# Wechsler Test of Adult Reading in Parkinson's: stable premorbid measure despite cognitive decline

Kyla-Louise Horne, PhD[1], Reza Shoorangiz, PhD[1,2,3], Daniel J. Myall, PhD[1], Tim J. Anderson, FRACP, MD[1,2,4], John C. Dalrymple-Alford, PhD[1,2,5], & Michael R. MacAskill, PhD[1,2]

[1] New Zealand Brain Research Institute, 66 Stewart St, Christchurch, New Zealand

[2] Department of Medicine, University of Otago, Christchurch; Christchurch, New Zealand

[3] Department of Electrical and Computer Engineering, University of Canterbury, Christchurch, New Zealand

[4] Department of Neurology, Christchurch Hospital, Christchurch, New Zealand

[5] School of Psychology, Speech, and Hearing, University of Canterbury, Christchurch, New Zealand

## Abstract

Background. The Wechsler Test of Adult Reading (WTAR) has been recommended as a measure of premorbid cognitive function and as a reference against which to assess current cognitive status in people with Parkinson's disease. To fulfil either role, however, it needs to be shown that WTAR scores remain stable in the presence of the substantial cognitive deterioration that can occur in Parkinson's. Objective. To assess long-term trajectories of WTAR scores in people with Parkinson's as a function of age. Methods. From 252 Parkinson's and 57 Control participants who had completed at least two WTARs, we analysed scores over time using latent class trajectory modelling. This allows for individual participants to be classified into data-driven clusters, depending on the shape of their longitudinal trajectory. Results. WTAR scores were reasonably stable within both Controls and Parkinson's participants, even for those who progressed to dementia. In both groups and regardless of current cognitive status, scores were higher than expected from population norms. Conclusion. The WTAR is stable in Parkinson's even when participants decline from normal cognitive function to dementia. Nonetheless, its apparent over-estimation of premorbid IQ and restriction to English speakers makes it poorly-suited for assessing current cognitive impairment in individuals with Parkinson's.


*Keywords:* Parkinson disease, cognitive impairment, dementia, neuropsychology, premorbid function

Word count: Abstract: 200 (of 250), main body: 2608 (of 3700)

# Wechsler Test of Adult Reading in Parkinson's: stable premorbid measure despite cognitive decline

## Introduction

In 2012, the International Parkinson and Movement Disorder Society published its Task Force guidelines for the diagnosis of mild cognitive impairment in Parkinson's disease (PD-MCI).[1] When the diagnosis is based on a comprehensive battery of multiple individual neuropsychological tests, rather than a scale of global cognitive ability, significant impairment may be determined in three ways: (i) current performance being below appropriate norms, (ii) decline on serial testing, or (iii) decline from estimated premorbid levels. The first method tends to predominate in current research practice.[2] The latter two approaches, however, could be advantageous in resolving the ambiguity that poor current cognitive function in a given individual might reflect long-standing low cognitive ability, rather than a recent disease-related deterioration. In practice, however, one seldom has reference to prior formal serial testing, and so a valid method of estimating premorbid function is appealing. The Wechsler Test of Adult Reading (WTAR)[3] was proposed in the guidelines as one of two recommended tests for this purpose (along with the National Adult Reading Test, NART).[3] Both tests are based on the ability to correctly pronounce phonetically-irregular English words, such as 'porpoise' or 'hyperbole'. This ability has been shown to be preserved in the face of both normal ageing and a number of brain insults.[3]

When the WTAR was initially normed, it was tested in Parkinson's as well as a number of other neurological disorders, such as Huntington's, schizophrenia, and traumatic brain injury. Only in Alzheimer's disease were scores significantly lower than in matched controls.[3] Those results were, however, published in 2001, well before the formal MDS guidelines on diagnosing PD-MCI[1] and Parkinson's dementia[4] were promulgated. It is therefore unknown to what extent the findings from that original sample remain valid in the face of the significant cognitive deterioration that often occurs in Parkinson's.

We therefore examined WTAR scores in a cognitively well-characterised sample, spanning the range from normal function through mild cognitive impairment and dementia. We restricted the sample to participants who had completed a minimum of two WTARs, so that the trajectory of the measure within individuals over time could be examined. To do this, we used latent class trajectory analysis[5]. This allows for detecting heterogeneous patterns of change over time, rather than needing to assign individuals to *a priori* sub-groups.

## Methods

### Participants

The New Zealand Parkinson's Progression Programme (NZP3) is a longitudinal study of a convenience prevalence sample of idiopathic Parkinson's participants, largely recruited from the specialist Movement Disorders Clinic at the New Zealand Brain Research Institute (NZBRI). Initial, and periodically updated, ethical approval was granted by the Health and Disability Ethics Committees of the New Zealand Ministry of Health. The study commenced in 2007, with recruitment ongoing, and hence covers patients ranging from the recently-diagnosed to those with advanced disease. Inclusion and exclusion criteria have been described previously.[6] For this analysis, we selected data from the 252 Parkinson's and 57 Control participants who had completed at least two WTARs (range 2 - 8, mean = 3.70, total number of measures = 1155). The period between successive WTAR assessments was multi-modal, mostly clustering around intervals of one and two years, due to varying follow-up periods over the course of the study (see Supplementary Material). All participants were classified as having normal cognition, mild cognitive impairment, or dementia via a comprehensive Level II neuropsychological battery,[6,7] administered in accordance with MDS guidelines.[1,4] For PD-MCI, this was on the basis of at least two test scores falling $\geq 1.5$ standard deviations below norms in at least one cognitive domain. For

PD dementia, a significant impairment ($\geq$ 2.0 standard deviations below normative data) in at least two cognitive domains was required, as well as evidence of significant impairment in everyday functioning.[6,7] The WTAR was not used in the cognitive diagnostic classification procedure. Characteristics of the sample are shown in Table 1. Of the 249 Parkinson's participants with a recorded ethnicity, 240 (96.4%) identified as Pākehā (New Zealand European), 2 as Māori, 1 as Samoan, 1 as Indian, and 5 were 'Other'. Of the Controls, 56 (98.2%) identified as Pākehā and 1 was 'Other'. All participants were speakers of New Zealand English, and were assessed against the US norms of the WTAR, corrected for age, sex, and years of education (the provided ethnicity classifications are not useful in a New Zealand context and all participants were assessed against norms for "Whites").

## Latent class trajectory modelling

We fitted latent class trajectory models using the *hlme* function from the *lcmm* package[5] (version 1.9.5), running in the R statistical environment[8] (version 4.1.2) and used the *tidyverse* constellation of packages[9] for data manipulation and visualisation. The dependent variable was WTAR-estimated premorbid IQ, initially modelled within each subject as a polynomial (cubic) function of age, to allow for non-linear trajectories. The resulting trajectories were close to straight lines (see Supplementary Material) so we simplified the models to be a linear function of age. The models were not informed by any other variables (such as categorical MCI or dementia diagnoses). We fitted five separate models, with the specified number of latent classes increasing from 1 to 5. The *gridsearch()* function of the *lcmm* package was used to run 500 departures from random initial values for each multi-class model, using the one-class model from which to generate the starting values. The maximum number of iterations within the *hlme()* function was set at 1000. Parameters corresponding to the best log-likelihood were used as initial values for the final estimation of the parameters[5]. The optimal model was selected on the basis of it converging and having the lowest BIC (Bayesian information criterion) value of the five

candidates. When reporting values from the chosen model, we formed the interval in brackets following the maximum likelihood estimate (MLE) by subtracting and adding 1.96 times the standard error given by *hlme* to the MLE.

### Reproducibility

The code and anonymised dataset extracts sufficient to reproduce the analyses and generate this manuscript are publicly available at github.com/nzbri/wtar-trajectory. There are a number of decisions that can affect the outcome of a latent class modelling analysis and therefore in the Supplementary Material we report against the 16-item checklist "Guidelines for Reporting on Latent Trajectory Studies".[10]

## Results

### Distribution of scores

The distributions of all WTAR-estimated IQ scores for the Parkinson's and Control participants are shown in Figure 1. The mean of each group's latest score is given in Table 1, with both well above the expected population norm mean IQ score of 100. The population distribution of IQ should be symmetrical about 100, yet in our sample, 87.0% of Parkinson's WTAR scores were greater than or equal to 100, as were 92.5% of Control scores.

### Longitudinal trajectories

The raw data showed that WTAR scores were relatively constant over time (Figure 2A). This was evident even in participants whose overall cognitive performance declined substantially, with progression to dementia (Figure 2B). We modelled WTAR scores as a linear function of age, with repeated measures within individuals, fitting candidate models ranging from one to five latent trajectory classes. We selected the two-class model on the

basis of it having the lowest BIC value. Full details on the modelling process are provided in the Supplementary Material.

Individuals are assigned to the class for which they have the highest posterior classification probability. The mean assigned probability was 0.91 (SD 0.14), with a value of 0.7 being regarded as acceptable model performance.[11] The two trajectories are depicted with the raw data superimposed in Figure 2A. We labelled the first trajectory 'Typical', as this class captured 78% of all participants (81% of the PD and 65% of Control participants). The intercept of the Typical trajectory, representing the estimated WTAR score of a person at the mean age of assessment (70.7 years) was 107.6 [105.9, 109.2]. Scores then gently declined at a rate of -1.2 [-1.9, -0.6] points per decade. The second trajectory we labelled 'High performers'. This class, capturing the remaining 22% of all participants, had an intercept of 121.0 [118.9, 123.1] and a slope of -0.2 [-1.2, 0.8] points per decade (effectively flat). That is, both trajectories were relatively stable over time, and were separated primarily due to the initial scores of their respective clusters of participants. We make no strong claims as to the number of classes. For example a single trajectory model, somewhat intermediate between the two reported here, would also describe the data relatively well, and a single trajectory was the optimal model when cubic rather than linear functions were used (see Supplementary Material). The key findings are that the trajectories are relatively flat, and do not separate performance between the Parkinson's and Control participants.

## Discussion

Unlike in Alzheimer's disease, we found that WTAR scores were relatively stable over time in Parkinson's, despite many of our participants undergoing substantial declines in overall cognitive performance over the course of the study (Figure 2B). Latent class analysis revealed that in our particular sample, people fell into one of two trajectory classes, but these appeared to be largely driven by two clusters of performance at baseline,

rather than being trajectories that substantially differed longitudinally. Those with very high initial WTAR scores (the 'High performers' class) maintained near-ceiling performance levels (Figure 2A), even if they subsequently developed dementia (Figure 2B). The remaining 78% of participants (the 'Typical' class) had a wider range of initial scores, but were again relatively stable, with an estimated decline of only -1.2 points per decade. If the WTAR was used to assess disparities between current cognitive performance and a premorbid level of function, a decadal decline of that magnitude would be unlikely to influence assignment to clinical categories like MCI and dementia.

Measures of premorbid function are closely tied to the concept of cognitive reserve.[14] For example, higher NART scores have been associated with better current neuropsychological test performance in people with Parkinson's,[15] indicating that higher cognitive reserve may be protective against subsequent cognitive decline. This can lead to circular reasoning if the premorbid measure is not reliable, because if the disease process reduces current performance on that measure, then inferring poorer cognitive reserve would not be valid. That is, due to current impairments, we would be estimating a lower level of premorbid function than the person actually had. Our findings, of stable WTAR scores despite substantial current cognitive decline, provide support for the WTAR being a stable indicator of cognitive reserve in Parkinson's.

There are reservations about the use of the WTAR, however. Although still in wide use, it is outdated, inasmuch as it was originally devised to estimate IQ scores from the WAIS-III (Wechsler Adult Intelligence Scale, released in 1997). The WAIS-IV (2008) is now in widespread use, and the WAIS-V is currently under development. The TOPF (Test of Premorbid Function) estimates WAIS-IV IQ and was designed as a successor to the WTAR, but has had limited uptake.[16] The age of an IQ-related test is important. The Flynn effect[17] is the robust finding that cohort IQ scores have increased substantially over time (by up to 1 standard deviation per generation). Our sample would (on average) have been born well after their age-matched WAIS-III normative sample group and that might

partly explain their above-average estimated IQ levels. Another possibility is that those who volunteer for research are unlikely to be representative of their respective populations. Such a self-selection effect should, however, apply more to the controls than to the Parkinson's group, particularly because our controls were excluded (at study entry only) if they showed cognitive impairment. This restriction did not apply to the Parkinson's participants, but again, as research volunteers, they may not be representative of the Parkinson's population at large.

To provide evidence of cognitive decline, scores on the tests from a neuropsychological test battery must fall below some reference value. If the WTAR was used as an individualised premorbid reference for those scores, rather than a per-test z-score norm of zero, we would expect to see inflated numbers of participants being classified as MCI and PDD, because of the over-estimation of premorbid function. The PD-MCI guidelines do not actually formalise how the results of premorbid function tests should be used in practice to demonstrate decline in cognitive functioning.[18,19] We are aware of only one study[20] that has compared the norms-based vs premorbid estimation approaches to establishing cognitive impairment. They converted the WTAR-estimated full-scale IQ to a z-score and used that as the reference against which to test each individual neuropsychological test z-score. Any test that was lower by more than 1.5 SD was deemed impaired. Of their sample of 139 non-demented people with Parkinson's, the WTAR method classified many more of them as having PD-MCI (79%) than did a norms-based Level II classification method (33%). Marras *et al.* interpreted this as evidence that most people with Parkinson's have undergone at least some decline relative to their premorbid function. Conversely it might support our contention that the WTAR simply over-estimates that premorbid level of function – just as in our sample, where the distribution of estimated IQ values was well above the supposed population mean of 100.

The WTAR (and similar tests such as the NART) are also limited in that they are inherently tied to peculiarities of written and spoken English. The orthography (writing

system) of a language can be described as deep/opaque when it contains many irregularities in its mappings of graphemes to phonemes, or shallow/transparent when the mappings are largely regular.[21] In a quantitative analysis of 17 orthographies,[22] the reading of English words stood out as particularly opaque, whereas by comparison Arabic, Finnish, Korean, Serbo-Croatian, and Turkish were highly transparent. Attempting to devise analogues of tests like the WTAR for speakers of languages with transparent orthographies is unlikely to be fruitful.

Even across varieties of English, both between and within countries, the US- and UK-based selection of words can be problematic. For example, the TOPF was found to not reliably predict WAIS-IV IQ in New Zealand English speakers of Māori ethnicity.[23] A word like 'porpoise', ranked in the TOPF as relatively simple (28 out of 70), resulted in more errors than a word like 'plethora' (ranked 42 out of 70), likely reflecting different word frequencies between North American and New Zealand English. After the commencement of our study, a New Zealand Adult Reading Test (NZART) was created[24], with a more culturally-relevant selection of words, but this will require further development to supplant the established international tests. Neuropsychological screening tests developed in the "Anglosphere" have been shown to have poor cross-cultural validity in Parkinson's.[25] This extends even to visuoperceptual and non-verbal executive tasks previously assumed to be "culture fair".[26] It is perhaps therefore not surprising that an inherently language-based test faces issues with cross-cultural applicability. In subsequent revisions of the PD-MCI diagnostic criteria, if premorbid measures continue to be advocated, more consideration could be given to alternative techniques[27] that could have improved validity in a wider international and cross-cultural context.

In summary, the WTAR is a measure that is quite stable in Parkinson's, even in the presence of a substantial decline in current cognitive performance. This makes it a valid research tool to probe cognitive reserve and premorbid function differences across groups. In absolute terms, however, it appears to overestimate premorbid IQ, making it unsuitable

to establish a decline in function within individuals. All such measures likely suffer from cross-cultural word choice issues even across different varieties of English. Analogous tests are not able to be created in languages that do not share the orthographic peculiarities of English. In future revisions of international MDS guidelines, we therefore believe that a test of premorbid function like the WTAR should no longer be recommended as one of the means of formally establishing cognitive decline at an individual level in Parkinson's.

## References

1

Litvan I, Goldman JG, Tröster AI, *et al.* Diagnostic criteria for mild cognitive impairment in Parkinson's disease: Movement Disorder Society Task Force guidelines. *Mov Disord* 2012;**27**:349–56. doi:10.1002/mds.24893

2

Aarsland D, Batzu L, Halliday GM, *et al.* Parkinson disease-associated cognitive impairment. *Nat Rev Dis Primers* 2021;**7**:47. doi:10.1038/s41572-021-00280-3

3

Strauss E, Sherman EMS, Spreen O. *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary.* 3rd ed. Oxford: : Oxford University Press 2006.

4

Dubois B, Burn D, Goetz C, *et al.* Diagnostic procedures for Parkinson's disease dementia: Recommendations from the Movement Disorder Society Task Force. *Mov Disord* 2007;**22**:2314–24.

5

Proust-Lima C, Philipps V, Liquet B. Estimation of extended mixed models using latent classes and latent processes: The R package lcmm. *J Stat Softw* 2017;**78**:1–56. doi:10.18637/jss.v078.i02

6

Wood K-L, Myall DJ, Livingston L, *et al.* Different PD-MCI criteria and risk of dementia in Parkinson's disease: 4-year longitudinal study. *npj Parkinson's Dis* 2016;**2**:15027. doi:10.1038/npjparkd.2015.27

7

Dalrymple-Alford JC, Livingston L, MacAskill MR, *et al.* Characterizing mild cognitive impairment in Parkinson's disease. *Mov Disord* 2011;**26**:629–36. doi:10.1002/mds.23592

8

R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: : R Foundation for Statistical Computing 2021. https://www.R-project.org/

9

Wickham H, Averick M, Bryan J, *et al.* Welcome to the Tidyverse. *Journal of Open Source Software* 2019;**4**:1686. doi:10.21105/joss.01686

10

Schoot R van de, Sijbrandij M, Winter SD, *et al.* The GRoLTS-checklist: Guidelines for reporting on latent trajectory studies. *Struct Equ Modeling* 2017;**24**:451–67. doi:10.1080/10705511.2016.1247646

11

Lennon H, Kelly S, Sperrin M, *et al.* Framework to construct and interpret latent class trajectory modelling. *BMJ Open* 2018;**8**:e020683. doi:10.1136/bmjopen-2017-020683

12

281 MacAskill MR. Distribution of Wechsler Test of Adult Reading (WTAR) estimated premorbid IQ scores in Parkinson's disease and controls. 2022. doi:10.6084/M9.FIGSHARE.19126538.V1
282
283 13
284 MacAskill MR. Longitudinal WTAR-estimated premorbid IQ score trajectories in Parkinson's disease and controls. 2022. doi:10.6084/M9.FIGSHARE.19126553
285
286 14
287 Stern Y. Cognitive reserve in ageing and Alzheimer's disease. *Lancet Neurol* 2012;**11**:1006–12. doi:10.1016/S1474-4422(12)70191-6
288
289 15
290 Koerts J, Tucha L, Lange KW, *et al.* The influence of cognitive reserve on cognition in Parkinson's disease. *Journal of Neural Transmission* 2013;**120**:593–6. doi:10.1007/s00702-012-0916-6
291
292 16
293 Bright P, Linde I van der. Comparison of methods for estimating premorbid intelligence. *Neuropsychol Rehabil* 2020;**30**:1–4. doi:10.1080/09602011.2018.1445650
294
295 17
296 Flynn JR. Searching for justice. The discovery of IQ gains over time. *Am Psychol* 1999;**54**:5–20. doi:10.1037/0003-066X.54.1.5
297
298 18
299 Marras C, Tröster AI, Kulisevsky J, *et al.* The tools of the trade: A state of the art 'How to Assess Cognition' in the patient with Parkinson's disease. *Mov Disord* 2014;**29**:584–96. doi:10.1002/mds.25874
300
301 19

302 Geurtsen GJ, Hoogland J, Goldman JG, *et al.* Parkinson's disease mild cognitive impairment: Application and validation of the criteria. *J Parkinson Dis* 2014;**4**:131–
303 7. doi:10.3233/JPD-130304

304 20

305 Marras C, Armstrong MJ, Meaney CA, *et al.* Measuring mild cognitive impairment in patients with Parkinson's disease: Measuring PD-MCI. *Mov Disord* 2013;**28**:626–
306 33. doi:10.1002/mds.25426

307 21

308 Frost R, Katz L, Bentin S. Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human*
309 *Perception and Performance* 1987;**13**:104–15. doi:10.1037/0096-1523.13.1.104

310 22

311 Marjou X. OTEANN: Estimating the transparency of orthographies with an artificial neural network. In: *Proceedings of the Third Workshop on Computational Typology and Multilingual NLP*. Online: : Association for Computational Linguistics 2021.
312 1–9. doi:10.18653/v1/2021.sigtyp-1.1

313 23

314 Dudley M, Scott K, Barker-Collo S. Is the test of premorbid functioning a valid
315 measure for Māori in New Zealand? *N Z J Psychol* 2017;**46**:8.

316 24

317 Starkey NJ, Halliday T. Development of the New Zealand Adult Reading Test
318 (NZART): Preliminary findings. *N Z J Psychol* 2011;**40**:13.

319 25

320 Statucka M, Cherian K, Fasano A, *et al.* Multiculturalism: A challenge for cognitive screeners in Parkinson's disease. *Mov Disord Clin Prac* 2021;**8**:733–42.
321 doi:10.1002/mdc3.13240

26

Statucka M, Cohn M. Origins matter: Culture impacts cognitive testing in Parkinson's disease. *Front Hum Neurosci* 2019;**13**:269. doi:10.3389/fnhum.2019.00269

27

Franzen MD, Burgess EJ, Smith-Seemiller L. Methods of estimating premorbid functioning. *Arch Clin Neuropsychol* 1997;**12**:711–38. doi:10.1093/arclin/12.8.711

Table 1

*Demographics of the Parkinson's and Control groups, and cognitive status at their latest assessment.*

| | Control | Parkinson's |
|---|---|---|
| n | 57 | 252 |
| Male | 63.2% | 69.4% |
| Age (SD) | 76.1 (7.9) | 72.7 (7.0) |
| Years of education (SD) | 13.6 (2.7) | 12.7 (2.6) |
| WTAR (SD) | 112.6 (9.5) | 108.9 (9.3) |
| Normal cognition | 94.7% | 41.3% |
| MCI | 3.5% | 34.9% |
| Dementia | 1.8% | 23.8% |

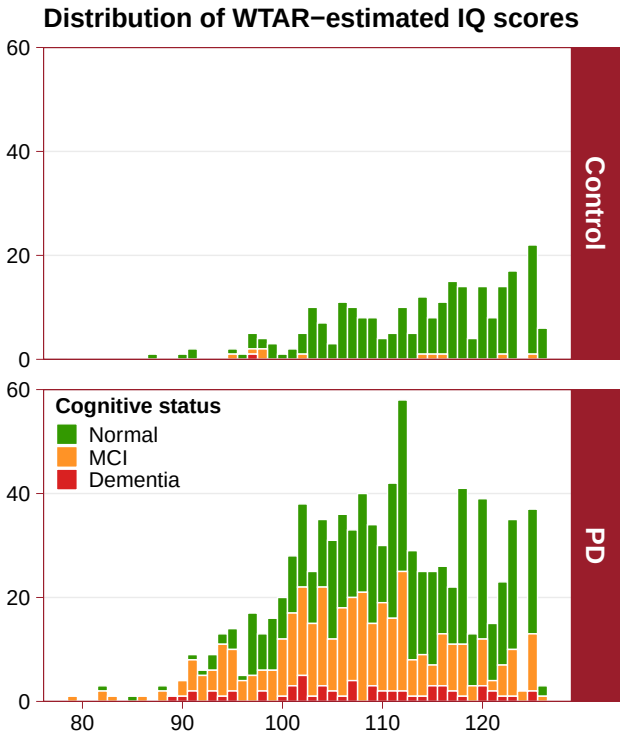


*Figure 1.* The distribution of WTAR-estimated IQ scores across all participants, containing at least two observations from each. Observations are colour-coded by their cognitive status at each session. Regardless of current cognitive status, WTAR scores span the full range. For each group, the distribution was skewed well above the population norm mean of 100. Figure available under an open CC-BY licence.[12]


*Figure 2.* ***(A)*** Longitudinal WTAR score trajectories for all individuals, divided into Controls, Parkinson's participants who remained dementia-free, and those who developed dementia. Superimposed are the two predicted latent class trajectories (blue ribbon = *High performers*, brown ribbon = *Typical*). Blue and brown connecting lines link observations from individuals from each class. ***(B)*** Mean z-score from multiple tests in a comprehensive

# Appendix

## Supplementary Material

Horne *et. al.*, *The Wechsler Test of Adult Reading in Parkinson's: stable premorbid measure despite cognitive decline*

Here we report the analysis of the above paper against the following checklist for reporting latent class trajectory analyses:

van de Schoot, R., Sijbrandij, M., Winter, S. D., Depaoli, S., & Vermunt, J. K. (2017). The GRoLTS-checklist: Guidelines for reporting on latent trajectory studies. *Structural Equation Modeling: A Multidisciplinary Journal, **24,*** 451–467. https://doi.org/10.1080/10705511.2016.1247646

### 1. Is the metric of time used in the statistical model reported?

We used age as the time metric (in years to one decimal place). This was to allow for direct comparison against controls (which would not be possible with disease duration). Follow-up time within the study was also not considered, as participants can enter the study at arbitrary points (in terms of both age and disease duration).

To avoid instabilities in the cubic models, age was standardised by subtracting the global mean age at assessment and dividing by ten. For comparability, this was also done for the linear models. In the visualisations, predictions were projected back into untransformed age in years.

### 2. Is information presented about the mean and variance of time within a wave?

The metric was age rather than intended study time points. The overall goal in the study was to follow-up participants two-yearly, however, the interval between follow-up times was multi-modal and varied both within and across participants (see Figure A1).

This variability was due to both changing study demands (such as recruiting participants for sub-studies) and participant health and availability.

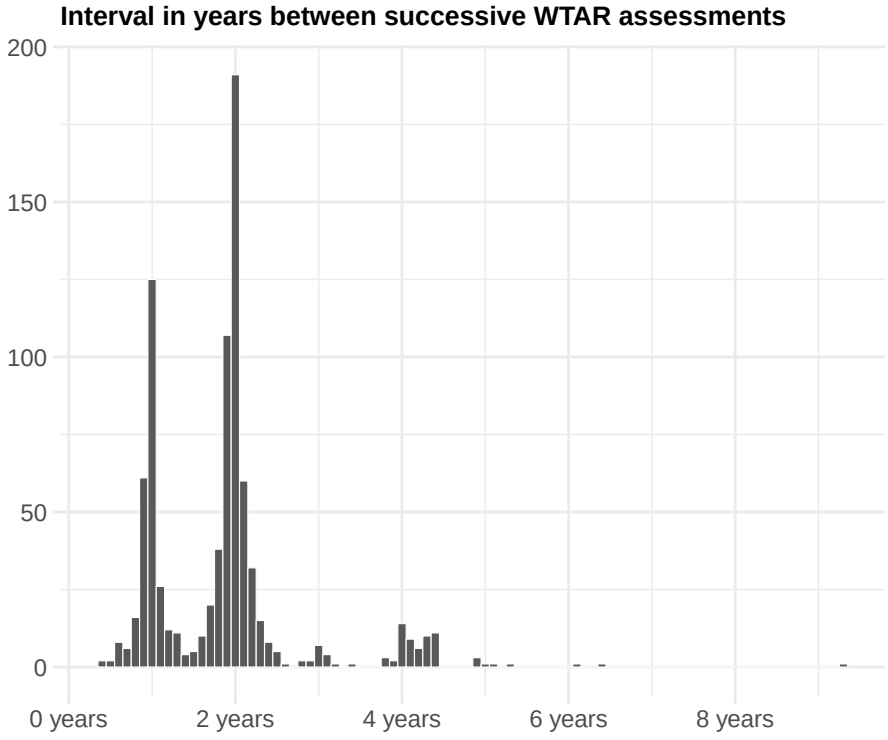


*Figure A1.* Frequencies of the interval between successive WTAR assessments.

**3a. Is the missing data mechanism reported?**

Data was not missing at random: participants are more likely to drop out of the study due to increasing disease severity, other health issues, and old age. As shown in Figure 2, however, many remained in the study until the end point of WTAR collection (the onset of dementia). Those participants did not show a decline in WTAR performance relative to other participants. Due to the findings of this study (that the WTAR is relatively stable throughout the disease course), differential drop-out is unlikely to affect the results.

**3b. Is a description provided of what variables are related to attrition/missing data?**

They are speculated upon in *3a* above.

**3c. Is a description provided of how missing data in the analyses were dealt with?**

As noted in *3a* above, missing data was not thought to be problematic in this analysis and was not included explicitly in the modeling process.

**4. Is information about the distribution of the observed variables included?**

In Figure 1 in the accompanying manuscript, we show the distribution of the WTAR-estimated IQ scores in each group, showing the cognitive status at each observation.

**5. Is the software mentioned?**

See the Methods section in the accompanying manuscript.

**6a. Are alternative specifications of within-class heterogeneity considered (e.g., LGCA vs. LGMM) and clearly documented? If not, was sufficient justification provided as to eliminate certain specifications from consideration?**

An LGMM which includes a varying intercept and slope by participant was used in this analysis and, given the nature of the within-subject variance in the data, is more appropriate than an LGCA.

**6b. Are alternative specifications of the between-class differences in variance–covariance matrix structure considered and clearly documented? If not, was sufficient justification provided as to eliminate certain specifications from consideration?**

The variance-covariance matrix was allowed to vary by a class-specific proportional parameter. The more restrictive case of restricting all classes to a common variance-covariance matrix structure was not considered.

**7. Are alternative shape/functional forms of the trajectories described?**

We initially modeled the trajectories using a cubic function, to allow for the depiction of any curvilinear decline in performance. The resulting trajectory in the one-class cubic model (which had the lowest BIC) was effectively flat, and hence we moved to linear modelling for the ease of interpretation of the model coefficients.

**8. If covariates have been used, can analyses still be replicated?**

Not applicable: age was the only predictor in the models.

**9. Is information reported about the number of random start values and final iterations included?**

The `gridsearch()` function of the `lcmm` package was used to run 500 departures from random initial values for each multi-class model, using the one-class model to generate those starting values from. The maximum number of iterations within the `hlme()` function was set at 1000.

Table A1

*Comparison of hlme models with G = 1 to 5 linear latent trajectory classes.*

| G | loglik | converged | BIC | entropy | %class1 | %class2 | %class3 | %class4 | %class5 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | -3,116.9 | yes | 6,268.2 | 1.00 | 100.0 | | | | |
| 2 | -3,102.4 | yes | 6,262.2 | 0.71 | 78.3 | 21.7 | | | |
| 3 | -3,092.6 | yes | 6,265.5 | 0.64 | 44.0 | 40.5 | 15.5 | | |
| 4 | -3,086.2 | yes | 6,275.5 | 0.69 | 30.4 | 9.7 | 46.9 | 12.9 | |
| 5 | -3,082.3 | yes | 6,290.8 | 0.68 | 27.5 | 48.9 | 15.5 | 2.3 | 5.8 |

**10. Are the model comparison (and selection) tools described from a statistical perspective?**

We selected the two-class linear model on the basis of it converging and having the lowest BIC value.

**11. Are the total number of fitted models reported, including a one-class solution?**

For both linear and cubic trajectories, models were fitted with one through five classes. See the model comparison tables in Tables A1 and A2.

**12. Are the number of cases per class reported for each model (absolute sample size, or proportion)?**

The number of cases within each class of the selected model is shown in Table A3. See Tables A1 and A2 for the percentage of cases within each class in all candidate models.

Table A2

*Comparison of hlme models with G = 1 to 5 cubic latent trajectory classes.*

| G | loglik | converged | BIC | entropy | %class1 | %class2 | %class3 | %class4 | %class5 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | -3,108.2 | yes | 6,302.4 | 1.00 | 100.0 | | | | |
| 2 | -3,093.6 | yes | 6,307.6 | 0.77 | 81.2 | 18.8 | | | |
| 3 | -3,083.0 | yes | 6,320.9 | 0.75 | 61.8 | 13.3 | 24.9 | | |
| 4 | -3,076.4 | yes | 6,342.0 | 0.68 | 23.6 | 23.9 | 22.7 | 29.8 | |
| 5 | -3,069.0 | yes | 6,361.7 | 0.71 | 21.7 | 3.2 | 18.8 | 24.6 | 31.7 |

**13. If classification of cases in a trajectory is the goal, is entropy reported?**

Classification of cases was not a goal of the analysis. However, entropy-related measures are shown in Tables A4 and A5.

**14a. Is a plot included with the estimated mean trajectories of the final solution?**

The trajectories are depicted within Figure 2 of the accompanying manuscript, and in isolation in Figure A2.

**14b. Are plots included with the estimated mean trajectories for each model?**

Figure A3 shows the predicted trajectories of all 5 candidate linear and cubic models, as tabulated in Tables A1 and A2.

Table A3

*Number within each class in the selected model.*

| | Typical | High |
|---|---|---|
| N | 242 | 67 |
| % | 78.32 | 21.68 |

Table A4

*Confidence of assignment to each class within the selected model.*

| | prob1 | prob2 |
|---|---|---|
| Typical | 0.96 | 0.04 |
| High | 0.29 | 0.71 |

Table A5

*Levels of classification confidence for each class within the selected model.*

| | Typical | High |
|---|---|---|
| prob>0.7 | 93.39 | 64.18 |
| prob>0.8 | 91.32 | 7.46 |
| prob>0.9 | 86.78 | 0.00 |

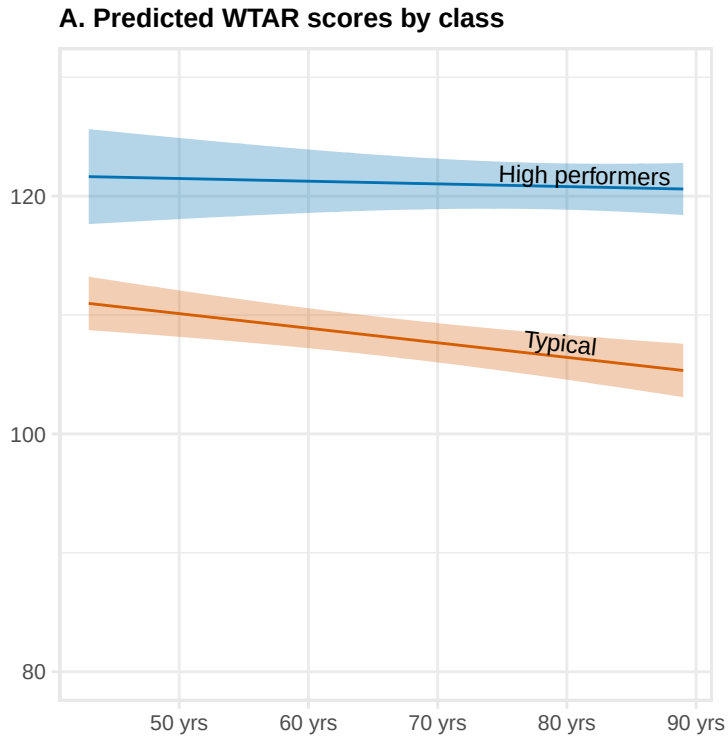


*Figure A2.* The predicted trajectories of the chosen two-class linear model.

**14c. Is a plot included of the combination of estimated means of the final model and the observed individual trajectories split out for each latent class?**

This is shown in Figure 2 of the accompanying manuscript.

**15.Are characteristics of the final class solution numerically described (i.e., means, SD/SE, n, CI, etc.)?**

Yes, see the Results section of the accompanying manuscript.

**16. Are the syntax files available (either in the appendix, supplementary materials, or from the authors)?**

Yes, they are publicly available at github.com/nzbri/wtar-trajectory.

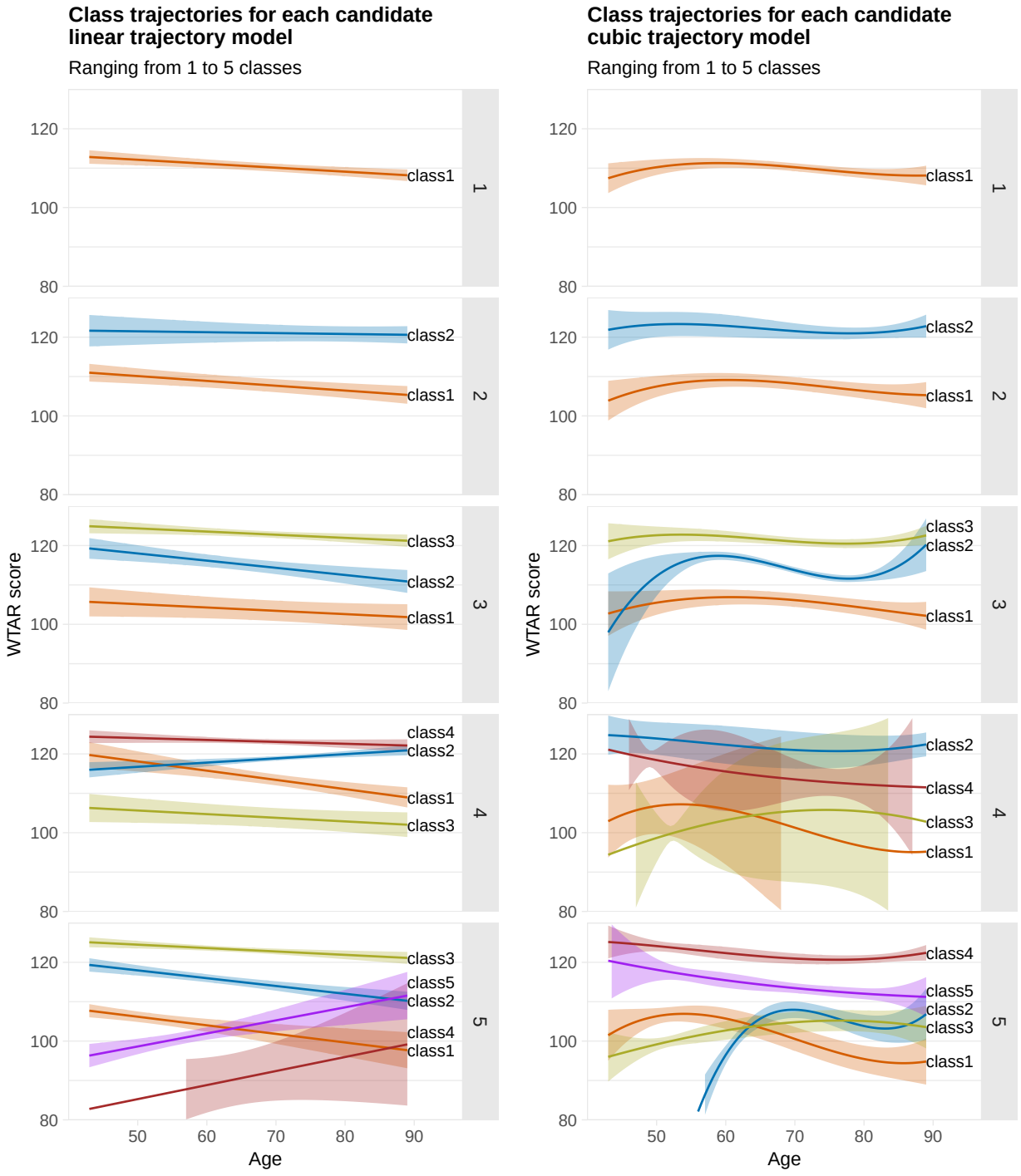


*Figure A3.* The predicted trajectories of all candidate latent trajectory models (the selected one was the two-class linear trajectory model, as shown in Figure A2). The percentage of participants assigned to each class in each model can be found in Tables A1 and A2.