Journal of the International Neuropsychological Society (2024), 1–7
doi:10.1017/S1355617724000420

Brief Communication

The prediction limits of the National Adult Reading Test and its abbreviated and international variants

Ian van der Linde[1,2] and Peter Bright[1,3]

[1]Cognition and Neuroscience Group, ARU Centre for Mind and Behaviour, Faculty of Science & Engineering, Anglia Ruskin University, Cambridge, UK, [2]School of Computing and Information Science, Faculty of Science & Engineering, Anglia Ruskin University, Cambridge, UK and [3]School of Psychology, Sport and Sensory Science, Faculty of Science & Engineering, Anglia Ruskin University, Cambridge, UK

Abstract

Objective: Premorbid tests estimate cognitive ability prior to neurological condition onset or brain injury. Tests requiring oral pronunciation of visually presented irregular words, such as the National Adult Reading Test (NART), are commonly used due to robust evidence that word familiarity is well-preserved across a range of neurological conditions and correlates highly with intelligence. Our aim is to examine the prediction limits of NART variants to assess their ability to accurately estimate premorbid IQ. **Method:** We examine the prediction limits of 13 NART variants, calculate which IQ classification system categories are reachable in principle, and consider the proportion of the adult population in the target country falling outside the predictable range. **Results:** Many NART variants cannot reach higher or lower IQ categories due to floor/ceiling effects and inherent limitations of linear regression (used to convert scores to predicted IQ), restricting clinical accuracy in evaluating premorbid ability (and thus the magnitude of impairment). For some variants this represents a sizeable proportion of the target population. **Conclusions:** Since both higher and lower IQ categories are unreachable in principle, we suggest that future NART variants consider polynomial or broken-stick fitting (or similar methods) and suggest that prediction limits should be routinely reported.

Keywords: Neuropsychological tests; premorbid intelligence; National Adult Reading Test; prediction limits; linear models; test development

(Received 12 March 2024; final revision 9 July 2024; accepted 12 August 2024)

Introduction

Comparison of premorbid IQ estimates against objective measures of current IQ enables the magnitude of cognitive impairment to be evaluated in neurological patients. This is useful for research, medicolegal, diagnostic and clinical management purposes. Premorbid IQ tests requiring the oral pronunciation of phonologically irregular words are commonly used due to robust evidence that single word pronunciation knowledge is preserved (held) across a wide range of conditions (Crawford, 1992; McGurn et al., 2004; O'Carroll, 1995; Sharpe & O'Carroll, 1991), and because the relationship between word reading and intelligence is largely independent of age and social class (Nelson, 1982). Alternative approaches that examine word familiarity independently of pronunciation include lexical decision tests like Spot-the-Word (Baddeley et al., 1993; Baddeley & Crawford, 2012; van der Linde et al., 2022), in which participants are asked to select real words rather than plausible non-word distractors. Lexical decision tests are particularly useful where speech production is impaired. However, since oral pronunciation tests are used most often, and are underpinned by a greater quantity of normative data, we focus on this approach.

The National Adult Reading Test (NART; Bright et al., 2018; Nelson & Willison, 1991; Nelson, 1982) is a free, fast, well-established and widely used word pronunciation-based premorbid IQ test. Evidence indicates equivalent or better predictive validity compared to using demographic data alone, using the best performing subtest from an IQ battery, or undertaking *hold* vs *no-hold* subtest comparisons (Bright et al., 2002; Bright and van der Linde, 2020). The most recent restandardization of the NART (Bright et al., 2018) enables estimation of full-scale IQ (FSIQ) on the current gold-standard Wechsler Adult Intelligence Scale – Fourth Edition (Wechsler, 2008).

Numerous variants of the original NART (Nelson, 1982) have been developed for revalidation against new revisions of IQ batteries (e.g., Bright et al., 2018; Nelson & Willison, 1991), abbreviation (e.g., Beardsall & Brayne, 1990 [Short NART]; Uttl, 2002 [NAART35]; McGrory et al., 2015 [mini-NART]; Mackinnon & Wooden, 2015; van der Linde & Bright, 2018 [NART17]), and internationalization (e.g., Blair & Spreen, 1989 [USA NART-R]; Schmand et al., 1991 [Dutch DART]; Grober et al., 1991 [USA AMNART]; Hennessy & Mackenzie, 1995 [Australian AUSNART]; Dalsgaard, 1998 [Danish DART];

Corresponding author: Ian van der Linde; Email: ian.vanderlinde@aru.ac.uk
Cite this article: van der Linde I. and Bright P. The prediction limits of the National Adult Reading Test and its abbreviated and international variants. *Journal of the International Neuropsychological Society*, 1–7, https://doi.org/10.1017/S1355617724000420

Mackinnon et al., 1999 [French fNART]; Vaskinn & Sundet, 2001 [Norwegian NART]; Matsuoka et al., 2006 [Japanese JART]; Rolstad et al., 2008 [Swedish NART-SWE]; Starkey & Halliday, 2011 [New Zealand NZART]; Watt, Ong & Crowe, 2016; Karakuła-Juchnowicz & Stecka, 2017 [Polish PART]; Yi et al., 2017 [Korean KART]). Some international variants provide new, population-appropriate regression equations to estimate premorbid IQ using the original word NART stimuli (e.g., Barker-Collo et al, 2011; Watt et al., 2018), some modify stimuli or grading rules to address differences in dialect/pronunciation (e.g., Hennessy & Mackenzie, 1995), while others propose entirely new sets of word stimuli in the local language (e.g., Krámská, 2014 [Czech Reading Test CRT]; Alves, Simões, & Martins, 2011 [Portuguese Irregular Word Reading Test TELPI]). However, most still provide a regression equation to estimate premorbid intelligence from reading test score.

In the development of the original NART and its variants calibration data were collected to calculate a straight line of best fit relating test score to the predicted variable (typically full-scale IQ, but sometimes constituent index scores). Clinicians use the resultant linear regression equation to obtain a premorbid IQ estimate, typically from the number of word pronunciation errors committed, although some provide conversion tables instead of, or in addition to, an equation. It is well-known that linear regression is less accurate for samples at the high and low end of a distribution (Basso et al., 2000; Graves et al., 1999; Griffin et al., 2002; Veiel & Koopman, 2001). In part, this is because fitting a straight line to normally distributed data (such as IQ scores) will lead to a poor fit at the tails of the distribution, along with general floor and ceiling effects.

The NART remains a popular and effective tool; however, its public domain status has led to a proliferation of variants for purposes such as those outlined above. These variants have never been systematically compared to assess their numerical prediction limits, or the reachability of IQ categories in standard classification systems. Such an evaluation is important since operating over a restricted IQ range will necessarily exclude a proportion of the target population (*viz.*, those who premorbidly possessed comparatively low or high IQ) from accurate clinical assessment, leading to suboptimal diagnosis and clinical management decisions.

In this article we review the specific numerical corollaries of these issues for all NART variants identified that give a regression equation to calculate FSIQ that does not require demographic variables, and where the test was not developed for a narrow clinical condition. We related the range of premorbid IQs that can be produced to categorical labels in common IQ classification systems and evaluate the proportion of the target population that falls outside the predictable range.

Method

A straight-line equation sets a NART score (or the number of errors committed), x, in the form of first-degree polynomial $y = mx + c$, where y is the premorbid IQ estimate, m is a coefficient of x (line equation gradient term, sometimes called the regression coefficient) and c is an additive constant (line equation intercept, sometimes called the regression constant). Using the regression equation provided with each NART variant (gradient and intercept are given in Table 1 which, since the line is strictly decreasing, would be used in the form $y = c - mx$) we calculated predicted IQ where a participant does not pronounce any test word correctly, i.e., maximizing the gradient term (m) and subtracting from the intercept (c). Using current population estimates, we then calculated the percentage of the target population that falls below that IQ score. We then calculated the highest attainable IQ score by supposing that no errors were committed, i.e., zeroed the gradient term (m) to leave only the intercept (c). Again, using current population estimates, we calculated the percentage of the target population that is above that IQ score. For each variant, we calculated the statistical range of IQ scores that are theoretically reachable, and the percentage of the target population for the respective test that falls outside that range. We then related the range of attainable scores to standard IQ classification systems.

Results

First, we present the upper and lower limits and range of each NART variant. Next, we evaluate which IQ class categories fall outside these limits. We then comment on clinical implications for patients with comparatively high or low premorbid intelligence.

Our main findings are presented in Table 1, showing that a significant proportion of the non-clinical population fall below the lowest predictable score. In the original NART (Nelson, 1982), Danish (Hjorthøj et al, 2013), Norwegian (Vaskinn & Sundet, 2001), and Polish variants (Karakuła-Juchnowicz & Stecka, 2017), this equates to approximately 1 in 5 (~20%) of the general population (Rain and Zaborowska, 2022). In the Australian (Hennessy & Mackenzie, 1995) and US (Blair and Spreen, 1989) variants it equates to approximately 1 in 10 (10%) of the general population (Rain and Zaborowska, 2022).

In standard IQ classification systems (Table 2) this would lead to widespread misclassification in the current WAIS-IV classification system (Wechsler, 2008); only Nelson & Willison (1991) can, barely, produce an IQ in the *Extremely Low* class (<70). Of the NART variants examined, six cannot produce an IQ<80 (Blair & Spreen, 1989; Hennessy & Mackenzie, 1995; Hjorthøj et al., 2013; Karakuła-Juchnowicz & Stecka, 2017; Nelson, 1982; Vaskinn & Sundet, 2001), which would cause all those in the *Borderline* or *Extremely Low* classes to be misclassified as *Low Average*. In the more granular Stanford-Binet Fifth Edition (SB5; Roid & Pomplun, 2012) classification system, none of the NART variants examined would be capable of producing IQs in the *Moderately Impaired or Delayed* range (which would be misclassified as *Borderline Impaired or Delayed*, or even *Low Average*), and only one of the NART variants examined (Nelson & Willison, 1991) can, barely, predict IQs in the *Mildly Impaired or Delayed* range. Six variants cannot produce an IQ below the *Low Average* range, missing the bottom three categories entirely. In the DAS-II classification system (Dumont et al., 2009), only two of the NART variants can, again barely, predict IQs in the *Very Low* class (Nelson & Wilison, 1991; Starkey & Halliday, 2011), which would be misclassified as *Low* or *Below Average*.

The same is true with high-performing patients whose score tends towards the top of the predictable range, with the French (Mackinnon et al., 1999), Japanese (Matsuoka et al., 2006), and New Zealand (Starkey & Halliday, 2011) variants of the NART unable to reach 1 in 20 (i.e., the top 5% of the population). This translates to millions of individuals (3.5 million from a 2022 French population of 67.5 million; 6.8 million from a 2022 Japanese population of 125.7 million; 0.27 million from a 2022 New Zealand population of 5.1 million).

In the Wechsler IQ classification system, only four of the NART variants examined can produce an IQ in the *Very Superior* (≥130)

https://doi.org/10.1017/S1355617724000420 Published online by Cambridge University Press

Table 1. Lowest and highest predictable IQ score, statistical range, and percentage of population falling below/above/within (percentiles from Rain and Zaborowska, 2022)

Publication	Region	Intercept	Gradient	Test words	Min predictable IQ	% population below	Max predictable IQ	% population above	IQ range	% population outside
Nelson, 1982	UK	127.7	0.826	50	86.40	18.23	127.70	3.24	41.30	21.47
Blair & Spreen, 1989	US and Canada	127.8	0.78	61	80.22	9.36	127.80	3.19	47.58	12.55
Nelson & Willison, 1991	UK	130.6	1.24	50	68.60	1.82	130.60	2.08	62.00	3.90
Mackinnon et al., 1999	France	124.44	1.54	33	73.62	3.93	124.44	5.16	50.82	9.09
Henessy & Mackenzie, 1995	Australia	135.27	0.822	64	82.66	12.40	135.27	0.94	52.61	13.34
Vaskinn and Sundet, 2001	Norway	121.2	0.68	50	87.20	19.67	121.20	7.88	34.00	27.55
Matsuoka et al., 2006	Japan	124.1	0.964	50	75.90	5.41	124.10	5.38	48.20	10.79
Starkey & Halliday, 2011	New Zealand	124.18	0.903	60	70.00	1.34	124.18	5.35	54.18	6.69
Hjorthøj et al., 2013	Denmark	128.5	0.84	50	86.50	18.41	128.50	2.87	42.00	21.28
Bright et al., 2018	UK	126.41	0.9775	50	77.54	6.72	126.41	3.91	48.88	10.63
Watt, Ong and Crowe, 2016	Australia	133.62	1.282	50	69.52	2.11	133.62	1.25	64.10	3.36
Karakuła-Juchnowicz & Stecka, 2017	Poland	126.72	0.7748	50	87.98	21.15	126.72	3.74	38.74	24.89
van der Linde & Bright, 2018	UK	132.71	3.4882	17	73.41	3.81	132.71	1.46	59.30	5.27

class (Hennessy & Mackenzie, 1995; Nelson & Willison, 1991; Watt et al., 2016; van der Linde & Bright, 2018), and one can, just barely, produce an IQ in the *Very High* class (Vaskinn & Sundet, 2001). In the SB5 classification system, no NART variant can detect an IQ in the *Very Gifted or Highly Advanced* class, and only four can detect an IQ in the *Gifted or Very Advanced* range. In the DAS-II classification system, only three variants can detect the *Very High* class.

Discussion

The compressed predictable IQ range stems from fitting a straight line to the datapoints of participants who have completed both the NART variant and, for calibration purposes, a full standard IQ test battery or (in some cases) a specific subtest. Perhaps counter-intuitively, where straight-line fitting is used, collecting more datapoints may not help: by definition, if participants across a wide range of ability levels are recruited, most will *not* be at the extrema and the gradient (m) and intercept (c) of the straight line will be unperturbed.

Similarly, developing tests of greater length cannot help: in terms of statistical range, the three highest-valued variants are the 50-word Australian restandardization (Starkey & Halliday, 2011) at 64.10, the first British restandardization (Nelson & Willison, 1991), also 50 words, at 62.00, but also the 17-word NART variant proposed in van der Linde and Bright (2018) at 59.30. Conversely, the three variants with the lowest ranges all have 50 words: Vaskinn & Sundet (2001) at 34.00; Karakuła-Juchnowicz & Stecka (2017) at 38.74; Nelson (1982) at 41.30.

The clinical significance of these issues is potentially large; they are poorly suited for use with patients who, prior to their neurological condition, would have fallen into the lower IQ classification ranges since the clinician's ability to accurately gauge the severity of their current impairment will be limited. Specifically, since premorbid IQ will be overestimated, a clinical evaluation will likewise overestimate the magnitude of impairment, on the assumption that current IQ will have fallen relative to the true pre-clinical IQ. For instance, a patient with pre-clinical IQ <70 may yield an overestimated premorbid IQ estimate of 80 due to floor effects, spuriously indicating an increase in cognitive ability. A measure of current IQ will produce a lower than pre-clinical score, and the difference between this and the estimated premorbid IQ will be larger than it should be, thereby causing the magnitude of the patient's impairment to be overestimated.

For patients who would have fallen into the higher IQ classification range, ceiling effects will cause premorbid IQ to be underestimated, and a clinical evaluation will underestimate the magnitude of impairment, based on the same assumption. For instance, a patient with pre-clinical IQ >140 may have their premorbid IQ estimated with NART at 130 due to ceiling effects, underestimating their pre-clinical ability. A measure of current IQ will produce a lower than pre-clinical score, likely bringing it closer to the premorbid IQ estimate ceiling, such that the difference between current IQ and premorbid estimate will be smaller than it should be, thereby causing the magnitude of the patient's impairment to be underestimated. Joseph et al. (2021) reported that the Test of Premorbid Functioning (TOPF; Wechsler, 2011), which is very similar to the NART, underestimated premorbid intelligence for around one third of their high-performing participants and was particularly poor for those falling into *Above Average* and *Superior* classes. This is despite the fact that the TOPF uses a third-degree polynomial rather than straight-line fit. Other work indicates that NART and its variants may estimate premorbid IQ more accurately than TOPF (Reale-Caldwell et al., 2021), perhaps because the specific polynomial used to fit TOPF calibration data is suboptimal.

In some neuropsychological tests, instructions suggest using a different line equation for scores above or below certain thresholds, to administer an alternative or abbreviated test, or simply to declare the prediction unreliable (which seems quite reasonable if the participant fails to respond correctly to nearly/all test words, rather than allocating a *Low Average* or *Borderline* IQ, as would be the case if some NART variants were used imprudently). For instance, in the original NART it is recommended that participants

https://doi.org/10.1017/S1355617724000420 Published online by Cambridge University Press

Table 2. Standard IQ classification systems with highest and lowest predictable IQ for each NART variant highlighted [note: also included are WTAR (Wechsler, 2001) and STW2 (Baddeley & Crawford, 2011) for comparison]

<table>
<tr><th colspan="2">IQ</th><th>≤40</th><th>41</th><th>42</th><th>43</th><th>44</th><th>45</th><th>46</th><th>47</th><th>48</th><th>49</th><th>50</th><th>51</th><th>52</th><th>53</th><th>54</th><th>55</th><th>56</th><th>57</th><th>58</th><th>59</th><th>60</th><th>61</th><th>62</th><th>63</th><th>64</th><th>65</th><th>66</th><th>67</th><th>68</th><th>69</th><th>70</th><th>71</th><th>72</th><th>73</th><th>74</th><th>75</th><th>76</th><th>77</th><th>78</th><th>79</th><th>80</th><th>81</th><th>82</th><th>83</th><th>84</th><th>85</th><th>86</th><th>87</th><th>88</th><th>89</th><th>90</th><th>91</th><th>92</th><th>93</th><th>94</th><th>95</th><th>96</th><th>97</th><th>98</th><th>99</th><th>100</th><th>101</th><th>102</th><th>103</th><th>104</th><th>105</th><th>106</th><th>107</th><th>108</th><th>109</th><th>110</th><th>111</th><th>112</th><th>113</th><th>114</th><th>115</th><th>116</th><th>117</th><th>118</th><th>119</th><th>120</th><th>121</th><th>122</th><th>123</th><th>124</th><th>125</th><th>126</th><th>127</th><th>128</th><th>129</th><th>130</th><th>131</th><th>132</th><th>133</th><th>134</th><th>135</th><th>136</th><th>137</th><th>138</th><th>138</th><th>140</th><th>141</th><th>142</th><th>143</th><th>144</th><th>145</th><th>146</th><th>147</th><th>148</th><th>149</th><th>150</th><th>151</th><th>152</th><th>153</th><th>154</th><th>155</th><th>156</th><th>157</th><th>158</th><th>159</th><th>160</th><th>161</th><th>162</th><th>163</th><th>164</th><th>165</th><th>166</th><th>167</th><th>168</th><th>169</th><th>170</th><th>171</th><th>172</th><th>173</th><th>174</th><th>175</th><th>176</th><th>177</th><th>178</th><th>179</th><th>180</th><th>181</th><th>182</th><th>≥183</th></tr>
<tr><td rowspan="9">CLASSIFICATION SYSTEM</td><td>WAIS-IV 2008</td><td colspan="30">Extremely Low</td><td colspan="10">Borderline</td><td colspan="10">Low Average</td><td colspan="20">Average</td><td colspan="10">High Average</td><td colspan="10">Superior</td><td colspan="54">Very Superior</td></tr>
<tr><td>WISC-V 2014</td><td colspan="30">Extremely Low</td><td colspan="10">Very Low</td><td colspan="10">Low Average</td><td colspan="20">Average</td><td colspan="10">High Average</td><td colspan="10">Very High</td><td colspan="54">Extremely High</td></tr>
<tr><td>SB5 2003</td><td colspan="15">Moderately Impaired or Delayed</td><td colspan="15">Mildly Impaired or Delayed</td><td colspan="10">Borderline Impaired or Delayed</td><td colspan="10">Low Average</td><td colspan="20">Average</td><td colspan="10">High Average</td><td colspan="10">Superior</td><td colspan="15">Gifted or Very Advanced</td><td colspan="39">Gifted or Very Advanced</td></tr>
<tr><td>WJ III 2007</td><td colspan="30">Very Low</td><td colspan="10">Low</td><td colspan="10">Low Average</td><td colspan="21">Average</td><td colspan="10">High Average</td><td colspan="10">Superior</td><td colspan="53">Very Superior</td></tr>
<tr><td>KAIT 1993</td><td colspan="30">Lower Extreme</td><td colspan="10">Well Below Average</td><td colspan="10">Below Average</td><td colspan="20">Average</td><td colspan="9">Above Average</td><td colspan="11">Well Above Average</td><td colspan="54">Upper Extreme</td></tr>
<tr><td>CAS 1997</td><td colspan="30">Well Below Average</td><td colspan="10">Below Average</td><td colspan="10">Low Average</td><td colspan="20">Average</td><td colspan="10">High Average</td><td colspan="10">Superior</td><td colspan="54">Very Superior</td></tr>
<tr><td>DAS-II 2007</td><td colspan="30">Very Low</td><td colspan="10">Low</td><td colspan="10">Below Average</td><td colspan="20">Average</td><td colspan="10">Above Average</td><td colspan="10">High</td><td colspan="54">Very High</td></tr>
<tr><td>RIAS 2003</td><td colspan="30">Significantly Below Average</td><td colspan="10">Moderately Below Average</td><td colspan="10">Below Average</td><td colspan="20">Average</td><td colspan="10">Above Average</td><td colspan="10">Moderately Above Average</td><td colspan="54">Significantly Above Average</td></tr>
<tr><td>Gross 2000</td><td colspan="76"></td><td colspan="14">Mildly Gifted</td><td colspan="15">Moderately Gifted</td><td colspan="15">Highly Gifted</td><td colspan="19">Exceptionally Gifted</td><td colspan="5">Profoundly Gifted</td></tr>
</table>

LOWER LIMIT:
- WTAR (Wechsler, 2001)
- STW2 (Baddeley & Crawford, 2012)
- Nelson & Wilison (1991); Starkey & Halliday (2011)
- Watt et al. (2016)
- van der Linde & Bright (2018)
- Jorge et al. (1991)
- Matsuoka et al. (2006)
- Bright et al. (2018)
- Blair & Spreen (1989)
- Nelson (1982)
- Vaskinn & Sundet (2001); Hjorthøj et al. (2013)
- Karakuła-Juchnowicz & Stecka (2017)

UPPER LIMIT:
- WTAR (Wechsler, 2001)
- Vaskinn & Sundet (2001)
- STW2 (Baddeley & Crawford, 2012)
- Jorge et al. (1991); Matsuoka et al. (2006)
- Bright et al. (2018)
- Karakuła-Juchnowicz & Stecka (2017)
- Nelson (1982); Blair & Spreen (1989)
- Starkey & Halliday (2011); Hjorthøj et al. (2013)
- Nelson & Wilison (1991)
- van der Linde & Bright (2018)
- Watt et al. (2016)

scoring<10 correct words (which are referred to as *poor readers*) take a second test (Schonell Graded Word Reading Test) and that a second regression equation incorporating both scores is used.

It is acknowledged in Nelson & Willison (1991) that a limitation of the NART is that it cannot detect IQs above 128. It is stated that this is less of a problem than it first seems because even those with IQs above 130 typically make one or more NART errors. However, this tacitly acknowledges prediction error and that artificially reduced IQ estimates are, in fact, potentially clinically disadvantageous.

In part, the method of obtaining a straight line of best fit to calibrate NART is used to keep the task of converting a NART score into a premorbid IQ score as simple as possible for the clinician, obviating the need for complex calculations, the application of an algorithm, or the use of computer software. In many cases, for convenience, conversion tables are also provided, so that the regression calculation need not be used in practice (perhaps removing one possible source of error, and speeding the assessment). However, most conversion tables simply provide the linear regression line calculated across the range of possible raw error scores. Despite this, conversion tables could just as easily be used to concretize a non-linear fit. Three possibilities are i. so-called *segmented* or *broken-stick regression,* in which multiple line segments are fit to different intervals of the observed calibration data, such as using a line for the main portion of the fit and two smaller lines for the tails; ii. fitting a cumulative distribution function; and iii. fitting a suitable higher-degree polynomial.

The issues discussed here also apply to tests that estimate constituent indices from the WAIS rather than (or in addition to) FSIQ (e.g., Grober et al, 1991), and to other reading tests, including the Wechsler Test of Adult Reading (WTAR; Wechsler, 2001), Cambridge Contextual Reading Test (CCRT; Beardsall, 1998), and numerous variants of the Word Accentuation Test (WAT; Del Ser et al., 1997 [WAT Spanish]; Burin et al., 2000 [WAT-Argentina]; Gil et al., 2019 [WAT-Brazil Portuguese]), Test Breve di Intelligenza (Colombo et al., 2002 [TIB-Italy]), and to lexical decision tests like Spot-the-Word (STW; Baddeley et al., 1993; Baddeley, & Crawford, 2012), the Swedish Lexical Decision Test (Almkvist et al., 2007), and German Mehrfachwahl-Wortschatz-Intelligenztest (MWT; Lehrl et al., 1995), among others. It has been suggested that the WTAR contains more readily recognized stimuli compared to the NART on average (Bright and van der Linde, 2020), so lower scores corresponding to lower IQ classifications may be even less likely to occur in practice.

The Hopkins Adult Reading Test (HART) provides only regression equations that require demographic information (Schretlen et al., 2009), so cannot be evaluated here. However, the authors of this test indicate that the HART is theoretically less constricted in the range of obtainable IQs than NART-R (Blair and Spreen, 1989), in part *because* of the inclusion of other variables in the regression equation. Whilst true, it is the case that demographic information, such as age and years of education, may not always be available (e.g., in the case of unidentified patient or those with dementia). Demographic information is similarly required in the USA (NAART) revision proposed by Uttl (2002), the New Zealand (NZ-NART) proposed by Barker-Collo et al (2011), and the Korean language KART (Yi et al., 2017). However, it has also been found that demographic information explains relatively little additional variance (e.g., Bright and van der Linde, 2018; Bright et al., 2002). NART-SWE (Rolstad et al., 2008) could not be evaluated due to the test and regression equation being kept private for commercial purposes. It is also the case that even the use of demographic variables in a multi-term first-degree polynomial does not solve the problems outlined above, since they will still produce a straight line and therefore incur poor fit at the distribution tails.

As a consequence of (mostly) being in the public domain, all variants of the NART are *unofficial* in the sense that no standard approval process or quality control mechanisms, beyond academic peer review, are in place. In many cases, publications describing new NART variants include thorough evaluations, including for the difficulty and predictive contribution of individual words, internal consistency and reliability (Osburn, 2000), test-retest reliability (Davidshofer & Murphy, 2005; Smith et al., 1998), inter-rater reliability (Saal et al., 1980), etc. However, what would seem like a critical factor, the upper and lower prediction limits and range of detectable IQs, are not commonly reported, nor is the corollary issue of the in-principle reachability of IQ categories in standard classification systems and the proportion of the target population that falls into these categories. It is also the case that some NART variants are orphaned, in the sense that they have not been recalibrated on the latest revisions on IQ batteries, which may cause their predictive accuracy to drift over time due to the Flynn effect (Flynn, 1987) and variations in word usage. It would seem reasonable to propose that the numerical issues explored here are examined and reported upon in future test variants, and to suggest that current tests are interpreted with caution for patients who are suspected to have had particularly high or low premorbid IQ.

Acknowledgements. None.

Funding statement. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interests. The authors have no conflicts of interest to declare.

References

Almkvist, O., Adveen, M., Henning, L., & Tallberg, I. M. (2007). Estimation of premorbid cognitive function based on word knowledge: The Swedish Lexical Decision Test (SLDT). *Scandinavian Journal of Psychology*, *48*(3), 271–279.

Alves, L., Simões, M. R., & Martins, C. (2012). The estimation of premorbid intelligence levels among Portuguese speakers: The Irregular Word Reading Test (TeLPI). *Archives of Clinical Neuropsychology*, *27*(1), 58–68.

Baddeley, A., & Crawford, J. (2012). *Spot the word*. Pearson Assessment.

Baddeley, A., Emslie, H., & 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*(1), 55–65.

Barker-Collo, S., Thomas, K., Riddick, E., & de Jager, A. (2011). A New Zealand regression formula for premorbid estimation using the National Adult Reading Test. *New Zealand Journal of Psychology*, *40*(2), 47–55.

Basso, M., Bornstein, R., Roper, B., & McCoy, V. (2000). Limited accuracy of premorbid intelligence estimators: A demonstration of regression to the mean. *Clinical Neuropsychologist*, *14*(3), 325–340.

Beardsall, L. (1998). Development of the Cambridge Contextual Reading Test for improving the estimation of premorbid verbal intelligence in older persons with dementia. *British Journal of Clinical Psychology*, *37*(2), 229–240.

Beardsall, L., & Brayne, C. (1990). Estimation of verbal intelligence in an elderly community: A prediction analysis using a shortened NART. *British Journal of Clinical Psychology*, *29*(1), 83–90.

Blair, J. R., & Spreen, O. (1989). Predicting premorbid IQ: A revision of the National Adult Reading Test. *The Clinical Neuropsychologist*, *3*(2), 129–136.

Bright, P., Hale, E., Gooch, V. J., Myhill, T., & van der Linde, I. (2018). The National Adult Reading Test: Restandardisation against the Wechsler adult intelligence scale—fourth edition. *Neuropsychological Rehabilitation*, *28*(6), 1019–1027.

Bright, P., Jaldow, E. L. I., & Kopelman, M. D. (2002). The National Adult Reading Test as a measure of premorbid intelligence: A comparison with estimates derived from demographic variables. *Journal of the International Neuropsychological Society*, *8*(6), 847–854.

Bright, P., & van der Linde, I. (2020). Comparison of methods for estimating premorbid intelligence. *Neuropsychological Rehabilitation*, *30*(1), 1–14.

Burin, D. I., Jorge, R. E., Arizaga, R. A., & Paulsen, J. S. (2000). Estimation of premorbid intelligence: The Word Accentuation Test-Buenos Aires version. *Journal of Clinical and Experimental Neuropsychology*, *22*(5), 677–685.

Colombo, L., Sartori, G., & Brivio, C. (2002). Stima del quoziente intellettivo tramite l'applicazione del TIB (Test Breve di Intelligenza). *Giornale Italiano di Psicologia*, *3*, 613–637.

Crawford, J. R. (1992). Current and premorbid intelligence measures in neuropsychological assessment. In J. R. Crawford, D. M. Parker, & W. McKinlay (Eds.), *A handbook of neuropsychological assessment* (pp. 21–49). Erlbaum.

Dalsgaard, I. (1998). Danish Adult Reading Test (DART). *World Psychiatry*, *6*(1), 38–31.

Davidshofer, K. R., & Murphy, C. O. (2005). *Psychological testing: Principles and applications* (6thed.). Prentice Hall.

Del Ser, T., González-Montalvo, J. I., Martınez-Espinosa, S., Delgado-Villapalos, C., & Bermejo, F. (1997). Estimation of premorbid intelligence in Spanish people with the Word Accentuation Test and its application to the diagnosis of dementia. *Brain and Cognition*, *33*(3), 343–356.

Dumont, R., Willis, J. O., & Elliot, C. D. (2009). *Essentials of DAS-II® Assessment* (pp. 126). Wiley, -

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, *101*(2), 171–191.

Gil, G., Magaldi, R. M., Busse, A. L., Ribeiro, E. S., Brucki, S. M. D., Yassuda, M. S., Jacob-Filho, W., & Apolinario, D. (2019). Development of a word accentuation test for predicting cognitive performance in Portuguese-Speaking populations. *Arquivos de Neuro-Psiquiatria*, *77*(8), 560–567.

Graves, R. E., Carswell, L. M., & Snow, W. G. (1999). An evaluation of the sensitivity of premorbid IQ estimators for detecting cognitive decline. *Psychological Assessment*, *11*, 29–38.

Griffin, S. L., Mindt, M. R., Rankin, E. J., Ritchie, A. J., & Scott, J. G. (2002). Estimating premorbid intelligence: Comparison of traditional and contemporary methods across the intelligence continuum. *Archives of Clinical Neuropsychology*, *17*(5), 497–507.

Grober, E., Sliwinski, M., & Korey, S. R. (1991). Development and validation of a model for estimating premorbid verbal intelligence in the elderly. *Journal of Clinical and Experimental Neuropsychology*, *13*(6), 933–949.

Hennessy, M., & Mackenzie, B. AUSNART: The development of an Australian version of the NART. In: *Treatment issues and long-term outcomes: Proceedings of the 18th Annual Brain Impairment Conference*. Australian Academic Press, 1995, pp. 183–188.

Hjorthøj, C. R., Vesterager, L., & Nordentoft, M. (2013). Test-retest reliability of the Danish Adult Reading Test in patients with comorbid psychosis and cannabis-use disorder. *Nordic Journal of Psychiatry*, *67*(3), 159–163.

Joseph, A. L. C., Lippa, S. M., McNally, S. M., Garcia, K. M., Leary, J. B., Dsurney, J., & Chan, L. (2021). Estimating premorbid intelligence in persons with traumatic brain injury: An examination of the Test of Premorbid Functioning. *Applied Neuropsychology: Adult*, *28*(5), 535–543.

Karakuła-Juchnowicz, H., & Stecka, M. (2017). Polish Adult Reading Test (PART)-construction of Polish test for estimating the level of premorbid intelligence in schizophrenia. *Psychiatria Polska*, *51*(4), 673–685.

Krámská, L. (2014). Český Test Čtení Slov. Propsyco; Otrokovice, Czech Republic. Hodnocení premorbidního intelektu v neuropsychologii.

Lehrl, S., Triebig, G., & Fischer, B. A. N. S. (1995). Multiple choice vocabulary test MWT as a valid and short test to estimate premorbid intelligence. *Acta Neurologica Scandinavica*, *91*(5), 335–345.

Mackinnon, A., Ritchie, K., & Mulligan, R. (1999). The measurement properties of a French language adaptation of the National Adult Reading Test. *International Journal of Methods in Psychiatric Research*, *8*(1), 27–38.

Mackinnon, A., & Wooden, M. (2015). A short form of the National Adult Reading Test for use in epidemiological surveys. *Personality and Individual Differences*, *86*, 101–107.

Matsuoka, K., Uno, M., Kasai, K., Koyama, K., & 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*(3), 332–339.

McGrory, S., Austin, E. J., Shenkin, S. D., Starr, J. M., & Deary, I. J. (2015). From "aisle" to "labile": A hierarchical National Adult Reading Test scale revealed by Mokken scaling. *Psychological Assessment*, *27*(3), 932–943.

McGurn, B., Starr, J. M., Topfer, J. A., Pattie, A., Whiteman, M. C., Lemmon, H. A., Whalley, L. J., & Deary, I. J. (2004). Pronunciation of irregular words is preserved in dementia, validating premorbid IQ estimation. *Neurology*, *62*(7), 1184–1186.

Nelson, H. E. (1982). *National Adult Reading Test (NART): For the assessment of premorbid intelligence in patients with dementia: Test manual*. NFER-Nelson.

Nelson, H. E., & McKenna, P. A. T. (1975). The use of current reading ability in the assessment of dementia. *British Journal of Social and Clinical Psychology*, *14*(3), 259–267.

Nelson, H. E., & Willison, J. (1991). *National Adult Reading Test (NART)* (pp. 1–26). Nfer-Nelson.

O'Carroll, R. (1995). The assessment of premorbid ability: A critical review. *Neurocase*, *1*(1), 8389–89.

Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods*, *5*(3), 343–355.

Rain, R., & Zaborowska, L. IQ percentile calculator 2022. https://www.omnicalculator.com/health/iq-percentile.

Reale-Caldwell, A., Osborn, K. E., Soble, J. R., Kamper, J. E., Rum, R., & Schoenberg, M. R. (2021). Comparing the North American Adult Reading Test (NAART) and the Test of Premorbid Functioning (TOPF) to estimate premorbid Wechsler Adult Intelligence Scale-FSIQ in a clinical sample with epilepsy. *Applied Neuropsychology: Adult*, *28*(5), 564–572.

Roid, G. H., & Pomplun, M. (2012). The Stanford-Binet intelligence scales (5th ed.). In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 249–268). The Guilford Press.

Rolstad, S., Nordlund, A., Gustavsson, M. H., Eckerström, C., Klang, O., Hansen, S., & Wallin, A. (2008). The Swedish National Adult Reading Test (NART-SWE): A test of premorbid IQ. *Scandinavian Journal of Psychology*, *49*(6), 577–582.

Saal, F. E., Downey, R. G., & Lahey, M. A. (1980). Rating the ratings: Assessment the psychometric quality of rating data. *Psychological Bulletin*, *88*(2), 413–428.

Schmand, B., Bakker, D., Saan, R., & Louman, J. (1991). The Dutch Reading Test for Adults: A measure of premorbid intelligence level. *Tijdschr Gerontol Geriatr*, *22*(1), 15–19.

Schretlen, D. J., Winicki, J. M., Meyer, S. M., Testa, S. M., Pearlson, G. D., & Gordon, B. (2009). Development, psychometric properties, and validity of the Hopkins Adult Reading Test (HART). *The Clinical Neuropsychologist*, *23*(6), 926–943.

Sharpe, K., & 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*(4), 381–384.

Smith, D., Roberts, S., Brewer, W., & Pantelis, C. (1998). Test-retest reliability of the National Adult Reading Test (NART) as an estimate of premorbid IQ in patients with schizophrenia. *Pathologies of Body, Self and Space*, *3*, 71–80.

Starkey, N. J., & Halliday, T. (2011). Development of the New Zealand Adult Reading Test (NZART): Preliminary findings. *New Zealand Journal of Psychology*, *40*(3), 129–141.

Uttl, B. (2002). North American Reading Test: Age norms, reliability, and validity. *Journal of Clinical and Experimental Neuropsychology*, *24*(8), 1123–1137.

van der Linde, I., Bright, P., Forloni, G. (2018). A genetic algorithm to find optimal reading test word subsets for estimating full-scale IQ. *PLoS One*, *13*(10), e0205754.

van der Linde, I., Horsman, L., & Bright, P. (2022). The validity of abbreviated forms of the National Adult Reading Test and Spot-the-Word 2 for estimating full-scale IQ. *Neuropsychological Rehabilitation*, *32*(10), 2534–2543.

https://doi.org/10.1017/S1355617724000420 Published online by Cambridge University Press

Vaskinn, A., & Sundet, K. (2001). Estimering av premorbid IQ: En norsk versjon av National Adult Reading Test/Estimating premorbid IQ: A Norwegian version of National Adult Reading Test. *Tidsskrift for Norsk Psykologforening*, *38*(12), 1133–1140.

Veiel, H. O. F., & Koopman, R. F. (2001). The bias in regression-based indices of premorbid IQ. *Psychological Assessment*, *13*(3), 356–368.

Watt, S., Ong, B., & Crowe, S. F. (2018). Developing a regression equation for predicting premorbid functioning in an Australian sample using the National Adult Reading Test. *Australian Journal of Psychology*, *70*(2), 186–195.

Wechsler, D. (2001). *WTAR: Wechsler test of adult reading*. Psychological Corporation.

Wechsler, D. (2008). *Wechsler adult intelligence scale* (4th edn). Pearson Assessment.

Wechsler, D. (2011). *Test of premorbid functioning. UK version (TOPF UK)*. Pearson Assessment.

Yi, D., Seo, E. H., Han, J. Y., Sohn, B. K., Byun, M. S., Lee, J. H., Choe, Y. M., Ahn, S., Woo, J. I., Jun, J., Lee, D. Y., Forloni, G. (2017). Development of the Korean Adult Reading Test (KART) to estimate premorbid intelligence in dementia patients. *PLoS ONE*, *12*(7), e0181523.

https://doi.org/10.1017/S1355617724000420 Published online by Cambridge University Press