

REVIEW

Estimating premorbid intelligence in people living with dementia: a systematic review

Margot Juliëtte Overman,^{1,*}  Shona Leeworthy,¹ and Tomas J. Welsh^{1,2,3}

¹Research Institute for the Care of Older People (RICE), Bath, UK

²Institute of Clinical Neurosciences, University of Bristol, Bristol, UK

³Royal United Hospital, Bath NHS Foundation Trust, Bath, UK

ABSTRACT

Objectives: In diagnosing dementia, estimating premorbid functioning is critical for accurate detection of the presence and severity of cognitive decline. However, which assessments of premorbid intelligence are most suitable for use in clinical practice is not well established. Here, we systematically evaluate the validity of instruments for measuring premorbid intelligence in people living with dementia.

Design and setting: In this systematic review, electronic databases (EMBASE, PsycINFO, MEDLINE, CINAHL, and AMED) were searched to identify studies reporting on objective measures of premorbid intelligence in dementia. Participants from included studies were recruited from local communities and clinical settings.

Participants: A total of 1082 patients with dementia and 2587 healthy controls were included in the review.

Measurements: The literature search resulted in 13 eligible studies describing 19 different instruments. The majority of instruments ($n = 14$) consisted of language-based measures, with versions of the National Adult Reading Test (NART) being most commonly investigated.

Results: Preliminary evidence suggested comparable performance of patients with mild dementia and healthy controls on word reading tasks in English, Portuguese, Swedish, and Japanese. In moderate dementia, however, the performance was significantly impaired on most verbal tasks. There was a lack of reliability and validity testing of available instruments, with only one of the included studies reporting psychometric properties within the patient group.

Conclusions: The results demonstrate that there is a wide range of tools available for estimating premorbid intelligence in dementia, with cautious support for the potential of word reading tasks across different languages in individuals with mild dementia. However, the review highlights the urgent need for extensive assessments of the psychometric properties of these tasks in dementia. We propose that further longitudinal research and assessments of nonverbal measures are necessary to validate these instruments and enhance diagnostic procedures for people living with dementia worldwide.

Key words: dementia, premorbid intelligence, systematic review, diagnosis, cognitive decline, cognitive assessment

Introduction

Dementia is a highly prevalent neurodegenerative disorder and a leading global cause of disability and mortality (Vos *et al.*, 2017; Nichols *et al.*, 2019). It has been estimated that 46.8 million people were living

with dementia worldwide in 2015, with numbers expected to rise to 131.5 million in 2050 (Alzheimer's Disease International, 2015). Recent global research prioritization initiatives and policies aiming to reduce the burden of dementia have highlighted the importance of early and accurate diagnosis (Alzheimer's Disease International, 2012; Shah *et al.*, 2016; World Health Organization, 2017). Among others, a timely diagnosis can result in earlier interventions such as prescription of acetylcholine inhibitors to maintain function, enhanced advance care planning for patients and their families, and identification of relevant agencies and support networks (Dubois *et al.*, 2016; Robinson *et al.*, 2015). To establish a diagnosis of

Correspondence should be addressed to: Margot Juliëtte Overman, MRC Cognition and Brain Sciences Unit, University of Cambridge, 15 Chaucer Road, Cambridge CB2 7EF, UK. Phone: +44 (0)1223 769440. E-mail: margot.overman@mrc-cbu.cam.ac.uk Received 10 Sep 2020; revision requested 20 Oct 2020; revised version received 19 Dec 2020; accepted 11 Feb 2021. First published online 05 April 2021.

*Author's current affiliation is the MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK

dementia, it must be ascertained that a decline in cognition compared to previous levels of functioning has occurred. Determining whether a change in cognitive ability has taken place can be challenging in clinical practice, however, as previously obtained measures of cognition are rarely available at clinical presentation.

To remedy this problem, various assessments have been developed to estimate premorbid intelligence. Three of the most commonly used approaches are demographic regression equations (Barona *et al.*, 1984; Wilson *et al.*, 1978), irregular word reading tasks (Nelson and O'Connell, 1978; Wechsler, 2001; Wechsler, 2011), and lexical decision-making tasks (Baddeley *et al.*, 1993; Yuspeh and Vanderploeg, 2000). The first method computes an estimated intelligence quotient (IQ) based on variables such as education, geographic residence, and occupation. A major advantage of regression equations is that demographic details are independent of current levels of functioning, and therefore are inherently unaffected by cognitive decline due to dementia. In practice, however, reliability may be hampered by difficulties in obtaining accurate information from patients and/or limited access to demographic records. Furthermore, it has been suggested that regression equations tend to provide inaccurate estimates for people with an IQ outside the average range (Goldstein *et al.*, 1986; Veiel and Koopman, 2001; Griffin *et al.*, 2002), and can only predict approximately 50% of the variance in measured intelligence (O'Carroll, 1995).

Irregular word reading and lexical decision-making tasks, on the other hand, rely on current performance rather than self-reported information. Scores on these tasks are strongly correlated with general intelligence assessments in healthy adults (Crawford *et al.*, 1989a; Yuspeh and Vanderploeg, 2000). Their use as a measure of *premorbid* intelligence is based on the assumption that the ability to pronounce irregularly spelled words or differentiate real words from pseudo-words is relatively resistant to cognitive decline. However, although some early studies supported the view that performance on these tasks is stable in dementia (O'Carroll *et al.*, 1987; Nelson and McKenna, 1975; Sharpe and O'Carroll, 1991; Crawford *et al.*, 1988a), others reported significantly lower scores in patients compared with healthy controls (Stebbins *et al.*, 1990; Patterson *et al.*, 1994; O'Carroll *et al.*, 1995; Schmand *et al.*, 1998). In addition, while good test-retest and inter-rater reliability have been demonstrated for several instruments in healthy adults (Crawford *et al.*, 1989a; O'Carroll, 1987; Dykiert and Deary, 2013), it is unclear whether these psychometric properties extend to clinical populations. Furthermore, the use of these measures across

cultures and languages has not been systematically evaluated.

Considering their widespread use in clinical settings, a better understanding of the validity of measures for estimating premorbid intelligence in dementia is vital. Accurate estimates are key to correct interpretations of scores on cognitive screening tests, and consequently accurate diagnosis of dementia. The present systematic review focuses on the following questions: (1) what assessments are currently available for the measurement of premorbid intelligence in dementia?; (2) do estimated premorbid intelligence scores on these instruments remain stable in dementia? That is, are task scores similar for healthy adults and patients in cross-sectional comparisons, and/or are patient scores constant over time?; and (3) what are the psychometric properties of the identified tools in people living with dementia? The main objective of this review is to clarify which measures of premorbid intelligence may be most suitable for assessing patients with dementia in global clinical practice.

Methods

Search strategy

The predefined protocol for this systematic review was registered with PROSPERO (CRD42019133499) and was based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher *et al.*, 2015). Databases (EMBASE, PsycINFO, MEDLINE, CINAHL, and AMED) were searched from 1999 until May 2019 using the NICE Healthcare Databases Advanced Search. Papers were identified through Boolean operators using keywords for dementia (“dementia” OR “Alzheimer”) and premorbid intelligence (“premorbid” AND [“intelligence” OR “intellect”]) with the thesaurus “explode” function. The search was restricted to papers published in the English language. References in the selected journal papers and previous reviews were screened manually to supplement the main search methods.

Paper selection

Titles and abstracts and, where appropriate, full text of identified citations were independently screened by two authors (MJO and SL). Any disagreements were resolved by consensus and a third author (TJW) was consulted when needed. The following criteria had to be met for inclusion in the review:

- 1) Published as a journal paper or letter.
- 2) Participants had a diagnosis of any type and severity of dementia, except for dementia secondary to acquired brain injury or non-neurological disease.

- 3) Diagnosis of a dementia was determined using standardised criteria (e.g. *Diagnostic and Statistical Manual of Mental Disorders, International Classification of Diseases*).
- 4) Performance on an objective assessment of premorbid intelligence was a primary or secondary outcome, and its relation to diagnosis or severity of dementia was investigated with statistical analyses.

Data extraction

The following details were extracted independently by two authors (MJO and SL) from each study using a structured form: study characteristics (study design, sample size, recruitment site, and diagnostic criteria used), participant demographics (mean age, years of education, type and severity of dementia, percentage of female participants), assessment scale of premorbid intelligence (name, type, language, and scores of patient and control groups), psychometric properties in the patient group (test–retest reliability and inter-rater reliability), and results (statistically significant findings at $p < 0.05$, unless otherwise determined by the authors). Disease severity was based on the range and mean of Mini-Mental State Examination (MMSE) scores reported in each study, with cognitive impairment being recorded as mild (MMSE 21–26), moderate (MMSE 14–20), moderately severe (MMSE 10–14), or severe (MMSE <10) in line with NICE guidance (National Institute For Health and Care Excellence, 2011). Where clearly identified, patients with a diagnosis of mild cognitive impairment (MCI) were excluded from the reported values and outcomes. The rationale for this exclusion is that MCI is a highly heterogeneous syndrome, which in some cases progresses to dementia but can also remain stable or even reverse over time (Gauthier *et al.*, 2006). Where possible, Cohen's d was estimated for the core comparisons by computing the difference between reported group means divided by their pooled standard deviation. The pooled standard deviation was calculated as per Cohen (1988):

$$\text{Pooled } SD = \sqrt{\frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2}{n_1 + n_2 - 2}}$$

with SD_1 and SD_2 denoting the standard deviations for each group and n_1 and n_2 referring to their respective sample sizes. Effect sizes were interpreted as small ($d = 0.2$), medium ($d = 0.5$), or large ($d = 0.8$).

Quality assessment and data synthesis

Study quality and risk of bias were evaluated according to the AXIS tool (Downes *et al.*, 2016), a checklist comprised of 20 items designed for quality assessment of observational studies. Quality was appraised according to the number of items for

which a “yes” response was recorded and rated as “high quality” (15–20), “moderate quality” (8–14), or “low quality” (0–7). A list of instruments for the assessment of premorbid intelligence in dementia was generated from the selected papers. The key outcomes of the identified studies and the psychometric properties of the instruments, where available, are presented in a narrative synthesis.

Results

Study selection and characteristics

Titles and abstracts of all identified papers after removal of duplicates ($n = 304$) were screened, with 13 studies meeting the stipulated eligibility criteria after full-text review. A flow diagram of the identification and attrition of studies is provided in Figure 1. Study design and demographic details of participants were recorded for all studies (see Table 1). Ten of the studies applied a cross-sectional design, with three studies investigating the change in premorbid intelligence scores over time. A total of 3669 participants were included with a mean age of 71.72 years ($SD \pm 7.06$) across the studies. Of the 1082 patients with dementia, 929 participants were diagnosed with Alzheimer's disease, 40 participants had vascular dementia, and dementia subtype was unspecified for 113 participants. Four of the studies (30.8%) only included patients with mild dementia, five studies (38.5%) involved patients with mild to moderate dementia, and four studies (30.8%) included patients from the mild to severe range. The studies were conducted in eight different countries (Australia, Brazil, Germany, Japan, Portugal, Sweden, the UK, and the USA).

Identified instruments

Nineteen objective measures of premorbid intelligence were identified, including revisions, parallel versions, and variants in different languages. The most commonly investigated tools were word reading tasks (47.4%), followed by lexical decision tasks (21.1%), visuospatial reasoning tasks (15.8%), demographic equations (10.5%), and a word description task (5.3%). The majority of assessments were conducted in English (47.4%) and Portuguese (31.6%), with the remaining tasks carried out in German (10.5%), Swedish (5.3%), and Japanese (5.3%). The key outcomes from cross-sectional and longitudinal studies for each of the identified instruments are summarized in Tables 2 and 3, respectively.

WORD READING

A total of nine studies investigated the performance of dementia patients on word reading tasks (see Tables 2

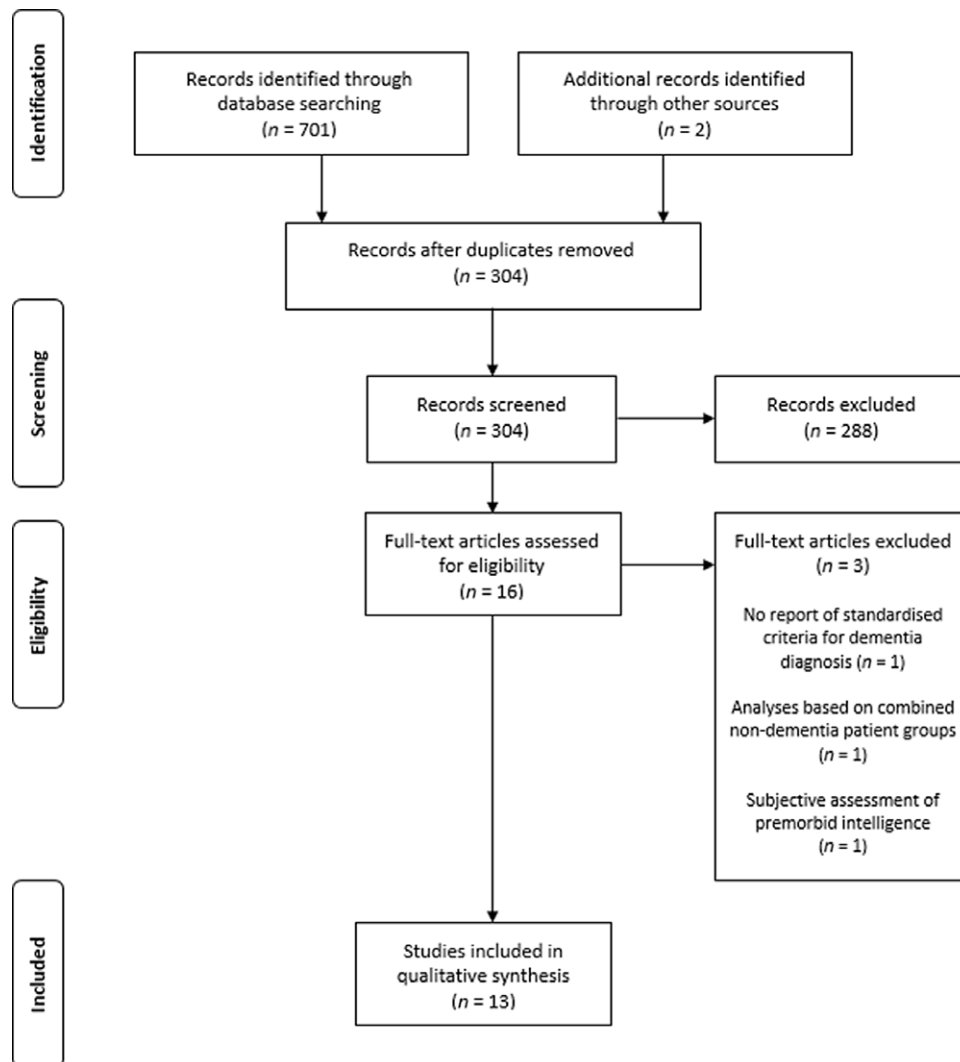


Figure 1. Flowchart of the systematic search process.

and 3). English assessments included the original and revised versions of the National Adult Reading Task (NART and NART-R), the Wechsler Test of Adult Reading (WTAR), version III and the revised version of the Wide Range Achievement Test (WRAT-III and WRAT-R), and the Cambridge Contextual Reading Test (CCRT). In the NART and the WTAR, participants are asked to read a list of 50 words which have irregular grapheme–phoneme correspondences (Nelson, 1982; Nelson and Willison, 1991; Wechsler, 2001). The WRAT differs from these two measures through its inclusion of words following regular spelling rules (Jastak and Wilkinson, 1984). Finally, the CCRT is comprised of the same words as the NART, but provides greater syntactic and semantic context by presenting each word within a sentence (Beardsall and Huppert, 1997). The aim of this adaptation is to facilitate recognition of the word and thereby improve task performance.

In the mild stage of dementia, no significant differences were observed between healthy adults and patients on any of these tasks in two cross-sectional studies (McCarthy *et al.*, 2005; McFarlane *et al.*, 2006). Direct task comparisons indicated that the performance of patients with mild dementia was better on the CCRT than the NART, suggesting that embedding words within a sentence may improve scores in dementia patients (McFarlane *et al.*, 2006). When disease severity was moderate, one study reported similar scores for healthy adults and patients on the WRAT-R and WRAT-III (McCarthy *et al.*, 2005). In contrast, performance on the NART (McGurn *et al.*, 2004; McFarlane *et al.*, 2006), NART-R (McCarthy *et al.*, 2005), WTAR (McFarlane *et al.*, 2006), and CCRT (McFarlane *et al.*, 2006) was significantly lower for patients with moderate dementia than control participants. A small effect size for the group

Table 1. Study design and demographic details

STUDY	DESIGN	COUNTRY	N		DISEASE SEVERITY	MEAN AGE (SD)	GENDER (% FEMALE)	EDUCATION (YEARS)
			PATIENTS	CONTROLS				
Alves <i>et al.</i> (2013)	Cross-sectional	Portugal	51	51	Mild-moderate	72.8 (8.0)	61.5%	10.3 (4.4)
Ashendorf <i>et al.</i> (2009)	Longitudinal	USA	41	200	Mild	72.7 (8.4)	61.8%	16.0 (3.0)
Binkau <i>et al.</i> (2014)	Cross-sectional	Germany	287	534	Mild-severe	71.9 (8.9)	52.9%	12.3 (2.8)
Cockburn <i>et al.</i> (2000)	Longitudinal	UK	78	0	Mild-severe	78.6 (6.7)	NR	NR [‡]
De Oliveira <i>et al.</i> (2014)	Cross-sectional	Brazil	45	61	Mild	72.7 (6.4)	59.0%	7.7 (5.0)
Hessler <i>et al.</i> (2013)	Cross-sectional	Germany	104	300	Mild-severe	78.4 (9.3)	100%	NR
Matsuoka <i>et al.</i> (2006)	Cross-sectional	Japan	74	106	Mild-moderate	70.2 (5.8)	64.6%	11.7 (2.5)
McCarthy <i>et al.</i> (2005)	Cross-sectional	USA	60	30	Mild-moderate	76.1 (5.3)	NR [†]	12.9 (2.7)
McFarlane <i>et al.</i> (2006)	Cross-sectional	UK	66	32	Mild-moderate	73.2 (10.7)	52.0%	11.6 (2.7)
McGurn <i>et al.</i> (2004)	Cross-sectional	UK	45	464	Mild-moderate	79.1 (0.7)	NR	NR
Rolstad <i>et al.</i> (2008)	Cross-sectional	Sweden	43	53	Mild	67.0 (7.0)	63.5%	11.8 (3.5)
Serrao <i>et al.</i> (2015)	Cross-sectional	Brazil	26	38	Mild	69.8 (6.2)	NR	10.4 (5.1)
Weinborn <i>et al.</i> (2018)	Longitudinal	Australia	162	718	Mild-severe	71.7 (7.8)	58.5%	NR [§]

NR, Not reported.

[†]Only reported for patient group.[‡]Reported as age left education.[§]Reported as categorical variable.

differences in word reading scores was found for the NART by McFarlane *et al.* (2006), while the remaining studies observed a medium effect size on the NART, NART-R, WTAR, and CCRT.

In a longitudinal study, WRAT-III reading scores were significantly higher for control participants than patients with Alzheimer's disease in baseline assessments, but raw scores did not decline significantly in either patients or controls over a 1-year period (Ashendorf *et al.*, 2009). In studies with a longer follow-up time of 3 years, however, steeper declines in performance were observed for patients than controls on the NART (Cockburn *et al.*, 2000), and lower MMSE scores were systematically associated with a greater decline on the WTAR (Weinborn *et al.*, 2018). In the latter, the authors noted that a large proportion of the recruited patients were lost to follow-up due to death (18.5%) or withdrawal (41.4%). Moreover, only 75 of the remaining 132 patients were able to complete the WTAR on follow-up, with participants who could not carry out the assessment being more cognitively impaired at baseline (Weinborn *et al.*, 2018). The degree of decline on the WTAR may therefore be larger than estimated in the assessed patient group.

In addition to these English tasks, the systematic search identified adaptations of the NART into Portuguese (TeLPI; Alves *et al.*, 2013), Swedish (NART-SWE; Rolstad *et al.*, 2008), and Japanese (JART; Matsuoka *et al.*, 2006). Due to language differences, the NART-SWE items consisted of loan words rather than irregular Swedish words (Rolstad *et al.*, 2008). The JART was based on *Kanji* characters, an ideographic script which is used to represent lexical morphemes. Many Japanese words are compounds comprised of multiple *Kanji*, and the pronunciation of an individual character can vary across different words. The authors propose that the JART provides a suitable adaptation of English irregular word reading tasks as it similarly requires word-specific translations from orthography to phonology (Matsuoka *et al.*, 2006). For all three tasks (the TeLPI, NART-SWE, and JART), comparable task performance was observed in patient and healthy control groups (Rolstad *et al.*, 2008; Matsuoka *et al.*, 2006; Alves *et al.*, 2013). It should be noted that all studies focused on patients with mild dementia. Overall, these findings suggest that word reading tasks may have potential as an assessment of premorbid functioning across different languages in early dementia.

LEXICAL DECISION-MAKING

The four studies assessing lexical decision-making tasks all focused on the Spot-the-Word (STW) task or adaptations of this instrument. In the original STW task, participants have to select which of a

Table 2. Overview of instruments and outcomes of identified cross-sectional studies

ASSESSMENT TYPE	DISEASE SEVERITY	INSTRUMENT	LANGUAGE	STUDY	SCORE		COHEN'S <i>d</i>	ASSOCIATION COGNITIVE IMPAIRMENT – TASK SCORE
					PATIENTS	CONTROLS		
Demographic regression equation	Mild	Barona Index	English	McCarthy <i>et al.</i> (2005)	106.5 (8.9) [†]	109.5 (7.8) [†]	– 0.359	No
	Mild–moderate	Crawford & Allan's (1997) regression equation	English	McFarlane <i>et al.</i> (2006)	NR	NR	–	No
Lexical decision	Moderate	Barona Index	English	McCarthy <i>et al.</i> (2005)	107.8 (7.5) [†]	109.5 (7.8) [†]	– 0.222	No
	Mild	STW	English	McFarlane <i>et al.</i> (2006)	17.4 (9.4) [‡]	17.3 (10.5) [‡]	0.010	No
	Moderate	LDT	Portuguese	Serrao <i>et al.</i> (2015)	49.0 (1.1)	52.48 (1.2)	–	No
		MWT-A	German	Binkau <i>et al.</i> (2014)	27.4 (5.0)	30.0 (3.4)	– 0.608	Yes
		MWT-B	German	Hessler <i>et al.</i> (2013)	17.8 (8.2)	28.5 (5.1)	– 1.567	Yes
		STW	English	McFarlane <i>et al.</i> (2006)	19.8 (6.7) [‡]	17.3 (10.5) [‡]	0.284	No
		MWT-B	German	Hessler <i>et al.</i> (2013)	17.8 (8.2)	28.5 (5.1)	– 1.567	Yes
		Moderate–severe	MWT-A	German	Binkau <i>et al.</i> (2014)	22.1 (7.1)	30.0 (3.4)	– 1.419
Visuospatial reasoning	Severe	MWT-B	German	Hessler <i>et al.</i> (2013)	14.4 (8.8)	28.5 (5.1)	– 1.961	Yes
	Mild	WAIS-III Colored Matrices	Portuguese	De Oliveira <i>et al.</i> (2014)	17.8 (6.0)	25.8 (6.2)	– 1.311	Yes
		WAIS-III Matrix Reasoning	Portuguese	Serrao <i>et al.</i> (2015)	9.8 (0.9)	13.1 (1.0)	– 3.469	Yes
		WAIS-III Block Design	Portuguese	De Oliveira <i>et al.</i> (2014)	14.2 (6.5)	25.4 (8.9)	– 1.437	Yes
Word description	Mild	WAIS-III Vocabulary	Portuguese	De Oliveira <i>et al.</i> (2014)	23.2 (9.2)	26.5 (9.8)	– 0.347	No
				Serrao <i>et al.</i> (2015)	42.0 (2.20)	43.1 (2.4)	– 0.478	No
Word reading	Mild	CCRT	English	McFarlane <i>et al.</i> (2006)	16.2 (8.0) [‡]	15.7 (7.3) [‡]	0.065	No
		NART	English	McFarlane <i>et al.</i> (2006)	17.2 (9.0) [‡]	16.6 (9.1) [‡]	0.066	No
		NART-R	English	McCarthy <i>et al.</i> (2005)	108.5 (8.4) [†]	110.0 (9.4) [†]	– 0.168	No
		WRAT-III	English	McCarthy <i>et al.</i> (2005)	105.1 (8.0) [†]	105.0 (9.7) [†]	0.011	No
		WRAT-R	English	McCarthy <i>et al.</i> (2005)	104.3 (9.4) [†]	105.7 (11.9) [†]	– 0.131	No
		WTAR	English	McFarlane <i>et al.</i> (2006)	11.0 (10.3) [‡]	10.0 (11.0) [‡]	0.094	No

Table 2. Continued

ASSESSMENT TYPE	DISEASE SEVERITY	INSTRUMENT	LANGUAGE	STUDY	SCORE			ASSOCIATION COGNITIVE IMPAIRMENT – TASK SCORE
					PATIENTS	CONTROLS	COHEN'S <i>d</i>	
Mild-moderate		NART	English	McGurn <i>et al.</i> (2004)	28.2 (9.8)	34.5 (8.0)	-0.704	Yes
		NART-SWE	Swedish	Rolstad <i>et al.</i> (2008)	36.8 (7.4)	38.2 (6.6)	-0.200	No
		JART	Japanese	Matsuoka <i>et al.</i> (2006)	102.0 (12.0)†	102.1 (12.1)†	-0.008	No
		TeLPI	Portuguese	Alves <i>et al.</i> (2013)	16.0 (9.9)‡	15.5(10.1)‡	0.050	No
Moderate		CCRT	English	McFarlane <i>et al.</i> (2006)	20.1 (7.5)‡	15.7 (7.3)‡	0.595	Yes
		NART	English	McFarlane <i>et al.</i> (2006)	24.1 (8.1)‡	16.6 (9.1)‡	0.420	Yes
		NART-R	English	McCarthy <i>et al.</i> (2005)	104.8 (8.2)†	110.0 (9.4)†	-0.590	Yes
		WRAT-III	English	McCarthy <i>et al.</i> (2005)	102.7 (8.3)†	105.0 (9.7)†	-0.255	No
		WRAT-R	English	McCarthy <i>et al.</i> (2005)	100.5 (11.2)†	105.7 (11.9)†	-0.450	No
		WTAR	English	McFarlane <i>et al.</i> (2006)	16.0 (11.3)‡	10.0 (11.0)‡	0.538	Yes

CCRT, Cambridge Contextual Reading Test; JART, Japanese Adult Reading Test; LDT, Lexical Decision Test; MWT-A, Mehrfachwahl-Wortschatz-Test – Version A; MWT-B, Mehrfachwahl-Wortschatz-Test – Version B; NART, National Adult Reading Test; NART-R, National Adult Reading Test – Revised; NART-SWE, Swedish National Adult Reading Test; STW, Spot-the-Word; TeLPI, Teste de Letture de Palavras Irregulares; WAIS-III, Wechsler Adult Intelligence Scale III; WRAT-III, Wide Range Achievement Test III – Reading subtest; WRAT-R, Wide Range Achievement Test Revised – Reading subtest; WTAR, Wechsler Test of Adult Reading

†Predicted FSIQ.

‡Number of errors.

word pair is the real word versus a pseudo-word (Wechsler, 2011). The assessment comprises 60 pairs of real words and pseudo-words of varying word length and frequency. An equivalent version of this lexical decision-making task (LDT) was developed in Portuguese by Serrao and colleagues (2015). Two German adaptations used the same general principle as the STW, but required participants to identify the real word among four pseudo-words rather than presenting word pairs (Mehrfachwahl-Wortschatz-Test A and B, MWT-A and MWT-B; Binkau *et al.*, 2014; Hessler *et al.*, 2013).

In the English STW task, test scores were comparable for controls and patients with both mild and moderate dementia (McFarlane *et al.*, 2006). In the Portuguese LDT, mild dementia was associated with numerically lower scores than healthy adults, but this difference was not significant in general linear models (Serrao *et al.*, 2015). For the German adaptations of the task, on the other hand, scores were significantly lower for patients than controls on both the MWT-A (Binkau *et al.*, 2014) and MWT-B (Hessler *et al.*, 2013). Impaired performance was observed across the entire spectrum of disease severity, from mild to severe dementia. In addition, effect sizes for these differences were estimated to be medium to large.

DEMOGRAPHIC REGRESSION EQUATIONS

The Barona Index (Barona *et al.*, 1984) and a demographic regression equation based on Crawford and colleagues' (1989b) work were assessed in two cross-sectional studies (McCarthy *et al.*, 2005; McFarlane *et al.*, 2006). The Barona Index is based on age, sex, race, education, occupation, and geographical residence, whereas the latter equation includes the variables age, total years of education, and social class. In both studies, estimated premorbid IQ scores were similar for the patient and control groups. As would be expected given the task's reliance on stable demographic characteristics as opposed to current performance, results were similar across patients with mild and moderate cognitive impairments.

OTHER ASSESSMENTS

Three less frequently used assessments of premorbid intelligence were identified in the present review. First, a Portuguese version of the Vocabulary subtask of the Wechsler Adult Intelligence Scale III (WAIS-III) was investigated in two cross-sectional studies (Serrao *et al.*, 2015; De Oliveira *et al.*, 2014). In this task, participants are asked to provide definitions of a list of words. The two studies demonstrated that Vocabulary task performance was similar in patients with mild dementia and controls with either normal (Serrao *et al.*, 2015) or low levels of education (De Oliveira *et al.*, 2014).

Finally, the remaining two assessments focused on performance in the visuospatial domain rather

Table 3. Overview of instruments and outcomes of identified longitudinal studies

ASSESSMENT TYPE	INSTRUMENT	LANGUAGE	STUDY	FOLLOW-UP TIME	MEAN SCORE DIFFERENCE (SD)		CHANGE OVER TIME
					PATIENTS	CONTROLS	
Word reading	NART	English	Cockburn <i>et al.</i> (2000)	3 years	NR	NR	Yes
	WRAT-III	English	Ashendorf <i>et al.</i> (2009)	1 year	-0.8 (5.8)	-0.5 (3.3)	No
	WTAR	English	Weinborn <i>et al.</i> <i>et al.</i> (2018)	3 years	NR	NR	Yes

NART, National Adult Reading Test; WRAT-III, Wide Range Achievement Test III; WTAR, Wechsler Test of Adult Reading

than verbal abilities. Two studies conducted in Portuguese (Serrao *et al.*, 2015; De Oliveira *et al.*, 2014) considered participants' scores on a Matrix Reasoning task derived from the Wechsler Adult Intelligence Scale III (WAIS-III) (Wechsler, 1997). In addition, one of these studies assessed people's performance on a Block Design task from the WAIS-III (De Oliveira *et al.*, 2014). Both Matrix Reasoning and Block Design draw on visuospatial problem-solving skills rather than knowledge acquired through past learning. It was found that participants with mild dementia scored significantly lower on all of these tasks compared with healthy individuals (Serrao *et al.*, 2015; De Oliveira *et al.*, 2014). The effect sizes of these group differences were large on all visuospatial tasks. There was thus no evidence to support the use of perceptual problem-solving tasks to estimate premorbid IQ in dementia.

Psychometric properties

Of the 13 studies evaluated here, only 1 reported reliability measures within the dementia patient group. Ashendorf and colleagues (2009) found that test-retest reliability for the irregular word reading WRAT-III task was .90 in the subgroup with Alzheimer's disease, indicating high stability of test scores across multiple measurements. No further statistics for test-retest or inter-rater reliability were provided for any of the other measures.

Quality assessment and risk of bias

All included studies were deemed to be of moderate ($n = 11$) to high quality ($n = 2$) as assessed with the AXIS tool (see Table 4). The criteria least frequently met were justification for the sample size ($n = 13$), representative participant selection ($n = 13$), and addressing and categorizing nonresponders ($n = 12$). These findings suggest that there was a risk of selection and nonresponse bias in the majority of the studies reported here.

Discussion

The main aims of the present review were to identify and evaluate instruments for estimating premorbid intelligence in people living with dementia. Our findings suggest that while a wide range of tools has been assessed for this purpose, evidence for their validity in patients with dementia is rather mixed. Furthermore, the lack of reliability testing across studies highlights the need for further information regarding the psychometric properties of the identified instruments. We will discuss the core findings and their implications for the assessment and diagnosis of dementia in clinical practice, and propose several directions for future research.

Stability of verbal task performance in early dementia

Of the 19 tools for estimating premorbid intelligence evaluated here, the vast majority consisted of verbal assessments. While a number of studies indicated that performance on word reading, lexical decision-making, or vocabulary tasks was unaffected by a diagnosis and/or severity of dementia, others reported significant differences between scores of healthy adults and patients groups or declining scores over time.

Word reading tasks were most frequently investigated, with a total of nine different instruments being identified. However, it should be noted that the majority of the findings on word reading performance stem from only two studies, which both examined several different tasks (McCarthy *et al.*, 2005; McFarlane *et al.*, 2006). McCarthy *et al.* (2005) found that performance on the NART-R was reduced in moderate but not mild dementia, whereas there was no evidence for impairment on the WRAT-III and WRAT-R in either mild or moderate dementia. McFarlane and colleagues (2006) indicated that the performance of patients

Table 4. Quality assessment with the AXIS tool per item (Y = yes; N = no)

STUDY	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	TOTAL	RATING
Alves <i>et al.</i> (2013)	Y	Y	N	Y	Y	N	N	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	15	High
Ashendorf <i>et al.</i> (2009)	Y	Y	N	Y	Y	N	N	Y	Y	Y	N	Y	N	N	Y	Y	Y	N	Y	Y	13	Moderate
Binkau <i>et al.</i> (2014)	Y	Y	N	Y	N	N	N	Y	Y	Y	N	Y	N	N	Y	Y	Y	N	Y	N	11	Moderate
Cockburn <i>et al.</i> (2000)	Y	Y	N	Y	N	N	N	Y	Y	N	N	Y	N	N	Y	N	N	N	Y	N	8	Moderate
De Oliveira <i>et al.</i> (2014)	Y	Y	N	Y	N	N	N	Y	Y	Y	Y	N	N	N	Y	Y	N	N	Y	Y	11	Moderate
Hessler <i>et al.</i> (2013)	Y	Y	N	Y	N	N	N	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N	N	Y	13	Moderate
Matsuoka <i>et al.</i> (2006)	Y	Y	N	Y	N	N	N	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	N	Y	12	Moderate
McCarthy <i>et al.</i> (2005)	Y	Y	N	Y	Y	N	N	Y	Y	N	N	N	N	N	Y	Y	N	N	N	N	8	Moderate
McFarlane <i>et al.</i> (2006)	Y	Y	N	Y	Y	N	N	Y	Y	Y	Y	Y	N	N	Y	Y	N	N	N	N	12	Moderate
McGurn <i>et al.</i> (2004)	Y	Y	N	Y	Y	N	N	Y	Y	Y	N	Y	N	N	Y	Y	N	N	N	N	10	Moderate
Rolstad <i>et al.</i> (2008)	Y	Y	N	Y	N	N	N	Y	N	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	13	Moderate
Serrao <i>et al.</i> (2015)	Y	Y	N	Y	N	N	N	Y	N	Y	N	Y	N	N	Y	N	N	N	Y	Y	9	Moderate
Weinborn <i>et al.</i> (2018)	Y	Y	N	Y	Y	N	N	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	15	High

with moderate (though not mild) dementia was impaired on three different measures: the NART, WTAR, and CCRT. The first explanation for the diverging findings could be that there were differences in sampling method or participant characteristics across the two studies. For example, in the study by McFarlane *et al.* (2006), participants with moderate dementia on average reported significantly fewer years of education than those with mild dementia and healthy controls. Furthermore, the total years of education were numerically higher for patients in the study by McCarthy and colleagues (2005). However, McFarlane *et al.* (2006) highlight that inclusion of education in the statistical analyses did not alter the pattern of results. The second possible reason for these conflicting findings is that, while some verbal abilities are presumed to be *relatively* resistant to dementia, it is improbable that they are *entirely* impervious to the condition. At the more severe end of the spectrum, we might therefore observe greater difficulties in completing verbal tasks. In line with this hypothesis, high performance on word reading tasks tended to be maintained in mild dementia, whereas patients with moderate cognitive impairments scored lower than controls on the NART, NART-R, WTAR, and CCRT (McCarthy *et al.*, 2005; McFarlane *et al.*, 2006; McGurn *et al.*, 2004). In the two studies investigating Vocabulary tasks, no differences between patients and healthy adults were found (De Oliveira *et al.*, 2014; Serrao *et al.*, 2015). Importantly, only individuals with mild dementia were included in these studies, leaving open the question of whether the performance would be similarly preserved in patients with more severe cognitive impairments. Further testing of verbal tasks across the full range of disease severity is therefore essential to determine whether such assessments are suitable beyond the early stages of dementia.

Finally, in addition to the potential impact of disease severity on task performance, we posit that variability in task demands may contribute to diverging results across verbal tasks. For example, while scores on tasks such as the NART, NART-R, WTAR, and CCRT were affected in moderate dementia and declined over time (McCarthy *et al.*, 2005; McFarlane *et al.*, 2006; Cockburn *et al.*, 2000; Weinborn *et al.*, 2018), performance on the WRAT-III and WRAT-R appeared to be relatively stable (McCarthy *et al.*, 2005; Ashendorf *et al.*, 2009). Interestingly, the WRAT-III and WRAT-R are the only word reading tasks which include phonetically regular items. As mistakes are presumably more likely to be made in the pronunciation of irregular than regular words, the WRAT instruments might be easier and less sensitive to

cognitive impairment compared to tasks which are comprised exclusively of irregular words. In lexical decision tasks, we hypothesize that a similar effect of difficulty may be responsible for inconsistent findings across tasks. Specifically, whereas patient and control scores were similar in the original and Portuguese version of the STW task (McFarlane *et al.*, 2006; Serrao *et al.*, 2015), significant impairments were observed even in mild dementia on the German MWT-A and MWT-B (Binkau *et al.*, 2014; Hessler *et al.*, 2013). Although this difference could be due to cultural or linguistic differences between the task versions, it is not clear why this should particularly affect German but not Portuguese adaptations. We propose that an alternative explanation could be that, whereas the English and Portuguese tasks asked participants to choose the real word from a word pair, the German versions required individuals to select the real word from a total of five words. The greater number of options likely increased the cognitive demands of the task, as well as reduced the chance of guessing words correctly.

Taken together, these findings are thus suggestive of an influence of disease severity and specific task demands on verbal premorbid intelligence scores in people living with dementia. In future research, it would be worth testing this hypothesis explicitly by (1) including participants with a wide range of scores on the MMSE or similar screening measures and (2) directly contrasting tasks which are based on a similar approach but may vary in difficulty, such as the NART and the WRAT. Crucially, in studies reporting significant differences between patients and controls on verbal tasks, the effect size tended to be medium to large. This suggests that inappropriate use of these tasks could lead to substantial underestimation of prior cognitive function, which would hamper the interpretation of neuropsychological assessments and consequently accurate diagnosis of dementia. Clarifying the cause of differences in performance across verbal tasks and establishing more firmly whether these measures are valid only in early dementia is therefore a critical next step in optimizing assessments of premorbid intelligence in patient groups.

Impact of language differences

Cultural and linguistic differences between the populations and tasks should also be taken into account when interpreting findings across studies. In this review, translations of English verbal tasks into German, Portuguese, Swedish, and Japanese were identified. For some of the word reading tasks, translating instruments which were originally developed in English was a nontrivial issue. Specifically,

an instrument developed in Sweden had to rely on loan words due to an absence of irregular words (which form the basis of the NART, WTAR, and WRAT) in the Swedish language (Rolstad *et al.*, 2008), and a Japanese adaptation used a different writing system (*Kanji*) (Matsuoka *et al.*, 2006). As a consequence, it is possible that various translations of word reading tasks relied on different cognitive processes compared to the original English versions. For example, *Kanji* characters are perceptually highly complex and tend to have fewer phonemic factors than English written words. As a consequence, reading *Kanji* by guessing is difficult when the reader is not familiar with the word, and may rely on semantic processing to a greater extent than the reading of English irregular words (Matsuoka *et al.*, 2006). Nevertheless, the absence of group differences between patients and healthy adults in any of these adaptations is encouraging, and suggests that such word reading tasks may hold promise as a measure of premorbid intelligence across a range of languages.

Nonverbal measures

Overall, there is thus preliminary evidence that language-based assessments may be suitable for estimating premorbid intelligence in dementia, although further research is needed to clarify the effects of disease severity and specific task differences. In addition, it should be noted that such measures are likely to be of limited use for people presenting with language variants of dementia (e.g. semantic dementia or primary progressive aphasia), as well as learning difficulties such as dyslexia. It is therefore worth considering the use of alternative, nonlinguistic assessments. However, research on such measures to date appears to be very limited. While two studies investigated patients' performance on visuospatial reasoning tasks (De Oliveira *et al.*, 2014; Serrao *et al.*, 2015), the authors highlight that these tasks were specifically included to demonstrate the deterioration of fluid intelligence in dementia compared with "crystallised" abilities such as lexical tasks. It was therefore unsurprising that impaired performance on these tasks was observed in people with dementia. In the future, it may be of interest to explore whether there are visual abilities that tend to be preserved in dementia and are good predictors of intelligence which can be exploited to devise suitable assessments for people with language difficulties.

Alternatively, it would be possible to utilize demographic equations, which are entirely independent of people's current performance. Preliminary findings suggest that such equations tend to result in similar estimates of premorbid IQ in healthy adults

and people with dementia (McCarthy *et al.*, 2005; McFarlane *et al.*, 2006). However, as only two studies were identified in the present review which used this method, we cannot make any strong claims regarding their global utility in the diagnosis of dementia. Furthermore, concerns raised in previous studies regarding limitations of this approach in accurately estimating high and low ranges of IQ (Goldstein *et al.*, 1986; Griffin *et al.*, 2002; Veiel and Koopman, 2001) remain to be addressed.

Implications and recommendations for clinical practice

As studies have rarely performed direct comparisons of the different measures presented here, an outstanding question is which of the various tasks is most suitable for application in people living with dementia. One study which contrasted performance on the NART, WTAR, CCRT, and STW suggested that the lexical decision-making STW task was the only measure on which no significant differences between groups were observed (McFarlane *et al.*, 2006). The word reading tasks were all found to result in lower scores in people with mild Alzheimer's disease compared to healthy controls, although embedding the words within sentences (as in the CCRT) was associated with better performance in the mild patient group than presenting a list of words (as in the NART). However, given that no other studies identified in this review have investigated the English STW, replication is needed to confirm the superiority of the lexical decision-making task over irregular word reading measures in estimating premorbid intelligence. Further comparisons of word reading tasks were conducted by McCarthy and colleagues (2005), who investigated differences in scores on the NART-R, WRAT-R, and WRAT-III as well as estimates derived from the Barona demographic equation. While the authors suggest that all four measures showed similar stability relative to Full-Scale IQ scores obtained from the Wechsler Adult Intelligence Scale – Revised (WAIS-R), no statistical analyses were conducted to directly compare the performance of the individual instruments. Additional studies contrasting the efficacy of different measures are therefore needed to identify the most suitable measure for assessments of dementia.

For clinical practice, the utility of different tools not only depends on their accuracy in estimating cognitive decline, but also the resources it requires in terms of financial costs, time, and expertise. From a pragmatic point of view, versions of the language-based NART have been investigated most extensively and are currently being used by many health

professionals, with the NART-R having recently been re-standardized for the WAIS-IV (Bright *et al.*, 2018). The NART-R may thus represent an up-to-date measure which can easily be implemented in clinical settings. However, as some cross-sectional and longitudinal studies have disputed the stability of NART scores in dementia (Cockburn *et al.*, 2000; McFarlane *et al.*, 2006), task performance should be interpreted with caution. In particular, the NART may not be appropriate when assessing patients with moderately impaired cognition, as a number of studies indicated that NART performance may be affected when dementia has progressed beyond the mild stage. As the specific advantages and limitations of the reviewed word reading tasks remain to be established, we propose that other sources of information should ideally be taken into account when assessing premorbid functioning. One potentially promising approach may be to combine several tools in order to increase confidence in estimations of premorbid intelligence. For instance, demographic equations which are independent of current abilities and take little time to complete could likely complement word reading assessments. This method may offer a sensible provisional solution for clinical settings while further evidence for the validity of currently available measures is being acquired.

Methodological considerations

There are several methodological limitations which should be considered in relation to both the present review and the included papers. First, we were unable to carry out a meta-analysis to formally compare findings from the different studies due to the heterogeneity of tasks and populations assessed. In addition, most of the studies evaluated here applied a cross-sectional design, which can only offer limited insight into the presence and rate of decline on specific tasks associated with dementia. Moreover, the longitudinal studies included tended to focus on people who had already been diagnosed with dementia at the time of the first assessment. While such research can provide information regarding performance changes with the progression of dementia, it does not capture performance prior to disease onset. Additional longitudinal studies, particularly those following participants before onset of dementia, would be useful for improving our understanding of the validity of different instruments for measuring premorbid intelligence.

A strength of this review is the inclusion of both English and translated versions of premorbid intelligence assessments, which were investigated in eight different countries. However, nearly all of

the included studies were conducted in countries with strong economies and educational systems. It was previously demonstrated that education is highly predictive of word reading, lexical decision, and vocabulary scores (Crawford *et al.*, 1988b; Kosmidis *et al.*, 2006; Starr *et al.*, 1992; Walker *et al.*, 2009), and it has been suggested that reading scores can be used as a proxy for quality of education (Manly *et al.*, 2002). The validity of using verbal tasks to assess premorbid intelligence in dementia, however, has rarely been investigated in countries with fewer socioeconomic or educational resources. Only one study in the present review, which was based in Brazil, focused on participants with low levels of education (De Oliveira *et al.*, 2014). Here, it was found that there were no significant differences between healthy adults and patients with Alzheimer's disease on a Vocabulary task. An untested possibility, however, is that these verbal tasks may underestimate cognitive abilities which are less strongly associated with educational background. Furthermore, it is unclear whether other, more frequently employed task types (e.g. irregular word reading) are useful for estimating premorbid intelligence in individuals with limited access to high-quality education. There is thus a clear need for more extensive testing of premorbid intelligence measures in low resource countries.

In addition, illiteracy presents a particularly important issue for the use of word reading assessments, as this inherently prevents accurate task performance even if cognition is unimpaired. According to recent estimates, approximately 750 million adults worldwide lack basic reading and writing skills (UNESCO Institute for Statistics, 2017). Exclusively relying on reading tasks as a measure of premorbid intelligence could therefore negatively affect many people across the world. In the broader neuropsychometry literature, some measures focusing on nonverbal skills have specifically been developed to measure intelligence in adults with low literacy (e.g. Ryan *et al.*, 2008). However, such instruments are scarce and have not yet been validated in patients with dementia. As an alternative, it has been suggested that informant-based questionnaires may be useful for estimating premorbid intelligence in individuals with low educational levels (Apolinario *et al.*, 2013). A drawback of this approach is that the estimated abilities are dependent on the accuracy of the information provided by the informant. As well as re-evaluating existing instruments across countries, it would therefore be valuable to develop novel objective measures which are less dependent on education and literacy.

Assessments of study quality and risk of bias indicated that the majority of studies did not satisfy criteria for sample size justification, participant

selection, and nonresponse. It is therefore possible that results were influenced by lack of power and/or a selection bias. This particularly complicates interpreting findings of instruments which were only assessed in one study. Finally, in addition to assessing the validity of premorbid intelligence tasks in dementia, this review set out to collate information regarding the psychometric properties of available instruments. However, only one of the studies reviewed here, which investigated performance on the WRAT-III, reported test-retest reliability within the dementia patient group (Ashendorf *et al.*, 2009). Reliability testing for other frequently employed tasks is therefore urgently needed.

Conclusion

Early detection and treatment of dementia are highly dependent on accurate information regarding premorbid functioning. The studies reviewed here demonstrate that there is a large number of tasks available for estimating premorbid intelligence, which are predominantly language-based. These verbal tasks appear to hold some promise for the assessment of people with mild dementia, but maybe unsuitable for individuals presenting with more severe cognitive impairments. Conclusions are limited by the fact that few tools have been investigated across multiple studies and direct comparisons of different instruments are rare. In addition, while there is some evidence supporting the use of verbal assessments across different languages, more extensive testing is needed to determine whether such measures are suitable for use in countries with lower socioeconomic and educational resources. We propose that, in clinical practice, it may be sensible to combine tools based on different mechanisms (e.g. word reading and demographic equations) in order to improve estimates of intelligence. In addition, longitudinal studies contrasting different measures would be valuable to confirm the validity of premorbid intelligence measures, and could thereby contribute to enhancing diagnostic procedures for people living with dementia worldwide.

Conflict of interest

None.

Description of authors' roles

M. Overman designed the study, collected and analyzed the data, and wrote the paper. S. Lee-worthy assisted with data collection and analysis. T. Welsh was responsible for the supervision of the study and revising drafts of the paper.

References

- Alves, L., Simões, M. R., Martins, C., Freitas, S. and Santana, I. (2013). TeLPI performance in subjects with mild cognitive impairment and Alzheimer disease: a validation study. *Alzheimer's Disease & Associated Disorders*, 27, 324–329. doi: [10.1097/WAD.0b013e31827bdc8c](https://doi.org/10.1097/WAD.0b013e31827bdc8c).
- Alzheimer's Disease International. (2012). National Alzheimer and dementia plans, planned policies and activities: Early detection, access to diagnosis, and stigma reduction. Available at: <https://www.alz.co.uk/sites/default/files/national-alzheimer-and-dementia-plans.pdf>; last accessed July 29, 2020.
- Alzheimer's Disease International. (2015). The global impact of dementia: an analysis of prevalence, incidence, cost and trends. Available at: <https://www.alz.co.uk/research/world-report-2015>; last accessed July 29, 2020.
- Apolinario, D. et al. (2013). Estimating premorbid cognitive abilities in low- educated populations. *PLoS One*, 8, e60084. doi: [10.1371/journal.pone.0060084](https://doi.org/10.1371/journal.pone.0060084).
- Ashendorf, L., Jefferson, A. L., Green, R. C. and Stern, R. A. (2009). Test-retest stability on the WRAT-3 reading subtest in geriatric cognitive evaluations. *Journal of Clinical and Experimental Neuropsychology*, 31, 605–610. doi: [10.1080/13803390802375557](https://doi.org/10.1080/13803390802375557).
- Baddeley, A., Emslie, H. and Nimmo-Smith, I. (1993). The Spot-the-Word test: a robust estimate of verbal intelligence based on lexical decision. *British Journal of Clinical Psychology*, 32, 55–65. doi: [10.1111/j.2044-8260.1993.tb01027.x](https://doi.org/10.1111/j.2044-8260.1993.tb01027.x).
- Barona, A., Reynolds, C. R., and Chastain, R. (1984). A demographically based index of premorbid intelligence for the WAIS-R. *Journal of Consulting and Clinical Psychology*, 52, 885–887. doi: [10.1037/0022-006X.52.5.885](https://doi.org/10.1037/0022-006X.52.5.885).
- Beardsall, L. and Huppert, F. (1997). Short NART, CCRT and Spot-the-Word: comparisons in older and demented persons. *British Journal of Clinical Psychology*, 36, 619–622. doi: [10.1111/j.2044-8260.1997.tb01266.x](https://doi.org/10.1111/j.2044-8260.1997.tb01266.x).
- Binkau, S., Berwig, M., Jänichen, J. and Gertz, H.-J. (2014). Is the MWT-A suitable for the estimation of premorbid intelligence level? *GeroPsych*, 27, 33–39. doi: [10.1024/1662-9647/a000098](https://doi.org/10.1024/1662-9647/a000098).
- Bright, P., Hale, E., Gooch, V. J., Myhill, T. and van der Linde, I. (2018). The National Adult Reading Test: restandardisation against the Wechsler Adult Intelligence Scale - Fourth edition. *Neuropsychological Rehabilitation*, 28, 1019–1027. doi: [10.1080/09602011.2016.1231121](https://doi.org/10.1080/09602011.2016.1231121).
- Cockburn, J., Keene, J., Hope, T. and Smith, P. (2000). Progressive decline in NART score with increasing dementia severity. *Journal of Clinical and Experimental Neuropsychology*, 22, 508–517. doi: [10.1076/1380-3395\(200008\)22:4;1-0;FT508](https://doi.org/10.1076/1380-3395(200008)22:4;1-0;FT508).
- Cohen, J. (1988). *Statistical power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic.
- Crawford, J. R. and Allan, K. M. (1997). Estimating premorbid WAIS-R IQ with demographic variables: Regression equations derived from a UK sample. *The Clinical Neuropsychologist*, 11, 192–197. doi: [10.1080/13854049708407050](https://doi.org/10.1080/13854049708407050).
- Crawford, J. R., Parker, D. M. and Besson, J. A. (1988a). Estimation of premorbid intelligence in organic conditions. *British Journal of Psychiatry*, 153, 178–181. doi: [10.1192/bjp.153.2.178](https://doi.org/10.1192/bjp.153.2.178).
- Crawford, J. R., Parker, D. M., Stewart, L. E., Besson, J. A. O. and De Lacey, G. (1989a). Prediction of WAIS IQ with the National Adult Reading Test: cross-validation and extension. *British Journal of Clinical Psychology*, 28, 267–273. doi: [10.1111/j.2044-8260.1989.tb01376.x](https://doi.org/10.1111/j.2044-8260.1989.tb01376.x).
- Crawford, J. R., Stewart, L. E., Cochrane, R. H., Foulds, J. A., Besson, J. A. and Parker, D. M. (1989b). Estimating premorbid IQ from demographic variables: regression equations derived from a UK sample. *British Journal of Clinical Psychology*, 28, 275–278. doi: [10.1111/j.2044-8260.1989.tb01377.x](https://doi.org/10.1111/j.2044-8260.1989.tb01377.x).
- Crawford, J. R., Stewart, L. E., Garthwaite, P. H., Parker, D. M. and Besson, J. A. O. (1988b). The relationship between demographic variables and NART performance in normal subjects. *British Journal of Clinical Psychology*, 27, 181–182. doi: [10.1111/j.2044-8260.1988.tb0077-x](https://doi.org/10.1111/j.2044-8260.1988.tb0077-x).
- De Oliveira, M. O., Nitrini, R., Yassuda, M. S. and Brucki, S. M. (2014). Vocabulary is an appropriate measure of premorbid intelligence in a sample with heterogeneous educational level in Brazil. *Behavioural Neurology*, 875960. doi: [10.1155/2014/875960](https://doi.org/10.1155/2014/875960).
- Downes, M. J., Brennan, M. L., Williams, H. C. and Dean, R. S. (2016). Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ Open*, 6, e011458. doi: [10.1136/bmjopen-2016-011458](https://doi.org/10.1136/bmjopen-2016-011458).
- Dubois, B., Padovani, A., Scheltens, P., Rossi, A. and Dell'Agnello, G. (2016). Timely diagnosis for Alzheimer's disease: a literature review on benefits and challenges. *Journal of Alzheimer's Disease*, 49, 617–631. doi: [10.3233/JAD-150692](https://doi.org/10.3233/JAD-150692).
- Dykiert, D. and Deary, I. J. (2013). Retrospective validation of WTAR and NART scores as estimators of prior cognitive ability using the Lothian Birth Cohort 1936. *Psychological Assessment*, 25, 1361–1366. doi: [10.1037/a0033623](https://doi.org/10.1037/a0033623).
- Gauthier, S. et al. (2006). Mild cognitive impairment. *Lancet*, 367, 1262–1270. doi: [10.1016/S0140-6736\(06\)68542-5](https://doi.org/10.1016/S0140-6736(06)68542-5).
- Goldstein, F. C., Gary, H. E. and Levin, H. S. (1986). Assessment of the accuracy of regression equations proposed for estimating premorbid intellectual functioning on the Wechsler Adult Intelligence Scale. *Journal of Clinical and Experimental Neuropsychology*, 8, 405–412. doi: [10.1080/01688638608401330](https://doi.org/10.1080/01688638608401330).
- Griffin, S. L., Mindt, M. R., Rankin, E. J., Ritchie, A. J. and Scott J. G. (2002). Estimating premorbid intelligence: comparison of traditional and contemporary methods across the intelligence continuum. *Archives of Clinical Neuropsychology*, 17, 497–507. doi: [10.1093/arclin/17.5.497](https://doi.org/10.1093/arclin/17.5.497).
- Hessler, J., Jahn, T., Kurz, A. and Bickel, H. (2013). The MWT-B as an estimator of premorbid intelligence in MCI and dementia. *Zeitschrift für Neuropsychologie*, 24, 129–136. doi: [10.1024/1016-264X/a000099](https://doi.org/10.1024/1016-264X/a000099).
- Jastak, S. and Wilkinson, G. (1984). *The Wide Range Achievement Test: Manual of instructions*. Wilmington, DE: Jastak Associates.
- Kosmidis, M. H., Taspkini, K. and Folia, V. (2006). Lexical processing in illiteracy: effect of literacy or education? *Cortex*, 42, 1021–1027. doi: [10.1016/S0010-9452\(08\)70208-9](https://doi.org/10.1016/S0010-9452(08)70208-9).

- Manly, J. J., Jacobs, D. M., Touradji, P., Small, S. A. and Stern, Y.** (2002). Reading level attenuates differences in neuropsychological test performance between African American and White elders. *Journal of the International Neuropsychological Society*, 8, 341–348. doi: [10.1017/S135561770102015X](https://doi.org/10.1017/S135561770102015X).
- Matsuoka, K., Uno, M., Kasai, K., Koyama, K. and Kim, Y.** (2006). Estimation of premorbid IQ in individuals with Alzheimer's disease using Japanese ideographic script (Kanji) compound words: Japanese version of National Adult Reading Test. *Psychiatry and Clinical Neurosciences*, 60, 332–339. doi: [10.1111/j.1440-1819.2006.01510.x](https://doi.org/10.1111/j.1440-1819.2006.01510.x).
- McCarthy, F., Burns, W. J. and Sellers, A. H.** (2005). Discrepancies between premorbid and current IQ as a function of progressive mental deterioration. *Perceptual and Motor Skills*, 100, 69–76. doi: [10.2466/pms.100.1.69-76](https://doi.org/10.2466/pms.100.1.69-76).
- McFarlane, J., Welch, J. and Rodgers, J.** (2006). Severity of Alzheimer's disease and effect on premorbid measures of intelligence. *British Journal of Clinical Psychology*, 45, 453–463. doi: [10.1348/014466505X71245](https://doi.org/10.1348/014466505X71245).
- McGurn, B. et al.** (2004). Pronunciation of irregular words is preserved in dementia, validating premorbid IQ estimation. *Neurology*, 62, 1184–1186. doi: [10.1212/01.WNL.0000103169.80910.8B](https://doi.org/10.1212/01.WNL.0000103169.80910.8B).
- Moher, D. et al.** (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4, 1.
- National Institute for Health and Care Excellence** (2011). Donepezil, galantamine, rivastigmine and memantine for the treatment of Alzheimer's disease. Available at: <https://www.nice.org.uk/guidance/ta217>; last accessed 29 July, 2020.
- Nelson, H. E.** (1982). *National Adult Reading Test*. Windsor, UK: NFER-Nelson.
- Nelson, H. E. and McKenna, P.** (1975). The use of current reading ability in the assessment of dementia. *British Journal of Social and Clinical Psychology*, 14, 259–267. doi: [10.1111/j.2044-8260.1975.tb00178.x](https://doi.org/10.1111/j.2044-8260.1975.tb00178.x).
- Nelson, H. E. and O'Connell, A.** (1978). Dementia: the estimation of premorbid intelligence levels using the New Adult Reading Test. *Cortex*, 14, 234–244. doi: [10.1016/S0010-9452\(78\)80049-5](https://doi.org/10.1016/S0010-9452(78)80049-5).
- Nelson, H. E. and Willison, J. R.** (1991). *The Revised National Adult Reading Test - Test Manual*. Windsor: NFER-Nelson.
- Nichols, E. et al.** (2019). Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurology*, 18, 88–106. doi: [10.1016/S1474-4422\(18\)30403-4](https://doi.org/10.1016/S1474-4422(18)30403-4).
- O'Carroll, R.** (1995). The assessment of premorbid ability: a critical review. *Neurocase*, 1, 83–89. doi: [10.1080/13554799508402350](https://doi.org/10.1080/13554799508402350).
- O'Carroll, R. E.** (1987). The inter-rater reliability of the National Adult Reading Test (NART): a pilot study. *British Journal of Clinical Psychology*, 26, 229–230. doi: [10.1111/j.2044-8260.1987.tb01352.x](https://doi.org/10.1111/j.2044-8260.1987.tb01352.x).
- O'Carroll, R. E., Baikie, E. M. and Whittick, J. E.** (1987). Does the National Adult Reading Test hold in dementia? *British Journal of Clinical Psychology*, 26, 315–316. doi: [10.1111/j.2044-8260.1987.tb01367.x](https://doi.org/10.1111/j.2044-8260.1987.tb01367.x).
- O'Carroll, R. E., Prentice, N., Murray, C., van Beck, M., Ebmeier, K. P. and Goodwin, G. M.** (1995). Further evidence that reading ability is not preserved in Alzheimer's disease. *British Journal of Psychiatry*, 167, 659–662. doi: [10.1192/bjp.167.5.659](https://doi.org/10.1192/bjp.167.5.659).
- Patterson, K., Graham, N. and Hodges, J. R.** (1994). Reading in dementia of the Alzheimer type: a preserved ability? *Neuropsychology*, 8, 395–407. doi: [10.1037/0894-4105.8.3.395](https://doi.org/10.1037/0894-4105.8.3.395).
- Robinson, L., Tang, E. and Taylor, J. P.** (2015). Dementia: timely diagnosis and early intervention. *BMJ*, 350, h3029. doi: [10.1136/bmj.h3029](https://doi.org/10.1136/bmj.h3029).
- Rolstad, S. et al.** (2008). The Swedish National Adult Reading Test (NART-SWE): a test of premorbid IQ. *Scandinavian Journal of Psychology*, 49, 577–582. doi: [10.1111/j.1467-9450.2008.00677.x](https://doi.org/10.1111/j.1467-9450.2008.00677.x).
- Ryan, E. L., Byrd, D., Mindt, M. R., Rausch, W. J. and Morgello, S.** (2008). Understanding the neuropsychological profile of HIV + participants with low literacy: role of the general ability measure for adults (GAMA). *Clinical Neuropsychology*, 22, 1018–1034. doi: [10.1080/13854040701750883](https://doi.org/10.1080/13854040701750883).
- Schmand, B., Geerlings, M. I., Jonker, C. and Lindeboom, J.** (1998). Reading ability as an estimator of premorbid intelligence: does it remain stable in emergent dementia? *Journal of Clinical and Experimental Neuropsychology*, 20, 42–51. doi: [10.1076/jcen.20.1.42.1485](https://doi.org/10.1076/jcen.20.1.42.1485).
- Serrao, V. T., Brucki, S. M. D., Campanholo, K. R., Mansur, L. L., Nitrini, R. and Miotto, E. C.** (2015). Performance of a sample of patients with Mild Cognitive Impairment (MCI), Alzheimer's Disease (AD) and healthy elderly on a lexical decision test (LDT) as a measure of pre-morbid intelligence. *Dementia & Neuropsychologia*, 9, 265–269. doi: [10.1590/1980-57642015dn93000009](https://doi.org/10.1590/1980-57642015dn93000009).
- Shah, H. et al.** (2016). Research priorities to reduce the global burden of dementia by 2025. *Lancet Neurology*, 15, 1285–1294. doi: [10.1016/S1474-4422\(16\)30235-6](https://doi.org/10.1016/S1474-4422(16)30235-6).
- Sharpe, K. and O'Carroll, R.** (1991). Estimating premorbid intellectual level in dementia using the National Adult Reading Test: a Canadian study. *British Journal of Clinical Psychology*, 30, 381–384. doi: [10.1111/j.2044-8260.1991.tb00962.x](https://doi.org/10.1111/j.2044-8260.1991.tb00962.x).
- Starr, J. M., Whalley, L. J., Inch, S. and Shering, P. A.** (1992). The quantification of the relative effects of age and NART-predicted IQ on cognitive function in healthy old people. *International Journal of Geriatric Psychiatry*, 7, 153–157. doi: [10.1002/gps.930070303](https://doi.org/10.1002/gps.930070303).
- Stebbins, G. T., Wilson, R. S., Gilley, D. W., Bernard, B. A. and Fox, J. H.** (1990). Use of the national adult reading test to estimate premorbid iq in dementia. *Clinical Neuropsychology*, 4, 18–24. doi: [10.1080/13854049008401493](https://doi.org/10.1080/13854049008401493).
- UNESCO Institute for Statistics** (2017). Literacy rates continue to rise from one generation to the next. *Fact Sheet*, 45. http://uis.unesco.org/sites/default/files/documents/fs45-literacy-rates-continue-rise-generation-to-next-en-2017_0.pdf
- Veiel, H. O. and Koopman, R. F.** (2001). The bias in regression-based indices of premorbid IQ. *Psychological Assessment*, 13, 356–368. doi: [10.1037/1040-3590.13.3.356](https://doi.org/10.1037/1040-3590.13.3.356).

- Vos, T. et al.** (2017). Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*, 390, 1211–1259. doi: [10.1016/S0140-6736\(17\)32154-2](https://doi.org/10.1016/S0140-6736(17)32154-2).
- Walker, A. J., Batchelor, J. and Shores, A.** (2009). Effects of education and cultural background on performance on WAIS-III, WMS-III, WAIS-R and WMS-R measures: systematic review. *Australian Psychologist*, 44, 216–223. doi: [10.1080/00050060902833469](https://doi.org/10.1080/00050060902833469).
- Wechsler, D.** (1997). *WAIS-III administration and scoring manual*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D.** (2001). *Wechsler Test of Adult Reading (WTAR)*. San Antonio, TX: Psychological Corporation.
- Wechsler, D.** (2011). *The Test of Premorbid Functioning - UK Version (TOPF UK) Manual*. San Antonio, TX: Psychological Corporation.
- Weinborn, M. et al.** (2018). Trajectories of irregular word reading ability as a proxy for premorbid intelligence in Alzheimer's disease, mild cognitive impairment, and healthy aging: a longitudinal study. *Psychological Assessment*, 30, 1308–1316. doi: [10.1037/pas0000565](https://doi.org/10.1037/pas0000565).
- Wilson, R. S., Rosenbaum, G., Brown, G., Rourke, D., Whitman, D. and Grisell, J.** (1978). An index of premorbid intelligence. *Journal of Consulting and Clinical Psychology*, 46, 1554–1555. doi: [10.1037/0022-006X.46.6.1554](https://doi.org/10.1037/0022-006X.46.6.1554).
- World Health Organization** (2017). Global action plan on the public health response to dementia 2017–2025. Available at: https://www.who.int/mental_health/neurology/dementia/action_plan_2017_2025/en/; last accessed July 29, 2020.
- Yuspeh, R. L. and Vanderploeg, R. D.** (2000). Spot-the-Word: a measure for estimating premorbid intellectual functioning. *Archives of Clinical Neuropsychology*, 15, 319–326. doi: [10.1093/arclin/15.4.319](https://doi.org/10.1093/arclin/15.4.319).