., hen and rooster), or an animal and its offspring (e.g., horse and foal), words were only scored as one. The participants did not receive points for naming a subcategory if they also gave specific examples of it [e.g., in the case of food items: fruit (0 points), apple (1 point), pear (1 point)].

https://doi.org/10.1017/S1355617721001454 Published online by Cambridge University Press

Table 2. List and definitions of the temporal parameters

Temporal fluency parameters	Description
Silent pause parameters	
Total number of silent pauses (count)	Number of silent segments
Average length of silent pauses (s)	Average length of silent segments
Total length of silent pauses (s)	Total length of silent segments
Hesitation parameters	
Total number of hesitations (count)	Total number of filled pauses (e.g., 'hmm', 'umm')
Average length of hesitations (s)	Average length of filled pauses (e.g., 'hmm', 'umm')
Total length of hesitations (s)	Total length of filled pauses (e.g., 'hmm', 'umm')
Irrelevant utterances parameters	
Total number of irrelevant utterances (count)	Total number of filler words and comment blocks (including articles and conjunctions)
Average length of irrelevant utterances (s)	Average length of filler words and comment blocks (including articles and conjunctions)
Total length of irrelevant utterances (s)	Total length of filler words and comment blocks (including articles and conjunctions)
Average word transition time (s)	Mean period of time between two consecutive 'task-oriented' words

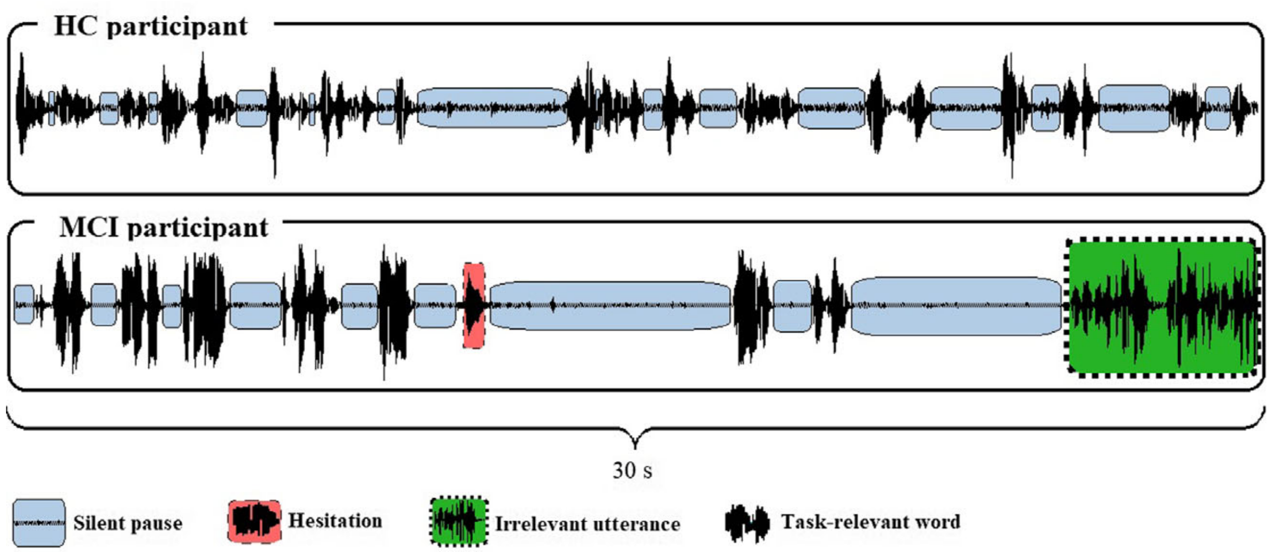

Fig. 2. Waveforms extracted from the food item fluency recordings of two participants. (Extracted from Praat. HC: healthy control; MCI: mild cognitive impairment).

Statistical Analysis

Descriptive statistical analysis was used to examine the demographic features, the neuropsychological test scores, and the fluency measures of the participants. The assumption of normality was not met according to the results of the Shapiro–Wilk test in more than two-thirds of the cases, therefore, in order to obtain comparable statistical measures, comparisons between the HC and the MCI groups were executed using the Mann–Whitney *U* test. Categorical variables were compared using the Chi-square test. Effect sizes were calculated using the Pearson correlation coefficient ($r = \frac{z}{\sqrt{N}}$) (Rosenthal, 1991).

Receiver operating characteristic (ROC) analysis was applied to assess the classification abilities of the temporal parameters and the traditional scores. Sensitivity and specificity were calculated using threshold values that yielded the highest possible sensitivity (while keeping specificity at a minimum of 50%). For the comparison of classification abilities, the differences between the area under the curve variables (AUCs) were compared based on the method of DeLong, DeLong, and Clarke-Pearson (1988).

For all statistical comparisons, the level of significance was set at $p < 0.05$. All analyses were performed using SPSS v.24 (IBM SPSS Statistics for Windows, 2016), except for the comparison of AUCs, for which the MedCalc Statistical Software v.19.6. (MedCalc Software, 2020) was utilized.

RESULTS

Temporal Parameters of Verbal Fluency Performance

Considering the phonemic fluency tasks, in the 'a' fluency, the average length and the total length of irrelevant utterances

https://doi.org/10.1017/S1355617721001454 Published online by Cambridge University Press

were significantly higher in the MCI group, while none of the temporal parameters differed between the two groups in the case of the 'k' and 't' phonemic fluencies (Table 3). Regarding the three semantic fluencies, the total number of silent pauses were significantly higher in the HC group in the animal and action fluency tasks, whereas the average length of silent pauses and the average word transition time were significantly higher in the MCI group throughout all of the three tasks (Table 4).

Traditional Word Count Measures of Verbal Fluency Performance

In the three phonemic fluency tasks, no statistically significant difference was found between the groups regarding the number of correct words and the number of repetitions or perseverations. However, in the 'a' phonemic fluency task, participants from the MCI group produced more errors than participants from the HC group (Table 5). As for the semantic fluency tests, participants from the HC group had a significantly higher number of correct words in the case of all three (animals, food items, and actions) tasks. In the number of repetitions or perseverations, there was no statistically significant difference between the two study groups (Table 6).

ROC Analysis of the Significant Temporal Parameters

ROC analysis of the temporal parameters was carried out in the case of the five parameters that, based on the previously conducted comparative tests, showed significant differences between the HC and MCI groups.

The analysis revealed that the average length and the total length of irrelevant utterances had a significant classification ability in the case of the 'a' phonemic fluency, with the same sensitivity (80%) and specificity (52%) for both parameters. In the semantic fluency tests, the number of silent pauses had significant classification ability in both the animal and action fluency tests, while the average length of silent pauses and the average word transition time was shown to be able to discriminate between the groups in the case of all three semantic fluency tests. Sensitivity was the highest in the case of the average word transition time in the animal fluency test (sensitivity: 96.0%; specificity: 62.5%). Accuracy measures of the temporal parameters that differed between the groups are given in Table 7. For every ROC analysis, sensitivity and specificity were determined using threshold values optimal for early screening, i.e., maximizing the sensitivity, while keeping specificity greater than or equal to 50%.

ROC Analysis of the Significant Traditional Measures

ROC analysis was also executed on the traditional measures that showed significant differences between the HC and MCI groups, to determine the classification ability of these measures. The analysis revealed that the number of errors in the 'a' phonemic fluency test had no significant classification ability. With respect to semantic fluency tests, the number of correct words showed significant classification abilities in the case of the animal, the food item, and the action fluencies. The animal naming fluency showed the highest sensitivity of 100% (specificity: 56%). Accuracy measures of the traditional fluency scores that showed significant differences between the groups are given in Table 8.

Comparison of the Temporal and Traditional Measures Regarding their Classification Ability

Pairwise comparisons of AUCs were executed to compare the classification ability of the three significant temporal parameters (total number of silent pauses, average length of silent pauses, average word transition time) and the significant traditional measure (number of correct words) in the semantic fluency tasks. In the animal category fluency, the results indicated no significant differences regarding AUCs between the number of correct words and the total number of silent pauses ($Z = 1.433$, $p = 0.151$) or the average word transition time ($Z = 1.579$, $p = 0.114$), however, the classification ability of the average length of silent pauses was smaller ($Z = 2.043$, $p = 0.041$) compared to the correct word count. In the case of the food item fluency, no difference was found between the AUCs of the number of correct words and the average length of silent pauses ($Z = 0.978$, $p = 0.328$), and the average word transition time ($Z = 0.662$, $p = 0.508$). Furthermore, in action fluency, the classification ability of correct word count did not differ from either the total number of silent pauses ($Z = 0.267$, $p = 0.789$), the average length of silent pauses ($Z = 0.056$, $p = 0.954$) or the average word transition time ($Z = 0.046$, $p = 0.962$).

DISCUSSION

Main Findings

This study presents a new practical framework for verbal fluency analysis. To the best of our knowledge, we are the first to report on verbal fluency performance beyond the recalled words, focusing on the pauses and task-irrelevant content of speech in the fluency recordings. We quantitatively analyzed a number of temporal parameters that were calculated based on silent pauses, hesitations, and irrelevant speech segments annotated in the recordings. Our main finding is that in the case of semantic fluency tests, some of the temporal parameters based on silent pauses can discriminate between individuals with cognitive impairment and individuals with healthy cognition. These results suggest that the analysis of these temporal parameters may complement or even substitute the widely applied, but more time-consuming and labor-intensive traditional word scoring method, while still providing comparable classification ability.

Table 3. Descriptive measures and statistical comparison of the temporal parameters in the phonemic fluency tasks

Phonemic fluency tasks	HC	MCI	Mann–Whitey *U* test			Effect size[r]
Temporal parameters	*M* (*SD*)		*U*	*Z*	*p*	*r*
Letter 'k'	$n = 25$	$n = 24^{*}$				
Total number of silent pauses (count)	19.040 (4.485)	17.291 (4.591)	230.500	-1.394	0.163	0.19
Average length of silent pauses (s)	2.438 (0.941)	2.767 (1.031)	244.000	-1.120	0.263	0.16
Total length of silent pauses (s)	42.569 (6.571)	43.532 (5.688)	278.000	-0.440	0.660	0.06
Total number of hesitations (count)	2.000 (2.645)	1.708 (2.095)	281.500	-0.382	0.702	0.05
Average length of hesitations (s)	0.482 (0.448)	0.398 (0.382)	279.500	-0.421	0.674	0.06
Total length of hesitations (s)	1.350 (1.737)	1.137 (1.334)	283.000	-0.349	0.727	0.05
Total number of irrelevant utterances (count)	3.280 (4.559)	4.333 (3.818)	225.000	-1.517	0.129	0.22
Average length of irrelevant utterances (s)	1.021 (0.666)	1.242 (0.851)	274.000	-0.520	0.603	0.07
Total length of irrelevant utterances (s)	4.283 (7.149)	4.889 (3.609)	213.000	-1.742	0.082	0.25
Average word transition time (s)	4.505 (2.687)	5.159 (3.979)	230.000	-1.400	0.162	0.20
Letter 't'	$n = 25$	$n = 25$				
Total number of silent pauses (count)	18.320 (4.269)	16.920 (5.259)	257.000	-1.081	0.280	0.15
Average length of silent pauses (s)	2.521 (0.879)	2.993 (1.542)	261.000	-0.999	0.318	0.14
Total length of silent pauses (s)	42.847 (5.770)	43.834 (5.666)	278.000	-0.669	0.503	0.07
Total number of hesitations (count)	1.480 (2.023)	1.720 (2.051)	290.000	-0.455	0.649	0.06
Average length of hesitations (s)	0.520 (0.509)	0.443 (0.348)	293.500	-0.375	0.708	0.05
Total length of hesitations (s)	1.128 (1.504)	1.069 (1.326)	312.500	0.000	1.000	0.00
Total number of irrelevant utterances (count)	3.240 (3.562)	3.720 (2.806)	256.500	-1.097	0.273	0.16
Average length of irrelevant utterances (s)	0.967 (0.580)	1.228 (0.616)	231.500	-1.573	0.116	0.21
Total length of irrelevant utterances (s)	4.154 (5.656)	4.825 (3.379)	234.500	-1.515	0.130	0.21
Average word transition time (s)	3.816 (1.739)	4.944 (3.045)	250.000	-1.213	0.225	0.17
Letter 'a'	$n = 25$	$n = 25$				
Total number of silent pauses (count)	13.920 (3.639)	14.120 (4.876)	298.000	-0.283	0.778	0.04
Average length of silent pauses (s)	3.636 (1.446)	3.853 (2.834)	268.000	-0.863	0.388	0.12
Total length of silent pauses (s)	45.881 (5.219)	43.042 (7.551)	235.00	-1.504	0.133	0.01
Total number of hesitations (count)	1.040 (1.059)	1.200 (1.354)	311.000	-0.031	0.976	0.00
Average length of hesitations (s)	0.640 (0.565)	0.462 (0.485)	263.500	-0.974	0.330	0.14
Total length of hesitations (s)	0.973 (1.316)	0.985 (1.254)	301.500	-0.219	0.827	0.03
Total number of irrelevant utterances (count)	3.480 (4.154)	4.560 (3.292)	214.500	-1.918	0.055	0.27
Average length of irrelevant utterances (s)	1.065 (0.701)	1.630 (0.725)	180.000	-2.572	**0.010**	0.36
Total length of irrelevant utterances (s)	4.637 (5.286)	7.160 (5.322)	204.000	-2.106	**0.035**	0.30
Average word transition time (s)	5.115 (2.651)	5.224 (2.839)	286.000	-0.514	0.607	0.07

M: mean; *SD*: standard deviation; HC: healthy control; MCI: mild cognitive impairment.
Significant *p*-values ($p < 0.05$) are in **bold**.
*One fluency voice recording was unsuitable for transcription.
[r] Effect size is calculated as Pearson's *r*, expressed in absolute value.
Strength of association: 0.1 to 0.3: small, 0.3 to 0.5: medium, 0.5 to 1.0: large (Cohen, 1988).

Three temporal parameters (total number of silent pauses, average length of silent pauses, and average word transition time) consistently differed between the HC and MCI groups in the case of the semantic (animal, food item, and action) fluency tests. In the phonemic fluency tests, differences could only be observed in the case of the 'a' phonemic fluency, where the average and total lengths of irrelevant utterances showed significant differences.

It should be noted that the direction of differences in the silence-based parameters might seem inconsistent: the average lengths of the silent pauses and the average word transition times were longer in the MCI group, whereas HC participants had a higher number of silent pauses in the case of the semantic tasks. Since silent pauses were defined as the absence of speech/sound regardless of length, every detectable silent segment found in the recordings was annotated as a silent pause, including even the brief transitions between words. Therefore, the number of silent pauses was increased by the number of words uttered by the participant. Since the HC group produced significantly more correct words in semantic fluency tasks, the number of silent pauses was also significantly higher in this group.

Table 4. Descriptive measures and statistical comparison of the temporal parameters in the semantic fluency tasks

Semantic fluency tasks	HC	MCI	Mann–Whitney *U* test			Effect size[r]
Temporal parameters	*M* (*SD*)		*U*	*Z*	*p*	*r*
Animals	$n = 24^*$	$n = 25$				
Total number of silent pauses (count)	25.666 (4.603)	21.760 (4.968)	156.000	-2.890	**0.004**	0.41
Average length of silent pauses (s)	1.437 (0.445)	1.883 (0.718)	179.000	-2.420	**0.016**	0.34
Total length of silent pauses (s)	35.489 (6.485)	37.982 (8.193)	229.000	-1.420	0.156	0.20
Total number of hesitations (count)	3.166 (2.371)	3.240 (3.620)	271.000	-0.586	0.558	0.08
Average length of hesitations (s)	0.564 (0.290)	0.460 (0.358)	237.500	1.255	0.209	0.18
Total length of hesitations (s)	2.195 (1.982)	2.139 (2.820)	264.500	-0.713	0.476	0.10
Total number of irrelevant utterances (count)	3.333 (3.595)	5.120 (4.850)	231.500	-1.380	0.167	0.20
Average length of irrelevant utterances (s)	1.019 (0.641)	1.146 (0.727)	277.000	-0.461	0.645	0.07
Total length of irrelevant utterances (s)	4.379 (6.116)	6.562 (5.647)	220.000	-1.603	0.109	0.23
Average word transition time (s)	2.021 (0.756)	2.852 (0.841)	128.000	-3.440	**0.001**	0.49
Food items	$n = 25$	$n = 25$				
Total number of silent pauses (count)	25.400 (6.062)	21.720 (5.926)	216.000	-1.877	0.061	0.26
Average length of silent pauses (s)	1.395 (0.504)	1.888 (0.937)	201.000	-2.163	**0.031**	0.30
Total length of silent pauses (s)	33.192 (6.464)	36.368 (7.200)	242.000	-1.368	0.171	0.19
Total number of hesitations (count)	2.600 (2.432)	2.600 (2.661)	307.000	-0.109	0.913	0.02
Average length of hesitations (s)	0.444 (0.348)	0.494 (0.435)	306.000	-0.128	0.898	0.02
Total length of hesitations (s)	1.636 (1.544)	1.855 (2.015)	302.000	-0.207	0.836	0.03
Total number of irrelevant utterances (count)	3.600 (3.905)	4.360 (4.733)	294.000	-0.362	0.717	0.05
Average length of irrelevant utterances (s)	0.772 (0.581)	1.051 (1.028)	273.000	-0.770	0.441	0.11
Total length of irrelevant utterances (s)	3.716 (4.898)	5.210 (5.353)	259.000	-1.044	0.297	0.15
Average word transition time (s)	1.755 (0.770)	2.630 (1.356)	171.000	-2.746	**0.006**	0.40
Actions	$n = 25$	$n = 25$				
Total number of silent pauses (count)	24.240 (6.332)	19.080 (5.597)	184.000	-2.502	**0.012**	0.35
Average length of silent pauses (s)	1.600 (0.565)	2.373 (1.439)	192.000	-2.338	**0.019**	0.33
Total length of silent pauses (s)	35.898 (5.605)	38.524 (7.485)	230.000	-1.601	0.109	0.22
Total number of hesitations (count)	2.720 (2.282)	2.840 (2.511)	309.000	-0.069	0.945	0.01
Average length of hesitations (s)	0.547 (0.362)	0.554 (0.477)	292.000	-0.401	0.689	0.06
Total length of hesitations (s)	1.963 (1.741)	2.096 (2.290)	302.000	-0.205	0.837	0.03
Total number of irrelevant utterances (count)	4.040 (3.920)	4.160 (3.681)	307.500	-0.098	0.922	0.01
Average length of irrelevant utterances (s)	1.069 (0.626)	1.153 (0.760)	290.500	-0.427	0.669	0.06
Total length of irrelevant utterances (s)	4.302 (4.600)	5.188 (4.351)	273.500	-0.757	0.449	0.11
Average word transition time (s)	2.258 (0.996)	2.989 (1.199)	196.000	-2.260	**0.024**	0.32

M: mean; *SD*: standard deviation; HC: healthy control; MCI: mild cognitive impairment.
Significant *p*-values ($p < 0.05$) are in **bold**.
*One fluency voice recording was unsuitable for transcription.
[r] Effect size is calculated as Pearson's *r*, expressed in absolute value.
Strength of association: 0.1 to 0.3: small, 0.3 to 0.5: medium, 0.5 to 1.0: large (Cohen, 1988).

The average word transition time parameter also had a direct influence on the number of correct words. Since this parameter contains every task-irrelevant segment, the increase of the average word transition time by definition led to the decrease of the number of recalled words, therefore it could be viewed that these two parameters were somewhat inversely proportional. The average length of silent pauses parameter also affected the number of correctly recalled words. However, this is less of a general phenomenon, since the average length of silent pauses does not have a sole effect on the number of recalled words – it can be also significantly influenced by other task-irrelevant contents of speech (e.g., loud hesitations).

The importance of silent pauses has also been highlighted in the area of connected speech analysis: studies have shown that compared to HC subjects, participants with MCI produce more and longer silent pauses in their speech (Sluis et al., 2020; Toth et al., 2018). Even though spontaneous speech samples provide ecologically valid data, utilizing verbal fluency tests for the analysis of speech can be even more advantageous, as it can be combined with already standardized qualitative approaches. To be able to compare the results

https://doi.org/10.1017/S1355617721001454 Published online by Cambridge University Press

Table 5. Descriptive measures and statistical comparison of the traditional fluency scores in the phonemic fluency tests

	HC	MCI				
	M (*SD*)		Mann–Whitney *U* test			Effect size[r]
Traditional fluency scores of the phonemic fluency tasks	*n* = 25	*n* = 25	*U*	*Z*	*p*	*r*
Letter 'k'						
Correct words	13.68 (4.571)	11.52 (4.700)	227.000	-1.667	0.096	0.24
Errors	0.04 (0.200)	0.16 (0.374)	275.000	-1.400	0.162	0.20
Repetitions/perseverations	0.16 (0.374)	0.32 (0.690)	294.000	-0.537	0.591	0.08
Letter 't'						
Correct words	12.88 (4.314)	10.76 (4.371)	233.000	-1.547	0.122	0.22
Errors	0.20 (0.408)	0.28 (0.614)	307.500	-0.139	0.889	0.02
Repetitions/perseverations	0.48 (0.653)	0.28 (0.678)	248.500	-1.577	0.115	0.22
Letter 'a'						
Correct words	8.68 (3.424)	7.32 (3.987)	240.000	-1.416	0.157	0.20
Errors	0.12 (0.332)	0.72 (1.208)	231.500	-2.106	**0.035**	0.30
Repetitions/perseverations	0.20 (0.577)	0.20 (0.408)	292.500	-0.609	0.542	0.09

M: mean; *SD*: standard deviation; HC: healthy control; MCI: mild cognitive impairment).
Significant *p*-values ($p < 0.05$) are in **bold**.
[r] Effect size calculated as Pearson's *r*, expressed in absolute value. Strength of association: 0.1 to 0.3: small, 0.3 to 0.5: medium, 0.5 to 1.0: large (Cohen, 1988).

Table 6. Descriptive measures and statistical comparison of the traditional fluency scores in the semantic fluency tests

	HC	MCI				
	M (*SD*)		Mann–Whitney *U* test			Effect size[r]
Traditional fluency scores of the semantic fluency tasks	*n* = 25	*n* = 25	*U*	*Z*	*p*	*r*
Animals						
Correct words	20.54 (4.412)	14.76 (3.358)	99.000	-4.154	**0.000**	0.59
Errors	0.00 (0.000)	0.04 (0.200)	300.000	-1.000	0.317	0.14
Repetitions/perseverations	0.42 (0.584)	0.48 (0.963)	298.000	-0.343	0.731	0.05
Food items						
Correct words	22.72 (6.073)	17.16 (5.249)	156.500	-3.034	**0.002**	0.43
Errors	0.04 (0.200)	0.04 (0.200)	312.500	0.000	1.000	0.00
Repetitions/perseverations	0.28 (0.458)	0.40 (0.764)	311.000	-0.038	0.970	0.01
Actions						
Correct words	18.72 (6.175)	14.40 (4.916)	194.500	-2.293	**0.022**	0.32
Errors	0.04 (0.200)	0.04 (0.200)	312.500	0.000	1.000	0.00
Repetitions/perseverations	0.40 (0.764)	0.48 (0.918)	308.500	-0.098	0.922	0.01

M: mean; *SD*: standard deviation; HC: healthy control; MCI: mild cognitive impairment.
Significant *p*-values ($p < 0.05$) are in **bold**.
[r] Effect size calculated as Pearson's *r*, expressed in absolute value.
Strength of association: 0.1 to 0.3: small, 0.3 to 0.5: medium, 0.5 to 1.0: large (Cohen, 1988).

of these two types of study, it is important to note the difference between connected (spontaneous) speech and verbal fluency performances. Compared to connected speech, where pauses appear more randomly, in the fluency recordings silent pauses (with varying lengths) appear between every word, therefore producing a 'word-pause-word-pause'-like sequence. Because of these distinct characteristics, the number of silent pauses needs to be interpreted based on the methodology of the specific study.

Most recent approaches to verbal fluency analysis usually focus on the semantic content when evaluating fluency performance (Tröger et al., 2019; Woods et al., 2016). In contrast, this work focused on the examination of more easily quantifiable, objective variables; nevertheless, we were able to achieve classification abilities comparable to those reported in previous studies [AUC: 0.758 (König et al., 2018), AUC: 0.77 (Chen et al., 2020)]. The significant classification ability of the silent pause parameters in our study suggests that differentiation between HC and MCI patients' semantic verbal fluency performance may be possible by examining only the silent pauses in their speech. This can be achieved, for example, by dividing the voice recordings into voiced and unvoiced segments (Lopez-de-Ipina et al., 2015).

https://doi.org/10.1017/S1355617721001454 Published online by Cambridge University Press

Table 7. Accuracy measures of those temporal parameters that significantly differed between the two groups based on the previous comparative statistic tests

		Accuracy measures					
Fluency tasks	Temporal parameters	*p*	AUC	95% CI-	95% CI+	Sensitivity (%)	Specificity (%)
Letter ‘a’	Average length of irrelevant utterances (s)	**0.010**	0.712	0.569	0.855	80.0	52.0
	Total length of irrelevant utterances (s)	**0.035**	0.674	0.523	0.824	80.0	52.0
Animals	Total number of silent pauses (count)	**0.004**	0.740	0.598	0.882	76.0	50.0
	Average length of silent pauses (s)	**0.016**	0.702	0.549	0–855	72.0	50.0
	Average word transition time (s)	**0.001**	0.787	0.651	0.922	96.0	62.5
Food items	Average length of silent pauses (s)	**0.031**	0.678	0.528	0.828	68.0	52.0
	Average word transition time (s)	**0.006**	0.726	0.587	0.866	76.0	52.0
Actions	Total number of silent pauses (count)	**0.013**	0.706	0.562	0.849	72.0	52.0
	Average length of silent pauses (s)	**0.019**	0.693	0.544	0.841	72.0	52.0
	Average word transition time (s)	**0.024**	0.686	0.536	0.837	80.0	52.0

AUC: area under the curve; CI: confidence interval.
Significant *p*-values (***p*** **< 0.05**) indicate that the measure is significantly better than chance at discriminating individuals of the two groups.

Table 8. Accuracy measures of those traditional fluency measures that significantly differed between the two groups based on the previous comparative statistic tests

		Accuracy measures					
Fluency tasks	Traditional measures	*p*	AUC	95% CI-	95% CI+	Sensitivity (%)	Specificity (%)
Letter ‘a’	Number of errors	0.116	0.630	0.474	0.785	36.0	88.0
Animals	Number of correct words	**0.000**	0.842	0.734	0.949	100	56.0
Food items	Number of correct words	**0.002**	0.750	0.616	0.884	76.0	64.0
Actions	Number of correct words	**0.022**	0.689	0.543	0.834	68.0	52.0

AUC: area under the curve; CI: confidence interval.
Significant *p*-values (***p*** **< 0.05**) indicate that the measure is significantly better than chance at discriminating individuals of the two groups.

Therefore, the described method would not require additional time-consuming steps, such as the manual transcription and preparation of the answers, nor their identification as correct words, errors, repetitions, or clusters, as opposed to the majority of fluency analysis techniques. This could make the analysis procedure considerably faster and easier. However, since this method does not provide any semantic information, it can be viewed for example as an alternative, inverse approach of the traditional analyses based on word count, because instead of considering the number of recalled words, this method focuses on the silent pauses between the words.

Our results confirmed the advantage of semantic fluency in the detection of MCI. In all three semantic fluency tests (animal, food item, and action), the same three temporal parameters (number of silent pauses, average length of silent pauses, average word transition time), and one of the traditional measures (correct word count) showed differences between the two groups. In contrast, regarding the phonemic fluency tests, differences were only observed in the case of the ‘a’ phonemic fluency, where two temporal parameters (the average and total length of irrelevant utterances) and one of the traditional measures (incorrect words) showed significant difference. These results are consistent with those of earlier studies, confirming that semantic fluency tasks may be more appropriate for detecting the cognitive changes that occur in MCI (McDonnell et al., 2020; Nikolai et al., 2018). Furthermore, when compared to other subtypes of semantic fluency tests (plants, clothes, vehicles), the animal fluency test has previously shown the highest sensitivity (98.8%) in discriminating between HC and MCI participants (García-Herranz et al., 2020). In agreement with the results of García-Herranz et al., animal fluency achieved the best accuracy scores in the present study as well, not only with the traditional scoring method but also when examining the temporal parameters.

Limitations

The significant age difference between the HC and MCI groups may be noted as a limitation of this study, although elderly age itself is a primary risk factor of MCI. However, it has been also suggested that age has a significant influence on verbal fluency abilities (Kempler, Teng, Dick, Taussig, & Davis, 1998; Rodriguez-Aranda & Martinussen, 2006). Thus, we cannot rule out the possibility that the age of the participants might have affected their verbal fluency performance regardless of their cognitive state. Nevertheless, this

https://doi.org/10.1017/S1355617721001454 Published online by Cambridge University Press

sample would closely represent the affected population in case of a potential real-life application.

When interpreting the results, it is important to take into consideration that because of the exploratory nature of this pilot study, corrections for multiple comparisons were not applied during the statistical analysis. As one of the main goals of this study was to investigate and identify all temporal fluency parameters that are able to differentiate between the groups, confirmatory studies are required to further attest the discriminatory ability and clinical utility of these significant temporal parameters.

This study established the main characteristics of a novel verbal fluency analysis, thus, further projects should be focused on the collection of more and higher quality data in order to define precise reference values for the amount of silent pauses associated with MCI. In the future, this would allow for the development of an automated tool for MCI screening, based on the analysis of temporal speech parameters. In addition, it remains to be determined whether combining this method of temporal parameter analysis with automated clustering analysis (reported earlier, e.g., König et al., 2018) could provide additional value with respect to classification.

CONCLUSION

In this study, we offered an alternative method of fluency analysis, and demonstrated the discriminatory ability of silent pause parameters in the case of semantic verbal fluency tests. Silence-related parameters can be extracted and calculated from fluency voice recordings using computerized methods. Therefore, this approach to fluency analysis seems to show promising potential, and, building on these results, the next step would be to construct an automated instrument capable of identifying MCI patients based on their speech/silence ratio. The development of remote, automated tools is especially important, seeing that the necessity and significance of medical consultations based on telemedicine are becoming common practice due to the current COVID-19 pandemic. Considering the high burden on the healthcare systems, an automated and cost-effective telemedical tool, based on the recognition of silent segments of speech, would be a valuable addition to practice, and it would likely improve the detection rates of MCI.

ACKNOWLEDGMENTS

The authors wish to thank all participants for their cooperation.

FINANCIAL SUPPORT

This work was supported by the Faculty of Medicine, University of Szeged (R.B. and I. N., grant number EFOP-3.6.3-VEKOP-16-2017-00009); the János Bolyai Research Scholarship of the Hungarian Academy of Sciences (G.G., grant number BO/00653/19); the Hungarian Ministry of Innovation and Technology (G.G., grant numbers ÚNKP-21-5, NKFIH-1279-2/2020); and the Ministry of Innovation and Technology NRDI Office (G.G., grant number NKFIH FK-124413) within the framework of the Artificial Intelligence National Laboratory Program (MILAB).

CONFLICTS OF INTEREST

The authors declare that there is no conflict of interest.

ETHICAL STANDARDS

Participation in the study was voluntary. All participants were informed about the aims of the study and gave their written consent. The experiment was conducted according to the ethical principles of the Declaration of Helsinki, and it was approved by the Regional Human Biomedical Research Ethics Committee of the University of Szeged (Reference No. 231/2017-SZTE).

REFERENCES

Boersma, P. & Weenink, D. (2020). *Praat: Doing Phonetics by Computer [Computer Program] (6.1.28) [Computer Software]*. Amsterdam, The Netherlands: Phonetic Sciences.

Borkowski, J.G., Benton, A.L., & Spreen, O. (1967). Word fluency and brain damage. *Neuropsychologia*, *5*, 135–140.

Chen, L., Asgari, M., Gale, R., Wild, K., Dodge, H., & Kaye, J. (2020). Improving the assessment of mild cognitive impairment in advanced age with a novel multi-feature automated speech and language analysis of verbal fluency. *Frontiers in Psychology*, *11*, 535.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Daneman, M. & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.

DeLong, E.R., DeLong, D.M., & Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, *44*, 837–845.

Demetriou, E. & Holtzer, R. (2017). Mild cognitive impairments moderate the effect of time on verbal fluency performance. *Journal of the International Neuropsychological Society*, *23*, 44–55.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Folstein, M.F., Folstein, S.E., & Mchugh, R. (1975). “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.

García-Herranz, S., Díaz-Mardomingo, M.C., Venero, C., & Peraita, H. (2020). Accuracy of verbal fluency tests in the discrimination of mild cognitive impairment and probable Alzheimer’s disease in older Spanish monolingual individuals. *Aging, Neuropsychology, and Cognition*, *27*, 826–840.

Gathercole, S.E., Willis, C.S., Baddeley, A.D., & Emslie, H. (1994). The children's test of nonword repetition: a test of phonological working memory. *Memory*, *2*, 103–127. https://doi.org/10.1080/09658219408258940

Gosztolya, G., Vincze, V., Tóth, L., Pákáski, M., Kálmán, J., & Hoffmann, I. (2019). Identifying mild cognitive impairment and mild Alzheimer's disease based on spontaneous speech using ASR and linguistic features. *Computer Speech & Language*, *53*, 181–197.

Hahn, E.A. & Andel, R. (2011). Nonpharmacological therapies for behavioral and cognitive symptoms of mild cognitive impairment. *Journal of Aging and Health*, *23*, 1223–1245.

Henry, J.D., Crawford, J.R., & Phillips, L.H. (2004). Verbal fluency performance in dementia of the Alzheimer's type: a meta-analysis. *Neuropsychologia*, *42*, 1212–1222.

Holmlund, T.B., Cheng, J., Foltz, P.W., Cohen, A.S., & Elvevåg, B. (2019). Updating verbal fluency analysis for the 21st century: applications for psychiatry. *Psychiatry Research*, *273*, 767–769.

IBM Corp. (2016). *SPSS Statistics for Windows (24.0) [Computer Software]*. Armonk, NY: IMB Corporation.

Jacobs, S., Mercuri, G., & Holtzer, R. (2021). Assessing within-task verbal fluency performance: the utility of individual time intervals in predicting incident mild cognitive impairment. *Aging, Neuropsychology, and Cognition*, *28*, 733–747.

Kempler, D., Teng, E.L., Dick, M., Taussig, I.M., & Davis, D.S. (1998). The effects of age, education, and ethnicity on verbal fluency. *Journal of the International Neuropsychological Society*, *4*, 531–538.

Knopman, D.S. & Petersen, R.C. (2014). Mild cognitive impairment and mild dementia: a clinical perspective. *Mayo Clinic Proceedings*, *89*, 1452–1459.

König, A., Linz, N., Tröger, J., Wolters, M., Alexandersson, J., & Robert, P. (2018). Fully automatic speech-based analysis of the semantic verbal fluency task. *Dementia and Geriatric Cognitive Disorders*, *45*, 198–209.

Ledoux, K., Vannorsdall, T.D., Pickett, E.J., Bosley, L.V., Gordon, B., & Schretlen, D.J. (2014). Capturing additional information about the organization of entries in the lexicon from verbal fluency productions. *Journal of Clinical and Experimental Neuropsychology*, *36*, 205–220.

Lezak, M.D. (2012). *Neuropsychological Assessment* (5th ed.). New York: Oxford University Press.

Lopez-de-Ipina, K., Martinez-de-Lizarduy, U., Barroso, N., Ecay-Torres, M., Martinez-Lage, P., Torres, F., & Faundez-Zanuy, M. (2015). *Automatic analysis of categorical verbal fluency for mild cognitive impartment detection: a non-linear language independent approach*. *2015 4th International Work Conference on Bioinspired Intelligence (IWOBI)*, San Sebastian, Spain, 101–104.

McDonnell, M., Dill, L., Panos, S., Amano, S., Brown, W., Giurgius, S., … Miller, K. (2020). Verbal fluency as a screening tool for mild cognitive impairment. *International Psychogeriatrics*, *32*, 1055–1062.

MedCalc Software Ltd (2020). *MedCalc Software (19.6.1.) [Computer Software]*. Ostend, Belgium: MedCalc Software Ltd.

Newcomb, F. (1969). *Missile Wounds of the Brain*. Oxford: Oxford University Press.

Nikolai, T., Bezdicek, O., Markova, H., Stepankova, H., Michalec, J., Kopecek, M., … Vyhnalek, M. (2018). Semantic verbal fluency impairment is detectable in patients with subjective cognitive decline. *Applied Neuropsychology: Adult*, *25*, 448–457.

Östberg, P., Fernaeus, S.-E., Hellström, Å., Bogdanović, N., & Wahlund, L.-O. (2005). Impaired verb fluency: a sign of mild cognitive impairment. *Brain and Language*, *95*, 273–279.

Pakhomov, S.V.S. & Hemmy, L.S. (2014). A computational linguistic measure of clustering behavior on semantic verbal fluency task predicts risk of future dementia in the Nun study. *Cortex*, *55*, 97–106.

Petersen, R.C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, *256*, 183–194.

Piatt, A.L., Fields, J.A., Paolo, A.M., & Tröster, A.I. (1999). Action (verb naming) fluency as an executive function measure: convergent and divergent evidence of validity. *Neuropsychologia*, *37*, 1499–1503.

Roberts, R.O. & Knopman, D.S. (2013). Classification and epidemiology of MCI. *Clinics in Geriatric Medicine*, *29*, 753–772.

Roberts, R.O., Knopman, D.S., Mielke, M.M., Cha, R.H., Pankratz, V.S., Christianson, T.J.H., … Petersen, R.C. (2014). Higher risk of progression to dementia in mild cognitive impairment cases who revert to normal. *Neurology*, *82*, 317–325.

Rodriguez-Aranda, C. & Martinussen, M. (2006). Age-related differences in performance of phonemic verbal fluency measured by controlled oral word association task (COWAT): a meta-analytic study. *Developmental Neuropsychology*, *30*, 697–717.

Rosen, W.G., Mohs, R.C., & Davis, K.L. (1984). A new rating scale for Alzheimer's disease. *American Journal of Psychiatry*, *141*, 1356–1364.

Rosenthal, R. (1991). *Meta-Analytic Procedures for Social Research*. Newbury Park: Sage.

Shao, Z., Janse, E., Visser, K., & Meyer, A.S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*, 772.

Shulman, K.I., Shedletsky, R., & Silver, I.L. (1986). The challenge of time: clock-drawing and cognitive function in the elderly. *International Journal of Geriatric Psychiatry*, *1*, 135–140.

Sluis, R.A., Angus, D., Wiles, J., Back, A., Gibson, T., Liddle, J., … Angwin, A.J. (2020). An automated approach to examining pausing in the speech of people with dementia. *American Journal of Alzheimer's Disease & Other Dementias*, *35*, 153331752093977.

Taler, V., Johns, B.T., & Jones, M.N. (2020). A large-scale semantic analysis of verbal fluency across the aging spectrum: data from the Canadian longitudinal study on aging. *The Journal of Gerontology Series B: Psychological Science and Social Science*, *75*, e221–e230.

Teng, E., Leone-Friedman, J., Lee, G.J., Woo, S., Apostolova, L.G., Harrell, S., … Lu, P.H. (2013). Similar verbal fluency patterns in amnestic mild cognitive impairment and Alzheimer's disease. *Archives of Clinical Neuropsychology*, *28*, 400–410.

Toth, L., Hoffmann, I., Gosztolya, G., Vincze, V., Szatloczki, G., Banreti, Z., … Kalman, J. (2018). A speech recognition-based solution for the automatic detection of mild cognitive impairment from spontaneous speech. *Current Alzheimer Research*, *15*, 130–138.

Tröger, J., Linz, N., König, A., Robert, P., Alexandersson, J., Peter, J., & Kray, J. (2019). Exploitation versus exploration – computational temporal and semantic analysis explains semantic verbal fluency impairment in Alzheimer's disease. *Neuropsychologia*, *131*, 53–61.

Troyer, A.K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: evidence

from younger and older healthy adults. *Neuropsychology*, *11*, 138–146.
Wechsler, D. (1981). *The Wechsler Adult Intelligence Scale – Revised*. New York: The Psychological Corporation.
Woods, D.L., Wyma, J.M., Herron, T.J., & Yund, E.W. (2016). Computerized analysis of verbal fluency: normative data and the effects of repeated testing, simulated malingering, and traumatic brain injury. *PLoS One*, *11*, e0166439.
Yesavage, J.A. & Sheikh, J.I. (1986). 9/Geriatric depression scale (GDS): recent evidence and development of a shorter version. *Clinical Gerontologist*, *5*, 165–173. https://doi.org/10.1300/J018v05n01_09